Modeling approaches to the analysis of trade policy: computable general equilibrium and gravity models

Olena Ivus and Aaron Strong

Introduction

Gravity models and computable general equilibrium (CGE) models are the most commonly used analytical techniques to perform a quantitative analysis in the area of trade policy. These models provide a consistent economy-wide picture that can be very beneficial to policy makers. Both CGE and gravity models have the advantages of general equilibrium approaches in examining a great variety of questions. In the partial equilibrium models of international trade the focus is on one sector of the economy with the cross-sector effects being disregarded. General equilibrium modeling, in turn, takes explicit account of the consequences that a policy change in one sector has on other sectors of the economy.

The gravity model is a popular empirical approach to trade that has been used widely for analyzing the impact of different trade policy issues on bilateral trade flows between different geographical entities. This model takes an ex-post approach to perform trade policy analysis. Gravity models measure the effect on trade flows of a past trade policy. By contrast, CGE modeling takes an ex-ante approach, which involves quantifying the future effects of a new policy. In addition, gravity models only explain the pattern of bilateral trade and do not provide direct estimates of welfare costs. CGE models, on the other hand, are generally used to quantify the impact of a change in trade policy on countries' welfare levels and the distribution of income across multi-country regions.

The main goal of this chapter is to provide an introduction to both CGE and gravity models. The next section discusses the gravity model. First, the basic gravity equation is introduced. Second, the theoretical foundations of gravity models are provided. Third, issues and caveats concerning the empirical estimation of gravity models are considered. The section concludes with a selective review of studies based on the gravity model. The third section presents the CGE model. First, the advantages and disadvantages of CGE modeling are discussed. Second, the basic structure of CGE models is presented. Third, the steps used to construct a CGE model are described. The section concludes with a discussion of the Harrison/Rutherford/Tarr Multi-Regional Global Trade Model.

Gravity models

Tinnbergen (1962) and Pöyhönen (1963) were the first economists to use a gravity-style equation to analyze international trade flows. Since these foundations were laid, the gravity model has become a popular empirical trade approach that has been used widely for analyzing the impact of different trade policy issues on bilateral trade flows between different geographical entities. It has also been applied to a wide range of other questions, such as foreign direct investment, tourism, migration, commuting, and so on.
**The basic gravity equation**

The name of gravity model was derived from Newton’s ‘law of universal gravitation’. In this version of ‘gravity’, the extent of a trade flow between two countries is equal to the product of their masses or economic sizes divided by a resistance or distance factor. A justification for the gravity models of Tinbergen (1962) and Pöyhönen (1963) can be based on Walrasian general equilibrium theory. The gravity equation is viewed as a representation of demand and supply forces. In this case, aggregate income of the importer proxies the level of demand in the destination region and aggregate income of the exporter proxies the level of its supply. Distance is used as a proxy for transport costs.

In greater detail, the general gravity model specifies that the bilateral trade between countries \(i\) and \(j\) in year \(t\) is positively related the economic sizes of the two countries, proxied by GDP, and negatively related to the trade costs, proxied by distance between the two countries’ capital cities. The baseline specification of the gravity model with exports as a dependent variable is summarized in Table 5.1.

In the importing country, a higher level of income should imply greater imports. In the exporting country, a higher level of income will give rise to a greater level of overall production and this, in turn, will increase the availability of goods for export. Distance drives a wedge between demand and supply, resulting in a lower equilibrium export flows. The model may be estimated for a single year, as a so-called cross-section of trading countries, or pooled over several years.

In order to account for as many extraneous factors as possible, it is common to augment the basic gravity equation with a number of extra conditioning variables that affect trade. Many authors estimate gravity equations with the per capita incomes of the exporter and importer as an additional measure of country size. Some authors have included a country ‘remoteness’ variable in the gravity equation. The effect of this variable measured by its estimated coefficient is expected to be positive, since the less remote a country is, the more sources of imports it has and, as a result, the smaller share of its imports comes from each particular source. Anderson and van Wincoop (2003), in turn, stress the importance of introducing ‘multilateral resistance’ terms – measured by the average trade barriers of the exporting and importing countries – into the estimated equation. They argue that adding a remoteness index is in discordance with theory, because it is a function of distance and it does not capture any other barriers to trade. Anderson and van Wincoop (2004) note that a multilateral resistance index may be replaced with importer and exporter country-specific effects. Rose (2004) applies an empirical strategy to control for as many ‘natural’

---

**Table 5.1 The baseline specification of the gravity model\(^a\)**

<table>
<thead>
<tr>
<th>(X_{ijt})</th>
<th>exports from country (i) to country (j) in year (t) depend on:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Y_{it})</td>
<td>GDP of the exporter (i) at time (t)</td>
</tr>
<tr>
<td>(Y_{jt})</td>
<td>GDP of the importer (j) at time (t)</td>
</tr>
<tr>
<td>(D_{ij})</td>
<td>distance between the trading regions (i) and (j)</td>
</tr>
</tbody>
</table>

**Note:**

\(^a\) In the estimation procedure a log-linear form (that is taking the natural logarithms of the variables) is often applied. In this case the specification will be the following:

\[
\log X_{ijt} = \alpha + \beta_1 \log Y_{it} + \beta_2 \log Y_{jt} + \beta_3 \log D_{ij} + \epsilon_{ijt}
\]
causes of trade as possible. In this case, the gravity equation includes variables for sharing a common land border, speaking a common language, ever having been colonized, using the same currency at time $t$, and so on.

**The theoretical foundations of gravity models**

Despite the gravity model’s considerable empirical success (for example, its high explanatory power), it was long criticized for lacking strong theoretical foundations. More recently, different theories have been developed to establish rigorous theoretical underpinnings of the gravity model. Anderson (1979), Bergstrand (1985, 1990), Deardorff (1998), and Eaton and Kortum (2002) have developed micro-foundations for the gravity model. Anderson (1979) provided a theoretical basis for the gravity model by assuming constant elasticity of substitution (CES) preferences and goods that are differentiated by country of origin. Bergstrand (1990) derived a gravity equation from a monopolistic competition trade model in which the countries are completely specialized in different product varieties. In this case, each country is exports one variety of a differentiated product to other countries. Deardorff (1998) has shown that the gravity model can arise from the Heckscher-Ohlin model, which explains trade based on relative differences in factor endowments across countries. Eaton and Kortum (2002) obtained a gravity equation from a Ricardian type of model, which explains trade based on relative differences in technology across countries.

Feenstra (2004) notes that the conventional gravity model assumes identical prices across countries. Therefore, price is not included in the gravity equation as a variable that affects bilateral trade flows. Under the micro-foundations approach this results in misspecification of the gravity model. It is important to allow for differing prices due to trade barriers between the countries. The gravity equation with so called ‘price effects’ was derived by Anderson (1979). Feenstra (2004) suggests three approaches to estimating this equation. First, the price effects may be measured by price indexes, as in Bergstrand (1985) and Baier and Bergstrand (2001). Second, estimated border effects may be used as an alternative measure, as in Anderson and van Wincoop (2003). Third, a fixed-effects approach, which allows each country to be different, may be applied as in Redding and Venables (2000) and Rose and van Wincoop (2001).

**Estimation of gravity models**

The gravity model is a conventional method used to estimate the impact of various types of trade-related policies on international trade flows. Many gravity-model papers, for example, examine the effect of the formation of regional trade areas (RTAs). In this case, the gravity equation is extended using regional dummy variables, which indicate whether or not a pair of countries is in the same region. However, one needs to be careful when interpreting the estimated coefficient, which describes the empirical effect of this dummy variable. Ideally, the intent is to measure the impact of the RTA, but other effects may also inadvertently be captured due to measurement problems. For example, the true costs of bilateral trade may be partially subsumed by the variables related to trade agreements because the distance between capital cities does not fully reflect the trade costs. Of course, simple distance measures are flawed in many ways; most countries have multiple economic centers and other features matter such as infrastructure quality and border waiting times.
Regional dummy variables, which are intended to show the impact of RTAs, may also catch the effects of any variables that are not included in the gravity regression. The baseline gravity model approach makes an assumption that the level of bilateral trade depends only on the included economic features of a given pair of countries. However, the level of export from country \( i \) to countries \( j \) and \( k \) may be different even if countries \( j \) and \( k \) have the same GDP levels and they are equally distant from country \( i \). These differences can be explained by political factors, historical links, cultural similarities, and so on, that are correlated with levels of bilateral trade and with the baseline gravity variables. Omission or misspecification of these variables will lead to omitted-variables bias or the so called heterogeneity bias.

As demonstrated by Cheng and Wall (2005), standard estimation methods overestimate trade between low-trade countries and underestimated trade between high-trade countries. They argue that the inability of the standard cross-section estimation to account for the pairwise heterogeneity of bilateral trade relations is the principal cause of the bias. To eliminate the heterogeneity bias Cheng and Wall (2005) adopted a model which includes both the country-pair and year specific effects. The term which is common to all years, but specific to the country pairs is used to take into account the specific country-pair effects between the trading partners, such as distance, border, language, culture, and so on. The term, which is common to all pairs, but specific to each year, is included to capture the year-specific fixed effects. They will catch all the omitted factors that affect bilateral trade, are constant across trading pairs and vary over time. Alternatively, Mátyás (1997) has emphasized: ‘Unfortunately, none of the applications of this model bothered to take into account the local, target and time effects, which means that all practitioners were imposing the unnecessary restrictions . . . These are unlikely to be correct.’ In this way, Mátyás (1997) argues that the correct econometric specification of the gravity model should include the importer (target) country effects, the exporter (local) country effects, and the time effects.

In summary, the main benefit of both of the above extensions of the gravity model is that they help to control for omitted variables that are unobservable or difficult to quantify. Adopting either or both extensions will provide a robustness check, which will help avoid misinterpretation of empirical results from the simpler formulations of the model.

**Applications: The impact of membership in the WTO and RTAs on trade**

Several studies have applied gravity equations to provide an empirical examination of the impact of multilateral or regional trade agreements on international trade and, thus, have contributed to the refinement of the gravity model approach. This section provides a selective overview of some of these applications.

A recent study by Rose (2004) provides a comprehensive econometric study that analyzes the effect of the World Trade Organization (WTO), the Generalized Agreement on Tariffs and Trade (GATT), and the General System of Preferences (GSP). The augmented gravity model studied real bilateral trade flows between trading countries for the period from 1948 to 1999. The author concluded that membership in WTO/GATT did not imply an increase in trade intensities because the volume of bilateral trade between members and non-members was not significantly different.

The results of Rose (2004) have been strongly questioned by Subramanian and Wei (2003) who contend that the analysis needs to take into account liberalization asymmetries
arising from the WTO: between developed and developing countries; between developing countries that joined the WTO before and after the Uruguay Round; and between trade sectors. Once the econometric specification incorporates these types of unevenness in the patterns of trade liberalization, the impact of the WTO on promoting world trade appears to be strong and positive, although uneven. On the other hand, the authors suggest that Rose’s gravity model specification needs to include country fixed effects to capture the impact of multilateral resistance.²

The economic implications of RTA formation for international trade have been examined in many empirical papers. Bayoumi and Eichengreen (1995) make an important distinction between trade creation effects, leading to increases in the intensity of trade between members of RTAs, and trade diversion effects, leading to decreases in trade with third countries.³ The authors apply the gravity model with the addition of specific dummy variables to capture the impact of participation in various RTAs. A positive coefficient on the dummy variable indicating both countries of a bilateral pair are members of the RTA, suggests that they trade more with one another than is predicted by their incomes and distance, and, so, would provide evidence of a trade creation effect. A negative coefficient on a dummy variable indicating that one country in a bilateral pair is a member of an RTA but the other is not, would suggest trade diversion vis-à-vis the rest of the world.

In order to identify differences over time in the trade creating and trade diverting effects of EU integration, successive cross-sections were analyzed for a sample of 21 industrial countries and a period of 1953–1992. The authors found that the European Free Trade Association (EFTA) resulted in trade creation, while the European Economic Community (EEC) promoted trade within the region through the combination of trade creation and trade diversion. It is also noted that the accession of Spain and Portugal resulted in almost no trade diversion.

Christie (2002) presents a classical approach to the problem of quantifying potential trade levels, with a specific emphasis on trade flows with and within southeast Europe. After applying the gravity model approach and using panel data from 1996 to 1999 the author found that EU members trade 122 percent more than non-member countries, however the cross-section estimation for 1999 indicated that this effect decreased to 98 percent. The author notes: ‘Overall, regional variables appear as significant in the panel model, but on separate cross-sections regression their significance deteriorates with the years’.

Chang and Winters (1999) have shown that regional integration, on average, has significant adverse effect on non-member countries exporting to the integrated market. Even if regional integration does not increase external barriers to trade, the excluded countries may still be negatively affected. This, for example, could arise from excluded exporters decreasing their prices to meet the competition from suppliers within the RTA.

Gravity models have also been widely applied to examine the link between trade and exchange rate volatility, trade and currency unions, trade and environment, and trade and growth. However, it is important to be cautious about drawing inferences from the results of gravity model estimation, particularly if only one model is being considered. As Piermartini and Teh (2005) point out, gravity-model results generally depend on a number of estimation choices, such as the use of aggregated or disaggregated data, the sample of countries, the length of time period, the specification of a gravity model, the use of country-specific or country-pair-specific effects, and so on.
Computable general equilibrium models
Although the gravity model provides a nice framework for considering *ex-post* analyses of trade policy, it is also desirable to have a tool for examining changes in trade policy prior to implementation in so-called *ex-ante* analysis. Computable/calibrated general equilibrium (CGE) models provide one such framework. One of the main motivations underlying the use of CGE models is to be able to consider large scale policy changes using the present economy as a benchmark. This aspect of large scale policy changes separates the ability of the policy analyst to evaluate policies using a simple theoretical model or a back of the envelope calculation to gauge the impacts of policies. Trade policy is an inherently large scale problem. Even a scale change in a single industry has the potential to cause drastic and unexpected consequences given backward and forward linkages within the economy. These interdependencies between industries need to be considered in order to analyze the full impact of policy changes.

Quantifying the impact of policies had its beginnings with Leontief (1951, 1953) in which he developed the structure of input–output models for economies. These models, still popular today, tend to focus on inter-industry connections to meet final demand and not necessarily an integration of production with consumption along with factor ownership. The theoretical underpinnings for analyzing a general or economy-wide equilibrium lie first with Walras (1896) who represented the economy with a system of simultaneous equations that describe supply and demand equilibrium or market clearing through a set of prices for goods and factors. Nobel laureates Kenneth Arrow and Gerard Debreu (1954) and Debreu (1959) extended the ideas of Walras to incorporate the conditions for which a competitive equilibrium exists. Further, they established the link between a market equilibrium and welfare. The first real CGE model can probably be attributed to Johansen (1960) in which a linear model of the economy is used to identify the sources of economic growth in Norway. The first rigorous treatment of the numerical algorithms involved to compute non-linear models can be attributed to Scarf (1969). More recently with the advent of improved software and computational power, there has been a steady increase in the use of computational methods to explore issues of interest not only to academics but also policy makers at all levels.

*Why use CGE models?*
In considering questions of trade policy, at one level we care about the qualitative impact, or about which production sectors and consumer groups will be positively and negatively affected by the policy change. Economic theory can provide us with these results in a variety of circumstances. Potentially more interesting is the degree to which groups are affected. CGE and other quantitative methods allow us to estimate these effects. Through simulation of the economy we will not only be able to know who are the winners and losers but also how big these gains and losses are in order to have a better sense of the economic and social impact of policy change. In addition to being able to know how big the gains and losses are, in order to make theoretical analyses tractable, a certain level of aggregation is needed. This aggregation loses the detail that many policy makers would like to have in order to make better informed choices. Through the use of computational methods, this curse of dimensionality may be weakened. The limits to the dimension of the problem are no longer driven by tractability but by data availability.
Trade policy changes and trade negotiations are usually not single dimensional, and there is usually a ‘give and take’ in different sectors as well as by different parties involved. Additionally, a set of policy changes may have both positive and negative effects on the same sector or group. Theoretical considerations do not allow us to know the magnitude of these impacts and to be able to compare them in a meaningful manner. Through the use of computational methods multiple and/or ambiguous policy changes may be analyzed. Along these same lines, much of economics is concerned with efficiency whereas many policy makers may be additionally concerned with equity implications of policy changes. We know that different consumer groups will be affected differently by the same policy change. Just as we may disaggregate sectors fairly finely, we may also disaggregate consumer groups. Again, the limit to which we may disaggregate is driven by data availability and not the tractability of the problem.

Structure of CGE models
One of the main advantages of modeling an economy and especially modeling a detailed economy is that we must truly understand the structure of the economy. In general, there are really five main aspects that need to be considered when trying to accomplish the goal of modeling an economy:

1. How do goods and factors flow through the economy?
2. In each sector, how does production take place?
3. In each industry, what does the market structure look like?
4. At the consumer level, how does consumption take place?
5. Finally, who owns which factors of production and firms?

In general, the structure of a CGE model may be described using an open economy circular flow model, which illustrates the linkages between different sectors of the economy. As illustrated in Figure 5.1, firms purchase (demand) intermediate goods from other domestic or foreign firms and primary factors from households. They produce final output and sell (supply) it to households, government and the investment sector or export it to the rest of the world (ROW). In addition, some final and intermediate products are purchased (demanded) from abroad. The aggregate output in the economy is distributed across households, governments and the investment sector. Households own factors of production, sell (supply) them to firms and get a reward for using these factors. Rent for land, wages for labour, interest for capital and profit for entrepreneurship are used as income to demand consumer goods. The role of a government sector is to collect taxes on domestic and imported goods, pay subsidies, buy goods and provide public goods and services. In summary, the circular flow diagram divides the economy into two sectors: one concerned with producing goods and services, and the other with consuming them.

Production in a CGE model takes place under the assumption of profit maximizing firms. These firms take prices of factors and goods as signals and make decision about output and input mixes. Recently there has been a move to incorporate not just production through factors but also through the use of intermediate inputs in the production technology. The production technology is usually modeled using a production function that is a second order approximation of the data. That is, the production
technology will use a benchmark set of prices and quantities for a given year and then estimate the substitutability of the inputs from either the rest of the economic literature or through direct estimation using either time series or panel estimation. The other assumption that needs to be considered at the production level is that of market structure of the sector. Most commonly, models take perfect competition as the assumption. When considering trade policy, this may not be the case. Recently within the trade literature, both theoretically and numerically, there has been a move to imperfect competition either through an oligopoly or monopolistically competitive assumption.

Following the structure of the firms, consumer behavior has a similar structure. Analogous to firm behavior, consumers are assumed to maximize wellbeing, while taking good and factor prices as given. What distinguishes general equilibrium models from partial equilibrium is the ability to track the income of consumers. Consumers are assumed to own the factors of production as well as the firms. Many CGE models assume that there is a single representative economic agent that stands in for all economic agents. However, this need not be the case. If the focus of consideration for the policy change is on how different consumer groups are affected, this representative agent may be disaggregated to many such agents with different endowments of factors as well as firm ownership.

Finally, since firms are maximizing profit and consumers’ wellbeing, we need to understand how these two sides of the economy interact. First, there is the assumption of full factor utilization. On the labor side, this means that there is full employment within the economy. Second, all markets, whether factor or good, must clear. Prices within the economy will adjust such that the supply of each good and factor will equal the demand. All of these assumptions come together into three types of conditions that will define a CGE model: market structure (zero profit) conditions, market clearing, and property rights (income balance) conditions.
Operationalizing CGE models
Following Markusen (2002), there are six steps to consider when constructing a CGE model of an economy. These are discussed in more detail below:

1. specify the dimension of the model;
2. choose functional forms for production, transformation and utility;
3. construct a micro-consistent dataset;
4. Calibrate the model to the data;
5. Replication of benchmark; and
6. run counter-factual experiments.

As a starting point to any modeling exercise, we must know the level of detail that the model must contain in order to analyze the economy at the appropriate scale. We must choose the dimension of the model or number of goods, factors, consumers, countries and active markets. Once the level of detail is chosen, the next set of assumptions that must be chosen is the structure or functional form of the production and consumption sectors. As discussed previously, this is usually chosen to allow the most flexibility to fit the data. Most commonly, a two level, nested constant elasticity of substitution is chosen. This allows the modeler to have the flexibility described above while not having to make too many additional assumptions regarding the parameters of the functions.

Once the structure of the model is chosen, data must be found to correctly calibrate the model to the appropriate level of detail. The data are usually constructed using two main types of data sources. First, a micro-consistent input–output matrix needs to be either constructed or obtained. Three common sources for this data are the Global Trade Analysis Project (GTAP), the Minnesota IMPLAN Group, Inc., and national level statistical agencies. Additionally, the International Food Policy Research Institute has a variety of input–output matrices for developing countries. If not already present in the data, the consumer side of the economy must be constructed. At the most basic level, this simply involves the final demand quantities, which are consumed and exported. In a more detailed model with heterogeneous consumer groups, consumer expenditure surveys may be used to obtain information on preferences and factor and firm ownership. From these data sources, a micro-consistent social accounting matrix (SAM) is constructed. The SAM is a generalization of the original work of Leontief (1951) that contains more detail about consumption and ownership and allows the modeler to have a benchmark equilibrium with which to calibrate the model. Combining the assumptions of the model with that of functional forms, the data are used to calibrate the parameters to replicate the benchmark equilibrium. Assuming that the economy presently satisfies the assumptions of market structure, market clearing and property rights in the economy, the model should reproduce the present economy. The ultimate goal of most CGE modeling exercises is to answer questions about how the economy responds to policy changes. Thus, the model should be able to replicate the present policy. Once a model is constructed to replicate the present economy, alternative counter-factual policies may be considered.

Application: the Harrison/Rutherford/Tarr Multi-Regional Global Trade Model
Uruguay Round are quantified for 24 regions and 22 production sectors in each region in the four following areas: (a) tariff reductions in manufactured products; (b) replacement of non-tariff barriers in agriculture by the equivalent tariffs and obligatory commitments to decrease the level of agricultural protection; (c) the reduction of export and production subsidies in agriculture; (d) the removal of Voluntary Export Restraints (VERs) and Multi-Fibre Arrangements (MFAs). All distortions, such as taxes, tariffs, subsidies, VERs, and non-tariff barriers, are modeled as ad-valorem price wedges. The data employed in the model come from the GTAP database for 1992 (Version 2). The results of the ‘base’ constant returns to scale and perfect competition static model suggest that the world welfare gains from the Uruguay Round would be $92.9 billion per year, out of which $18.8 billion are from manufacturing sector reforms, $58.3 billion are from agricultural reform, and $16 billion are from MFA reform.

Variations of the Multi-Regional Global Trade Model may be applied to analyze a wide range of trade issues. For example, it is possible to assess the impact of global free trade on the countries’ welfare and the distribution of income across the regions. In addition, the impact of a country’s accession to the WTO may be quantitatively assessed. Further, the model may be extended to analyze the effects of regional trade agreements on the member countries as well as on those excluded from membership regions. Incorporating the short-run effects of changes in the trade policy in addition to the long run effects will allow evaluation of the transaction costs associated with a policy change, which must be taken into account by policy-makers.

Conclusions
This chapter introduces the most commonly used analytical techniques for a quantitative analysis in the area of trade policy, namely computable general equilibrium and gravity models. Gravity models are popular empirical trade devices that have been used widely for analyzing the impact of different trade policy issues on bilateral trade flows between different geographical entities. CGE models are generally used to quantify the impact of a change in trade policy on the countries’ welfare and the distribution of income across countries or multi-country regions. Through the use of computational methods, it is possible to analyze multiple policy changes and/or policy changes with ambiguous effects where theory is silent. Both types of models provide a consistent economy-wide picture that is beneficial to policy makers.

Notes
1. The standard estimation method restricts the intercept of the gravity equation to be the same for all trading partners.
2. This issue was discussed further in Anderson and van Wincoop (2003).
3. Trade creation will occur if formation of a regional trade association results in the transfer of production from a high-cost source in a home country to the low-cost source in a partner country, because tariffs have been removed from the trade between these countries. Trade diversion, in contrast, will occur if production is transferred from a low-cost source in a third country to a higher-cost source in a partner country, because tariffs are no longer imposed on the goods from the latter.
4. Important caveats concerning the conventional practice of ad valorem equivalent modeling are discussed in Chapter 18.

References
Walras, L. (1896), Élémens d’économie politique pure; ou, Théorie de la richesse sociale, Ed. 3, Lausanne: Rouge.