Federal Reserve Bank of Minneapolis
Research Department Staff Report 537

September 2016

Quantitative Trade Models: Developments and Challenges*

Timothy J. Kehoe

University of Minnesota,
Federal Reserve Bank of Minneapolis,
and National Bureau of Economic Research

Pau S. Pujolàs

McMaster University

Jack Rossbach

Georgetown University Qatar

ABSTRACT

Applied general equilibrium (AGE) models, which feature multiple countries or regions, multiple sectors, and input-output linkages across sectors in a Walrasian general equilibrium framework, have been the dominant tool for evaluating the impact of trade liberalization since the 1980s. We provide an overview of the historical development of AGE models and a guide as to how they are used to perform policy analysis. We then review and document shortcomings in the performance of AGE models in predicting the sectoral effects of past trade reforms, that is, we show that AGE models often perform poorly. We provide suggestive evidence that incorporating some of the recent advances in quantitative trade theory in AGE models can improve their predictive ability.

JEL Codes: F11, F13, F14, F17

Keywords: applied general equilibrium; trade liberalization; input-output linkages; extensive margin.

*This paper has been prepared for publication in the *Annual Review of Economics*. We thank Kim Ruhl for extensive discussions of the issues addressed in this paper. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System. An online data appendix is available at http://www.econ.umn.edu/~tkehoe

1. Introduction

Since rising to dominance in the 1980s, applied general equilibrium (AGE) models (sometimes referred to as computable general equilibrium, or CGE, models) have served as the tool of choice for addressing policy questions in international trade. Originally defined as any general equilibrium model that researchers solve numerically after calibrating its parameters to data, AGE models now distinguish themselves from other common quantitative trade models — most of which now fit the prior definition — through their multisector, multiregion nature and through their focus on input-output linkages across sectors. Since policy discussions surrounding trade reforms typically focus on changes at the industry level, these features explain why AGE models are considered the most appropriate tool when policymakers evaluate the desirability of different trade policies. While AGE models have retained their prominence in policy work, the theoretical advancement of AGE models has slowed significantly as the academic trade literature has shifted its attention to firm-level data and models that focus on them. We intend this paper as a guide to why and where AGE models demand more attention among researchers. We have organized the paper around three themes. In Sections 2 and 3, we review the development and use of AGE models and provide a practical outline for how researchers can use these models (and quantitative trade models more broadly) to evaluate the economic impact of trade policy reforms. In Section 4, we review and expand on research that evaluates the performance of AGE models for predicting the effects of past trade reforms, casting doubt on their accuracy and reliability. In the remaining sections, we discuss recent developments in the academic trade literature and evaluate the extent to which incorporating these advances into AGE models can overcome the shortcomings identified in Section 4. We intend the message of our paper to be hopeful: while we still have some way to go before AGE models can be fully trusted for trade policy analysis, we have come a long way, and we provide suggestive evidence that we are on a path with promising leads in front of us.

2. The Development and Use of Applied General Equilibrium Models

As Kehoe & Prescott (1995) explain, "Applied general equilibrium analysis is defined to be the numerical implementation of general equilibrium models calibrated to data: An applied GE model is a computer representation of a national economy or a group of national economies, each of which consists of consumers, producers, and possibly a government." In international trade, AGE models are used extensively to provide quantitative estimates on the economic impact of policy reforms, particularly trade liberalization. Dervis et al. (1982) and Shoven & Whalley (1984) provide surveys of the early AGE literature.

A typical AGE model of international trade — for example, as described in Kehoe & Kehoe (1994a) — consists of multiple countries that trade with each other; each country consisting of multiple sectors, all of which are linked through an input-output structure. In each sector, capital and labor is combined in a production function, often Cobb-Douglas, to produce the sector's output good. A representative household in each country maximizes a utility function, usually homothetic, defined over the consumption of goods from each sector, as well as over public consumption and investment goods. The household purchases goods using the income received from renting its labor and capital to firms and from the resources generated by tariffs on imported goods. An AGE modeler calibrates the parameters of the model so that the equilibrium of the model resembles the data of the countries of interest.

We discuss the calibration of AGE models in detail in the next section, but it is important to note that AGE models are designed from the ground up so that their parameters can be calibrated to the data. Introducing investment goods into the utility function, for example, makes a static model consistent with the investment that appears in the input-output tables. Similarly, imposing exogenous aggregate trade imbalances is a standard procedure that makes the equilibria of the models consistent with observed trade surpluses and trade deficits. The composition of this imbalance, however, is left to be endogenously determined, as imposing exogenous bilateral imbalances between all country pairs would restrict the flexibility of the model to respond to changes in policy. Likewise, models must be congruous with the fact that goods in most sectors are both imported and exported simultaneously; introducing individual preferences and production function specifications consistent with the Armington specification (after Armington (1969)) allows the model to deliver this desired pattern; moreover, this specification is tractable enough to allow for differing degrees of home bias and elasticity of substitutions across sectors. Finally, models have market clearing conditions that pin down all wages and prices in the model.

AGE models enjoyed a golden age in the academic international trade literature that lasted from the early 1980s until the mid-1980s. Whalley (1985), Shoven & Whalley (1992), and Kehoe & Kehoe (1995) provide guides on how AGE models were used during this period to evaluate the impact of trade liberalizations. Originally featuring perfectly competitive environments, AGE models expanded as trade theory evolved to analyze the role of imperfect competition and scale economies, as in Krugman (1980) and Markusen (1981), and intra-industry trade as documented by Grubel & Lloyd (1971). Imperfect competition and scale economies were adapted into AGE models by Harris (1984) in a small open economy of Canada and where later expanded by Smith & Venables (1988) to study the impact of removing trade barriers in the European Community in a preliminary assessment of what would end up being the Single Market in Europe. Another important advancement in trade theory was the introduction of varieties in a monopolistically competitive framework using Dixit & Stiglitz (1977) preferences, which provided a convenient way of modeling intra-industry trade. The first AGE model to incorporate these preferences was by Brown & Stern (1989), who use the model to evaluate the impact of the U.S.-Canada Free Trade Agreement (CUSFTA) on consumers' welfare, focusing their analysis on the expanded set of varieties available to consumers through CUSFTA. CUSFTA expanded to include Mexico, in what would become the largest free trade area in the world, when the North American Free Trade Agreement (NAFTA) was enacted in 1994. Many articles evaluated the impact of NAFTA on the welfare of consumers. Among the more influential studies are Brown et al. (1992), Cox & Harris (1992), Hunter et al. (1995), and Sobarzo (1995), which are discussed and compared by Kehoe & Kehoe (1994b). Following NAFTA, the number of articles written by academic researchers using AGE models to assess the impact of important trade reforms has diminished, but AGE models have remained heavily used in policy work. Li & Whalley (2014) study the impact of the Trans-Pacific Partnership (TPP) for different scenarios concerning the entry of China, and Böhringer & Löschel (2006) investigate the use of AGE models in the designing of environmental policy. One of the most widely known projects for facilitating the use of AGE models for policymaking is the Global Trade Analysis Project (GTAP), a global network of researchers who quantitatively analyze the impact of various trade policies. See Hertel (2013) for a description of the GTAP and its applications. Recently, Narayanan et al. (2016) review a number of trade agreements, showing that AGE models remain the dominant paradigm that policymakers use to evaluate the impact of a wide range of policy reforms.

One area in which AGE models have thrived in the academic literature is environmental economics. Hazilla & Kopp (1990) pioneer the literature by performing a cost-benefit analysis using an AGE model of the United States to assess the divergence between social and private costs of regulations mandated by the Clean Air Act and the Clean Water Act. Grossman & Krueger (1994) extend a standard AGE model to address environmental issues to assess the environmental impact of NAFTA. Since then, a number of AGE models of international trade and the environment have been developed to study the impact of various environmental reforms, especially multilateral environmental agreements involving many countries. Burniaux & Truong (2002) expand the GTAP model to incorporate energy and environmental considerations into this tool for policymakers, for example; Böhringer & Vogt (2003) develop an AGE model of international trade and the environment to study the impact of the Kyoto Protocol; and Peterson et al. (2011) do a similar analysis to capture the effects of the Copenhagen Accord. Thus, while AGE models have seen less attention in the trade literature compared with their golden age, they continue to receive robust attention in other areas owing to their flexibility and wide applicability.

3. Calibration

The standard procedure for calibrating the parameters of an AGE model exploits the equilibrium relationships that hold in the model, using these relationships to pin down parameter values from the data. Many of the central parameters can be estimated in a straightforward way from the national accounts and input-output data. Total supplies of factors are determined by normalizing base year prices of each factor and then setting total supplies equal to inflation-deflated factor payments to labor, capital, and intermediate inputs from the National Accounts. Utility functions and production functions are frequently assumed to have Cobb-Douglas or fixed-coefficient form, which makes it easy to calibrate factor and demand intensities directly from input-output tables. For a simple example, Kehoe & Kehoe (1994a) show how to calibrate a small AGE model using a three-sector input-output table.

The Armington specification of the model, according to which output in each industry is differentiated by country of origin (cars produced in Germany are different from cars produced in Japan), ensures that both domestic and foreign goods are used in production as intermediate inputs and consumed as final output. Output from each source country is combined according to an Armington aggregator, which is a constant elasticity of substitution (CES) function:

$$y_{jk} = \theta_{jk} \left(\sum_{i=1}^{m} a_{ijk} y_{ijk}^{\rho_k} \right)^{1/\rho_k} \tag{1}$$

Here y_{ijk} is the imports by country j of the output of industry k from country i, and y_{jk} is the aggregate amount of output of industry k available in country j for consumption or use as an intermediate input in production. The elasticity of this CES function, $1/(1-\rho_k)$, referred to as the Armington elasticity, determines the degree of differentiation across origins and plays a crucial role in determining the response of trade flows and consumption patterns to changes in productivity parameters and trade costs. Thus, the Armington elasticity estimates play important roles in determining the effects of policy reforms in the model. As discussed by Feenstra et al. (2014), however, these estimates can be sensitive to the choice of calibration strategy. As we discuss later in this paper, the last decade has seen significant advances in understanding how these elasticity estimates affect the welfare interpretations of observed changes in trade flows and consumption patterns.

It is important to note that input-output data are typically not able to distinguish between countries of origin on an industry-use basis. To give an example, we know how much machinery is imported from Canada in the United States, but we do not know in which industries this machinery is then used. To get around this limitation, and to reduce the scale of the model, we assume that the Armington aggregation takes place "at the border," or before the inputs are used in the production of other industry outputs. This means that U.S. imports of machinery from Canada are combined with machinery from all the other countries via an Armington aggregator to get a U.S. machinery aggregate, which is then used in the production of output for other industries in the United States according to the production functions specified earlier. Under the Armington assumption, observed foreign and domestic expenditure shares are used to calibrate the Armington elasticities, which are exogenous parameters in the model. A robust empirical observation is that countries tend to disproportionately consume domestic output compared with foreign varieties of the same goods, as well as disproportionately use domestic intermediate inputs in production. While this "home bias" is commonly thought to arise because of trade costs and barriers that limit international trade (see Obstfeld & Rogoff (2001)), these elements are often not modeled explicitly in AGE models. Standard AGE models instead rationalize observed home bias through the share parameters a_{ijk} of the Armington aggregators in the model, effectively assuming home bias exists solely because of preferences and production technologies. Two recent and influential papers discussing the home bias puzzle are Hanson & Xiang (2004), who present a model of endogenous product differentiation where home bias is generated by variation in transportation costs across industries, and Yi (2010), who investigates the role of vertical linkages and fragmentation of production in generating home bias. Wolf (2000) points out that home bias (or local bias) also exists on a subnational level. As home bias has become better understood, researchers have continued to shift away from using the Armington specification to capture home bias, instead opting to explicitly model the trade costs that lead to a home bias pattern. Corsetti et al. (2007) provide an example in which incorporating home bias through trade costs is quantitatively important for the transmission and welfare impact of macroeconomic shocks.

As we have mentioned, calibrating the production parameters in AGE models requires the use of input-output tables. Traditional sources for input-output tables include individual countries' national accounts (the Bureau of Economic Analysis produces a set of input-output tables for the United States, for example) and the Organisation for Economic Co-operation and Development Structural Analysis (STAN) Input-Output Database for OECD member countries. More recently, the World Input-Output Database (WIOD) described by Timmer et al. (2015) and the Global Trade Analysis Project (GTAP) described by Hertel (2013) have greatly increased the number of standardized input-output tables available to researchers. As of this publication, the WIOD covers 40 countries and 35 sectors, while the GTAP 9 covers 140 regions and 57 sectors. Since most AGE models are static models, aggregate trade imbalances are imposed as an exogenous parameter to match observed aggregate trade imbalances. It is important to note that while trade imbalances are exogenously imposed, the sectoral and bilateral composition of these trade imbalances arises endogenously. That is to say, an AGE model of the United States would impose an overall trade deficit, but it would not impose with which individual countries the United States has bilateral deficits nor in which sectors net exports are negative.

The final set of parameters that must be calibrated are the policy parameters. These parameters are perhaps the most important ones: changing these parameters is what allows the model to perform counterfactuals. While in principle any of the exogenous parameters can be changed, in the context of international trade the most common choices of policy parameters are taxes, trade costs, and sectoral productivity parameters. An example of a policy reform that falls outside of those choices is analyzed by Kehoe et al. (2013), who evaluate a policy reform in which the United States trade deficit is the policy parameter itself. In the AGE models that are used to evaluate the potential impacts of trade reform, the central policy parameters are typically tariffs. Common trade reforms studied in AGE frameworks include the uniform lowering of all tariffs by a set amount, the complete removal of all tariffs, and, when available, changing tariffs from calibrated initial pre-reform levels to proposed post-reform levels.

It is important to note that although changes in trade policy are taken as exogenous in AGE models, there is a large literature studying the endogeneity of tariffs and trade policy, as well as the political economy of international organizations such as the World Trade Organization (WTO). Bagwell & Staiger (1999) show how two key elements of the WTO, reciprocity and non-discrimination, can act to offset the terms-of-trade externality and resulting prisoner's dilemma that causes high tariff rates to arise endogenously when nations set tariffs unilaterally. There is also a large literature examining how tariffs and subsidies interact with the industrial structure of countries. Broda et al. (2008) show that market power and concentration affect optimal tariffs for a country; Costinot et al. (2015) show that optimal tariffs are affected by comparative advantage; and Demidova & Rodriguez-Clare (2009) show that export subsidies generate productivity increases. These and related papers suggest that observed tariffs contain information beyond just the tariff rates themselves. It is important to note, however, that AGE models rarely incorporate this type of information, which is a likely shortcoming that future research on AGE models should seek to address.

There are two standard ways of calibrating tariff parameters in AGE models. The first is to use observed tariff revenues, as in Kehoe & Kehoe (1994a), to back out implied tariff rates. This has the benefit of generating tariff revenues that are consistent with the national accounts. Despite a large literature predicting that countries should impose tariffs uniformly across products (see Opp (2010) for a recent example and Costinot et al. (2015) for a notable exception), in practice, tariff rates are rarely set uniformly. To account for this empirical regularity, we can alternatively calibrate the tariff rates in the model to match reported tariff rates in the data. The United Nations Conference on Trade and Development (UNCTAD) Trade Analysis Information System Tariffs (TRAINS) database contains reported bilateral tariff rate information for products categorized at the 6-digit Harmonized System (HS) level and constructs estimated tariff rates for other classification systems and levels of aggregation. Tariff agreements organized through the WTO and regional trade agreements typically take the form of setting caps on the maximum level of tariff rates that can be applied. In practice, these caps are often nonbinding. Maggi & Rodríguez-Clare (2007) rationalize that trade agreements set caps instead of tariff rates themselves, since this allows for large immediate cuts in tariff rates followed by further gradual reductions. When performing calibration for an AGE model, instead of calibrating tariff rates in the model to these tariff caps, we use what the TRAINS database reports as "effectively-applied" tariff rates.

Since these effectively applied tariffs are reported as advalorem equivalents, these rates can be inserted directly into the AGE model in the form of tax wedges for consumers and producers.

Most countries are now members of the WTO, and there have been significant gains in lowering tariff rates over the past 30 years. Because of this decline in tariffs, quantitative trade models have moved away from using tariffs as the object of policy reform and focused more generally on trade costs. Iceberg trade costs, as modeled by Samuelson (1954), enter into the AGE framework identically to tariffs, with the exception that tariffs produce revenue that is rebated to consumers or used to fund government expenditures, whereas iceberg trade costs are lost completely. This means that there are larger gains from reducing trade costs than there are from reducing tariffs, an implication that is explored quantitatively by Felbermayr et al. (2015).

Relative to tariffs, trade costs are a less well-defined concept and more difficult to observe directly in the data. One exception is data on freight costs charged by shipping providers to transport goods internationally. Recent papers have begun making use of this data; see , for example, Adao et al. (2015) and Shapiro (Forthcoming). Other papers, however (e.g., McCallum (1995) and Wolf (2000), show that freight costs do not appear to make up a majority of observed trade costs. In fact, recent research has highlighted that freight costs are endogenous and has shown that there is little correlation between traditional trade cost measures, using distance as a proxy, and freight costs themselves (see Kleinert & Spies (2011) and Asturias (2016)).

Additional sources of trade costs include non-tariff trade barriers and regulatory policies (Dean et al. (2009), Goldberg & Pavcnik (2016)), marketing costs involved in serving additional markets (Arkolakis (2010)), transportation costs (Limão & Venables (2001)), transportation time (Hummels & Schaur (2013)), exchange rate hedging costs (Allayannis & Ofek (2001)), and fixed costs associated with gaining access to foreign markets (Krugman (1980), Melitz (2003)). Other costs are less obviously trade costs, but nonetheless can function as such in trade models; recently proposed examples are information frictions (Allen (2014)) and credit constraints (Manova (2013);, Leibovici (2015); and Kohn et al. (2016)). The trade costs mentioned here are far from exhaustive — Anderson & van Wincoop (2004) provide a survey of the literature surrounding trade costs — yet they make our point that constructing nontariff trade costs directly from data is a challenging endeavor, prohibitively so for many early AGE applications.

For this reason, rather than as a constructive approach to calibrating trade costs based on direct data on these costs, trade costs are most often calibrated using the same model-implied equilibrium relationships used to estimate demand elasticities and production function intensities. Anderson & van Wincoop (2003) show that with Armington models, implied advalorem trade costs can be recovered from a gravity regression. Structurally connecting the results from the gravity regression to the Armington framework, they estimated the implied trade costs generated by the border between the United States and Canada, and showed that the naïve gravity regression estimates in McCallum (1995) were biased and significantly overstated the trade costs implied by the border. Using model-implied trade costs rather than direct data on observed trade costs has allowed researchers to better understand the effects of economic integration, as surveyed by Donaldson (2015). For recent examples of this approach, Comerford & Rodriguez-Mora (2015) apply this strategy to study the costs of regional independence in Scotland and in Catalonia, while Ottaviano et al. (2014) evaluate the potential impact on the United Kingdom of leaving the European Union.

Over the past decade, there have been significant advances in the ability of researchers to calibrate model-implied trade costs and elasticities through gravity regressions. Projects like CEPII's Gravity Database, described by Head et al. (2010), have increased the availability of much of the data necessary for running these gravity regressions. Simultaneously, there have been advances in understanding the proper way to estimate the model-implied equilibrium equations. Egger (2000) shows that these gravity relationships are best estimated using panel data, while Santos Santos Silva & Tenreyro (2006) provide an alternative estimation procedure that estimates the relationship without taking logs and avoids the bias arising from Jensen's inequality. Head & Mayer (2014) provide a survey of the advances related to estimating gravity regressions. While the typical approach is to derive implied trade costs using aggregate trade flows, Irarrazabal et al. (2015) provide micro-level estimates of trade costs using firm-level trade data while exploring the quantitative implications of modeling additive trade costs as multiplicative trade costs. Additionally, while bilateral symmetry in trade costs is often imposed, Hummels et al. (2009) and Waugh (2010) show that trade costs are in fact asymmetric and vary systematically depending on the income of the trading partners.

Economic historians have shown that trade costs have declined significantly over time. Model-implied trade costs, however, have remained substantial even as tariffs have decreased to near zero for many countries and products. Jacks et al. (2008) provide estimates of bilateral trade costs between 1870 and 2000, and Jacks et al. (2011) show that changes in trade costs over this period correlate strongly with aggregate changes in trade flows. Estevadeordal et al. (2003) investigate the extent to which transportation costs, tariffs, and the gold standard explain the rise and subsequent decline in world trade from 1870 to1939. More recently, Donaldson (Forthcoming) studies the reduction of transportation costs associated with railroad construction in colonial India.

While significant advances have been made in the calibration techniques available for using AGE models in an evaluative sense, challenges remain with using AGE models in a predictive manner. Following a policy implementation, post-reform trade costs can be recovered using the gravity approach and inserted into the model to evaluate how well the framework can capture observed changes. It may not have been possible, however, to predict the resulting changes in trade costs prior to the policy reform. While some changes in trade costs are easily predictable, such as changes in tariff schedules, and other trade costs can be inferred by predictable changes in gravity regression variables, such as whether countries are entered into a trade agreement (Baier & Bergstrand (2007)), much work remains to be done to adapt the calibration methods described above to be used in AGE models to yield the predictions that policy questions require.

4. Evaluating the Performance of AGE Models

For AGE models to be useful as predictive tools for policy analysis, an essential requirement is that the models be able to forecast the effects of policy reforms with some degree of accuracy. To evaluate the accuracy of AGE models, we can look at the pre-reform predictions that researchers produced using AGE models, or the predictions that AGE models would have yielded had they been used pre-reform, and compare them with what actually happened post-reform. While there are complications in evaluating models this way, particularly because many changes occur post-reform that were not considered in the model, it can still be informative to observe the correlation between the predictions of the models and what actually happened and use this to understand where the models succeeded, where they fell short, and how they can be improved.

A natural test case for the AGE framework is NAFTA because it was a significant policy reform between countries that continue to trade heavily with each other and it attracted a large amount of attention from economists both pre- and post- reform. Between 1993 and 2005, the average effectively applied tariff rate between NAFTA countries fell to near zero while each of the NAFTA countries remained as the top three trading partners of the other NAFTA countries.

Burfisher et al. (2001) provide a post-NAFTA review and evaluation of the arguments for and against NAFTA. NAFTA is also an ideal test case, since AGE models were the primary models used to inform policymakers on how NAFTA would affect the United States, Canada, and Mexico.

Fox (1999), Kehoe (2005), Shikher (2012), and Kehoe et al. (2015)evaluate how the AGE models originally used to predict the effects of NAFTA did in matching actual changes in bilateral industry-level trade flows, and they show that the models did poorly. The models did so poorly, in fact, that the predictions of the models were often negatively correlated with the actual changes observed post-NAFTA. This finding is deeply concerning, since if AGE models cannot get NAFTA — one of the largest and most significant trade reforms in recent history — correct, then there may be reasons to doubt the reliability of the AGE models currently being used for evaluating trade policy. It is worth noting that the failure of AGE models is potentially unique to their application to evaluating the impact of trade reforms. AGE models have been found to perform well for certain other policy reforms; for example, Kehoe et al. (1988) predicted the impact of a tax reform in Spain, and Kehoe et al. (1995) show that the predictions performed well overall.

The natural questions of why the models did so poorly in predicting the effects of trade liberalization and of where the models went wrong can be evaluated empirically to a large extent by examining where AGE models consistently produced inaccurate predictions and what features are observed in the data but not incorporated in the AGE frameworks. We argue that one reason for the models' poor perfomance is due to not accounting for product-level trade within industries, in particular, measures of extensive margin growth. Several papers hint at the importance of accounting for trade at more disaggregated levels. Using an AGE framework as an evaluative tool, Romalis (2007) shows that a significant amount of growth in trade between NAFTA countries occurred by displacing trade from non-NAFTA countries. While Brown et al. (1992) and related AGE models incorporated nonmember countries as a rest-of-the-world (ROW) aggregate, it may be the case that a more disaggregated approach is necessary to fully predict displacement at the product level. Trefler (2004) shows that plant-level productivity rose significantly in the industries experiencing the largest tariff cuts following the Canada-U.S. Free Trade Agreement signed in 1988, while low productivity plants reduced employment following the trade liberalization. Hillberry & McDaniel (2002) show that a significant portion of growth in trade between the United States and Mexico was in products that were previously not traded. Kehoe et al. (2015) show that using the extensive margin could have led to better predictions than the models originally used, a finding we explore further in this paper. One reason why extensive margin growth was not initially built into AGE models is due to a historical lack of theoretical underpinnings. As we discuss in the next sections, the significant amount of advances over the past 15 years has allowed researchers to better understand how the extensive margin is affected by trade policy. Much progress has also been made in understanding how changes in the extensive margin can be mapped into elasticities that allow AGE models to capture these effects.

Before jumping to the conclusion that current AGE models perform badly because AGE failed to predict the effects of NAFTA, it is fair to ask whether this inaccuracy was a shortcoming only with the particular models used at the time and whether the AGE models currently used in policy evaluation have overcome this shortcoming. AGE models may perform better now, for example, because of improvements in the econometric foundations underlying how parameters for AGE models are estimated. McKitrick (1998) raises econometric concerns over how functional form assumptions were imposed by the early applications of AGE models, whereas Hertel et al. (2007) discuss improvements in the AGE modeling framework that partially or fully alleviate several related econometric concerns. To evaluate whether AGE models currently used for policy analysis have improved, we use the GTAP model and database described in Hertel (2013) to evaluate several recent bilateral trade agreements and compare the predictions from the GTAP AGE framework with observed changes. We choose the GTAP framework since it is widely used for policy work and targeted toward researchers in policy-oriented institutions such as the WTO and World Bank.

We evaluate the GTAP framework for the following bilateral trade agreements (the year in which the trade agreement began implementation is in parentheses): United States–Australia (2005), United States–Chile (2004), China–Chile (2006), and China–New Zealand (2008). For all of the trade agreements, average effectively applied tariff rates between the partner countries were near zero by five years from implementation. It is worth mentioning that the GTAP framework and its extensions were, in fact, used by policymakers to evaluate the potential impact of many of these free trade agreements prior to implementation (see, for example, US International Trade Commission (2003) and New Zealand Ministry of Foreign Affairs and Trade and China Ministry of Commerce (2004)), and they continue to be used by both member and non-member nations to evaluate bilateral and multilateral trade agreements such as the Trans-Pacific Partnership (Strutt et al. (2015), Thorstensen & Ferraz (2014); Burfisher et al. (2014); and Strutt et al. 2015;). To allow for a consistent evaluation of the GTAP framework for all of the above trade agreements, we generate predictions for the impact of the policy reform in a standardized way for each reform using the GTAP 9 database and standard GTAP model described in Hertel (2013). The standard version of the GTAP model is an AGE with perfect competition and constant returns to scale, and makes use of multisector, multiregion, production, tariff, and input-output data to calibrate the model. The base GTAP 9 database features 57 industries and 140 regions. For each trade reform, we aggregate the base data into three regions: the two partner countries involved in the bilateral trade agreement and a ROW aggregate. To generate counterfactual predictions for the trade reforms, we use 2004 as our reference year for calibration, and we consider the impact of setting tariffs to zero for all commodities traded between the two partner countries while leaving the tariff rates between each country and the ROW aggregate unchanged. A minor shortcoming of this approach is that in practice, despite average tariff rates approaching zero following the implementation of the free trade agreements, there remain a few commodities in which tariffs have not yet been completely eliminated. A more significant concern is that many of these countries signed multiple bilateral trade agreements over the same period, and so holding tariffs fixed with the ROW is an unrealistic assumption and may be the source of inaccuracies. As we discuss in Section 6, however, including all actual tariff changes for the ROW lead to only minor improvements in accuracy for the NAFTA AGE models.

To evaluate the accuracy of the GTAP models, we compare the counterfactual predictions of the model with actual growth observed in the data. The GTAP counterfactuals are generated using runGTAP with the standard closure applied to input-output aggregations constructed from the GTAP 9 database. The GTAP 9 database contains production and input-output data for 57 industries and 140 regions in 2004, although for each trade agreement counterfactual, we combine the nonpartner country regions into a ROW aggregate. The results from the GTAP model are reported as the percentage change in trade value from base year value.

To compare these counterfactuals with the actual data, we use trade data collected from UN Comtrade at the 6-digit Harmonized System (HS2002) level and aggregate this product-level trade into trade flows for GTAP industries using a concordance provided by GTAP. Trade data are available for 42 of the 57 GTAP industries, although not all industries report positive trade for all importer-exporter pairs and all years. When we compute the percentage changes for actual growth in the trade data, we deflate by exporter's GDP. Formally, actual growth, $z_{ijk}^{t,t'}$, for exports from country i to country j in industry k between periods t and t' is computed as

$$z_{ijk}^{t,t'} = 100 \times \left(\frac{x_{ijk}^{t'} / gdp_i^{t'}}{x_{ijk}^{t} / gdp_i^{t}} - 1 \right), \tag{2}$$

where x_{ijk}^{t} is exports from country i to country j in industry k in period t, and gdp_i^t is GDP for country i in period t, both reported in current price USD. Following Kehoe (2005), we set the base period as two years before our base period for tariffs, since there are often significant changes in trade flows prior to the actual implementation of trade agreements because of announcement effects.

To construct the weighted correlation, we compute the weighted average growth rates for both the actual changes in trade flows, $z_{ijk}^{t,t'}$, and the predicted changes in trade flows from the GTAP model counterfactuals, $\hat{z}_{ijk}^{t,t'}$, as well as their weighted variances and the covariance between them. For example, the formula for the weighted average growth rate for actual changes in trade flows is given by

$$\text{mean}(z)_{ijk}^{t,t'} = \sum_{k=1}^{57} \omega_{ijk}^{t,t'} z_{ijk}^{t,t'}, \tag{3}$$

and the weighted covariance between actual and predicted changes is given by

$$\text{cov}(z,\hat{z})_{ijk}^{t,t'} = \sum_{k=1}^{57} \omega_{ijk}^{t,t'} \left(z_{ikj}^{t,t'} - \text{mean}(z)_{ijk}^{t,t'} \right) \left(\hat{z}_{ikj}^{t,t'} - \text{mean}(z)_{ijk}^{t,t'} \right), \tag{4}$$

where the weight used for each industry is the industry's share of exports averaged across the base period ($t =$ 2004) and the end period ($t' =$ 2015). That is,

$$\omega_{ijk}^{t,t'} = \frac{\left(x_{ijk}^{t} + x_{ijk}^{t'} \right)}{\sum_{k'=1}^{57} \left(x_{ijk'}^{t} + x_{ijk'}^{t'} \right)}. \tag{5}$$

The weighted correlation is then given by

$$\rho(z,\hat{z})_{ijk}^{t,t'} = \frac{\text{cov}(z,\hat{z})_{ijk}^{t,t'}}{\sqrt{\text{cov}(z,z)_{ijk}^{t,t'}} \sqrt{\text{cov}(\hat{z},\hat{z})_{ijk}^{t,t'}}}. \tag{6}$$

These weights imply that, when actual growth deviates from predicted growth, the deviation factors more heavily into the weighted correlation for industries that are traded more heavily. The results we show, however, are qualitatively robust to several other weighting schemes based on pre- and post-reform trade values as well as to using the unweighted correlation that factors deviations the same even for industries that account for very little trade.

Table 1 reports the simple average (in percent) across 6-digit Harmonized System products of effectively applied tariff rates between each country pair in percent 2002 and 2015 (collected from TRAINS) under the columns "Average 2002 tariffs" and "Average 2015 tariffs." As we can see, there were significant declines in tariffs for each of the trade liberalizations, with post-reform rates near or at zero for all of the countries. Table 1 also reports the weighted correlation between the predicted and actual changes in trade flows between 2002 and 2015 using (6) under the column "Correlation of GTAP with data" (the last column, "Correlation of LTP with data," is discussed in a few paragraphs). When computing these weighted correlations, for each importer-exporter pair, industries that report zero trade in either 2002 or 2015 are dropped from the sample (replaced with zeros in (5)). We also exclude U.S. exports to Chile in the petroleum industry as an outlier. Petroleum grew from 1.9 percent of U.S. exports to Chile to 33.6 percent of U.S. exports to Chile between 2002 and 2015. This growth in exports was due to innovations in hydraulic fracturing techniques, which led to greatly increased oil production and exports in the United States and has little to do with changes in tariffs (there was little change in U.S. petroleum exports to Australia, and so it is not excluded as an outlier). Australian exports of cattle meat to the United States are also excluded as an outlier industry. Cattle meat was the top export from Australia to the United States in both 2002 and 2015; however, a tariff rate quota has remained in place even after the signing of the trade agreement. Australia has continually used 99 percent of its quota, which has increased by approximately 10 percent between 2002 and 2015. Growth in the trade value of cattle meat exports from Australia to the United States was driven primarily by increases in the price of beef, which was due in part to the spread of bovine spongiform encephalopathy, commonly known as mad cow disease. Mad cow disease was first discovered in the United States in 2003 (it has not been discovered in Australia), and between 2002 and 2015, the average price of cattle meat imports from all countries to the United States increased by 123 percent compared with just a 32 percent increase in the U.S. Consumer Price Index over the same time frame. In fact, cattle meat imports by net weight were actually lower throughout 2003–2014 than they were in 2002, returning to their pre mad cow disease levels only in 2015.

Table 1: Comparisons of GTAP and LTP predictions for recent trade liberalizations with data

Exporter[a]	Importer	Average 2002 tariffs	Average 2015 tariffs	Correlation of GTAP with data	Correlation of LTP with data
United States	Australia	4.47	0.00	0.27	0.55
Australia	United States	3.86	0.72	–0.14	0.53
United States	Chile	6.98	0.00	0.08	0.55
Chile	United States	2.83	0.07	0.03	0.48
China	Chile	7.00	0.13	0.14	0.61
Chile	China	11.68	0.49	0.04	0.07
China	New Zealand	4.06	0.04	–0.36	0.61
New Zealand	China	11.72	0.45	–0.09	0.48
Simple average		5.40	0.24	–0.00	0.49

[a]Outliers in data excluded: U.S. exports of petroleum to Chile, Australian export of beef to the United States.

As we can see from Table 1, the weighted correlations between the predicted changes from the GTAP model counterfactuals and the actual changes in growth are fairly low. A fair question to ask when evaluating the GTAP framework is whether any alternative models could have been expected to perform better. As suggestive evidence that we should be able to design AGE models that perform better, we adapt the methodology from Kehoe et al. (2015) to show that including information along only a single dimension — the share of least-traded products in each industry — can lead to predictions that outperform the AGE models that take into account information on taxes, production, and input-output linkages, yet lack this essential margin. This methodology is based on the insight from Kehoe & Ruhl (2013) that the products that experience the most growth following trade reform and growth episodes are the products that were previously traded, yet in small amounts. This contrasts with the traditional extensive margin, captured by Eaton & Kortum (2002) and Melitz (2003), of goods moving between traded and not traded and moving from being produced to not being produced at all.

To construct the share of least-traded products in each industry, we use the same 6-digit HS2002 data and concordance that we used to construct the actual changes in industry trade for each trade reform. We define the set of least traded products by sorting all the products by their average trade value over 2002−2004, starting with the products with the smallest average values of trade. For each product, we then compute the cumulative value of trade in 2002 of all products with less trade over the 2002−2004 period, and we classify as the set of least-traded products the set of goods that accounts for exactly 10 percent of trade in 2002. Note that products that are not traded in 2002 do not enter into the share of least-traded products, and so the margin we are focusing on is indeed the set of products that are traded in positive, but very small, amounts. We sort over multiple years so that products that are traded in large amounts, but experience lumpy trade, are not classified as least-traded products. To meet the 10 percent cutoff exactly, we count only a fraction of the cutoff product toward the growth in least-traded products. As Kehoe et al. (2015) show, our methodology is robust to using alternative cutoffs — for example, 5 percent or 20 percent of trade instead of 10 percent, as long as the cutoff is not so small that it omits much of the margin we want to focus on or so large that it completely dilutes our margin.

After constructing the set of least-traded products, for each industry we compute the share of trade in the base period that is composed of least-traded products, s^t_{ijk}. Note that while overall the share of least-traded products is 10 percent of total trade flows, in any particular industry this share may be more or less than 10 percent, since the set of least-traded products is computed ignoring industries. Kehoe & Ruhl (2013) show that least-traded products grow much faster than non-least-traded products following trade liberalization. The hypothesis for this exercise is that the industries that should experience the most growth are those that are disproportionately composed of least-traded products. Kehoe et al. (2015) show that the share of least-traded products in each industry can be combined with the change in tariffs and the trade elasticity from a simple gravity regression to construct level predictions for growth. Further, they show that for the case of NAFTA, these predictions outperformed the predictions of the AGE models originally used to predict the effects of NAFTA and inform public policy. Note that the level predictions constructed following Kehoe et al. (2015) are perfectly correlated with the share of least-traded products in each industry for each importer-exporter pair; therefore, the share of least-traded products in each industry is sufficient if we are only interested in the weighted correlations between predicted and actual changes. The reason for this result is that the level predictions are linear functions of the share of least-traded products in each industry. While the slope and intercept of these functions are crucial for determining the predicted changes, as long as the slope is positive (least-traded products grow more than non-least-traded products), then the correlation is independent of the actual values of the slope and intercept.

The weighted correlations between the share of least-traded products (LTP) and actual changes in industry-level trade are reported under the “Correlation of LTP with data” column of

Table 1. These weighted correlations are produced by substituting the share of least-traded products, s_{ijk}^{t}, for $\hat{z}_{ijk}^{t,t'}$ in equations (4)–(6). The LTP-based predictions outperform the GTAP predictions in terms of matching actual changes in post-reform trade flows for each of the 6 importer-exporter pairs. The GTAP predictions perform the best for U.S. exports to Chile following the U.S.-Chilean free trade agreement, however, even in this case it performs worse than the average weighted correlations between actual changes and the LTP-based predictions of 0.41, which is much higher than the average weighted correlation of 0.05 for the GTAP-based predictions.

These results indicate that the problems with AGE models have not been fully overcome, at least in standard applications. Our results in this section show that the AGE models still commonly used to predict trade reform likely suffer from the same problems that plagued the models that originally performed so poorly for NAFTA. Despite this apparent lack of progress, we believe that there have been several advances in trade theory over the past 15 years that have not fully made it into multi-sector AGE models, and that these advances have the potential to significantly improve the reliability of AGE models for policy evaluation.

5. The Extensive Margin Revolution in International Trade

While it is not clear why AGE models became less prominent within the academic literature following the mid-nineties, it is clear that the emergence of detailed datasets on plants and firms in a number of countries played a pivotal role in the refocusing of the international trade literature on firm heterogeneity. See Tybout et al. (1991) and Pavcnik (2002) for an analysis of data on Chilean plants; Roberts & Tybout (1997), Fernandes (2007), Eslava et al. (2013), and Kealey et al. (2016) for Colombia; Abowd et al. (1999), Eaton et al. (2004), and Eaton et al. (2011) for France; Feenstra et al. (1999) for South Korea, Taiwan, and Japan; Bernard et al. (2003) and Bernard et al. (2007) for the United States; Amiti & Konings (2007) for Indonesia; Goldberg et al. (2010a) and Goldberg et al. (2010b) for India; Bustos (2011) for Argentina; and Manova & Zhang (2012) and Feenstra et al. (2014) for China. These datasets have proven useful for uncovering many important patterns within international trade; for instance, even for heavily export-oriented economies, only a small fraction of firms are exporters. Furthermore, the firms who export tend to be larger, employ more educated workers, and be more productive than firms that do not export. To analyze these findings, the trade literature developed theoretical models that are consistent with these patterns: most notably, the Melitz (2003) and Chaney (2008) models of firm heterogeneity, love-for-variety, and trade — expanded along various dimensions by Yeaple (2005), Ghironi & Melitz (2005), Demidova (2008), Melitz & Ottaviano (2008), and Arkolakis (2010) among many others — and the Eaton & Kortum (2002) model of perfect competition and international trade — used by Alvarez & Lucas Jr. (2007) to measure the welfare gains from free trade for many countries.

A common aspect in many of these novel models of international trade is driven by the following key finding: after a country undergoes a trade liberalization, a large number of firms start exporting, and a large number of firms shut down production. This "extensive margin" is achieved theoretically in trade models either by introducing differences in the marginal cost of monopolistically competitive firms and a fixed cost to export, as in Melitz (2003), or by introducing differences in the marginal cost of perfectly competitive firms that face different transportation costs, as in Eaton & Kortum (2002). Neither of these mechanisms exist in the AGE models discussed in previous sections, even though evidence has shown the extensive margin to be hugely important for international trade. Helpman et al. (2008) show the robustness of the extensive margin across several estimation techniques, whereas Bernard et al. (2009) provide transaction-level measures of the extensive margin in trade for firms in the United States. Hummels & Klenow (2005) measure the extensive margin for 126 exporting countries and find that it accounts for around 60 percent of the greater exports of larger economies, Hillberry & McDaniel (2002) find that most of the increase in trade in the United States after NAFTA consists of new varieties coming from Mexico, and Kehoe & Ruhl (2013) provide a product-level decomposition of the extensive margin across all importer-exporter pairs, showing that large changes in trade are disproportionately driven by changes in the extensive margin. Given the importance of the extensive margin, it seems likely that there will be gains from merging these developments into the AGE framework.

6. Plugging Advances in Trade Theory into AGE Models

Much of the models in the trade literature on the extensive margin do not distinguish firms by industry, but, as discussed in the previous section, they have been successful at capturing many empirical regularities observed in firm-level data. The fundamental goal of AGE models is to provide useful estimates of the welfare impact of changes to trade policy. As we discuss in Section

7, Arkolakis et al. (2012) have shown that a wide class of models featuring a few common characteristics — this class of models is often referred to as gravity models, since they generate trade flows that are consistent with the gravity regressions discussed in Section 3 — predict identical welfare gains from trade liberalization. In fact, to capture the welfare impact of trade reform, we require only the change in trade that the reform produces and the trade elasticity. The basic formula derived in their paper is

$$\hat{w} = \hat{\lambda}^{1/\varepsilon} \ , \tag{7}$$

where $\hat{\lambda}$ is the change in domestic expenditure share (or self-trade share) produced by the reform, ε is the trade elasticity (how much imports respond to changes in trade costs), and $\hat{w}$ is the resulting change in welfare implied by this class of models. This result has been widely used in recent papers, since it allows for changes in welfare to be computed easily, especially when comparing gains from moving away from autarky when the self-trade share is 1. Costinot & Rodríguez-Clare (2014) use this result to quantitatively evaluate the welfare implications of globalization. This basic formula computes changes in welfare resulting from changes in iceberg trade costs, rather than in tariffs, and as Felbermayr et al. (2015) point out, this distinction has welfare implications, since tariff revenue is redistributed back to consumers. As Goldberg & Pavcnik (2016) point out, tariffs are now relatively low for most of the world; however, there is significant room for trade policy to address trade costs associated with nontariff barriers that may more closely resemble iceberg trade costs.

Why, then, do we still need multisector AGE models? For one, even if the models were able to correctly forecast changes in aggregate bilateral trade, the composition of that trade matters for welfare gains because the aggregate trade elasticity depends on the sectoral composition of trade. Ossa (2015) shows that accounting for differences in elasticities across sectors greatly increases estimates for welfare gains of trade compared with single sector models. Countries experience large welfare gains from importing goods in low-elasticity industries, such as automobiles, in which they are not efficient at producing domestically. French (2016) further shows how the pattern of comparative advantage across sectors affects welfare gains as well as for the aggregate impact of trade barriers, the insight being that multisector models are needed even if the goal is only to capture changes in aggregate trade flows and not in welfare or disaggregated trade flows. Levchenko & Zhang (2016) find evidence of cross-country convergence in relative productivities across sectors and show that this has affected world aggregate trade flows by lowering them, compared with equivalent growth without convergence.

A distinguishing feature of AGE models is their focus on the input-output structure of economies. Taking the input-output structure into account is essential for understanding the nature and impact of trade flows, since trade in intermediate goods makes up a large fraction of international trade. As Yi (2003) and Ramanarayanan (2012) show, the impact of tariffs and other trade barriers is amplified when there is trade in intermediate goods, as the trade costs apply both directly to trade in final goods, as well as indirectly through their embodied impact on intermediate goods — a result similar to the concept of double marginalization in the industrial organization literature. While trade in intermediate goods can be accounted for in single sector models, as it is by Eaton & Kortum (2002), incorporating input-output linkages allows for the richer interactions that are necessary to capture how tariff reductions in one sector affect trade and production in other sectors.

While multiple sectors and input-output linkages are not incompatible with the advances discussed in Section 5, there has been relatively little work merging the two literatures, which we attribute to the relative infancy of the new theoretical literature and to the lack of awareness with regard to the shortcomings of AGE models. Two recent exceptions are papers by Caliendo & Parro (2015), henceforth CP, and Heerman et al. (2015). These papers embed the framework of Eaton & Kortum (2002) into a multisector environment featuring input-output linkages. These papers depart from standard AGE models in that trade costs, as opposed to variation in preferences or production functions, are the driving factor behind home bias and expenditure shares. One of the key innovations of CP has to do with their calibration of trade elasticities, which they are able to estimate at the sectoral level. These elasticities are essential for determining the counterfactual trade response to changes in trade costs and tariffs and for translating this response into welfare predictions. A benefit of merging AGE models with the Eaton & Kortum (2002) framework is that the models deliver recognizable gravity-type equations and transparent mappings to welfare predictions, which helps circumvent the common criticism that AGE models act like "black boxes."

One question is whether these new models outperform the standard AGE models, such as GTAP and the models originally used to predict NAFTA. To answer this question, we can look at the model predictions from CP. CP calibrate their model using pre-NAFTA data on trade flows, production and input-output linkages, and tariffs for the United States, Canada, Mexico, and 28 other regions. They then use their models to generate counterfactual predictions resulting from tariffs changing to their post-NAFTA levels. While the focus of their paper is on using their model to disentangle the welfare implications of NAFTA, we can also evaluate the accuracy of the model in matching actual changes in trade flows following the implementation of NAFTA.

We repeat the evaluation exercise from Section 4, computing the industry-weighted correlation between changes in observed industry-level trade flows for several variations of the CP model and for the share of LTP in each industry. Compared with when we evaluated the GTAP models, there are a few changes. The industries used in the CP model differ from the industries defined in the GTAP. CP have 20 traded and 20 nontraded industries in their model, and we focus on the 20 traded industries, since we are evaluating only the accuracy in predicting changes in trade flows. CP provide a full description of the industries and a concordance between their industries and 2-digit ISIC Rev. 3 industry codes in their paper. To compute actual changes in trade flows, we download trade data at the 6-digit HS1988/1992 level, which we map into 2-digit ISIC Rev. 3 industry codes using a concordance from the World Bank's World Integrated Trade Solution (WITS) database[1], and then into our final industries using the concordance from CP. Unlike the GTAP simulations, which produce percent changes, CP consider changes to be in terms of logdifferences (the results are similar regardless of whether logdifferences or percentage changes are used; however, we use logdifferences to be consistent with their original framework). Note also that unlike the GTAP models, CP consider NAFTA as their policy reform when computing their counterfactuals; therefore, instead of setting up and solving an equivalent model, we are able to use the predicted changes generated by their own code, available in their data appendix (details in our own online data appendix).

To evaluate how well the CP predictions perform, we now compute actual changes in exports, $z_{ijk}^{t,t'}$, from country i to country j for each industry k between t=1991 and t'= 2006 using the log approximation of equation (2):

$$z_{ijk}^{t,t'} = 100 \times \left(\log\left(x_{ijk}^{t'} / gdp_i^{t'} \right) - \log\left(x_{ijk}^{t} / gdp_i^{t} \right) \right), \tag{8}$$

[1] http://wits.worldbank.org/data/public/concordance/Concordance_H0_to_I3.zip

where x_{ijk}^t is exports from country i to country j in industry k in period t, and gdp_i^t is GDP for country i in period t, both reported in current price USD. We provide an appendix that shows that our results are robust to calculated changes in exports using either method and to excluding or including outliers. For the least-traded exercise, the share of least-traded products in each industry is computed in the same way as in Section 4, where, for each importer-exporter pair, we sort products by their average trade value between 1991 and 1993 and then count products as least-traded until they cumulatively account for exactly 10 percent of trade in 1991.

We report the results of these exercises in Table 2. The column "CP correlation with data" refers to the weighted correlation between actual changes in trade flows and the predicted changes of the full CP model, taking into account all tariff changes prior to 2006. This is computed using equations (4)–(6), with the predicted changes for each industry taken from CP and the actual changes computed using (8). The column "LTP correlation with data" provides the same comparison benchmark of the weighted correlation between actual changes in trade flows and the share of least-traded products in any industry, which we compute following the same methodology as we did for Table 1.

So does the CP framework outperform standard AGE models? The answer is: it depends. The average correlation across country pairs is near zero and slightly lower than the GTAP correlations in Table 1; therefore, it may appear that there are little to no gains from incorporating recent advances into AGE models. In contrast to the GTAP results in Table 1, however, the CP framework outperforms the LTP methodology for half of the country pairs, whereas the GTAP framework is uniformly beaten by the LTP methodology. Similarly, the correlations between actual and CP predicted exports from the United States to Canada and Mexico are high and significantly higher than any of the correlations between actual changes and predicted changes for the GTAP model. This, to us, represents the success of the CP framework and the improvements possible when incorporating trade flows. While it may initially appear unfair to compare the accuracy of the CP predictions, which take into account all tariff changes, with the GTAP predictions, which take into account only changes in tariffs between the member countries of the free trade agreement, CP also compute counterfactuals taking into account only NAFTA tariff changes. The column "CP correlation with data (only NAFTA tariffs)" in Table 2 shows that the results are nearly unchanged if tariff changes in non-NAFTA countries are disregarded.

Table 2: Comparisons of CP and LTP predictions for NAFTA with data

Exporter	Importer	CP correlation with data	CP correlation with data (only NAFTA tariffs)	CP correlation with data (no IO structure)	LTP correlation with data
Canada	Mexico	−0.46	−0.46	−0.50	0.25
Canada	United States	0.36	0.32	0.36	0.19
Mexico	Canada	−0.68	−0.66	−0.71	0.83
Mexico	United States	−0.17	−0.12	−0.21	0.33
United States	Canada	0.35	0.05	0.14	0.28
United States	Mexico	0.54	0.53	0.64	0.16
Simple average		−0.01	−0.06	−0.05	0.33

While the CP model shows considerable success for some country pairs, it also shows considerable failure in producing accurate predictions for other country pairs, particularly trade between Canada and Mexico. Why does the model perform so poorly in these cases? We hypothesize that it is because the CP methodology lacks the least-traded margin. Note that trade between Canada and Mexico is exactly where the LTP methodology performs best. Indeed, we find that for exports from Mexico to Canada, least-traded products grew by 206.0 percent more than GDP between 1992 and 2006, whereas non-least-traded products grew only 55.8 percent. Why do least-traded products grow so much more than non-least-traded products? The baseline models of Eaton & Kortum (2002) and Melitz (2003) are incapable of reproducing this observation, at least as they are usually parameterized. One model that is capable of capturing the Kehoe & Ruhl (2013) extensive margin is that of Arkolakis (2010), in which acquiring new customers is costly, and non-linear marketing costs can explain why small firms grow faster than large firms. Potentially in agreement with this theory, Ruhl & Willis (Forthcoming) show that exporters tend to start small and grow over time, whereas Schmeiser (2012) shows that the expansion of exporters to new markets occurs slowly.

The column "CP correlation with data (no IO structure)" in Table 2 reports the weighted correlation between actual changes in trade flows and the counterfactuals CP produce using their framework for NAFTA tariff changes only, but discarding the input-output structure of their model. It is disturbing that the CP framework performs nearly identically (and actually slightly better on average) when the input-output structure of the model is left out. Earlier, we argued that taking into account linkages across sectors is essential if we want to understand the sectoral impact of trade reforms. It seems likely that the fact that the input-output structure does not improve the performance of the CP model in general suggests that economists need to think more carefully about the way in which input-output relationships are being built into AGE trade models.

Overall, the results in Table 2 suggest that merging recent trade advances with AGE models shows promise in improving the performance of these models, but also that further improvements are needed. The fact that the LTP methodology performs best precisely when the CP model performs worse and vice versa hints at a likely pathway for improvement. That the LTP methodology performs better on average than the CP model, and uniformly outperforms the GTAP models in Section 4, suggests that the gains from incorporating the least-traded margin in AGE models are potentially large.

7. Future Directions for AGE Models in International Trade

As Shoven & Whalley (1984) put it, AGE models aim "to convert the Walrasian general-equilibrium structure […] from an abstract representation of an economy into realistic models of actual economies" to "use these models to evaluate policy options by specifying production and demand parameters and incorporating data reflective of real economies". Ultimately, AGE models are used to address the question of whether policy reforms are welfare improving or not. During their golden age, AGE models were built using state-of-the-art theory in the literature. The literature in international trade has increased enormously over the last two decades, and, with a few notable exceptions, such as the CP model, much of it has yet to be incorporated into AGE models. The purpose of this section is to draw attention to additional advances in the trade literature that we believe would be fruitful to incorporate into AGE models.

As we mentioned in the previous section, one of the biggest breakthroughs in the recent trade literature is the result from Arkolakis et al. (2012) for computing welfare gains from trade formula. They show that many models that are used in the academic trade literature share the same predictions for the welfare gains from trade and can easily be summarized by a simple formula combining how much trade a country has and the elasticity of trade flows to trade costs. This result can be taken to mean that to measure the effects of a trade reform, one simply needs to estimate how much trade will change, plug this change into a formula containing an elasticity, and get the predicted change in welfare. Building on this formula, Arkolakis et al. (2015) show that a similar version of the same formula arises when changing markups are considered. Costinot & Rodríguez-Clare (2014) provide a quantitative exploration of the welfare gains from trade by applying the welfare formula to trade data, and Adao et al. (2015) go beyond parametric specifications of functional forms to show how welfare gains from trade can be measured.

One shortcoming of these approaches is that the trade elasticity is typically assumed to be constant. A notable exception is by Brooks & Pujolàs (2014), who develop a generalized version of the welfare gains from trade formula for the case in which the demand system features a non-constant trade elasticity. The issue of having a constant trade elasticity and its effect on welfare has been studied by Fieler (2011), who shows the importance of having goods with different elasticities; Simonovska (2015), who highlights the importance of having a trade elasticity that changes with the income of the country; Simonovska & Waugh (2012), who show that similar models have different trade elasticities, so the intensity by which goods are desired has different implications for different models; Jung et al. (2015), who show that to properly match salient features of the data, we have to escape from the notion of having a constant trade elasticity; and Bertoletti et al. (2016), who incorporate the notion of "indirectly additive" preferences into the measurement of welfare gains. Similarly, as Ruhl (2008) points out, the macro-trade elasticity must differ from standard trade elasticities, since a perpetual reduction in trade costs is different from a momentary change in the terms of trade. Challenging the applicability of the welfare formula, Melitz & Redding (2015) provide an example showing the quantitative importance of the breakdown of the Arkolakis et al. (2012) welfare formula, which does not apply to models such as Melitz (2003) if the firm distribution is not Pareto, a point similar to di Giovanni & Levchenko (2013), who show that welfare implications differ significantly depending on whether the distribution follows Zipf's law or not. Ultimately, when evaluating the desirability of policies, economists often focus on the effect that the policy will have on the welfare of the population. For this reason, AGE models could benefit from a deeper understanding of how the results of the models map into welfare, and the advances over the past decade provide a road map on how to do this.

In this paper, as in much of the literature, we focused on static AGE models. While the focus of this paper is relatively narrow because of its nature, there has been increased usage of dynamic AGE models in policy-related applications;, for example, by Diao et al. (1998) and also Ianchovichina (2012), who discuss how to estimate the GTAP model in a dynamic framework. Despite this progress, however, many features and implications of dynamic trade models have yet to be fully explored in AGE settings. Baldwin (1992) and Anderson et al. (2015) show that there are dynamic gains from trade by encouraging capital accumulation and human capital accumulation, which can be large. Building on this argument, Brooks & Pujolàs (2016) show that gains from trade driven by capital accumulation can be large, even though in the transition they are potentially negative. Studying the impact that trade reforms have on the creation of new firms, Alessandria et al. (2014) and Alessandria & Choi (2015) show that gains during the transition can be even larger. The potential impact of building dynamics into trade models is non-trivial. As Bajona & Kehoe (2010) show, a standard trade model such as the Heckscher-Ohlin model, when built into a dynamic framework, implies that standard results in closed-economy dynamic models, such as long-run convergence of income, disappear. No AGE model incorporates these features, although incorporating them will provide better estimates on the impact of opening up to trade by currently poor, closed economies, as well as the interaction between open economies that undergo large trade liberalizations.

The trade literature has also focused on analyzing nontraditional sources for gains from trade. Galdón-Sánchez & Schmitz Jr. (2002) and Schmitz Jr. (2005) show that the increased competitive pressure from foreign producers forced iron ore producers to increase their productivity. Broda et al. (2008) empirically show that trade reduces market power, but they go beyond to show that it also affects trade policy. Building on these insights, Atkeson & Burstein (2008) and Arkolakis et al. (2015) develop trade models showing how to capture the pro-competitive gains from trade — the lowering of markups that firms charge because of increased competition from foreign firms — which Asturias et al. (2015) apply to quantitatively evaluate the welfare impact of transportation reform for India. Although some work had been done to AGE models to embed pro-competitive effects (see Devarajan & Rodrik (1991)) expanding the literature along that dimension is likely to provide important insights into how to measure the impact of trade policy changes. A related yet different gain from trade is pointed out by Holmes & Schmitz Jr. (2001), who argue that increases in trade force unproductive entrepreneurs to become more productive.

8. Concluding Remarks

The purpose of this paper was to argue that AGE models still require the attention of academic researchers and that there remain many potential avenues for exploration. Given the ubiquity of AGE models in economic policymaking, it is important that we work to address the shortcomings of AGE models discussed throughout this paper. We believe expanding AGE models by incorporating recent advances in the trade literature remains an exciting area for future research and that economists who undertake this research should expect to find a large audience interested in the policy implications of their work.

Our survey suggests three major directions in which AGE analysis of trade policy can be improved: First, AGE modelers can better identify existing barriers to trade and better model how changes in trade policy reduce these barriers. Second, AGE modelers can experiment with formulations that incorporate the extensive margin identified by Kehoe & Ruhl (2013) into their models. The model of Arkolakis (2010) provides a possible direction in which to start. Third, AGE modelers should incorporate recent innovations in the theoretical trade literature into their models, evaluating whether or not these innovations are capable of improving the performance of their models in accounting for changes in trade patterns after changes in trade policy. Ideally, AGE analysis would regain the status in the literature on trade theory and its applications that it held during its golden age from the early 1980s through the mid-1990s when AGE modelers were frequent contributors to the academic literature. By carrying out performance evaluations of the sort that we do in Tables 1 and 2, AGE modelers can turn their models into testing grounds for innovations in trade theory.

Literature Cited

Abowd JM, Kramarz F, Margolis DN. 1999. High wage workers and high wage firms. *Econometrica* 67:251–333

Adao R, Costinot A, Donaldson D. 2015. *Nonparametric counterfactual predictions in neoclassical models of international trade*. NBER Work. Pap. 21401

Alessandria G, Choi H. 2015. *The dynamics of the trade balance and the real exchange rate: the J curve and trade costs?* SED 2015 Meet. Pap. 1413

Alessandria G, Choi H, Ruhl KJ. 2014. *Trade adjustment dynamics and the welfare gains from trade*. NBER Work. Pap. 20663

Allayannis G, Ofek E. 2001. Exchange rate exposure, hedging, and the use of foreign currency derivatives. *J. Int. Money Finance* 20:273–96

Allen T. 2014. Information frictions in trade. *Econometrica* 82:2041–83

Alvarez F, Lucas Jr. RE. 2007. General equilibrium analysis of the Eaton–Kortum model of international trade. *J. Monet. Econ.* 54:1726–68

Amiti M, Konings J. 2007. Trade liberalization, intermediate inputs, and productivity: evidence from Indonesia. *Am. Econ. Rev.* 97:1611–38

Anderson JE, Larch M, Yotov YV. 2015. *Growth and trade with frictions: a structural estimation framework*. NBER Work. Pap. 21377

Anderson JE, van Wincoop E. 2003. Gravity with gravitas: a solution to the border puzzle. *Am. Econ. Rev.* 93:170–92

Anderson JE, van Wincoop E. 2004. Trade costs. *J. Econ. Lit.* 42:691–751

Arkolakis C. 2010. Market penetration costs and the new consumers margin in international trade. *J. Polit. Econ.* 118:1151–99

Arkolakis C, Costinot A, Donaldson D, Rodríguez-Clare A. 2015. *The elusive pro-competitive effects of trade*. NBER Work. Pap. 21370

Arkolakis C, Costinot A, Rodríguez-Clare A. 2012. New Trade Models, Same Old Gains? *Am. Econ. Rev.* 102:94–130

Armington PS. 1969. A theory of demand for products distinguished by place of production. *IMF Staff Pap.* 16:159–78

Asturias J. 2016. *Endogenous transportation costs*. Work. Pap., Georgetown Univ.

Asturias J, Garcia-Santana MJ, Ramos R. 2015. *Competition and the welfare gains from transportation infrastructure: evidence from the golden quadrilateral of India*. Work. Pap., Georgetown Univ.

Atkeson A, Burstein A. 2008. Pricing-to-market, trade costs, and international relative prices. *Am. Econ. Rev.* 98:1998–2031

Bagwell K, Staiger RW. 1999. An economic theory of GATT. *Am. Econ. Rev.* 89:215–48

Baier SL, Bergstrand JH. 2007. Do free trade agreements actually increase members' international trade? *J. Int. Econ.* 71:72–95

Bajona C, Kehoe TJ. 2010. Trade, growth, and convergence in a dynamic Heckscher-Ohlin model. *Rev. Econ. Dyn.* 13:487–513

Baldwin RE. 1992. Measurable dynamic gains from trade. *J. Polit. Econ.* 100:162–74

Bernard AB, Eaton J, Jensen JB, Kortum S. 2003. Plants and productivity in international trade. *Am. Econ. Rev.* 93:1268–90

Bernard AB, Jensen JB, Redding SJ, Schott PK. 2007. Firms in international trade. *J. Econ. Perspect.* 21:105–30

Bernard AB, Jensen JB, Redding SJ, Schott PK. 2009. The margins of U.S. trade. *Am. Econ. Rev.* 99:487–93

Bertoletti P, Etro F, Simonovska I. 2016. *International trade with indirect additivity*. NBER Work. Pap. 21984

Böhringer C, Löschel A. 2006. Computable general equilibrium models for sustainability impact assessment: status quo and prospects. *Ecol. Econ.* 60:49–64

Böhringer C, Vogt C. 2003. Economic and environmental impacts of the Kyoto Protocol. *Can. J. Econ. Can. Déconomique* 36:475–96

Broda C, Limao N, Weinstein DE. 2008. Optimal tariffs and market power: the evidence. *Am. Econ. Rev.* 98:2032–65

Brooks WJ, Pujolàs PS. 2014. *Nonlinear gravity*. McMaster Univ. Work. Pap. 2014-15

Brooks WJ, Pujolàs PS. 2016. *Capital accumulation and the welfare gains from trade*. McMaster Univ. Work. Pap. 2016-03

Brown DK, Deardorff AV, Stern RM. 1992. A North American Free Trade Agreement: analytical issues and a computational assessment. *World Econ.* 15:11–30

Brown DK, Stern RM. 1989. U.S.-Canada bilateral tariff elimination: the role of product differentiation and market structure. In *Trade Policies for International Competitiveness*, ed. RC Feenstra, pp. 217–54. Chicago: University of Chicago Press

Burfisher ME, Dyck J, Meade B, Mitchell L, Wainio JT, et al. 2014. *Agriculture in the Trans-Pacific Partnership*. USDA-ERS ERR 176

Burfisher ME, Robinson S, Thierfelder K. 2001. The Impact of NAFTA on the United States. *J. Econ. Perspect.* 15:125–44

Burniaux J-M, Truong TP. 2002. *GTAO-E: an energy-environmental version of the GTAP model*. GTAP Tech. Pap. 16

Bustos P. 2011. Trade liberalization, exports, and technology upgrading: evidence on the impact of MERCOSUR on Argentinian firms. *Am. Econ. Rev.* 101:304–40

Caliendo L, Parro F. 2015. Estimates of the trade and welfare effects of NAFTA. *Rev. Econ. Stud.* 82:1–44

Chaney T. 2008. Distorted gravity: the intensive and extensive margins of international trade. *Am. Econ. Rev.* 98:1707–21

Comerford D, Rodriguez-Mora JV. 2015. *The gains from economic integration*. SED 2015 Meet. Pap. 569

Corsetti G, Martin P, Pesenti P. 2007. Productivity, terms of trade and the "home market effect." *J. Int. Econ.* 73:99–127

Costinot A, Donaldson D, Vogel J, Werning I. 2015. Comparative advantage and optimal trade policy. *Q. J. Econ.* 130:659–702

Costinot A, Rodríguez-Clare A. 2014. Trade Theory with Numbers: Quantifying the Consequences of Globalization. In *Handbook of International Economics*, Vol. 4, ed. G Gopinath, E Helpman, and K Rogoff, pp. 197–261. Amsterdam: Elsevier

Cox D, Harris RG. 1992. North American free trade and its implications for Canada: results from a CGE model of North American trade. *World Econ.* 15:31–44

Dean JM, Signoret JE, Feinberg RM, Ludema RD, Ferrantino MJ. 2009. Estimating the price effects of non-tariff barriers. *BE J. Econ. Anal. Policy* 9:

Demidova S. 2008. Productivity improvements and falling trade costs: boon or bane? *Int. Econ. Rev.* 49:1437–62

Demidova S, Rodriguez-Clare A. 2009. Trade policy under firm-level heterogeneity in a small economy. *J. Int. Econ.* 78:100–112

Dervis K, de Melo J, Robinson S. 1982. *General equilibrium models for development policy*. Cambridge, England: Cambridge University Press

Devarajan S, Rodrik D. 1991. Pro-competitive effects of trade reform: results from a CGE model of Cameroon. *Eur. Econ. Rev.* 35:1157–84

di Giovanni J, Levchenko AA. 2013. Firm entry, trade, and welfare in Zipf's world. *J. Int. Econ.* 89:283–96

Diao X, Roe TL, Yeldan E. 1998. A simple dynamic applied general equilibrium model of a small open economy: transitional dynamics and trade policy. *J. Econ. Dev.* 23:77–101

Dixit AK, Stiglitz JE. 1977. Monopolistic competition and optimum product diversity. *Am. Econ. Rev.* 67:297–308

Donaldson D. Forthcoming. Railroads of the Raj: estimating the impact of transportation infrastructure. *Am. Econ. Rev.*

Donaldson D. 2015. The Gains from Market Integration. *Annu. Rev. Econ.* 7:619–47

Eaton J, Kortum S. 2002. Technology, geography, and trade. *Econometrica* 70:1741–79

Eaton J, Kortum S, Kramarz F. 2004. Dissecting trade: firms, industries, and export destinations. *Am. Econ. Rev.* 94:150–54

Eaton J, Kortum S, Kramarz F. 2011. An anatomy of international trade: evidence from French firms. *Econometrica* 79:1453–98

Egger P. 2000. A note on the proper econometric specification of the gravity equation. *Econ. Lett.* 66:25–31

Eslava M, Haltiwanger J, Kugler A, Kugler M. 2013. Trade and market selection: evidence from manufacturing plants in Colombia. *Rev. Econ. Dyn.* 16:135–58

Estevadeordal A, Frantz B, Taylor AM. 2003. The rise and fall of world trade, 1870–1939. *Q. J. Econ.* 118:359–407

Feenstra RC, Li Z, Yu M. 2014. Exports and credit constraints under incomplete information: theory and evidence from China. *Rev. Econ. Stat.* 96:729–44

Feenstra RC, Yang T-H, Hamilton GG. 1999. Business groups and product variety in trade: evidence from South Korea, Taiwan and Japan. *J. Int. Econ.* 48:71–100

Felbermayr G, Jung B, Larch M. 2015. The welfare consequences of import tariffs: a quantitative perspective. *J. Int. Econ.* 97:295–309

Fernandes AM. 2007. Trade policy, trade volumes and plant-level productivity in Colombian manufacturing industries. *J. Int. Econ.* 71:52–71

Fieler AC. 2011. Nonhomotheticity and bilateral trade: evidence and a quantitative explanation. *Econometrica* 79:1069–1101

Fox AK. 1999. *Evaluating the success of a CGE model of the Canada-U.S. Free Trade Agreement*. Unpublished Manuscript, Univ. Michigan

French S. 2016. The composition of trade flows and the aggregate effects of trade barriers. *J. Int. Econ.* 98:114–37

Galdón-Sánchez JE, Schmitz Jr. JA. 2002. Competitive pressure and labor productivity: world iron-ore markets in the 1980's. *Am. Econ. Rev.* 92:1222–35

Ghironi F, Melitz MJ. 2005. International trade and macroeconomic dynamics with heterogeneous firms. *Q. J. Econ.* 120:865–915

Goldberg PK, Khandelwal AK, Pavcnik N, Topalova P. 2010a. Imported intermediate inputs and domestic product growth: evidence from India. *Q. J. Econ.* 125:1727–67

Goldberg PK, Khandelwal AK, Pavcnik N, Topalova P. 2010b. Multiproduct firms and product turnover in the developing world: evidence from India. *Rev. Econ. Stat.* 92:1042–49

Goldberg PK, Pavcnik N. 2016. *The effects of trade policy*. NBER Work. Pap. 21957

Grossman GM, Krueger AB. 1994. Environmental impacts of a North American Free Trade Agreement. In *The Mexico-U.S. Free Trade Agreement*, ed. PM Garber, pp. 13–56. Cambridge, MA: MIT Press

Grubel HG, Lloyd PJ. 1971. The empirical measurement of intra- industry trade. *Econ. Rec.* 47:494–517

Hanson GH, Xiang C. 2004. The home-market effect and bilateral trade patterns. *Am. Econ. Rev.* 94:1108–29

Harris R. 1984. Applied general equilibrium analysis of small open economies with scale economies and imperfect competition. *Am. Econ. Rev.* 74:1016–32

Hazilla M, Kopp RJ. 1990. Social cost of environmental quality regulations: a general equilibrium analysis. *J. Polit. Econ.* 98:853–73

Head K, Mayer T. 2014. Gravity equations: workhorse, toolkit, and cookbook. In *Handbook of International Economics*, Vol. 4, ed. G Gopinath, E Helpman, K Rogoff, pp. 131–95. Amsterdam: Elsevier

Head K, Mayer T, Ries J. 2010. The erosion of colonial trade linkages after independence. *J. Int. Econ.* 81:1–14

Heerman KER, Arita S, Gopinath M. 2015. Asia-Pacific integration with China versus the United States: examining trade patterns under heterogeneous agricultural sectors. *Am. J. Agric. Econ.* 97:1324–44

Helpman E, Melitz M, Rubinstein Y. 2008. Estimating trade flows: trading partners and trading volumes. *Q. J. Econ.* 123:441–87

Hertel T. 2013. Global applied general equilibrium analysis using the Global Trade Analysis Project framework. In *Handbook of Computable General Equilibrium Modeling*, Vol. 1, ed. PB Dixon, D Jorgenson, pp. 815–76. Amsterdam: Elsevier

Hertel T, Hummels D, Ivanic M, Keeney R. 2007. How confident can we be of CGE-based assessments of free trade agreements? *Econ. Model.* 24:611–35

Hillberry RH, McDaniel CA. 2002. A decomposition of North American trade growth since NAFTA. *Int. Econ. Rev.* USITC Pub. No. 3527:1–6

Holmes TJ, Schmitz Jr. JA. 2001. A gain from trade: from unproductive to productive entrepreneurship. *J. Monet. Econ.* 47:417–46

Hummels D, Klenow PJ. 2005. The variety and quality of a nation's exports. *Am. Econ. Rev.* 95:704–23

Hummels D, Lugovskyy V, Skiba A. 2009. The trade reducing effects of market power in international shipping. *J. Dev. Econ.* 89:84–97

Hummels DL, Schaur G. 2013. Time as a trade barrier. *Am. Econ. Rev.* 103:2935–59

Hunter L, Markusen JR, Rutherford TF. 1995. Trade liberalization in a multinational-dominated industry. *J. Int. Econ.* 38:95–117

Ianchovichina E. 2012. *Dynamic modeling and applications for global economic analysis*. Cambridge, England: Cambridge University Press

Irarrazabal A, Moxnes A, Opromolla LD. 2015. The tip of the iceberg: a quantitative framework for estimating trade costs. *Rev. Econ. Stat.* 97:777–92

Jacks DS, Meissner CM, Novy D. 2008. Trade costs, 1870–2000. *Am. Econ. Rev.* 98:529–34

Jacks DS, Meissner CM, Novy D. 2011. Trade booms, trade busts, and trade costs. *J. Int. Econ.* 83:185–201

Jung JW, Simonovska I, Weinberger A. 2015. *Exporter heterogeneity and price discrimination: a quantitative view*. NBER Work. Pap. 21408

Kealey J, Pujolàs PS, Sosa-Padilla C. 2016. *Trade policy reform and firm-level productivity growth: does the choice of production function matter?* McMaster Univ. Work. Pap. 2016-08

Kehoe PJ, Kehoe TJ. 1994a. A primer on static applied general equilibrium models. *Fed. Reserve Bank Minneap. Q. Rev.* 18:2–16

Kehoe PJ, Kehoe TJ. 1994b. Capturing NAFTA's impact with applied general equilibrium models. *Fed. Reserve Bank Minneap. Q. Rev.* 18:17–34

Kehoe TJ. 2005. An evaluation of the performance of applied general equilibrium models on the impact of NAFTA. In *Frontiers in Applied General Equilibrium Modeling*, ed. TJ Kehoe, TN Srinivasan, J Whalley, pp. 341–77. Cambridge, England: Cambridge University Press

Kehoe TJ, Noyola PJ, Manresa A, Polo C, Sancho F. 1988. A general equilibrium analysis of the 1986 tax reform in Spain. *Eur. Econ. Rev.* 32:334–42

Kehoe TJ, Polo C, Sancho F. 1995. An evaluation of the performance of an applied general equilibrium model of the Spanish economy. *Econ. Theory* 6:115–41

Kehoe TJ, Prescott EC. 1995. Introduction to the symposium: the discipline of applied general equilibrium. *Econ. Theory* 6:1–11

Kehoe TJ, Rossbach J, Ruhl KJ. 2015. Using the new products margin to predict the industry-level impact of trade reform. *J. Int. Econ.* 96:289–97

Kehoe TJ, Ruhl KJ. 2013. How important is the new goods margin in international trade? *J. Polit. Econ.* 121:358–92

Kehoe TJ, Ruhl KJ, Steinberg JB. 2013. *Global imbalances and structural change in the United States. 19339*, NBER Work. Pap. 19339

Kleinert J, Spies J. 2011. *Endogenous transport costs in international trade*. IAW Disc. Pap. 74

Kohn D, Leibovici F, Szkup M. 2016. Financial frictions and new exporter dynamics. *Int. Econ. Rev.* 57:453–86

Krugman P. 1980. Scale economies, product differentiation, and the pattern of trade. *Am. Econ. Rev.* 70:950–59

Leibovici F. 2015. *Financial development and international trade*. York Univ. Work. Pap. 2015–3

Levchenko AA, Zhang J. 2016. The evolution of comparative advantage: measurement and welfare implications. *J. Monet. Econ.* 78:96–111

Li C, Whalley J. 2014. China and the Trans-Pacific Partnership: a numerical simulation assessment of the effects involved. *World Econ.* 37:169–92

Limão N, Venables AJ. 2001. Infrastructure, geographical disadvantage, transport costs, and trade. *World Bank Econ. Rev.* 15:451–79

Maggi G, Rodríguez-Clare A. 2007. A political-economy theory of trade agreements. *Am. Econ. Rev.* 97:1374–1406

Manova K. 2013. Credit constraints, heterogeneous firms, and international trade. *Rev. Econ. Stud.* 80:711–44

Manova K, Zhang Z. 2012. Export prices across firms and destinations. *Q. J. Econ.* 127:379–436

Markusen JR. 1981. Trade and the gains from trade with imperfect competition. *J. Int. Econ.* 11:531–51

McCallum J. 1995. National borders matter: Canada-U.S. regional trade patterns. *Am. Econ. Rev.* 85:615–23

McKitrick RR. 1998. The econometric critique of computable general equilibrium modeling: the role of functional forms. *Econ. Model.* 15:543–73

Melitz MJ. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71:1695–1725

Melitz MJ, Ottaviano GIP. 2008. Market size, trade, and productivity. *Rev. Econ. Stud.* 75:295–316

Melitz MJ, Redding SJ. 2015. New trade models, new welfare implications. *Am. Econ. Rev.* 105:1105–46

Narayanan GB, Ciuriak D, Singh HV. 2016. Quantifying the mega-regional trade agreements: a review of the models. In *TPP and India: Implications of Mega-regionals for Developing Economies*, ed. HV Singh, pp. 93–131. New Delhi, India: Wisdom Tree

New Zealand Ministry of Foreign Affairs and Trade and China Ministry of Commerce. 2004. *A joint study report on a free trade agreement between China and New Zealand.*

Obstfeld M, Rogoff K. 2001. The six major puzzles in international macroeconomics: is there a common cause? In *NBER Macroeconomics Annual 2000*, Vol. 15, ed. BS Bernanke, K Rogoff, pp. 339–412. Cambridge, MA: MIT Press

Opp MM. 2010. Tariff wars in the ricardian model with a continuum of goods. *J. Int. Econ.* 80:212–25

Ossa R. 2015. Why trade matters after all. *J. Int. Econ.* 97:266–77

Ottaviano GIP, Pessoa JP, Sampson T, van Reenen J. 2014. *The costs and benefits of leaving the EU*. CFS Work. Pap. 472

Pavcnik N. 2002. Trade liberalization, exit, and productivity improvements: evidence from Chilean plants. *Rev. Econ. Stud.* 69:245–76

Peterson EB, Schleich J, Duscha V. 2011. Environmental and economic effects of the copenhagen pledges and more ambitious emission reduction targets. *Energy Policy* 39:3697–3708

Ramanarayanan A. 2012. *Imported inputs and the gains from trade*. SED 2012 Meet. Pap. 612

Roberts MJ, Tybout JR. 1997. The decision to export in Colombia: an empirical model of entry with sunk costs. *Am. Econ. Rev.* 87:545–64

Romalis J. 2007. NAFTA's and CUSFTA's impact on international trade. *Rev. Econ. Stat.* 89:416–35

Ruhl KJ. 2008. *The international elasticity puzzle*. Work. Pap., Penn State Univ.

Ruhl KJ, Willis JL. Forthcoming. New exporter dynamics. *Int. Econ. Rev.*

Samuelson PA. 1954. The transfer problem and transport costs, ii: analysis of effects of trade impediments. *Econ. J.* 64:264–89

Santos Silva JMC, Tenreyro S. 2006. The log of gravity. *Rev. Econ. Stat.* 88:641–58

Schmeiser KN. 2012. Learning to export: export growth and the destination decision of firms. *J. Int. Econ.* 87:89–97

Schmitz Jr. JA. 2005. What determines productivity? Lessons from the dramatic recovery of the U.S. and Canadian iron ore industries following their early 1980s crisis. *J. Polit. Econ.* 113:582–625

Shikher S. 2012. Predicting the effects of NAFTA: now we can do it better!. *J. Int. Glob. Econ. Stud.* 5:32–59

Shoven JB, Whalley J. 1984. Applied general-equilibrium models of taxation and international trade: an introduction and survey. *J. Econ. Lit.* 22:1007–51

Shoven JB, Whalley J. 1992. *Applying general equilibrium*. Cambridge, England: Cambridge University Press

Simonovska I. 2015. Income differences and prices of tradables: insights from an online retailer. *Rev. Econ. Stud.* 82:1612–56

Simonovska I, Waugh M. 2012. *Different trade models, different trade elasticities?* SED 2012 Meet. Pap. 618

Smith A, Venables AJ. 1988. Completing the internal market in the European Community: some industry simulations. *Eur. Econ. Rev.* 32:1501–25

Sobarzo HE. 1995. A general equilibrium analysis of the gains from NAFTA for the Mexican economy. In *Modeling North American Economic Integration*, ed. PJ Kehoe, TJ Kehoe, pp. 91–115. Boston: Kluwer Academic

Strutt A, Minor P, Rae A. 2015. *A dynamic computable general equilibrium (CGE) analysis of the Trans-Pacific Partnership agreement: potential impacts on the New Zealand economy*. New Zealand MFAT Rep.

Thorstensen V, Ferraz L. 2014. *The impacts of TTIP and TPP on Brazil*. Fundação Getúlio Vargas

Timmer MP, Dietzenbacher E, Los B, Stehrer R, de Vries GJ. 2015. An illustrated user guide to the World Input–Output Database: the case of global automotive production. *Rev. Int. Econ.* 23:575–605

Trefler D. 2004. The long and short of the Canada-U.S. Free Trade Agreement. *Am. Econ. Rev.* 94:870–95

Tybout J, de Melo J, Corbo V. 1991. The effects of trade reforms on scale and technical efficiency: new evidence from Chile. *J. Int. Econ.* 31:231–50

US International Trade Commission. 2003. *U.S.-Chile free trade agreement: potential economy-wide and selected sectoral effects*. USITC Pub. 3605

Whalley J. 1985. *Trade liberalization among major world trading areas*. Cambridge, MA: MIT Press

Wolf HC. 2000. Intranational home bias in trade. *Rev. Econ. Stat.* 82:555–63

Yeaple SR. 2005. A simple model of firm heterogeneity, international trade, and wages. *J. Int. Econ.* 65:1–20

Yi K-M. 2003. Can vertical specialization explain the growth of world trade? *J. Polit. Econ.* 111:52–102

Yi K-M. 2010. Can multistage production explain the home bias in trade? *Am. Econ. Rev.* 100:364–93

A. Appendix

Below is a selection from our online appendix, which explores the robustness of our results.

A.1 Log Differences versus Percent Changes

When evaluating the accuracy of the GTAP predictions in Table 1, we define growth in terms of percent changes. When evaluating the accuracy of the CP predictions in Table 2, we defined growth in terms of log differences. We chose to use a different definition for growth between the two tables for two reasons. First, we chose each definition to be consistent with the definition used by the original papers. GTAP papers use percent changes, whereas CP use log differences. Second, the choice shows that the overall shortcomings of AGE models do not depend on exactly how we define growth (percent changes or log differences).

To explore the robustness of our results, we show that Tables 1 and 2 are similar overall when the definitions of growth are switched between them. Note that we can move between percent changes and log differences with the following equations:

$$\text{log diff} = 100 * \log\left(\frac{\text{percent change}}{100} + 1\right)$$

$$\text{percent change} = 100 * \left(\exp\left(\frac{\text{log diff}}{100}\right) - 1\right).$$

Tablew A.1 and A.2 show the results of this exercise.

Table A.1: Comparisons of GTAP and LTP predictions of recent trade liberalizations with data (log differences)

Exporter	Importer	GTAP correlation with data	LTP correlation with data
United States	Australia	0.64	0.49
Australia	United States	–0.18	0.45
United States	Chile	0.33	0.59
Chile	United States	0.15	0.44
China	Chile	0.51	0.78
Chile	China	–0.09	–0.08
China	New Zealand	–0.38	0.56
New Zealand	China	–0.17	0.30
Simple average		0.10	0.44

Overall, the LTP correlation in Table A.1 stays almost the same as in Table 1 in the paper. The biggest changes are that GTAP appears to perform better for China exports to Chile and for

U.S. exports to Australia when expressed in log differences, whereas Chile exports to China perform worse. Overall, the average LTP correlation is similar across the two tables (0.44 compared with 0.49 in Table 1). The average GTAP correlation is slightly higher with log differences (0.10 compared with –0.00). It still performs significantly worse than the LTP methodology with both definitions, however.

Table A.2: Comparisons of CP and LTP predictions of NAFTA with data (percent changes)

Exporter	Importer	CP correlation with data	LTP correlation with data
Canada	Mexico	−0.47	0.19
Canada	United States	0.63	0.05
Mexico	Canada	−0.46	0.76
Mexico	United States	−0.11	0.14
United States	Canada	0.38	0.20
United States	Mexico	0.96	−0.02
Simple average		0.16	0.22

The results in Table A.2 are similar to Table 2, with the LTP methodology outperforming the CP predictions. By far the biggest change is for U.S. exports to Mexico, which exhibit a negative correlation for LTP and a near perfect correlation for CP. This result is driven entirely by growth in exports in the petroleum industry, which grew by over 1500 percent between 1991 and 2006 (the second highest growth was in chemicals, which grew by 200 percent). Despite this huge growth, it actually grew less than predicted. CP predicted an increase in growth of nearly 3000 percent (the second highest predicted growth was in electrical machinery, at nearly 200 percent). CP predict this large increase because of a much higher estimated trade elasticity for the petroleum industry than for any other sector (over 50; no other sector was over 20).

If the petroleum industry is dropped when evaluating the performance of each methodology, then the results are as shown in Table A.3.

Table A.3: Comparisons of CP and LTP predictions of NAFTA with data (percent changes, U.S. petroleum exports to Mexico excluded)

Exporter	Importer	CP correlation with data	LTP correlation with data
Canada	Mexico	−0.47	0.19
Canada	United States	0.63	0.05
Mexico	Canada	−0.46	0.76
Mexico	United States	−0.11	0.14
United States	Canada	0.38	0.20
United States	Mexico	−0.37	0.30
Simple average		−0.06	0.27

As we can see, simply excluding the petroleum industry changes the correlation between the predictions of the CP model and actual growth for U.S. exports to Mexico from 0.96 to −0.37. One of the effects of log differences is that it lessens the overall influence of extreme growth rates for individual industries compared with using percentage changes.

A.2 Effects of Outlier Observations

When evaluating the GTAP predictions, we have excluded two industries: oil exports from the United States to Chile and cattle meat exports from Australia to the United States. While excluding these industries does effect the correlations for each of those importer-exporter pairs, removing them has little effect on the overall performance of the GTAP predictions and LTP methodology. Table A.4 shows that LTP methodology performs substantially better than the GTAP predictions even when these outlier industries are not excluded from the analysis.

Table A.4: Comparisons of GTAP and LTP predictions of recent trade liberalizations with data (percent changes, outliers not excluded)

Exporter	Importer	GTAP correlation with data	LTP correlation with data
United States	Australia	0.27	0.55
Australia	United States	0.49	0.08
United States	Chile	0.05	–0.05
Chile	United States	0.03	0.48
China	Chile	0.14	0.61
Chile	China	0.04	0.07
China	New Zealand	–0.36	0.61
New Zealand	China	–0.09	0.48
Simple average		0.07	0.35

Note that including the cattle meat exports from Australia to the United States makes the GTAP predictions appear to perform much better for that pair. It is, however, a classic case of getting things right for the wrong reason. GTAP predicts a large increase in the quantity of beef exports because of a complete removal of the tariff quota in the simulation. In actuality, the trade agreement between Australia and the United States left the quota largely intact, and the increase in trade value was due to a worldwide increase in the price of beef. If we instead ran GTAP without simulating a complete elimination of all trade barriers (a complete removal is a reasonable approximation for the other industries and trade agreements we evaluate), it would be unable to capture the increased value of exports of cattle meat. Since it is questionable whether any trade model should be expected to capture such an increase in the worldwide price, we exclude the industry from our main analysis in the paper.

While our robustness appendix shows that our overall results are largely unchanged by outliers, base years, weighting schemes, and how we define growth, it also shows that for some individual country pairs, these choices substantially alter the apparent accuracy of various predictions and methodologies. Thus, it is important for researchers and policymakers to think carefully about how the performance of their models should be evaluated. Beyond that, it is important that we actually carry out such evaluations, so that shortcomings can be identified and our models can continue to improve.