CESifo Economic Studies, Vol. 59, 2/2013, 199–222 doi:10.1093/cesifo/ift004

Measuring the Effects of Endogenous Policies on Economic Integration

Jeffrey H. Bergstrand

Department of Finance, Mendoza College of Business, and Kellogg Institute for International Studies, University of Notre Dame, Notre Dame, IN 46556, USA, and CESifo Munich. e-mail: bergstrand.1@nd.edu.

Abstract

Despite widespread anecdotal evidence that lower trade barriers increase international trade, there is little firm quantitative evidence of the 'trade-cost elasticity' of trade flows, one of the two key aggregate statistics that have recently been identified as sufficient to quantify the economic welfare effects of trade-policy liberalizations and/or trade-cost reductions (the other statistic being the import-penetration ratio). In other words, most estimates of the trade-cost elasticity are imprecise and lack consistency. In this article, we discuss two issues that are critical in better assessing empirically the trade-flow and welfare effects of trade liberalizations (or trade-cost changes). The first issue is how to quantify the trade-cost elasticity when trade costs themselves are approximated imperfectly. The second issue is that typical empirical evaluations to estimate the impact of trade-policy liberalizations on trade flows use the 'gravity equation'. However, the self-selection of country pairs into such agreements introduces endogeneity bias in the estimation of the trade-cost elasticity in gravity equations, requiring better identification techniques.(JEL codes: F10; F12 and F13)

Keywords: International trade, economic integration agreements, gravity equations

1 Introduction

One of the most prominent aspects of the world's transformation over the last six decades has been increased globalization. Globalization is broader than just increased economic interactions; it also embraces increased exchanges of cultures, attitudes, and mores. In economic terms, increased globalization typically refers to increased international trade flows, investment flows, and migration flows. Of course, such flows are endogenous to changes in—what I term—'natural' and 'unnatural' costs (among other factors to be discussed later).

For tractability, let me define briefly what I mean by the terms 'natural' and 'unnatural' costs. I will refer to those costs associated with geography and technology as 'natural' costs. For instance, an obvious natural cost to shipping a good internationally is distance; larger distances raise the cost of transport. But such costs are also influenced by technology (which is, of course, man-made). However, in my discussion below, I will treat technology as exogenous and treat technological innovations that have increased globalization—such as containerization for the shipment of goods internationally—as reductions in 'natural' trade costs.

By contrast, I will refer to 'policy-based' (or 'man-made') costs as 'unnatural' costs. In most cases, these costs are created by government policies—such as tariffs on internationally traded goods—and these can be raised or lowered by policymakers. Despite enormous progress by policymakers worldwide in lowering these 'unnatural' trade, investment, and migration costs since World War II, policy-based impediments to world trade, investment, and migration still exist and remain substantive. As one piece of evidence that the world is not yet—as Friedman (2005) would characterize it—'flat', Eaton and Kortum (2002) in a seminal article show using a calibration exercise that the world is much closer to one of 'autarky'—that is, one where world trade flows are prohibitively expensive—rather than one of frictionless trade.

Surprisingly, despite the fact that international trade economists spend considerable time and effort explaining that reductions in (natural and unnatural) trade costs increase trade—and such augmented trade improves consumers' welfare—we actually know quite little *quantitatively* of the impact of international economic costs on international economic flows. Specifically, we lack precise and consistent quantitative knowledge of the impact of (bilateral) trade costs on (bilateral) trade flows, of investment costs on foreign direct and portfolio investment flows, and of migration costs on migration flows—much less the cross-impact of each of these costs on the other flows. In fact, we also lack firm systematic data on trade costs themselves!

In a recent influential article on international trade flows' determinants, Arkolakis et al. (2012) show that there is a wide class of 'Quantitative Trade Models' in the international trade literature for which the welfare effect of trade-cost reductions can be summarized with two aggregate statistics. One of these statistics is the 'import-penetration' ratio, which is simply one minus the share of aggregate national expenditures on domestically produced goods. The second is the 'trade-cost elasticity', which is the percentage change in trade flows in response to a 1% change in an *ad valorem* measure of 'trade costs' (natural or unnatural). While the first statistic can be estimated fairly precisely using national income accounts data, the second statistic has remained elusive empirically. Arkolakis et al. (2012) actually devote the last substantive section of their article articulating some of the issues related to estimating this trade-cost elasticity (which they denote ϵ). They argue that the principal remaining issues for estimating ϵ are econometric ones. In particular, they note—citing empirical estimates of ϵ in Baier and Bergstrand (2001)—that one of the key econometric issues is the 'standard orthogonality condition', that is, whether measures of trade costs are *exogenous* variables in typical gravity equations.

This article takes up where Arkolakis et al. (2012) left off in discussing econometric (or estimation) issues in measuring the trade-cost elasticity—as well as measuring trade costs themselves—not addressed in their article. In section 2, we address first methods to estimate the trade-cost elasticity in the presence of only imperfect measures of true *ad valorem* trade costs (denoted τ). The world is not so generous as to provide observable measures of true trade costs. In this section, however, we assume such imperfect trade-cost measures are orthogonal to the gravity equation's error term (i.e., no endogeneity bias). In section 3, we address methods to estimate the trade-cost elasticity when such (imperfect) trade-cost measures are not orthogonal to the error term, that is, when there is endogeneity bias. Section 4 concludes.

2 Estimating the Trade-Cost Elasticity with Imperfect Exogenous Trade-Cost Measures

2.1 Background

Arkolakis et al. (2012), henceforth ACRC, recently re-examined theoretically the key elements in measuring the economic welfare 'gains from trade'. Looking back over developments in the international trade literature over the past 30 years, ACRC found that the 'gains from trade' were basically measurable in a wide class of—what they termed—'Quantitative Trade Models' (QTMs). This broad class of models includes endowment-economies with Armington preferences (Anderson 1979; Anderson and van Wincoop 2003), monopolistic competition models with increasing returns to scale (Krugman 1980), Ricardian models with perfectly competitive firms with heterogeneous productivities (Eaton and Kortum 2002), and Melitz-type models with heterogeneous firms, monopolistic competition, and variable and fixed exporting costs (Melitz 2003; Chaney 2008; Redding 2011).

The principal conclusion from the ACRC article is that the welfare gains from trade in this broad class of QTMs could be quantified in terms of two key statistics. The first statistic is the share of aggregate national expenditures on domestically produced goods, which they denote λ_{jt} for country j in year t ($0 < \lambda_{jt} < 1$). The second statistic is the 'trade-cost elasticity', denoted ϵ (which they actually refer to as the 'trade elasticity'). The trade-cost elasticity is defined as $\epsilon = d\ln X_{ijt}/d\ln \tau_{ijt} < 0$, where X_{ijt} is the aggregate bilateral trade flow from country i to country j in year t and τ_{ijt} is the gross *ad valorem* bilateral trade cost associated with trade flow X_{ijt} (either natural or unnatural cost; $\tau_{ijt} > 1$). The main insight in ACRC is that—for the wide class of QTMs listed above—the welfare gain from a reduction in trade costs simplifies to:

$$d\ln W_{jt} = (1/\epsilon)d\ln \lambda_{jt} \tag{1}$$

where W_{jt} is country j's welfare (or real income of the representative consumer) in year t.

The economic intuition for this simple and common result for all these QTMs is that a reduction in trade costs for importing country j improves the country's terms of trade. This improvement in the terms of trade can be inferred from the changes in its relative import demands from various countries. The reason that this simple and 'general' result surfaces across a broad class of QTMs is that it relies on a very small set of feasible assumptions, common to this broad class of models. Each of the models listed above shares only four primitive assumptions: Dixit–Stiglitz preferences; one factor of production (typically, labor); linear cost functions; and perfect or monopolistic competition. Also, all the QTMs share three 'macro-level' restrictions: multilateral trade balance; aggregate profits are a constant share of aggregate revenue; and import demand systems are constant-elasticity-of-substitution (CES).

It turns out that all of the QTMs listed above share the fact that bilateral import flows can be described in equilibrium by a 'gravity equation'. For instance, in the Armington endowment-economy model in Anderson and van Wincoop (2003), the implied gravity equation is:

$$X_{ijt} = Y_{it} Y_{jt} \left(\frac{(\beta_{it} p_{it})^{-(\sigma-1)} \tau_{ijt}^{-(\sigma-1)}}{\sum_{k=1}^{K} Y_{kt} (\beta_{kt} p_{kt})^{-(\sigma-1)} \left(\tau_{kjt}\right)^{-(\sigma-1)}} \right) \tag{2}$$

where X_{ijt} is the trade flow from i to j in year t, Y_{it} (Y_{jt}) is gross domestic product (GDP) of i in t, β is an unobservable preference parameter for i's good, p_{it} is the price of i's good, and σ is the elasticity of substitution in consumption.[1] So here, $\epsilon = -(\sigma - 1)$.

In the context of CES preferences, increasing returns to scale, and monopolistic competition, the Krugman (1980) model (summarized in

[1] See also Anderson (1979) and Bergstrand (1985).

Baier and Bergstrand (2001) and Feenstra (2004)) yields a gravity equation:

$$X_{ijt} = N_{it} Y_{jt} \left(\frac{w_{it}^{-(\sigma-1)} \tau_{ijt}^{-(\sigma-1)}}{\sum_{k=1}^{K} N_{kt}(w_{kt})^{-(\sigma-1)} \left(\tau_{kjt}\right)^{-(\sigma-1)}} \right) \tag{3}$$

where N_{it} is the number of products/producers in i, and w_{it} is the wage rate for labor in i. Here also, $\epsilon = -(\sigma - 1)$.

In the context of a Ricardian model of international trade with perfect competition and heterogeneous firms/productivities, Eaton and Kortum (2002) generate a gravity equation:

$$X_{ijt} = T_{it} Y_{jt} \left(\frac{w_{it}^{-\gamma} \tau_{ijt}^{-\gamma}}{\sum_{k=1}^{K} T_{kt}(w_{kt})^{-\gamma} \left(\tau_{kjt}\right)^{-\gamma}} \right) \tag{4}$$

where T_{it} is the 'state of technology' in i, and γ is an index of the heterogeneity of firms' productivities (or comparative advantages). In the Eaton–Kortum model, $\epsilon = -\gamma$.

Finally, allowing for heterogeneous firms/productivities, monopolistic competition, increasing returns to scale, and fixed exporting costs, Melitz (2003) yields a gravity equation:

$$X_{ijt} = N_{it} Y_{jt} \left(\frac{w_{it}^{-\gamma} \tau_{ijt}^{-\gamma} f_{ijt}^{-[\gamma/(\sigma-1)-1]}}{\sum_{k=1}^{K} N_{kt}(w_{kt})^{-\gamma} \left(\tau_{kjt}\right)^{-\gamma} \left(f_{kjt}\right)^{-[\gamma/(\sigma-1)-1]}} \right) \tag{5}$$

where f_{ijt} is the fixed costs of exporting from i to j in t and, as noted in Chaney (2008), $\gamma = (\sigma - 1) + [\gamma - (\sigma - 1)]$. In this model's context, $(\sigma - 1)$ is the 'intensive margin' elasticity of trade with respect to variable trade costs (τ_{ij}), and $\gamma - (\sigma - 1)$ is the 'extensive margin' elasticity of trade with respect to variable trade costs. Here also, $\epsilon = -\gamma$.

Thus, while the primary goal of ACRC is to show that the welfare gains from trade liberalization across a broad range of QTMs are basically summarized by a simple function of two common key 'aggregate statistics', λ_{jt} and ϵ, a second major insight of the article is the importance for measuring $d \ln W_j$ of estimating the variable trade-cost elasticity (ϵ) with both *consistency* as well as *precision*.

ACRC argue in the final substantive section of their article (titled 'Estimating the Trade Elasticity') that the gravity equation provides a very useful approach to estimating ϵ. For all the models above (in the absence of zero trade flows), one can express the implied gravity equation as:

$$\ln X_{ijt} = A_{it} + B_{jt} + \epsilon \ln \tau_{ijt} + \nu_{ijt} \tag{6}$$

where A_{it} is an exporter-time (fixed) effect, B_{jt} is an importer-time (fixed) effect, and ν_{ijt} is an error term. However, a key issue they note is that proper and consistent estimation of ϵ requires the 'standard orthogonality condition' holding: independence of τ_{ijt} with the gravity equation's error term (ν_{ijt}). Yet, ACRC stop here, noting that this is principally an *econometric* issue for which they have 'little to contribute'.

Where ACRC end, we begin. In this article, we consider two issues relevant to estimating ϵ. First, we do not observe true values of τ_{ijt}; typically, we explain trade flows using proxies for τ such as bilateral distance, which obscures identification of ϵ because of the unknown relationship between bilateral distance ($dist_{ij}$) and unobservable true bilateral trade cost τ_{ijt}. In the remainder of this section, we address methods to infer ϵ when τ is unobserved. A second issue is selection bias. Suppose country pairs self-select into economic integration agreements (EIAs). For instance, trade between a pair of countries may be below its 'natural' level because of barriers to trade unobservable to the econometrician. If trade is low where such barriers exist, inducing a country pair to form an EIA to reduce this impediment, coefficient estimates on right hand side (RHS) dummy variables for these EIAs in gravity equations may be biased downward (referred to as 'negative selection'). We address this issue in section 3.

2.2 Estimation issues

In a recent article, Bergstrand et al. (2013) address a method for estimating the trade-cost elasticity in the presence of imperfect measures of true *ad valorem* trade costs. For instance, consider the Krugman (1980) model of international trade. As discussed in ACRC, this model satisfies the four primitive assumptions in the broad class of QTMs discussed above, as well as the three 'macro-level' restrictions. This model yields a gravity equation of the form in equation (3):

$$X_{ij} = \frac{Y_i Y_j}{Y^W} \left(\frac{(Y_i/L_i)^{-\sigma} \left(\tau_{ij}\right)^{-(\sigma-1)}}{\sum_{k=1}^{N} (Y_k/Y^W)(Y_k/L_k)^{-\sigma} \left(\tau_{kj}\right)^{-(\sigma-1)}} \right) u_{ij} \tag{7}$$

where we have replaced w_i with Y_i/L_i because with one factor, labor, $Y_i = w_iL_i$. Under the assumption of market clearing, this gravity equation can be estimated structurally by assuming K market-clearing equations:

$$Y_i = \sum_{j=1}^{K} X_{ij} \quad i = 1,...,K, \tag{8}$$

Assuming no zeros in trade flows, one can take the logarithm of equation (7) and obtain:

$$\begin{aligned} \ln X_{ij} = &-\ln Y^W + \ln Y_i + \ln Y_j - \sigma \ln(Y_i/L_i) - (\sigma - 1)\ln \tau_{ij} \\ &- \ln\left(\sum_{k=1}^{N} (Y_k/Y^W)(Y_k/L_k)^{-\sigma}\left(\tau_{kj}\right)^{-(\sigma-1)}\right) + \nu_{ij} \end{aligned} \tag{9}$$

Bergstrand et al. (2013) suggest three possible methods to estimate σ. One possible method is to estimate σ based on the coefficient for $\ln(Y_i/L_i)$. However, $\ln(Y_i/L_i) = \ln w_i$. Since w_i is endogenous, it is likely correlated with the error term ν_{ij}; coefficient estimates would likely be biased. A second possibility is to estimate σ from the coefficient estimate for $\ln \tau_{ij}$. However, in reality, τ_{ij} is not observable. Typically, one uses proxies for trade costs such as bilateral distance and dummy variables for the presence or absence of an EIA between country pairs using:

$$\ln X_{ij} = A_i + B_j + \rho(1-\sigma)\ln dist_{ij} + \psi(1-\sigma)EIA_{ij} + \nu_{ij} \tag{10}$$

where $dist_{ij}$ is the bilateral distance between the economic centers of regions i and j, and EIA_{ij} is a dummy variable assuming a value of 1 (0) if the two regions share (do not share) an EIA. Bergstrand et al. (2013) suggest a third way to estimate σ. Estimating equation (10) using fixed-effects A_i and B_j and exponentiating yields a set of estimates $\widehat{dist_{ij}^{\rho(1-\sigma)}}$ and $\widehat{e^{\psi(1-\sigma)EIA_{ij}}}$. The product of these two estimates is an estimate of $\widehat{\tau_{ij}^{1-\sigma}}$. Bergstrand et al. (2013) show that one can substitute $\widehat{\tau_{ij}^{1-\sigma}}$ into equation (7) (which eliminates the u_{ij}), $\widehat{\tau_{kj}^{1-\sigma}}$ into the same equation's analogue for X_{kj}, take the ratio of the two equations for X_{ij} and X_{kj}, and solve for an estimate of σ ($\hat{\sigma}$).

Using an extensive set of Monte Carlo simulations, Bergstrand et al. (2013) show that this technique provides *consistent* and *precise* estimates of gravity equation parameters, the elasticity of substitution in consumption (σ)—which is the 'trade elasticity' in this context—*and* the welfare changes from increases or decreases in trade costs.

To illustrate the relevance of this technique to empirical applications, Bergstrand et al. (2013) apply the framework to the well-known

McCallum 'border-puzzle' issue. The McCallum border puzzle refers to McCallum (1995). McCallum (1995) applied a traditional gravity equation—ignoring the role of relative prices discussed earlier—to the trade flows among Canadian provinces, among US states, and between Canadian provinces and US states, including a dummy variable for the presence of the 'national border', i.e., the US–Canadian border. McCallum found an enormous quantitative effect of this national border. He found that two typical Canadian provinces had *22 times* more goods trade than a typical pairing of a Canadian province with a US state.

This 'mind-boggling' empirical finding spurred a cottage industry of papers. Of the more important subsequent articles investigating this finding further, Anderson and van Wincoop (2003) argued that there were two essential elements missing from the analysis of the border in McCallum (1995). First, they derived gravity equation (2) above, which emphasizes the importance of relative prices in influencing trade flows, and demonstrated that estimating a traditional gravity equation *ignoring* relative prices would lead to omitted variables bias in the gravity-equation parameter estimates. Second, they emphasized that a better estimate of the border effect would be a *general equilibrium* comparative static estimate, rather than simply the partial effect. Anderson and van Wincoop themselves chose to use a nonlinear least squares estimation procedure for estimating the parameters and then used a nonlinear program to estimate the general equilibrium comparative static effects.

However, one of the shortcomings of the Anderson and van Wincoop approach is that it cannot *estimate* the actual elasticity of substitution in consumption. That is a problem when trying to estimate welfare effects; to do any welfare calculations, they would have to assume some value of σ. This elasticity is essential for welfare calculations.

Bergstrand et al. (2013) replicated the Anderson and van Wincoop (2003) study and its empirical results, assuming as there symmetric effects of the national border on trade (i.e., a single dummy coefficient estimate, regardless of the direction of the trade flow between the two countries). Bergstrand et al. (2013) showed that—owing to correlated errors (say, owing to omitted variables bias)—estimation of gravity-equation parameters would likely yield consistent parameter estimates if exporter and importer fixed effects were used instead. However, for the same gravity-equation parameter estimates, the Bergstrand et al. (2013) approach generated an *estimated* value of the trade elasticity of (approximately) 7, whereas Anderson and van Wincoop (2003) considered the effects of various alternative elasticities ranging from 2 to 20 (choosing 5 as a representative elasticity of substitution). Bergstrand et al. (2013) showed that the welfare effects of trade-cost changes—here, the elimination of the national border—were not only quantitatively different between the two approaches, but also *qualitatively* different. To see the importance of consistent and precise estimation of the 'trade elasticity' as emphasized in Arkolakis et al. (2012), the approach in Anderson and van Wincoop (2003) using their preferred assumed elasticity of 5 implied a welfare gain for Canada but a welfare *loss* for the United States from eliminating the US–Canadian national border. By contrast, Bergstrand et al. (2013) found using the same *parameter* estimates—but using their *estimated* trade elasticity—a welfare gain for Canada *and* a welfare gain for the United States of eliminating the two countries' common border.

3 Estimating the Trade-Cost Elasticity with Imperfect Endogenous Trade-Policy Measures

The last substantive section of Arkolakis et al. (2012) argued that a useful methodology empirically to estimate the trade-cost elasticity was the gravity equation. However, the authors reminded the reader that estimates using the gravity equation must satisfy the 'standard orthogonality condition' to generate consistent parameter estimates. In the previous section of the article, we discussed the usefulness of recent gravity-equation estimates under a strong assumption: that the explanatory (or RHS) variables satisfied the standard orthogonality condition, that is, that they were 'exogenous'. In the case of bilateral distance and other natural trade costs, such an assumption is quite feasible. However, when the policy maker is interested in evaluating *ex post* the effects of measures of 'policy' on trade flows, a concern about the 'exogeneity' of the policy measures arises. In this section, we address recent methodologies to estimate with consistency and precision the effects of endogenous policies on trade flows or—in more general terms—the 'effects of endogenous policies on economic integration'.

Nearly 20 years have passed since Trefler (1993) showed that *ex post* empirical estimates of the effects of U.S. trade policies on U.S. imports were underestimated *considerably* by not accounting for the endogeneity of trade-policy measures.[2] The downward bias of such estimates was confirmed later in Lee and Swagel (1997) for a broad cross-section of country pairs' bilateral trade flows and trade-policy measures. More recently, Baier and Bergstrand (2007) argued that—in the context of EIAs and gravity equations—the endogeneity bias may well be attributed to *self-selection* bias. Using panel techniques to account for endogeneity bias,

[2] Trefler (1993) showed that—after accounting for endogeneity—the effect of trade liberalizations was 10 times that estimated otherwise.

Baier and Bergstrand (2007) showed—in the spirit of Trefler (1993) and Lee and Swagel (1997)—that previous estimates of the effects of EIAs on bilateral trade flows were underestimated considerably; the authors showed that after accounting for endogeneity, the effects of EIAs were five times that estimated otherwise. Several articles have shown—using instrumental variables or panel techniques—the downward bias of EIAs' effects when not accounting for endogeneity bias, including most recently Baier et al. (2011).

3.1 Sources of trade-policy endogeneity in gravity equations

In determining the potential correlation between the gravity equation's error term with the EIA dummy, one first needs to consider what determines the likelihood of a free trade agreement (FTA) between a pair of countries. Although trade economists have examined empirically for many years the determinants of tariff rates and nontariff barrier levels across industries and across countries, little empirical work has examined the determinants of EIAs until Mansfield and Reinhardt (2003) and Baier and Bergstrand (2004a). The former study examined political determinants of EIAs, whereas the latter presented a theoretical and empirical model of economic determinants of FTAs. Baier and Bergstrand (2004a) found strong empirical evidence that pairs of countries that have FTAs tend to share economic characteristics that their theory suggests should enhance economic welfare of the pairs' representative consumers. For instance, two countries tend to have an FTA the larger and more similar their GDPs, the closer they are to each other but the more remote the pair is from the rest-of-the-world (*ROW*), and the wider (narrower) the difference in their relative factor endowments with respect to each other (the *ROW*). But these include the same factors that tend to explain large trade flows. Thus, in terms of observable economic characteristics, countries with FTAs have 'chosen well', in the sense that most country pairs with FTAs tend to have the economic characteristics associated with considerable trade and with (in theory) welfare-enhancing net trade creation. Yet, the estimated probit functions in Baier and Bergstrand (2004a) have pseudo-R2 values of only 70%, still leaving considerable unobserved heterogeneity.

3.1.1 *Selection bias*

How does the unobserved heterogeneity in trade flow determinants matter? For instance, error term ν_{ij} in equation (10) may be representing unobservable (to the empirical researcher) policy-related barriers—tending to reduce trade between two countries—that are not accounted for by standard gravity equation RHS variables. As an example, suppose two countries have extensive unmeasurable domestic regulations (e.g., internal shipping regulations) that inhibit trade (causing X_{ij} to be low). The likelihood of the two countries' governments selecting into an FTA may be high if there is a large expected welfare gain from potential bilateral trade creation if the EIA deepens liberalization beyond tariff barriers into domestic regulations (and other nontariff barriers). Thus, EIA_{ij} and the intensity of domestic regulations may be positively correlated in a cross-section of data, but the gravity equation error term ν_{ij} and the intensity of domestic regulations may be negatively correlated. This reason suggests that EIA_{ij} and ν_{ij} are negatively correlated, and the EIA coefficient will tend to be underestimated.

In support of this argument, numerous authors have noted that one of the major benefits of regionalism is the potential for 'deeper integration'. Lawrence (1996, p. xvii) distinguishes between 'international policies' that deal with border barriers, such as tariffs, and 'domestic policies' that are concerned with everything 'behind the nation's borders, such as competition and antitrust rules, corporate governance, product standards, worker safety, regulation and supervision of financial institutions, environmental protection, tax codes ...' and other national issues. The General Agreement on Tariffs and Trade (GATT) and World Trade Organization (WTO) have been remarkably effective in the post-WWII era reducing border barriers. However, these institutions have been much less effective in liberalizing the domestic policies just named. As Lawrence states it, 'Once tariffs are removed, complex problems remain because of differing regulatory policies among nations (p. 7).' He argues that in many cases, EIA 'agreements are also meant to achieve deeper integration of international competition and investment' (p. 7). Gilpin (2000) echos this argument: 'Yet, the inability to agree on international rules or to increase international cooperation in this area has contributed to the development of both managed trade and regional arrangements' (p. 108; italics added). Preeg (1998) concludes:

> [Free] trade agreements over time, however, have tended to include a broader and broader scope of other trade-related policies. This trend is a reflection, in part, of the fact that as border restrictions [tariffs] are reduced or eliminated, other policies become relatively more important in influencing trade flows and thus need to be assimilated in the trade relationship (p. 50).

We believe this omitted variable (selection) bias is the major source of endogeneity facing estimation of FTA effects in gravity equations using cross-section data. Moreover, the arguments above suggest that policy makers' decisions to select into an EIA are likely related to the level of trade (relative to its potential level), and not to recent changes in trade levels. Thus, the determinants of FTA are likely to be cross-sectional in nature.

3.1.2 *Simultaneity bias*

Consider the potential endogeneity bias created by simultaneity. As discussed earlier, there exists a large empirical literature in international trade on the effects of multilateral tariff and nontariff barriers on multilateral trade volumes, and the simultaneous effects of these trade volumes on multilateral barriers using cross-industry and cross-country data for particular years, cf., Trefler (1993) and Lee and Swagel (1997). Simultaneity may be an issue for EIA_{ij} in cross-section gravity equations, motivated as in these two studies. For example, holding constant typical gravity equation RHS variables, two countries (say, the United States and China) that possibly trade more than their 'natural' level, as predicted by a typical gravity equation, may create political pressures to avoid trade liberalization or possibly raise trade barriers. This would cause a negative simultaneity bias in the EIA coefficient estimate. On the other hand, the governments of two countries that trade more than their gravity-equation-suggested 'natural' level might be induced to form an EIA because there might potentially be less 'trade diversion' due to their extensive trading relationship, suggesting a positive simultaneity bias. However, as just noted since the decisions to select into EIAs are likely influenced by the levels of trade relative to natural levels, recent changes in trade levels are not likely to influence EIA formations.

3.1.3 *Measurement error*

Measurement error in an explanatory variable, such as an EIA dummy, is generally associated with negative bias (in absolute terms) in the variable's coefficient. For instance, with the classical 'errors-in-variables' assumption, the 0-1 EIA dummy variable would be correlated positively with the measurement error if the true trade-policy variable (say, the tariff rate) was assumed uncorrelated with the measurement error. In equation (10)'s context, the correlation between EIA_{ij} and the error term ν_{ij} would be negative, leading to the classical 'attenuation bias' of *EIA*'s coefficient estimate toward zero. This may be part of the reason—but neither the entire, nor even the most important, reason—*EIA* coefficient estimates have been underestimated.

Of course, the best method for eliminating this bias is construction of a continuous variable that would more accurately measure the degree of trade liberalization from various EIAs. If EIAs only eliminated bilateral tariff rates, one would ideally measure this liberalization with a change in the *ad valorem* tariff rate (for which data are poor). However, EIAs liberalize trade well beyond the elimination of tariffs. Calculation of such measures is beyond the scope of this particular study, but is a useful direction for future research. Our goal rather is to discuss reliable estimates of the treatment effect of an EIA, similar to the 0-1 variable representing program participation in empirical labor economics. Thus, I constrain the discussion here to more accurate estimation of the *ex post* partial ‘average treatment effect’ (ATE) of an EIA dummy on trade flows, as has been used in the gravity equation literature for five decades.

3.2 Estimation using cross-section data for a single year

With cross-section data, standard econometric techniques to address omitted variables (and selection) bias include estimation using instrumental variables (IV) and Heckman control functions. Alternatively, with panel data, fixed effects and first differencing can be used to treat endogeneity bias; we discuss panel approaches in the next section.

Baier and Bergstrand (2002) was the first article to follow in the spirit of Trefler (1993) to apply IV to account for the endogeneity of EIAs in estimating their effect on trade flows. The first stage of the approach is to estimate the likelihood of a pair of countries having an EIA; this can be done using probit, logit, or linear probability functions. Several studies have used probit functions based upon Baier and Bergstrand (2004a) to estimate the probability of an EIA. Although most studies have found that the probit estimates provide ‘good’ predictions for the first stage of the estimation, the problem lies in ‘identification’ for the second stage (i.e., the ‘exclusion restriction’). As with any IV application, there needs to be at least one variable explaining the probability of a pair having an EIA that does *not* also influence trade flows. Herein lies the practical problem of using IV to alleviate the endogeneity bias.

Various authors have had alternative success in identification for the second-stage regression. Baier and Bergstrand (2002, 2004b) tried various IV and Heckman control function approaches to account for endogenous EIAs. Before trying IV, Baier and Bergstrand using ordinary least squares (OLS) found an ATE for EIAs (defined there as free trade agreements and deeper EIAs only) of 34% using trade flows for year 2000 (with 7302 observations). The authors then tried several two-stage procedures. One way to achieve identification in the second stage using probit in the first stage derives from the nonlinearity of the probit function. Baier and Bergstrand (2004b) found that with such identification an EIA had a negative and statistically significant effect on trade flows. Another IV estimate used a first-stage probit including the standard gravity-equation covariates and also a measure of remoteness, a measure of bilateral absolute differences in capital–labor ratios, and a measure of the difference of the pair’s capital–labor ratio relative to that of the *ROW*. The second-stage EIA coefficient estimate was economically and statistically insignificant. Moreover, the 'identifying' variables had often been included in gravity equations in the past and may be correlated with the gravity-equation error term. Baier and Bergstrand (2004b) also obtained first-stage probit estimates using political variables. The resulting second-stage EIA coefficient estimate was also economically and statistically insignificant.

Baier and Bergstrand (2004b) also tried linear probability models for the first-stage estimates. For the first-stage regression using the gravity covariates and economic identification variables (remoteness and two capital–labor variables), the second-stage EIA coefficient estimate of 2.51 implied that the partial ATE of an EIA was to increase bilateral trade by over *1100%*. Using political variables in the first-stage instead yielded in the second stage an ATE of 733%. One of the benefits of using a linear probability model is the econometrically feasible inclusion of fixed effects in the first stage. Using the previously described economic variables in the first stage with fixed effects led to a statistically insignificant second-stage EIA coefficient estimate of 0.41; however, this implied an economically plausible 51% EIA partial effect. By contrast, using the political variables in the first stage including fixed effects, the second-stage EIA coefficient estimate was a statistically significant −3.97, implying the EIA lowered trade by 92%. The vast differences in the second-stage EIA coefficients led Baier and Bergstrand (2004b, 2007) to conclude that IV approaches yielded unstable results, likely owing predominantly to the inability to find economic or political variables that satisfied the 'exclusion restriction' with confidence. Another author that has tried to use IV similarly with mixed results is Magee (2003).[3]

More recently, Egger et al. (2011) accounted for the endogeneity of EIAs in a cross-section analysis of EIAs while also accounting for zeros in trade flows and also estimating general equilibrium—alongside partial equilibrium (ATE)—effects, combining in one analysis the insights of Baier and Bergstrand (2002, 2007), Helpman et al. (2008), and Anderson and van Wincoop (2003). Using trade flows from year 2000 applying IV as above, they found that not accounting for self-selection of country pairs into trading and into EIAs (by using a first-stage bivariate probit model) led to a downward bias of 75% in the EIA's effect, with 45 percentage points of the bias attributed to the endogeneity of EIAs. The

[3] An alternative method for estimating the ATE of EIAs uses Heckman's control-function approach. Baier and Bergstrand (2002) estimated similar specifications using this alternative approach with qualitatively similar findings; the control-function approach does not solve the endogeneity bias issue either. The likely problem is this: the vast number of variables that are correlated cross-sectionally with the probability of having an FTA are also correlated with trade flows, preventing elimination of the endogeneity bias using cross-section data.

bias attributed to ignoring general equilibrium effects was minor, explaining only 4 percentage points of the downward bias.

3.3 Estimation using panel data

Given the problems associated with accounting for endogeneity of EIAs using instrumental variables and cross-section data, Baier and Bergstrand (2007), or BB, argued that a better approach to eliminate endogeneity bias of EIAs is to use panel techniques. In the context of the theory and endogenous self-selection of country pairs into EIAs, BB argue that one method to obtain consistent estimates of the ATE of EIAs is by fixed effects estimation of:

$$\ln X_{ijt} = \beta_0 + \beta_1 EIA_{ijt} + \eta_{ij} + \delta_{it} + \psi_{jt} + \epsilon_{ijt} \tag{11}$$

where η_{ij} is a country-pair fixed effect to capture all time-invariant unobservable bilateral factors influencing nominal trade flows, and δ_{it} and ψ_{jt} are exporter-time and importer-time fixed effects, respectively, to capture time-varying exporter and importer GDPs as well as all other time-varying country-specific unobservables in i and j influencing trade, *including* the exporter's and importers' 'multilateral price/resistance' terms (cf., Anderson and Wincoop 2003). We refer to this as the Fixed-Effects (FE) specification. It is important to note that, in most gravity-equation applications using a comprehensive set of RHS variables, the vast bulk of 'bilateral' trade-cost variables are time invariant, such as bilateral distance, common border, common language, etc. For instance, in Helpman et al. (2008), the only time-varying bilateral trade-cost variables are their EIA dummy and a dummy for the presence or absence of a currency union. As explained in BB, using panel data ATEs of EIAs (represented by β_1) estimated using equation (11) are likely to be consistent and precise. BB showed that the ATE of the typical EIA on nominal trade flows was 0.76, implying that the typical EIA increased bilateral trade by 114% after 10 years.

BB also used an alternative specification using first-differencing:

$$\Delta \ln X_{ij,t-(t-5)} = \beta_0 + \beta_1 \Delta EIA_{ij,t-(t-5)} + \delta_{i,t-(t-5)} + \psi_{j,t-(t-5)} + \upsilon_{ij,t-(t-5)} \tag{12}$$

We refer to this as the First-Difference (FD) specification. Note that the bilateral country-pair fixed effects are eliminated; however, the exporter-time and importer-time fixed effects are retained to capture time-varying exporter and importer GDPs and multilateral price terms. The latter effects were ignored in Foster et al. (2011), creating potential omitted variables bias.

Standard econometric discussions of treating endogeneity bias using panel data focus on a choice between estimation using the FE and FD specifications, cf., Wooldridge (2002, Ch. 10). As Wooldridge notes, when the number of time periods (T) exceeds two, the FE estimator is more efficient under the assumption of serially uncorrelated error terms ϵ_{ijt}. The FD estimator is more efficient (when $T > 2$) under the assumption that the error term ϵ_{ijt} follows a random walk (i.e., that the error term $\upsilon_{ij,t-(t-5)} = \epsilon_{ijt} - \epsilon_{ij,t-5}$ is white noise).[4]

First-differencing the panel data yields some potential advantages over fixed effects. First, it is quite plausible that the unobserved factors influencing the likelihood of an EIA (say, trade below its 'natural' level) are likely slow moving and hence serially correlated. If the ϵ_{ijt} are highly serially correlated, the inefficiency of FE is exacerbated as T gets large. This suggests that differencing the data will increase estimation efficiency for our large T panel. Second, aggregate trade flow data and real GDP data are likely 'close to' unit-root processes. Using FE is equivalent to differencing data around the mean (in BB's sample, year 1980); this may create a problem, as T is large in our panel. As Wooldridge (2000, p. 447) notes, if the data follow unit-root processes and T is large, the 'spurious regression problem' can arise in a panel using FE. FD yields data that deviates from the previous period of our panel, and thus is closer to a unit-root process. Consequently, the preferred estimation technique in BB and Baier et al. (2011) is the FD approach.

One of the other potential contributions of BB's panel methodology was to show that the full impact of EIAs on trade flows took 10–15 years. One reason is that most EIAs are 'phased-in' over 5–10 years. The second reason is the lagged effect of the trade-cost changes (such as terms-of-trade changes) on trade flows. As in BB, using a panel allows for differentiating the shorter-term effects (5 years) from the longer-term effects (5–15 years). Using the FD specification, BB found that the ATE of an EIA (FTA or higher degree of economic integration) was 0.61, imply an 84% increase in trade after 15 years.

While such positive ATE estimates for EIA dummy variables were interpreted in the context of either Armington or Krugman models as EIAs increasing trade volumes of existing homogeneous firms (the 'intensive margin'), consideration of zeros in bilateral trade, fixed export costs, and firm heterogeneity have led researchers more recently to examine various 'extensive margins' of trade. Such extensive margins fall under three general categories: country, goods, and firm. The existence of zeros in aggregate bilateral trade flows among many country pairs has

[4] When the number of time periods is exactly two ($T = 2$), estimation with FE and FD produce identical estimates and inferences; then, FD is easier to estimate. When $T > 2$, the choice depends upon the assumption one wants to make about the distribution of the error term ϵ_{ijt}.

led some researchers to explore the probability that a pair of countries trades at all; to the extent that an EIA affects this probability, this changes the *country* extensive margin of trade and potentially economic welfare.

However, the few empirical studies to date using gravity equations for a large number of country pairs and EIAs to examine extensive- and intensive-margin effects of EIAs have led to two puzzling results. First, two studies have used cross sections for a single year. Helpman et al. (2008), or HMR, found evidence using a cross-section and a two-stage estimator that EIAs influenced the country extensive margin, but had *no* significant effect on the intensive margin of trade (for existing firms). Egger et al. (2011) explored the country extensive and intensive margins also using a cross-section and a two-stage estimator and found in contrast a significant positive effect of EIAs on the intensive margin in their preferred specification, but no extensive margin effect.[5] The absence of country extensive-margin effects from an EIA suggests that trade liberalization does *not* lead to increases in varieties of goods from new trade partners, a potentially important source of welfare gains. The opposite EIA effect finding of the two articles is a puzzle.

A second margin is known as the 'goods' margin of trade. Hummels and Klenow (2005), or HK, introduced this notion by examining zeros in bilateral trade flows at *highly disaggregated product-category levels*.[6] The motivation for HK was to explore in a cross-section a fundamental question: Do large economies export more because they export larger quantities of a given good (intensive margin) *or* a wider set of goods (extensive margin)?[7] They found in their cross-section that 60% of larger exports of large economies was attributable to the extensive margin of 'goods' trade; specifically, as the exporter country's economic size grew, it exported a larger number of product categories (or 'goods') to more markets. The finding also that larger economies import more goods from more partners is important because it suggests an *improvement in welfare* owing to the consumption of more varieties; yet, 40% of the increased trade was still explained by the intensive margin in this cross-section. However, HK did not explore the relationship between trade liberalizations and the intensive and extensive goods margins of trade.

Only three empirical studies have explored the effect of trade liberalizations—and, in particular, EIAs—on the intensive and extensive goods

[5] They also evaluated whether their results were biased by omitting firm-heterogeneity, but concluded that firm-heterogeneity had no significant effect (also in contrast with HMR). Their preferred specification accounted for endogeneity of EIAs.

[6] By contrast, both HMR and Egger et al. (2011) used only bilateral *aggregate* trade flows to determine zeros in trade.

[7] Each 'good' was a 6-digit SITC category. They also explored the effects of country size and per capita GDP on the quality of goods exported, as well as the two margins.

margins of trade using the HK methodology. The earliest study using the HK decomposition to explore this issue is Hillberry and McDaniel (2002), focusing solely on the North American Free Trade Agreement (NAFTA). Although they do not attempt to establish causal effects from NAFTA to trade increases, they provide a decomposition of post-NAFTA trade among the three partners into goods intensive and extensive margins using 4-digit Standard International Trade Classification (SITC) data. They find evidence of both margins changing between 1993 and 2001. Kehoe and Ruhl (2009) examined NAFTA, the earlier Canada–US FTA trade liberalization, and some structural transformations using a modified version of the HK decomposition methodology and applied to a series of cross-sections. Similar to Hillberry and McDaniel (2002), they do not conduct an econometric analysis trying to explain the effect of NAFTA (or the Canada–US FTA) on trade flows conditional on other variables. They decompose actual goods extensive- and intensive-margin changes post-agreement also using 4-digit SITC data for goods categories from Feenstra et al. (2005). They find significant evidence of both extensive and intensive margin changes using their modified HK decomposition methodology. Both studies' evidence of goods intensive and extensive margins of trade expanding following the signing of NAFTA suggests the *need* for a comprehensive *econometric* analysis (conditional on other covariates) of the effects of EIAs in general on the goods intensive and extensive margins of trade, in the spirit of HK's original analysis of the effect of country size and per capita GDP on the two goods' margins.[8] However, the one panel study that did such a comprehensive analysis—Foster et al. (2011)—examined only short-run (5-year window) EIA effects motivated by a traditional gravity equation (ignoring multilateral price/resistance terms); they found economically small extensive margin effects and no intensive margin effects.[9]

In the context of the recent developments in the trade literature emphasizing intensive versus extensive margin effects, the panel approach in BB allows for differential timing of these trade-margin effects. In reality, one would expect that the intensive margin would be affected by a trade-cost change sooner than the extensive margin, because intensive margin changes in volumes do not require any start-up costs. Such costs—critical to the extensive margin—may delay the entry of new firms into exporting,

[8] Using a methodology similar to HK for estimating the goods extensive margin, Feenstra and Kee (2007) provided an econometric analysis of the effect of NAFTA on the extensive margin of Mexico's exports to the United States; they found a significant effect of NAFTA's reduction in tariff rates on this margin.

[9] With the short window, the authors could not address short-run versus long-run effects, likely missed phase-in and lagged terms-of-trade effects, and did not distinguish between alternative types of EIAs (in terms of depth of economic integration).

and thus we should expect the intensive margin to be influenced in the shorter term and the extensive margin in the longer term, as the results in Bernard et al. (2009) show. The panel data approach allows for evaluating this issue.[10]

BB did not estimate differential effects of various types of EIAs (in terms of depth of integration) on trade flows. Magee (2008) and Roy (2010) using the methodology of BB found that trade flows were impacted by larger amounts for customs unions relative to FTAs. However, no empirical study has until recently examined the differential impact of FTAs relative to deeper EIAs on the extensive *versus* intensive margins—much less the differential *timing* of such effects.[11]

Baier et al. (2011) recently addressed the effects of EIAs on the ‘goods’ extensive and intensive margins of trade. First, they extended Baier and Bergstrand (2007)’s panel econometric methodology for the (partial) effects of EIAs on aggregate trade flows using a gravity equation to examine in a large country-pair setting the effects of virtually all EIAs on the extensive and intensive goods margins, using the HK trade-margin-decomposition methodology. In the context of an econometric analysis, they are the first to find economically and statistically significant EIA effects on both the intensive *and* extensive (goods) margins in the context of a large number of country-pairs, EIAs, and years, in contrast to HMR and Egger et al. (2011).

Second, Baier et al. (2011) allowed for various *types* of EIAs—one-way preferential trade agreements (OWPTAs), two-way preferential trade

---

10 These differential timing effects were ignored in Foster et al. (2011). As discussed earlier, two recent theoretical articles suggest a reason for the low trade-cost elasticity of trade flows in macroeconomic analyses using time-series data and the relatively higher trade--cost elasticities of trade in cross-sectional trade analyses. Ruhl (2008) explains this puzzle by noting that the macroeconomic time-series approach is estimating the intensive-margin effect of trade, whereas the trade literature’s cross-sectional approach is capturing the extensive-margin effect, due to export fixed costs for new producers delaying trade effects and entry. In a complementary approach, Arkolakis et al. (2011) present a demand-oriented staggered-adjustment “Calvo-pricing” approach to explain the lower time-series elasticity in terms of solely an intensive-margin effect, and the higher long-run cross-section trade-cost elasticity capturing the longer-term extensive-margin elasticity as well.

11 It is useful to note here a parallel literature examining the effect of GATT and/or WTO membership on trade flows. For brevity, we note that there now appears little convincing evidence of substantive GATT/WTO effects on trade, once one accounts for EIA dummies, multilateral resistance, and unobserved country-pair fixed effects (as we do here). This is the conclusion of Eicher and Henn (2011) (though they found a nontrivial WTO ‘terms-of-trade’ effect) and of Felbermayr and Kohler (2010) who examined possible extensive-margin effects; Eicher and Henn (2011) ignored extensive- versus intensive-margin effects. We also note an issue raised in Martin and Ng (2004), which is the role of multilateral tariff reductions under the GATT/WTO. Most-Favored-Nation (MFN) tariff cuts could also be affecting results. However, such MFN tariff cuts by country would be accounted for by the exporter-time and importer-time fixed effects.

agreements (TWPTAs), FTAs, and a variable for customs unions, common markets, and economic unions (CUCMECU)—and they decomposed trade flows into extensive and intensive margins using the HK methodology.[12] While two recent studies have adapted the Baier–Bergstrand methodology for estimating the effect of differing 'types' of EIAs on bilateral aggregate trade flows, no econometric studies had examined the effect of various types of EIAs on the (goods) extensive and intensive margins of trade using a large number of country pairs and EIAs.[13] Neither HMR nor Egger et al. (2011) distinguished among various types of EIAs in their analyses of country intensive and extensive margins; also, Foster et al. (2011) used only a single EIA dummy. Baier et al. (2011) find not only that deeper EIAs have larger trade effects than FTAs, and the latter have larger effects than (partial) two-way and one-way PTAs, but they distinguished between these various trade effects at the extensive and intensive margins using a panel of (disaggregate) bilateral trade flows from 1962 to 2000 covering 98% of world exports.

Third, Bernard et al. (2009) is likely the only empirical study to date to explore the 'timing' of extensive and intensive margin responses to shocks. Using cross-sectional variation to examine long-run aspects, Bernard et al. (2009) find that variation in trade flows across country pairs is explained largely by the extensive margin, using firm-level data (the 'firm' margin); this result is consistent with HK using their 'goods' margin. But using *time-series* variation, Bernard et al. (2009) find that a larger proportion of trade variation can be explained by the intensive margin at short (5-year) time intervals. They show that, following the Asian financial crisis of 1997, virtually all of the variation in trade flows within 2–3 years could be explained by the *intensive* margin. This finding is consistent with two recent theoretical studies arguing that the low trade-cost elasticity found in macroeconomic analyses of business cycles should be associated with the intensive margin of trade compared with the relatively higher trade-cost elasticity found in international trade, which reflects extensive margin effects.[14] Allowing for differential 'timing' of EIA effects using panel data, Baier et al. (2011) find the first comprehensive empirical

12 Owing to few observations on common markets and economic unions, they combined these two types of 'deeper' EIAs with customs unions to form the variable CUCMECU, representing 'deep' EIAs.

13 The two studies that extended the Baier–Bergstrand framework to differing types of EIAs are Magee (2008) and Roy (2010); both found that customs unions had larger trade effects than FTAs. However, neither study examined extensive versus intensive margin issues.

14 Ruhl (2008) explains the delayed effect of the extensive-margin effects to fixed export costs on the supply side, while Arkolakis et al. (2011) explain the delayed effect of the extensive-margin effects to "Calvo-pricing" by consumers on the demand side.

evidence that the shorter-term effects of EIAs on trade flows are more at the (goods) intensive margin and longer-term effects are more at the extensive margin (the latter entailing either fixed export costs or staggered 'Calvo' pricing by consumers), consistent with recent theoretical studies and empirical results in Bernard et al. (2009).

While the articles discussed in this section have addressed how to better estimate with consistency and precision EIA coefficient estimates—such as $\psi(1-\sigma)$ in equation (10)—in the presence of endogeneity, they have not addressed how to identify σ specifically. However, future work may want to pursue a combination of the issues raised in sections 2 and 3 to better identify consistently and precisely σ in the presence of imperfect endogenous trade-policy measures.

4 Conclusion

Despite widespread anecdotal evidence that lower trade barriers increase international trade, there is little firm quantitative evidence of the 'trade-cost elasticity' of trade flows, one of two key aggregate statistics that have recently been identified as sufficient to quantify the economic welfare effects of trade-policy liberalizations and/or trade-cost reductions (the other statistic being the import-penetration ratio). In other words, most estimates of the trade-cost elasticity are imprecise and lack consistency. In this article, we discussed two issues that are critical in better assessing empirically the trade-flow and welfare effects of trade liberalizations (or trade-cost changes). The first issue was how to quantify the trade-cost elasticity when trade costs themselves are approximated imperfectly. Various articles have suggested methods to better estimate with consistency and precision the trade-cost elasticity when the RHS variables can be treated as 'exogenous'. The second issue was that typical empirical evaluations to estimate the impact of trade-policy liberalizations on trade flows use the 'gravity equation'. However, the self-selection of country pairs into such agreements introduces endogeneity bias in the estimation of the trade-cost elasticity in gravity equations, requiring better identification techniques. Various articles have suggested methods to better identify the effects of EIAs on trade flows, adjusting for self-selection bias. Future work may want to pursue a combination of these two methodological issues to better identify consistently and precisely σ in the presence of imperfect endogenous trade-policy measures.

Acknowledgments

I am grateful to Volker Nitsch and Nikolaus Wolf for the invitation to write this article and to CESIfo for their financial support. Much of the work cited in here is related to research with Scott Baier, Peter Egger, Michael Feng, and Mario Larch, and I am grateful for their collaborations.

References

Anderson, J. E. (1979), "A Theoretical Foundation for the Gravity Equation", *American Economic Review* **69**, 106–116.

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

Arkolakis, C., A. Costinot and A. Rodriguez-Clare (2012), "New Trade Models, Same Old Gains?", *American Economic Review* **102**, 94–130.

Arkolakis, C., J. Eaton and S. Kortum (2011), "Staggered Adjustment and Trade Dynamics", Working paper.

Baier, S. L. and J. H. Bergstrand (2001), "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity", *Journal of International Economics* **53**, 1–27.

Baier, S. L. and J. H. Bergstrand (2002), "On the Endogeneity of International Trade Flows and Free Trade Agreements, Manuscript (August)", University of Notre Dame.

Baier, S. L. and J. H. Bergstrand (2004a), "Economic Determinants of Free Trade Agreements", *Journal of International Economics* **64**, 29–63.

Baier, S. L. and J. H. Bergstrand (2004b), "Do Free Trade Agreements Actually Increase Members' International Trade? Manuscript (October)", University of Notre Dame.

Baier, S. L. and J. H. Bergstrand (2007), "Do Free Trade Agreements Actually Increase Members' International Trade?", *Journal of International Economics* **71**, 72–95.

Baier, S. L., J. H. Bergstrand and M. Feng (2011), "Economic Integration Agreements and the Margins of International Trade", Working paper.

Bergstrand, J. H. (1985), "The Gravity Equation in International Trade: Some Micro- Economic Foundations and Empirical Evidence", *Review of Economics and Statistics* **67**, 474–481.

Bergstrand, J. H., P. Egger and M. Larch (2013), "Gravity Redux: Estimation of Gravity-Equation Coefficients, Elasticities of Substitution, and General Equilibrium Comparative Statics Under Asymmetric Bilateral Trade Costs", *Journal of International Economics* **89**, 110–121.

Bernard, A. B., J. B. Jensen, S. J. Redding and P. K. Schott (2009), "The Margins of U.S. Trade", *American Economic Review* **99**, 487–493.

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

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

Egger, P., M. Larch, K. E. Staub and R. Winkelmann (2011), "The Trade Effects of Endogenous Preferential Trade Agreements", *American Economic Journal: Economic Policy* **3**, 113–143.

Eicher, T. and C. Henn (2011), "In Search of WTO Trade Effects: Preferential Trade Agreements Promote Trade Strongly, but Unevenly", *Journal of International Economics* **83**, 137–153.

Feenstra, R. (2004), *Advanced International Trade: Theory and Evidence*, Princeton University Press, Princeton, NJ.

Feenstra, R. C. and H. L. Kee (2007), "Trade Liberalization and Export Variety: A Comparison of Mexico and China", *The World Economy* **30**, 5–21.

Feenstra, R. C., R. E. Lipsey, H. Deng, A. C. Ma and H. Mo (2005), "World Trade Flows: 1962-2000, National Bureau of Economic Research", Working Paper No. 11040, January.

Felbermayr, G. and W. Kohler (2010), "Modelling the Extensive Margin of World Trade: New Evidence on GATT and WTO Membership", *The World Economy* **33**, 1430–1469.

Foster, N., J. Poeschl and R. Stehrer (2011), "The Impact of Preferential Trade Agreements on the Margins of International Trade", *Economic Systems* **35**, 84–97.

Friedman, T. (2005), *The World is Flat*, Farrar, Straus, and Giroux, New York.

Gilpin, R. (2000), *The Challenge of Global Capitalism*, Princeton University Press, Princeton, NJ.

Helpman, E., M. Melitz and Y. Rubinstein (2008), "Estimating Trade Flows: Trading Partners and Trading Volumes", *Quarterly Journal of Economics* **123**, 441–487.

Hillberry, R. and C. McDaniel (2002), "A Decomposition of North American Trade Growth since NAFTA, United States International Trade Commission", Working Paper 2002-12-A, July.

Hummels, D. and P. J. Klenow (2005), "The Variety and Quality of a Nation's Exports", *American Economic Review* **95**, 704–723.

Kehoe, T. J. and K. J. Ruhl (2009), "How Important is The New Goods Margin in International Trade?", Federal Reserve Bank of Minneapolis Staff Report 324 (June).

Krugman, P. R. (1980), "Scale Economies, Product Differentiation, and the Pattern of Trade", *American Economic Review* **70**, 950–959.

Lawrence, R. Z. (1996), *Regionalism, Multilateralism, and Deeper Integration*, Brookings Institution, Washington, DC.

Lee, J.-W. and P. Swagel (1997), "Trade Barriers and Trade Flows Across Countries and Industries", *Review of Economics and Statistics* **79**, 372–382.

Magee, C. S. P. (2008), "New Measures of Trade Creation and Trade Diversion", *Journal of International Economics* **75**, 349–362.

Mansfield, E. D. and E. Reinhardt (2003), "Multilateral Determinants of Regionalism: The Effects of GATT/WTO the Formation of Preferential Trading Arrangements", *International Organization* **57**, 829–862.

Martin, W. and F. Ng (2004), *A Note on Sources of Tariff Reductions in Developing Countries, 1983–2003*, World Bank, Washington, DC.

McCallum, J. (1995), "National Borders Matter: Canada-US Regional Trade Patterns", *American Economic Review* **85**, 615–623.

Melitz, M. J. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Profitability", *Econometrica* **71**, 1695–1725.

Preeg, E. H. (1998), *From Here to Free Trade*, University of Chicago Press, Chicago, IL.

Redding, S. J. (2011), "Theories of Heterogeneous Firms and Trade", *Annual Review of Economics* **3**, 77–105.

Roy, J. (2010), "Do Customs Union Members Engage in More Bilateral Trade than Free-Trade Agreement Members?", *Review of International Economics* **18**, 663–681.

Ruhl, K. (2008), "The International Elasticity Puzzle", Working paper.

Trefler, D. (1993), "Trade Liberalization and the Theory of Endogenous Protection: An Econometric Study of U.S. Import Policy", *Journal of Political Economy* **101**, 138–160.

Wooldridge, J. M. (2000), *Introductory Econometrics*, South-Western Publishing, Chicago, IL.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross-Section and Panel Data*, MIT Press, Cambridge, MA.