The Gains from Market Integration

Dave Donaldson

Department of Economics, Stanford University, Stanford, California 94305; email: ddonald@stanford.edu

Annu. Rev. Econ. 2015. 7:619–47

First published online as a Review in Advance on May 13, 2015

The *Annual Review of Economics* is online at economics.annualreviews.org

This article's doi:
10.1146/annurev-economics-080213-041015

JEL codes: F10, F60, O10, O18, O40, R10, R40

Keywords

international trade, intranational trade, trade costs, gravity models, transportation infrastructure

Abstract

How large are the gains from product market integration—or, equivalently, from a reduction in barriers to trade over space? This article surveys recent work on this question in the context of both international and intranational trade.

1. INTRODUCTION

The past two centuries have witnessed a dramatic change in the ability to trade goods—and even services—across and within national borders. Container megaships have replaced steamships, which replaced sailboats. Trucks on divided-lane expressways have replaced railroads, which replaced ox carts. E-mails have replaced telegrams, which replaced carrier pigeons. And now, unthinkably, in some locations the US Postal Service will even deliver packages on Sundays. Equally, waves of post–World War II multilateral and preferential trade agreements have eroded many of the tariff barriers that apply when trades cross international borders.

This liberalized mobility of goods and services across locations has given rise to a greater integration of the markets for these products at different points in space.[1] But what has been the resulting impact of this transformation on economic welfare? In this article, I describe recent work on this question. Researchers have made exciting progress in their ability to study how and why regions trade at all, the extent to which there are barriers that impede trade, and the gains from reductions in those barriers. This progress has advanced our understanding of the gains from market integration that have occurred in the past, as well as the potential for further gains in the future.

An important theme in this literature is that significant barriers to trade exist both between and within countries. Indeed, many of the technological innovations that have improved the ability to trade goods and services over the past 200 years could be argued to have affected intranational trade more than international trade. Much recent work exploits an essential symmetry between intra- and international trade to learn about the fundamental drivers of exchange among locations, whether those exchanges cross international borders or not. The review here therefore places equal emphasis on international and intranational settings. But care is typically required in extrapolating from one setting to another if the mobility of other determinants of trade, such as technology or factors of production, is differential within nations relative to across them.

I begin in Section 2 with a description of the gains from market integration as predicted by economic theory. Due to the complexities of modeling a multiregional trading environment in the presence of barriers to trade, the theoretical literature has converged on a series of modeling tricks to make progress. The resulting set of models, often referred to as gravity models, has become a standard workhorse in the fields of international and interregional economics. Recent decades have seen a sharpened understanding of the central features of gravity models, as well as how to estimate these models and apply them to answer counterfactual questions.

In Section 3, I describe some important applications of quantitative gravity models aimed at estimating the change in economic welfare that we should expect to see from improvements in market integration, or reductions in trade barriers. I also highlight some of the nascent work that explores potentially important departures from the assumptions inherent in gravity models.

Section 4 contrasts the lessons from quantitative trade models with the empirical estimates that emerge when researchers compare regions that have seen reductions in the trade barriers they face (often referred to as treatment regions) to other regions that did not (so-called control regions).[2] The use of such comparisons to isolate the resulting effects of the change in trade barriers is of course complicated by the presence of unobserved and yet potentially important determinants of the evolution of treatment and control regions. Recent research has therefore gone to great lengths

[1] These waves of globalization have seen similar progress, by and large, in the movement of people, capital, and technology, but owing to space constraints, the focus of this review is on product market integration.

[2] A related literature draws comparisons from the extent to which regions within a country (if factors are immobile across those regions, at least in the short run) are differentially exposed to a foreign shock (that affects industries differently) due to the differential industrial composition of these regions before the shock (see, e.g., Topalova 2010, Autor et al. 2013a, Kovak 2013).

to purge the variation, both across and within countries, of such confounding factors. However, a second potential concern is that the very act of treating the treatment region also affects the control region directly. This is especially worrying in the context of the questions posed here because the change under study—a reduction in trade barriers among a set of regions—is one that necessarily affects all of these regions, although perhaps differentially so, at the same time. As a result, the prospects for finding a truly unaffected group, which can act as a pure control, seem distant. The literature I describe in Section 4 is only beginning to assess the relevance of this potential concern. Finally, Section 5 concludes with a summary and thoughts on areas for potential future work.

Space constraints prevent me from providing a complete survey of all recent work in this vast area. Fortunately, this review is complemented by a number of outstanding recent surveys on related topics, such as gains from trade and quantitative gravity models (Harrison & Rodríguez-Clare 2010, Anderson 2011, Costinot & Rodríguez-Clare 2014, Head & Mayer 2014), the distributional consequences of trade openness (Winters et al. 2004, Goldberg & Pavcnik 2007, Harrison et al. 2011), and the consequences of market integration for firm and industry performance (de Loecker & Goldberg 2014), as well as the wider consequences of improvements in transportation infrastructure (Redding & Turner 2014). An additional related area concerns the estimation of the magnitude of barriers to trade, without particular emphasis on the effects of those barriers, as surveyed in Anderson & van Wincoop (2004) and Hummels (2007).[3]

2. GAINS FROM ECONOMIC INTEGRATION, IN THEORY

The potential for heterogeneous agents to gain from trading is at the heart of economic models of exchange. Models of international and interregional trade are then no different, with agents in distinct regions typically characterized by ex ante differences in their tastes, technology, or endowments.[4] I begin with a discussion in Section 2.1 of what is known about such gains in general economic environments and then move in Section 2.2 toward more tightly specified settings, so-called gravity models, that have proved useful for quantitative applied work.

2.1. General Results

A general approach to the question of gains from trade, at least as it concerns a neoclassical economy, is discussed in Dixit & Norman (1980). A central result is that in an economy that admits a representative agent, this agent is made weakly better off when a region leaves autarky and allows any amount of trade with other regions (and is made strictly better off if the agent has different preferences, technology, or relative endowments than agents have elsewhere).[5] This result

[3]Since the publication of Anderson & van Wincoop (2004), empirical progress has been made in understanding, for example, how spatial price gaps can be used to estimate trade costs (Simonovska & Waugh 2014a,b; Donaldson 2015), how information frictions can impede trade in a manner distinct from trade costs (Jensen 2007, Steinwender 2013, Allen 2014), how time delays act as a barrier to trade (Harrigan 2010, Hummels & Schaur 2013), how imperfect competition in the trading sector affects barriers to trade (Hummels et al. 2009, Mitra et al. 2013, Atkin & Donaldson 2014), and how novel aspects of the global trading environment such as containerization (Bernhofen et al. 2013) and piracy (Besley et al. 2015) can be important.

[4]There is also a rich tradition of models, synthesized in Helpman & Krugman (1985), in which trade and, typically, gains from trade occur even though agents are ex ante identical because increasing-returns-to-scale technologies allow agents to specialize ex post. Such models feature prominently below.

[5]Naturally, this focus on a representative agent is limiting and requires the applicability of the Kaldor-Hicks criterion (such that, within each region, those helped by integration compensate those harmed) (see, e.g., Dixit & Norman 1980). Owing to space constraints, I abstract from distributional considerations here, but readers are referred to the surveys described in Section 1 for a discussion of work in this area.

is a simple application of revealed preference arguments on both the consumption and production sides. If the barriers that impede trade are technological in nature (e.g., because these barriers result from a feature of the region's transportation infrastructure), then any reduction in these barriers enlarges the overall gains from trade, and this is likely to benefit all regions.[6] However, if the trade barriers in question are trade taxes that generate tax revenue for the taxing region, and the region in question is large enough that it can influence its terms of trade with outside regions, then, even though free trade is better than autarky, there is an intermediate positive value of the trade tax that is optimal for that region (while globally inefficient).[7]

The discussion above focuses on essentially bilateral considerations, in which there are only two regions. An important extension then is to the case of multilateral liberalization of trade barriers, which are the norm rather than the exception in many empirical applications (be they tariff liberalization or transportation infrastructure improvements). The complexities of multilateral environments have been the focus of a large literature on customs unions (a group of regions that share a common external trade policy and typically allow free trade within the union) and, more generally, preferential trade agreements (in which a region's trade policy discriminates between outside regions depending on whether they are a member of a common preferential trade agreement) (see, e.g., the recent survey in Freund & Ornelas 2010).

Two general lessons prevail from this literature, again with a distinction between trade barriers that generate revenue and those that do not. To simplify the discussion, consider three regions: A, B, and C. First, as concerns non-revenue-generating trade barriers (such as poor transportation infrastructure), in neoclassical settings, region A is unlikely to be made worse off by an improvement in κ_{AB} (letting κ_{ij} denote the cost of trading from region i to j induced by the non-revenue-generating barrier between i and j) or κ_{AC}, or indeed by κ_{BA} or κ_{CA}. The reason is that these improvements can be thought of as a strict enhancement in region A's technology set, and this cannot harm a neoclassical economy, holding terms-of-trade effects constant (and, as argued in footnote 6, a country's terms of trade are unlikely to deteriorate when its own transportation costs fall). However, an improvement in κ_{BC} or κ_{CB} could in principle harm region A through terms-of-trade effects.[8]

Second, consider now trade barriers that generate revenues and let t_{ij} denote such a barrier. Here there is no presumption that any reduction in, for example, t_{AB} would make region A better off. The reason is that, once again, there is no guarantee that the reduction in t_{AB} does not deteriorate region A's terms of trade. This point has important consequences for the empirical literature evaluating preferential trade agreements, such as NAFTA, in which the reduction in tariffs under study is discriminatory, as considered below.

This discussion of bilateral and multilateral settings suggests that not much can be said, in general, about the gains from reductions in trade barriers. That is, letting W_i indicate the level of welfare of the representative agent in region i, the sign of neither $\mathrm{d}W_i/\mathrm{d}\kappa_{jk}$ nor $\mathrm{d}W_i/\mathrm{d}t_{jk}$ is known

[6]The extreme setting in which a reduction in a region's transportation costs would harm that region, via adverse terms-of-trade effects, would require that region to be so large (relative to all other regions) that its increased trade due to lower transport costs would change the world price by more than the costs saved in transportation.

[7]Transport infrastructure generates revenue in the form of payments to the firms that provide transportation, but unlike tariff revenue, agents use real resources to produce these transportation services. Assuming that transportation is competitively provided, these payments equal the marginal social cost of providing the services. The worst type of trade barrier would therefore be one in which the barriers create rents and agents engage in rent-seeking behavior to attempt to capture those rents. Some recent work has explored this idea empirically (see, e.g., Schmitz 2012, Khandelwal et al. 2013, Sequeira 2014).

[8]For example, if regions A and B are identical economies apart from the fact that $\kappa_{AC} < \kappa_{BC}$, then region A benefits from trading with region C more than region B does, but then if τ_{BC} were to fall, region A would lose some of its advantage in market C relative to region B and would be made worse off as a result.

in general. This should not be surprising given that we have placed no restrictions on technology, preferences, endowments, or the level of trade barriers, other than a restriction to neoclassical economies featuring a representative agent. Reactions in the recent literature to this theoretical indeterminacy lie along a spectrum between two extremes. At one end, theoretical work has sought to place further (but ideally still realistic and testable) restrictions on the primitives of the economy under study. But, even then, the typical settings considered are sufficiently complicated that qualitative results cannot be obtained for most questions of interest, so researchers using this approach have sought to estimate the underlying structural parameters in their models and hence provide quantitative theoretical answers to counterfactual questions. I briefly discuss these models in the remainder of this section, as well as their quantitative applications in Sections 3.1 and 3.2. At the other end of the spectrum is an empirical body of work that, in response to theoretical indeterminacy, has aimed to estimate the causal effects of κ_{jk} or t_{jk} on W_i by seeking out quasi-experimental variation in these variables and simply letting the data speak. In practice, this empirical literature has been forced to deal with the high-dimensional complexities of multilateral settings by invoking theoretical restrictions of some sort or another. Section 4 describes a number of studies that have attempted to proceed along these lines.

2.2. An Introduction to Modern Gravity Models

As discussed above, it has proven difficult to study the gains from trade in general neoclassical environments, even those featuring a representative agent. For this reason, the bulk of work in this area has placed significant restrictions on tastes, technology, trade costs, and market structure so as to arrive at a model of a multiregion economy that is at once general and tractable. Usually, these models are so-called gravity models, and I review this approach briefly here.

Following Arkolakis et al. (2009), a modern working definition of a gravity model (in a one-sector economy) is one in which the equilibrium value of total trade from region o to region d (indicating the origin and destination of each flow) takes the form

$$\ln X_{od} = A_o(\mathbf{X}, \boldsymbol{\tau}) + B_d(\mathbf{X}, \boldsymbol{\tau}) - \theta \ln \tau_{od}. \tag{1}$$

In this expression, X_{od} indicates the value of trade from region o to region d, τ_{od} indicates the total (ad valorem equivalent) cost of trading goods from region o to region d, and the terms $A_o(\mathbf{X}, \boldsymbol{\tau})$ and $B_d(\mathbf{X}, \boldsymbol{\tau})$ refer to all determinants of trade between regions o and d that are specific to the exporter o and importer d, respectively.[9] The terms $\mathbf{X}$ and $\boldsymbol{\tau}$ indicate that the terms $A_o(\mathbf{X}, \boldsymbol{\tau})$ and $B_d(\mathbf{X}, \boldsymbol{\tau})$ could depend in an unrestricted manner on the endogenously determined total vector of trade flows and on the total vector of exogenous trade costs in the world. Finally, the parameter θ is known as the trade elasticity for the key role it plays in Equation 1.

To appreciate the simplification that has been achieved here, relative to the general environment referred to in Section 2.1, note that Equation 1 mandates that the only way that trade costs outside of the od pair can affect trade within the od pair—even in full general equilibrium—is via the terms $A_o(\mathbf{X}, \boldsymbol{\tau})$ and $B_d(\mathbf{X}, \boldsymbol{\tau})$. This implies that trade flows take an irrelevance of irrelevant alternatives form. A further simplifying restriction is that the partial effect [i.e., the effect holding the terms $A_o(\mathbf{X}, \boldsymbol{\tau})$ and $B_d(\mathbf{X}, \boldsymbol{\tau})$ constant] of bilateral trade costs τ_{od} on bilateral trade flows X_{od} takes on a constant elasticity form, with the same elasticity, θ, for all pairs of regions. Although

[9]In general, $\tau_{od} = f(\kappa_{od}, t_{od})$ where, as above, κ_{od} represents trade costs (e.g., transportation costs) that do not generate tax revenue and t_{od} denotes trade costs (e.g., tariffs) that do, and the function $f(\cdot)$ is monotone in both arguments. In practice, much of the literature assumes that $f(\kappa_{od}, t_{od}) = \kappa_{od} t_{od}$.

these are strong simplifying assumptions, they have allowed a vast body of work to study the gains from trade in this class of models, as I describe below.

Equation 1 describes an endogenous relationship among trade flows and trade costs that must hold, by definition, in a gravity model. But what economic primitives would give rise to such a relationship? For decades, this gravity equation lacked any microfoundations, but recent work has dramatically changed that. Costinot & Rodríguez-Clare (2014) summarize this work by providing a set of sufficient conditions for a multiregion model to be a gravity model. On the demand side, almost all known gravity model microfoundations begin with constant elasticity of substitution (CES) utility across a continuum of horizontally differentiated varieties (in which the trade flow X_{od} above is then the total expenditure on all such varieties made in region o and sold in region d).[10]

Research on the supply side has demonstrated a greater love of variety. Prominent models that satisfy the sufficient conditions in Costinot & Rodríguez-Clare (2014) include the following: under perfect competition, either the model of Armington (1969), in which each region is endowed with the ability to be the unique global producer of an exogenously specified set of varieties, or the model of Eaton & Kortum (2002), in which each region's ability to produce a given variety is drawn independently from a Fréchet distribution with a common shape parameter; under monopolistic competition, models with free entry and fixed costs of production in which firm productivities are exogenous as in Melitz (2003) and drawn independently from a Pareto distribution as in Chaney (2008) or Eaton et al. (2011)[11]; and under Bertrand competition, models with firm productivities drawn from a Fréchet distribution as in Bernard et al. (2003).

With the gravity model defined and delineated in how it simplifies the analysis of interregional trade flows, an important question is what the gravity model implies for the gains from market integration. That is, in a gravity model, what are the welfare consequences (for a representative agent) of counterfactual reductions in trade costs? To answer this question, Arkolakis et al. (2012b) show that as long as trade in goods is balanced (an assumption that is standard in one-sector models but, as I discuss further below, introduces complications in extending this result to a multisector setting), a very simple formula describes the change in welfare from any arbitrary change in trade costs, from $\boldsymbol{\tau}$ to $\boldsymbol{\tau}'$, between any pair of regions.[12] Remarkably, in this class of models, the proportional change in welfare in region d (denoted ΔW_d) can always be written as

$$\Delta W_d = \Delta \lambda_{dd}^{-1/\theta}, \tag{2}$$

where λ_{dd} denotes the share of expenditure that region d spends on goods produced in region d. That is, whatever was the underlying change in the trading environment, a sufficient statistic for the change in welfare (in addition to an estimate of θ) is simply the change in the extent to which a region trades with itself.

[10]An exception is by Arkolakis et al. (2012a), who work with a class of demand relations that encompasses an additively separable but non-CES utility function as in Krugman (1979), a quadratic but nonseparable utility function as in Melitz & Ottaviano (2008), and preferences deriving from a translog expenditure function as in Feenstra (2003). With additional restrictions on the supply side (no fixed exporting costs and Pareto-distributed firm productivities), trade flows in these non-CES settings nevertheless take a gravity equation form.

[11]This class therefore subsumes the model of Krugman (1980) in which the firm productivity distribution is degenerate.

[12]In this class of models, a reduction in region o's costs of exporting τ_{od} to all locations d (including o itself) is isomorphic to an improvement in region o's productivity level. Thus, one can similarly use the procedure here to study the effects of foreign productivity improvements on local welfare, as in Hsieh & Ossa (2011). An analogous remark applies to changes in preferences.

One limitation of the result in Equation 2 is that, as it stands, it can be used only for ex post analysis—that is, only if one knows the resulting change in λ_{dd} that resulted from a particular event. However, one counterfactual question for which this limitation does not bind concerns the welfare costs of a move from current levels of trade costs $\boldsymbol{\tau}$ to autarky (where $\tau'_{od} = \infty$ for all $o \neq d$). Although this is clearly an unrealistic and extreme scenario, it offers the advantage of delivering a very simple expression for the change in welfare. Because $\lambda_{dd} = 1$ in autarky, the move to autarky requires $\Delta\lambda_{dd} = 1/\lambda_{dd}$, so there is no ambiguity about the magnitude of this cost: It is simply proportional to the extent to which a region is currently trading, with the proportionality given by $1/\theta$.

A stronger ex ante result is possible, however. Building on a procedure pioneered by Dekle et al. (2007), Arkolakis et al. (2012b) show that by imposing one additional restriction—that any fixed costs of exporting are paid in the importing country—the effect of any change in trade costs on any endogenous variable (e.g., welfare in any region) can be computed as the solution to a simple system of nonlinear equations. Importantly, the exogenous elements of this system involve only data on all regions' current trade flows, and a given value for θ. This result is attractive for guiding quantitative work for several reasons. First, this procedure requires no knowledge of the underlying primitives of the model (e.g., region-specific taste shifters, productivity shifters, or fixed or variable trade costs). All that is required of the analyst about such variables is already embodied, within this class of models, in the observable data $\mathbf{X}$.[13] Second, this procedure does not require knowledge of which particular microfoundations of the gravity model are at work. Conditional on sharing the same value for θ, any gravity model would agree on the result of any particular counterfactual exercise of the sort described here.[14] However, an important point about this equivalence result is that it does not apply if the analyst wishes to compare the results of a similar counterfactual across two different models with different values of θ or to models that do not belong to the (inevitably restrictive) gravity model class.[15]

2.3. Extensions to the Basic Gravity Model

I now briefly summarize some of the extensions to the basic gravity model in Section 2.2 that have appeared in the literature. Many of these elements have proven essential when taking gravity models to the data.

First, it is natural to extend the above approach to settings featuring multiple sectors. In practice, this is attractive if one expects the underlying primitives (tastes, technology, or trade costs) to differ across broad groups of the economy. Most multisector versions of the gravity model specify an environment in which production is separable across sectors (just as it is across

[13]In practice, it is important to bear in mind that this full vector of trade flows between all regions, $\mathbf{X}$, includes elements that cover trade from a region to itself, X_{oo}. When data on gross output are available by region, this can be used to compute X_{oo} as the difference between gross output and total exports to all other regions. But such data are not always available (especially at the sectoral level or in historical settings).

[14]Reassuringly, the exclusion restrictions required to estimate θ from moments based on an empirical version of Equation 1—that is, on how log trade flows react to exogenous changes in log trade costs, conditional on exporter and importer fixed effects—would be the same for all models in the class of gravity models. That said, as argued by Simonovska & Waugh (2014b), some procedures that aim to estimate trade costs from underlying micro price variation across locations may estimate different levels of trade costs depending on the underlying gravity model microfoundations.

[15]Melitz & Redding (2015) provide interesting quantitative examples of such cases, such as a comparison between a Krugman (1980) model without firm heterogeneity and a Melitz (2003) model with Pareto-distributed firm heterogeneity (in which case θ is not the same in these two models) and a comparison between a model with Pareto-distributed heterogeneity and one with log-normally distributed heterogeneity (in which case the latter model is not a gravity model).

varieties) and in which preferences take a particular functional form that is separable across sectors (the Cobb-Douglas functional form has proven most prominent). Costinot & Rodríguez-Clare (2014) survey this approach and describe how much of the above logic extends to the multisector case. A robust result is that adding multiple sectors raises the gains from trade, but this is no surprise when cross-sectoral preferences are Cobb-Douglas, as this mechanically lowers the substitutability between some pairs of varieties relative to the one-sector case with CES elasticity greater than unity. However, one important result that does not extend is the equivalence between models featuring a fixed set of varieties [e.g., the Armington (1969) and Eaton & Kortum (2002) members of the gravity model class] and models featuring love-of-variety preferences and free entry [e.g., the Krugman (1980) and Melitz (2003) settings]. Intuitively, the reason for this nonequivalence is that, with multiple sectors, even if factors are in fixed supply in total, their supply to each sector is anything but fixed. And, as is well known (see, e.g., Helpman & Krugman 1985), models with love-of-variety preferences and free entry are isomorphic to models with external economies of scale, so in this class of models, a sector's scale affects its aggregate productivity, whereas in models with a fixed set of varieties, there is no such effect. Costinot & Rodríguez-Clare (2014) conduct various calculations concerning the losses of a move to autarky, in models with multiple sectors and input-output (IO) linkages, under the two alternative assumptions that (*a*) the set of varieties is fixed within each sector and (*b*) there is love of variety and free entry within each sector. These calculations (which are reviewed below) demonstrate how the gains from trade can differ across these two types of gravity models, but the literature to date does not clarify when we should expect the gains from trade to be larger in one type of model or the other, let alone when each type of gravity model is more appropriate.

Second, the recent literature on multisector gravity models has been extended to the case in which sectors interact through IO linkages, and those linkages occur both within and between regions. This important feature of reality has been embedded into quantitative trade models in a number of different ways. The most common method, from Krugman & Venables (1995), is to imagine that production requires a Cobb-Douglas aggregator of input bundles from other sectors, with those input bundles themselves being a CES aggregator (with the same elasticity of substitution as in consumer preferences) of the varieties from each sector.[16] Caliendo & Parro (2015) and Costinot & Rodríguez-Clare (2014) show that this modeling trick ensures that the introduction of IO linkages does not interfere with the above results about gravity models. As discussed further below, in general, the presence of these IO linkages means that the harm done by trade barriers is significantly larger. This is intuitive given that, in these models, imported goods affect consumers' consumption bundles both directly (as some imported goods are consumed) and indirectly (as some imported goods are used to make domestic production cheaper).

Finally, the gravity model settings described here have been applied to trade among regions, both between and within countries. The only distinction between inter- and intranational trade in this literature is whether other drivers of trade (factors of production, tastes, and technology) are mobile, as we would expect them to be relatively more so within countries. Recent work on quantitative gravity models with different aspects of nongoods mobility includes Allen & Arkolakis (2014) and Redding (2013) on the mobility of workers, Eaton & Kortum (2001) on the mobility of capital goods, Eaton & Kortum (1999) on the diffusion of technology, and Ramondo & Rodríguez-Clare (2013) and Tintelnot (2013) on the mobility of technology embodied in multiregional firms.

[16]Other approaches, such as those in Yi (2003) and Johnson & Moxnes (2013), yield new insights about how the presence of global supply chains affects what can be learned from trade flows (e.g., how much borders affect trade costs). But these approaches do not yield an aggregate gravity equation, nor has the focus of this work been on the gains from integration.

In summary, a rich body of recent work has explored the efficacy of gravity models for understanding the effects of market integration under a wide range of auxiliary assumptions. Although these models are certainly restrictive, especially against the backdrop of the general framework of Section 2.1, the tractability of these models has greatly enhanced the connection of theory to data in this area, as I now discuss.

3. ESTIMATES FROM QUANTITATIVE TRADE MODELS

The gravity model approach, outlined in Section 2.2, has provided a new methodology for measuring the effects of economic integration. I turn now to a discussion of some of the quantitative estimates appearing in this literature, be they applications of gravity models (Sections 3.1 and 3.2) or not (Section 3.3).

3.1. Ex Ante Counterfactuals Based on Gravity Models

Eaton & Kortum (2002) pioneered and popularized the use of ex ante counterfactual experiments in the fields of international and interregional trade. Similar ex ante counterfactuals have been performed in a variety of settings and with the use of other members of the gravity model class. Particularly prominent examples include those by Eaton et al. (2011) and Balistreri et al. (2011).

A key contribution of Costinot & Rodríguez-Clare (2014) was to not only survey the literature on quantitative gravity models, but to also perform ex ante counterfactual calculations on a consistent data set of countries, under virtually every permutation of modeling assumptions (within the gravity model class) that have appeared in the literature.[17] Under a set of assumptions that is relatively agnostic about the data (i.e., with multiple factors of production, multiple sectors with Cobb-Douglas upper-tier preferences, and IO linkages between sectors used in Cobb-Douglas production functions), these authors calculate that, for the average country in their data set, a move to autarky would reduce welfare substantially—by between 27% and 40% depending on the particular member of the gravity model class one considers.

Turning to the study of economic integration both between and within nations, a natural instinct for the spatially oriented economist is to consider a model in which small spatial units are the primitives and countries are simply larger regions that aggregate the finer spatial primitives. Although this idea has a long tradition in the context of Canadian provinces and US states (Anderson & van Wincoop 2003), only recently have researchers begun to construct quantitative gravity models with high spatial resolution.

An important new paper in this area is that by Allen & Arkolakis (2014). These authors begin by characterizing a one-sector gravity model with perfect factor mobility.[18] They then introduce a new concept of geography to the literature by considering instantaneous trade barriers that apply when goods pass through each region; the total cost of trading from region *o* to region *d* is then the sum of all of the instantaneous trade costs that must be paid along the shortest path between *o* and *d* (where, of course, the location of the shortest path itself depends on all of the instantaneous trade costs). For example, a region could lie on a highway, which would mean that the

[17]Costinot & Rodríguez-Clare (2014) use data from the World Input-Output Database in 2008, comprising 40 countries and 31 sectors.

[18]In addition, Allen & Arkolakis (2014) allow for certain forms of negative congestion externalities in which an agent's utility of living in a region is affected directly (even holding prices constant) by the presence of others in that region, as well as analogous positive agglomeration externalities in which a region's productivity depends directly on the number of agents living there.

instantaneous cost of traveling over such a region is presumably relatively small. Allen & Arkolakis (2014) then apply their model to the United States, using counties as the regions in the model. Because (publicly available) interregional trade data are not disaggregated within the United States at such a fine level, these authors' quantitative estimates follow a procedure that differs somewhat from that described in Section 2.2, but they show how one can nevertheless construct quantitative ex ante counterfactual experiments in this context using aggregate data. As an application, they calculate that removing the US interstate highway system would lead to a welfare loss of 1.1–1.4%, a figure that substantially exceeds the annual costs of maintaining the system. As spatially disaggregated data become increasingly more available, the tools developed by Allen & Arkolakis (2014) are sure to see many applications.[19]

Costinot et al. (2015) provide a second recent example of a high–spatial resolution approach to ex ante counterfactual simulations of the effects of trade barriers. They ask whether the impact of climate change on agricultural production will be mitigated by the fact that agents can alter their consumption and production decisions as climate change affects relative productivities around the world. To study this question, these authors draw on the Food and Agriculture Organization's (FAO's) Global Agro-Ecological Zones (GAEZ) model, which aims to simulate how the production of any major crop will fare given local conditions (e.g., climate, soil, topography) at any of approximately 1.7 million grid cells on Earth. Importantly, GAEZ provides these simulations both for current climatic conditions and for those predicted by climatologists to occur because of climate change. Costinot et al. (2015) then embed the GAEZ-predicted technologies of each of these 1.7 million grid cells into an otherwise standard gravity model and calculate that world total welfare will fall by about one-sixth of current crop value. This number would be three times larger if producers were unable to adjust their production patterns, suggesting that production reallocation is an important form of adjustment to climate change. By contrast, constraining countries such that their export patterns cannot adjust has almost no impact on the welfare cost of climate change, which suggests that international trade is an unimportant vehicle for adjustment.

3.2. Additional Gravity-Based Counterfactuals

The discussion above focuses on the type of counterfactual exercise that has been performed most commonly in the literature—one concerning an ex ante scenario that has no counterpart in the data. But recent research has seen additional applications of the gravity model to extended settings. I give three examples here.

First, a common obstacle when studying the impact of trade liberalization on a country is the challenging reality that most instances of tariff liberalization have occurred simultaneously with other exogenous shocks to the country (e.g., political changes or other market reforms, as discussed in Goldberg & Pavcnik 2007). Mexico's entry to NAFTA in 1994 was no exception. For this reason, Caliendo & Parro (2015) use a multisector gravity model, with rich intersectoral IO linkages, to quantify the impact of NAFTA on trade flows and welfare. These authors feed the actual time path of Mexico-US bilateral tariffs into their gravity model to isolate the effect, according to the model, that NAFTA would have had in the absence of all other confounding factors. The estimated pure NAFTA effect on Mexico was to raise welfare by 1.3%.

[19]One example is by Desmet et al. (2014), who build a high-resolution spatial growth model and use it to study the impact of cross-country migration restrictions (as well as the effects of coastal flooding, due to global warming, in the presence of such a restriction).

Second, Eaton et al. (2013) aim to decompose the decline in world trade flows and welfare during the global recession of 2008–2009, during which the ratio of trade to GDP fell by 20%, into their underlying causes due to changes in trade costs, demand, and supply. They do so by first filtering the data through the lens of a multiperiod gravity model, so as to isolate shocks to each of these three forces. Eaton et al. (2013) then conduct a series of counterfactual simulations in which each of the various shocks are eliminated. Surprisingly, changes in trade costs appear to have almost nothing to contribute to the reduction in global trade according to these estimates.

Finally, a third example draws on work by Broda & Weinstein (2006). These authors aim to isolate a component of the change in US consumer welfare from 1972 to 2001 that results from the importation of new product varieties (with a variety characterized as a unique narrowly defined product from a unique country). Drawing on the methodologies suggested by Feenstra (1994), Broda & Weinstein (2006) estimate elasticities of cross-variety substitution within hundreds of aggregate industrial groups.[20] They go on to use the resulting estimates to calculate the consumer surplus generated by the fact that US consumers have been steadily importing more products from more countries, holding constant other margins of consumer and producer adjustment (e.g., the number of varieties produced domestically). Accounting for new imported varieties reduces a continuing-variety US import price index by an additional 28% between 1972 and 2001.

3.3. Nongravity Model Approaches

The gravity equation in Equation 1 explains interregional trade flows with remarkable parsimony. This has obvious advantages, and these advantages have given rise to the rich body of work described above (as well as guidance for some of the empirical work described in Section 4). But the strong simplifying assumptions embodied in Equation 1—namely, that a particular substitution matrix of international trade flows is (up to the scalar, $-\theta$) the identity matrix—have prompted recent work on alternatives.

One reaction has been to model consumer preferences in a more flexible manner than the CES formulation that is a typical gravity microfoundation. For example, Novy (2013) allows for symmetric translog consumer preferences in a study of bilateral trade patterns (but not the gains from trade). Recent work by Fajgelbaum & Khandelwal (2014) allows for nonhomothetic preferences (of the AIDS sort, following Deaton & Muellbauer 1980) and nondegenerate income distributions within each country so as to study the possibility that a nation's income inequality determines its trade flows and to allow for unequal (across the income distribution within a country) gains from international trade. A second reaction has been to use newly available firm-level microdata to model the supply side more flexibly. Recent examples of this include work by Melitz & Redding (2015) and Head et al. (2014), who consider monopolistic competition models of heterogeneous firms (facing CES preferences) but with those firms' productivity levels drawn from a log-normal distribution.

Relatedly, Costinot & Donaldson (2014) consider a departure from gravity in a simple Ricardian model setting. As is well known, the gains from trade in a Ricardian model arise because of specialization: When regions specialize according to their comparative advantages, gains from trade are realized. The magnitude of these gains hinges on just how bad a region would be at producing the goods it does not produce (because of the fact that it can import those goods), but an analyst would never be able to observe such an outcome (outside of autarky). Costinot &

[20]The estimation procedure of Broda & Weinstein (2006) actually allows for both constant elasticity import demand and export supply relations, which means that the resulting model is not a gravity model (but the gravity model is nested).

Donaldson (2014) argue that, because of this selection problem, conventional data cannot hope to nonparametrically estimate the gains from trade in a Ricardian model.[21] But they propose a setting, agriculture, in which scientific knowledge about the production function—how any crop would be likely to grow under a whole range of growing conditions—can be used to inform estimates of the gains from trade.

These authors show how agronomic information (from the FAO's GAEZ project), when combined with standard aggregate production statistics about output by crop, can be used to infer the prices that farmers in two or more regions appear to be facing as well as the gains that would accrue to those regions if the regions were to face more similar prices, that is, if market integration were to occur. They apply this methodology to the case of US agricultural production from 1880 to 1997, for which the availability of agricultural output data for a panel of counties from the US Census is particularly rich. As one might expect, US counties appear to have been facing agricultural prices that became increasingly aligned across space throughout the 118 years under study. But surprisingly, the estimated US-wide efficiency gains that resulted from this market integration process are substantial: approximately 0.5–1.5% per year compounding (depending on the assumptions made about whether intercounty price gaps reflect revenue-generating distortions or purely resources used for transportation). This is of the same magnitude as the impressive rate of technological progress (within counties and crops) in the same data and time period. Costinot & Donaldson (2014) lack the intercounty trade data that would be necessary to compare their estimates to gravity model–based estimates from a similar setting, but the estimates here suggest that, at least in this setting, the (nonparametrically estimated) gains from market integration can be substantial.

4. EVIDENCE BASED ON EMPIRICAL COMPARISONS

Section 3 discusses the many challenges of speaking with confidence about the effects of market integration using theory alone. Although reductions in trade barriers are guaranteed to raise welfare for at least one region, the study of any given region in general neoclassical models is particularly complicated by the reality of multiregion interactions. As described in Section 2, most tractable attempts to cut through this complexity have involved gravity models, which offer a suite of powerful analytical tools that calibrate a model to match a base cross section of trade patterns and then simulate, with relative ease, the impact of any counterfactual change in trade barriers that is desired. The challenge faced by this approach is credibility—that is, the largely unanswered question of whether the numerous strong modeling assumptions result in a model that can make successful predictions about changes in trade barriers. And, as reported in Section 3.3, nongravity approaches to these issues are still in their infancy.

Because of these challenges, a prominent instinct in the literature has been to let the data speak as freely as possible to the question of how a region's welfare level W_i is affected when trade costs τ_{jk} (at home or elsewhere) change. In doing so, researchers have searched for settings in which measured welfare in regions exposed to low trade barriers can be compared, empirically, to that in other regions exposed to higher trade barriers. In the program evaluation literature (see Imbens & Rubin 2015 for a textbook treatment), these would be referred to as treatment and control regions, respectively.

[21]This nonidentification result relates closely to the literature on the Roy model (Heckman & Honoré 1990) (see Costinot & Vogel 2015 in this volume for further discussion of the similarities).

As is well known, such comparisons identify the causal effect of the observed difference in trade barriers on the measured welfare of an isolated region under two strong conditions. First, it must be the case that the only systematic difference between the treatment and control regions, in terms of underlying determinants of welfare and after controlling for observable determinants, is the difference in trade barriers. This is known as the unconfoundedness requirement (Rubin 1990).[22] Second, it must be the case that the two regions do not interact—that the true treatment effect on a region is independent of whether any other region is treated. This is known as the stable unit treatment value assumption (SUTVA) requirement, as coined by Rubin (1980).

The literature has devoted a great deal of creative energy to the search for comparisons that satisfy the unconfoundedness requirement. Although identifying the treatment region is usually straightforward, finding a suitable control region for the treatment region in question is typically a substantial challenge. As I discuss further below, comparisons across countries and across regions within a country (based either on changes in intranational trade barriers or on changes in international barriers interacting with heterogeneous characteristics of internal regions such as internal geography) have featured prominently in the literature.

By contrast, the SUTVA requirement for the identification of causal effects has received far less attention. This is especially strange in the present context given that the question of interest—the effects of market integration—by its very nature involves interactions between regions. If SUTVA-violating spillovers between regions are negative, then empirical comparisons, which identify the relative effect of the treatment on the two types of regions, will overstate the true treatment effect. Likewise, if the spillovers are positive, then standard estimators will understate the true effect. Unfortunately, direct tests for the size and direction of such spillovers, in the context of this review, are in short supply, and more work is needed to understand the potential for bias due to SUTVA violations here.

4.1. Comparisons Across Countries

Although the discussion above has been largely agnostic about whether the trade barriers under study exist between regions that represent nations or regions that represent subnational geographic units, the earliest work on the impact of trade barriers focused almost entirely on international barriers. Countries were therefore the natural unit of analysis in these studies. I thus begin with a description of some of this work that aims to learn about the effects of market integration by drawing comparisons across countries.

An extremely influential study in this context is by Frankel & Romer (1999). These authors begin by positing the existence of a stable statistical relationship between the log of a country's real GDP (intended as a proxy for economic welfare for a representative agent, as it would be in the models described in Section 2 above) and a measure of its openness to international trade:

$$\ln(\mathrm{GDP}_d) = \beta\,\mathrm{OPENNESS}_d + \varepsilon_d, \tag{3}$$

where $\mathrm{OPENNESS}_d \doteq (\mathrm{EXPORTS}_d + \mathrm{IMPORTS}_d)/\mathrm{GDP}_d$ for region d. The goal is then to estimate the parameter β, the causal impact of one additional unit of $\mathrm{OPENNESS}_d$ on the log of

[22]Many of the studies reviewed below apply IV procedures that isolate particular comparisons that identify a treatment effect. In such settings, the reduced-form effect of the IV on the outcome variable is identified under the unconfoundedness and SUTVA assumptions, so for expositional simplicity, I refer to these two assumptions throughout. However, in IV settings, the unconfoundedness assumption is more commonly referred to as an independence assumption (which itself requires that the IV is randomly assigned and that an exclusion restriction—that the IV has no direct effect on the outcome variable—is satisfied) (see Angrist et al. 1996).

GDP_d, real (purchasing power parity–adjusted) per capita GDP.[23] It is important to note that (as stressed in Frankel & Romer 1999) the relationship in Equation 3 has nothing direct to say about the impact of trade barriers on economic welfare. And although it is plausible that the trade barriers a researcher is interested in evaluating (e.g., a reduction in tariffs or an improvement in transportation infrastructure) would change OPENNESS_d, the relationship between trade barriers and OPENNESS_d is typically not simple, and the literature is far from reaching a concrete understanding of it. This is true both in gravity model and in more general settings (as discussed in Sections 2.1 and 2.2).

Although the Frankel & Romer (1999) openness measure has intuitive appeal, it is not the case that Equation 3 would emerge as a structural relationship in standard models of international trade. However, Equation 3 is closely related to Equation 1. To see this, note that if trade is balanced (i.e., the value of total exports equals that of total imports), we have $\text{OPENNESS}_d \doteq 2(\text{IMPORTS}_d/\text{GDP}_d)$. After integrating and adding an error term, ε_d, that subsumes, among other exogenous characteristics, the log of autarky-level welfare in location d, one can write Equation 2 as

$$\ln(\text{GDP}_d) = -\frac{1}{\theta}\ln(\lambda_{dd}) + \varepsilon_d = -\frac{1}{\theta}\ln\left(1-\frac{\text{IMPORTS}_d}{\text{GDP}_d}\right) + \varepsilon_d \approx \frac{1}{2\theta}\text{OPENNESS}_d + \varepsilon_d, \qquad (4)$$

where the last approximation holds under the condition that imports are a relatively small share of GDP, not a bad approximation for most countries. That is, the specification estimated by Frankel & Romer (1999), Equation 3, is not that different from the relationship one would expect to obtain in a simple, one-sector gravity model (approximated by Equation 4).

Frankel & Romer (1999) study a cross section of 150 countries in 1985. Hence, the comparisons being made here are across countries: More open countries are compared with less open countries. This comparison clearly calls into question the unconfoundedness requirement discussed above. For this reason, the key innovation of Frankel & Romer (1999) was to propose an instrumental variable (IV) that would isolate plausibly exogenous variation in OPENNESS—or, equivalently, to restrict attention to a subset of comparisons induced only by the IV. The IV used was based on a country's geographic position on Earth. Many empirical applications of the gravity equation in Equation 1 find that the distance between countries o and d is an important correlate of $\ln X_{od}$, and hence, it is often supposed, distance is an important shifter of the trade cost, τ_{od}. Frankel & Romer (1999) therefore construct an instrument for OPENNESS_d based on a two-step procedure in which they first estimate a gravity equation like Equation 1, with geodesic distances (among other plausible cost shifters) substituted for τ_{od} and then sum the fitted values for X_{od} in this regression over all countries $o \neq d$.[24] Essentially, the resulting IV captures country d's distance from economically large foreign countries. Frankel & Romer (1999) then assume that this IV is orthogonal (conditional on controls such as population and area) to the error term ε_d in Equation 3; that is, the comparisons induced by their IV satisfy the unconfoundedness requirement.

[23]In practice, there are substantial concerns about the ability to construct appropriate measures of real GDP in a multilateral cross-country context (Deaton & Heston 2010) and reasons to suspect that these concerns may particularly inhibit the ability of analysts to use standard measures of international real GDP to study the effects of market integration (Kehoe et al. 2008, Burstein & Cravino 2015).

[24]The resulting instrument has strong explanatory power—a high correlation in the IV first stage—but this is unsurprising given that the gravity equation is known to have a high R^2 and the procedure here simply sums, across all trading partners, the fitted values from a bilateral gravity regression to construct an instrument for OPENNESS_d, a variable that itself is essentially just the sum of trade across all trading partners.

Does the Frankel & Romer (1999) procedure satisfy the SUTVA requirement discussed above? This question hinges on whether the relationship in Equation 3 is stable. Reassuringly, and although this point is not discussed in Frankel & Romer (1999) or the ensuing literature described below, as shown in Equation 4, there do exist microfoundations in which Equation 3 is (approximately equal to) a stable structural relationship between the two endogenous variables, ln(GDP) and OPENNESS. Those microfoundations comprise the assumptions inherent to the gravity model class discussed in Section 2.2. So if we can be confident in these assumptions, then we can be confident that estimates of Equation 3 are identifying a meaningful parameter in the sense that the comparisons being drawn satisfy SUTVA. Unfortunately, as discussed in Section 3.3, the literature on testing those assumptions is in its infancy.

The results in Frankel & Romer (1999) are truly striking. Their preferred estimate of the coefficient in Equation 3 above is $\hat{\beta} = 1.97$. Although many regard this estimate as surprisingly large based on intuition, we can see from Equation 4 that it is also shocking from the perspective of the gravity models described in Section 2.2. A common estimate of the trade elasticity is $\hat{\theta} = 5$ (Head & Mayer 2014). Hence, Equation 4 suggests that a one-sector gravity model, without IO linkages, would predict $\beta \cong 1/2\theta$, or (using $\hat{\theta} = 5$) a $\hat{\beta}$ that is approximately 20 times smaller than the estimate in Frankel & Romer (1999). This gaping difference—between a leading estimate of the effects of international trade on GDP based on comparisons across countries and the estimate that would obtain in a wide body of simple models—is puzzling.

There are several potential responses to this puzzle. The most common in the literature, by a considerable margin, has been that the Frankel & Romer (1999) estimates are biased upward due to omitted variable bias (OVB) in the IV procedure—that the IV in question does not isolate comparisons that satisfy the unconfoundedness condition. I discuss this next. Three other possibilities have not, to my knowledge, been pursued in the literature. First, it is possible that the IV in question is identifying a particular latent average treatment effect, in the language of Imbens & Angrist (1994), that is particularly high. This would require that the countries induced to be more open to trade by their proximity to large countries (i.e., by the IV) have a higher marginal effect of openness. Yet, as seen in Equation 4 above, most tractable models for thinking about such issues belong to the gravity model class in which (at least with one sector) the marginal effect of OPENNESS on ln(GDP) is approximately constant. A second possibility is that the estimates in Frankel & Romer (1999) suffer from a violation of the SUTVA requirement. This would be the case if countries that enjoy their openness because of their relative proximity to other countries are made better off by this proximity, but countries that are relatively distant to other countries are actually harmed by their relative remoteness. In such a case, the IV estimate being identified by the empirical comparisons in Frankel & Romer (1999) is really the relative effect, or the difference between a positive effect on proximate countries and a negative effect on distant countries. Although I am unaware of tests for such negative impacts in the cross-country context, there is some evidence for this mechanism in the within-country context, discussed in Section 4.2. Third, it should be stressed that the comparison here to gravity models is to one-sector gravity models without IO linkages. As described in Section 2.2, the inclusion of such real-world features, when calibrated to reasonable parameter values, can allow the gains from integration in such models to grow considerably (and even without bound, in extreme cases). Unfortunately, those settings do not permit a simple comparison with estimates of β in Equation 3, but future work could easily explore this quantitatively.

Returning to the possibility of OVB in the Frankel & Romer (1999) estimates, a central concern is that the IV (based on a country's distance from major trading countries) is correlated with determinants of economic development that are treated as unobservables, ε_d, in Equation 3. Such unobserved determinants could be other variables that flow internationally and do so in relation to the distance involved, such as the flows of factors (capital, multinational firms, or migration) or the flow of technology. One reaction in the literature has hence been to attempt to control for such variables. A notable result in this regard is that from Ortega & Peri (2015), who consider the case of international migration. A second source of OVB would be some long-run determinant of GDP that happens to be spatially correlated; indeed, Rodriguez & Rodrik (2001) show that the coefficient β in Frankel & Romer (1999) falls from 1.97 to 0.34 when they simply control for distance to the equator.

An important response to the OVB concern in Frankel & Romer (1999) has recently been provided in two influential papers by Feyrer (2009a,b), who finds ways to identify time variation in the effective distance, between certain pairs of countries, that changes over time. Armed with such variation, Feyrer (2009a,b) can control, through the use of time-differencing procedures, for any time-invariant (and additively separable) unobservable determinants of a nation's GDP.

Feyrer (2009b) starts by considering that the share of international trade conducted by air, rather than by sea or land, has grown rapidly over the past 50 years. Some intercity (and hence intercountry) distances, such as that between Los Angeles and Tokyo, are the same whether the journey is by air or sea, but others, such as that between London and Tokyo, are significantly longer by sea than by air. Feyrer (2009b) calculates the shortest route by air (assumed to be the geodesic distance) and by sea between each pair of countries in the world. He then estimates, separately for each year from 1950 to 1997, a gravity equation but in which the log of each of these distance terms—the air distance and the sea distance—enters as an explanatory variable. The estimated coefficients, which are negative as expected, display the sensible property that the coefficient on air distance is rising over the time period (after being near zero in 1950), whereas that on sea distance is falling (such that it is near zero by 1985). This is consistent with the notion that, over the past 50 years, as the relative price of air shipments has fallen relative to that of sea shipments, the relevance of sea distance for explaining trade has fallen and the relevance of air distance has risen.

Feyrer (2009a) pursues an analogous idea, with another imaginative source of time variation in effective international distances, this time derived from the abrupt closing of the Suez Canal in 1967 and its reopening in 1975. In this case, the analysis focuses purely on sea distance (because air shipping was uncommon during that time period), as the shortest distance by which one could travel by sea between many pairs of cities (e.g., London and Tokyo) did rise and then fall as the canal closed and then reopened. Again it is found that the effect of sea distance on trade flows is negative, even when identified purely from the time variation driven by either the opening or closing of the canal.

Feyrer then goes on to use this variation in effective distance—be it from the rise of air-based trade (Feyrer 2009b) or the temporary closure of the Suez Canal (Feyrer 2009a)—as an IV for OPENNESS in a second-stage equation similar to Equation 3. Crucially, this IV regression can now be performed with country fixed effects. In addition, the OVB concern in Frankel & Romer (1999) is also plausibly weaker here. One source of bias might result from the many potential long-run determinants of economic development that are spatially correlated, yet it is exactly the long-run nature of such factors that would make them likely to be constant throughout Feyrer's sample period of 1950–1997. A second source of OVB would derive from other determinants of economic performance that flow between countries, such as factors of production or technology. Here the distinction between the relative air-sea distance IV in Feyrer (2009b) and the Suez Canal IV in Feyrer (2009a) is potentially useful because of the unique features of the latter case. In particular, although we might expect airplanes to facilitate the spread of capital, labor, multinational firms, and technology, the temporary closure of the Suez Canal occurred at a time when, Feyrer (2009a) argues, goods traveled by boat but factors and technology would have traveled by air.

Against the backdrop of this discussion, the results in Feyrer (2009a,b) are striking. His estimates of the (equivalent of the) parameter β in Equation 3 range from 1.15 to 5.65 (Feyrer 2009b), depending on the exact procedure used, such as long differences, fixed effects, or first differences, and from 0.31 to 0.53 (Feyrer 2009a), again, depending on the exact procedure, this time fixed effects with lagged effects in the first and second stages.[25] Notably, the estimates in Feyrer (2009a) are considerably lower than those in Frankel & Romer (1999), but still somewhat higher than would be predicted by a simple, one-sector gravity model (i.e., approximately $1/2\hat{\theta} \simeq 0.1$).

Two additional features of Feyrer's (2009a) results are remarkable because the events under study, the sudden and unanticipated closure and reopening of the Suez Canal, lend themselves to a deeper understanding of the time path of the relationship between trade and GDP. First, the effect of OPENNESS on GDP appears to take place within a relatively short time horizon (e.g., within five years) of the event in question. This contrasts with Frankel & Romer's (1999) results, which are cross sectional and hence plausibly long run in nature, and Feyrer's (2009b) results, which were estimated from the slow rise of air shipping relative to sea shipping. Feyrer's (2009a) estimates therefore suggest that market integration can affect a nation's living standards significantly and quickly. Second, Feyrer (2009a) documents a similar estimated effect from either the closing or the reopening of the Suez Canal. This is to be expected in standard, static trade models, such as those in Section 2.2, in which the effect of trade barriers on GDP is symmetric with regard to whether these barriers rise or fall. But in models in which trade contributes new capital equipment (Eaton & Kortum 2001) or new ideas (Alvarez et al. 2013), it is natural to expect (at least in the short run, before a long-run steady state is reached) that, although a reduction in trade barriers would raise GDP, an increase in trade barriers would not lower GDP.

Overall, the results of Frankel & Romer (1999) and Feyrer (2009a,b) offer a number of puzzles for work on the economic effects of market integration. How can we explain the magnitude of the effects in these papers even when they appear to arise purely from variation in transport costs, to arrive even in the short run, and to appear approximately symmetric with respect to whether rising trade barriers reduce GDP or falling trade barriers raise GDP? As discussed above, potential explanations include the following: the possibility that the IV techniques here are estimating a latent average treatment effect that is particularly large, the possibility that the empirical comparisons made here do not satisfy the SUTVA requirement, the possibility that even these results suffer from residual forms of OVB, or simply the possibility that the stylized and parametric quantitative gravity models—especially those with just one sector and no IO linkages—against which the empirical results here are being compared are too pessimistic about the size of the gains from trade. More work is needed to shed light on these potential explanations.

However, an intriguing counterpoint to the results discussed above has been recently provided by Pascali (2014). Similar to Feyrer (2009b), Pascali (2014) seeks to identify the effect of OPENNESS on GDP by using variation in shipping technologies over time. To do so, Pascali exploits the rise of steam-powered oceanic shipping, relative to wind-powered shipping, in the late nineteenth century. Steam shipping was not only faster per unit distance traveled, but differentially so between pairs of countries because of the meandering routes that needed to be taken, so as to

[25]Feyrer (2009a,b) actually estimates a slight variant of Equation 3: $\ln(\text{GDP}_d) = \gamma \ln(\text{IMPORTS}_d) + \varepsilon_d$. But with approximately balanced trade (i.e. $\text{IMPORTS}_d \simeq \text{OPENNESS}_d/2\text{GDP}_d$), this amounts to estimating $\ln(\text{GDP}_d) = \delta \ln(\text{OPENNESS}_d) + \varepsilon_d$, with $\delta \equiv \gamma/(1-\gamma)$, which is a log-linear version of Frankel & Romer's (1999) semilog specification in Equation 3. Feyrer (2009b) estimates $\hat{\gamma}$ ranging from 0.42 to 0.78 (i.e., $\hat{\delta}$ from 0.72 to 3.55), and the equivalent in Feyrer (2009a) ranges from 0.16 to 0.25 (i.e., $\hat{\delta}$ from 0.19 to 0.33). I convert Feyrer's elasticity of δ (from his reported estimates of γ) into Frankel & Romer's semilog coefficient β by using the level of OPENNESS for the median country in Frankel & Romer (1999, table A1).

maximize sail speeds, among certain pairs of ports. He uses maps of prevailing winds and a textbook model of sailing speed optimization (as a function of the prevailing wind speed and the angle of the boat to these prevailing winds) to predict sailing routes among all country pairs.[26] These predicted routes correlate extremely well with actual routes taken in the late sail era (1860–1860) during which such routes are known. He then constructs a predicted journey time, by sail and by steam, for each pair of countries and uses these two variables analogously to the air and sea distance variables employed by Feyrer (2009b). That is, Pascali (2014) finds that, as expected, the sail journey time variable predicts well bilateral trade early in his sample (1850) and less so over time, whereas the steam journey time variable predicts bilateral trade well at the end of the sample (1900) and increasingly so over time.

Pascali (2014) then goes on, as in Feyrer (2009b), to use these two variables (interacted with time dummies) as instruments for OPENNESS in a panel data regression version of Equation 3—in which measures of urbanization and total population are used, alongside incomplete data on GDP, as the outcome measure of interest. The results are striking, but in the opposite sense to those in the previous literature described above. For all outcome variables, the estimate of β is actually negative (and statistically significantly so at standard levels). However, Pascali also reports that the coefficient on an interaction term between OPENNESS and a measure of the quality of a country's institutions, in which the variation in this measure is instrumented for with the settler mortality instrument from Acemoglu et al. (2001), is positive and statistically significant. And although the sum of these coefficients is positive, the effect of OPENNESS on economic development does not appear to be statistically significantly greater than zero, even for countries with high-quality institutions.[27] These results therefore stand in stark contrast to those in Frankel & Romer (1999) and Feyrer (2009a,b). That is, although the modern literature has found evidence for strikingly large gains from trade, the historical setting from Pascali (2014) finds the opposite, at least to the extent that trade-driven changes in urbanization rates and population levels track living standards faithfully. This is a puzzle that future work needs to understand better.

All the work discussed in this section so far has aimed to study the effects of market integration as brought about by a reduction in transportation costs (or other costs correlated with distance). In standard models, the effect of transport costs on trade (and hence on the gains from trade) is no different from the effect of tariffs, after adjusting for the tax revenues that tariffs generate. However, Rodriguez & Rodrik (2001) argue that the effects of these two types of trade cost reductions may be different, even for the same reduction in trade costs and when holding tariff revenue constant. A relatively small literature has aimed to identify settings in which a country's external tariffs changed considerably in order to study market integration from this angle. However, as discussed by Wacziarg & Welch (2008, p. 206), "countries carrying out trade reforms often simultaneously adopt policies favoring domestic deregulation, privatization, and other microeconomic reforms and macroeconomic adjustments," which then means that empirical comparisons of trade-liberalizing countries to control-group countries may violate the unconfoundedness requirement.[28]

[26]This is done separately by season and also separately for the pre- and post-1860 periods as the Suez Canal opened (to steamships only) in 1860.

[27]Pascali (2014) estimates the interaction specifications when using urbanization rates and total population as the dependent variable, but not—presumably for lack of data—when using GDP as the outcome of interest.

[28]As discussed in Section 1, a larger literature has aimed to compare, within a country, industries or regions that were relatively more exposed to the trade liberalization episode.

An impressive study in this regard is by Bernhofen & Brown (2005), who consider Japan's (forced) emergence from near-complete autarky in 1853. These authors lack modern data on consumption and production, but they argue that an upper bound on the welfare benefits in Japan due to its departure from autarky can be bounded through the use of autarky price data and postautarky international trade data alone. Their preferred estimate of this upper bound is 9%. However, it is difficult to compare this estimate with those from the modern era because this was a change from autarky to trade without tariff barriers, but with a background level of transport costs in 1853 that was presumably quite high.

4.2. Comparisons Across Regions Within a Country

A common reaction to the cross-country comparisons that are unavoidable in the work described in Section 4.1 is that it is especially challenging to estimate meaningful treatment effects when using units of observation that are as heterogeneous as countries and that isolating comparisons that hold all equal across countries, apart from the treatment in question, is hopeless. For this reason, a large body of empirical work, across many fields, has turned to the possibility of identifying treatment effects by comparing regions within the same country.

The literature on the gains from economic integration is no exception. However, a central barrier to employing such a research design is that it has proven difficult to isolate plausibly exogenous variation in the barriers to trade across regions within a country.[29] One reason is that countries rarely permit policy barriers that prevent the intranational movement of goods, so research designs based on changes in such policies have been difficult to pursue.[30] Because of this, a recent literature has used variation in nonpolicy trade barriers, such as transportation infrastructure, to estimate the gains from intranational economic integration.

Before discussing that literature further, I note that it is important to keep in mind two factors that may make intranational settings different from international ones (from the perspective that concerns us here, the use of interregional comparisons within a country to estimate the gains from economic integration). First, it is natural to expect that the cost of moving nearly anything (factors, technology, or information) will be lower within countries than across international borders. This increases the likelihood that cross-regional comparisons fail to satisfy the SUTVA requirement discussed above. To take an extreme example, if workers are freely mobile across regions within a country, then any change in trade barriers affects all workers, in all locations, equally, and an analysis based on cross-regional comparisons would erroneously (because of the SUTVA violation) conclude that the change in trade barriers had no impact on worker welfare. Strategies for navigating this concern have varied depending on the data and setting, and I discuss some examples below. But one general strategy has been to work with geographical units of analysis that are sufficiently aggregated that it is plausible that spatial spillovers take place largely within each unit rather than across the units. For example, agglomeration externalities in consumption and production are typically thought to take place strongly within cities but only weakly across them (Ahlfeldt et al. 2014).

Second, it is possible that product market interactions across regions change in nature when those regions become increasingly smaller. For example, the gravity approaches discussed in

[29]As discussed in Section 1, a complementary strategy has been to use cross-regional variation in the exposure of subnational units to a change in international trade barriers, as I discuss further in Section 4.3.

[30]An exception is work by Young (2000), who studies a cluster of interregional trade conflicts in China during the 1980s and 1990s.

Section 2.2 predict that all regions will sell a strictly positive amount (within each sector) to every other region.[31] Yet this is clearly implausible as it is likely that the products made by neighboring regions become increasingly likely to be perfect substitutes as the definition of a region becomes smaller. This again suggests that the SUTVA requirement will become harder to sustain when drawing estimates from comparisons across increasingly disaggregated regions.

With these caveats in mind, we return to research that has sought to estimate the effects of economic integration from variation in the transportation infrastructure that connects regions within a country. Donaldson (2015) examines the case of the 67,247-km-long railroad network that was built in colonial India from 1853 to 1930, one of the most ambitious transportation projects in history. Using data on 235 districts, the smallest administrative unit for which annual data on output are available, Donaldson (2015) pursues two complementary avenues of empirical analysis. In the first, the evolution of districts before and after they gained access to the railroad network is compared to the evolution of districts that did not. The estimated treatment effect implies that access to railroads raised agricultural income by 16%. Reassuringly, estimates obtained from over 40,000 km of lines that were planned and surveyed, but then never built because of various plausibly idiosyncratic factors, are approximately zero. This suggests that railroad planners in this setting were unwilling or unable to allocate railroads in locations that were on different trends, in terms of agricultural output, from other locations. Surprisingly, there is little evidence of substantial SUTVA violations, at least due to neighboring districts—possibly because districts are already relatively large spatial units or because migration was famously low in colonial India.[32]

The second stream of analysis in Donaldson (2015) does more justice to the many complexities of a change in the trading environment such as that due to India's expanding railroad network. The idea is to estimate an empirical version of a gravity model of the sort described in Section 2.2. Armed with such a model, Donaldson (2015) solves for the equilibrium value of λ_{oo}, the share of expenditure a district o spent on its own goods, as in Equation 4, for each district, crop, and time period. Equation 4 states that in gravity models, the resulting variable λ_{oo} should act like a sufficient statistic for the impact of railroads on a measure of real living standards. In other words, the simple railroad indicator that was an important determinant of real incomes in Donaldson's initial analysis should become irrelevant once the sufficient statistic λ_{oo} is controlled for. This is roughly what is found. Furthermore, the coefficient on λ_{oo}, which would be equal to $-1/\theta$ in Equation 4, is approximately equal to the predicted value in this context. Put together, Donaldson (2015) therefore documents that India's railroads reduced trade barriers, and the resulting economic integration had important benefits (raising agricultural living standards by 16% on average) that are well accounted for by a gravity model.[33]

Another setting in which the transportation infrastructure was famously improved by the arrival of railroads is that of the United States in the late nineteenth century. This episode (1870–

[31]This can be seen from Equation 1 with finite trade costs τ_{od}. That said, there have been recent attempts (e.g., Helpman et al. 2008, Eaton et al. 2012) to build frameworks that explain the presence of many bilateral zeroes (even in the international trade data, aggregated across sectors) with finite trade costs.

[32]Another possibility, however, is that SUTVA violations occur across districts in a manner that need not respect the proximity of these districts in space. This highlights the central challenge in testing and correcting for spillovers: One needs a priori knowledge of the dimension(s) along which units of observation can be ordered in terms of their exposure to the spillover(s).

[33]To estimate how railroads reduced trade costs, Donaldson (2015) employs a spatial price gap approach similar to that discussed in footnote 1. Armed with estimates of trade costs, he then estimates the trade elasticity, θ, from a gravity equation similar to that in Equation 1. A final ingredient of the model is an estimate of a Ricardian production function, modeled as a simple function of rainfall (an important productivity shifter in this largely unirrigated agricultural setting).

1890 in particular) was the focus of Fogel's (1964) landmark analysis of the impact that railroads could have had on the US economy. Using a so-called social savings calculation, which argues that the benefits from railroads can be bounded above (in a competitive economy) by the additional amount it would cost to ship the postrailroad quantities at the prerailroad transport costs, Fogel (1964) finds that the railroads had a modest impact on agriculture (where he argues the largest effects would be seen), an upper bound of approximately 2.7% of GNP. The details of this calculation have been controversial, but the range of estimates in the social savings literature is relatively tight (Fogel 1979).

Donaldson & Hornbeck (2013) take a different approach to the evaluation of the gains from market integration brought about by US railroads, one that builds on the gravity model literature discussed above.[34] Given that the setting is one in which the movement and mobility of the population were relatively high, and that the US railroads enabled entire regions of the country to see substantial settlement for the first time, Donaldson & Hornbeck (2013) argue that a preferable outcome variable to study is the value of a fixed factor—land in this context—in each location. Drawing on decadal US Census of Agriculture data, which tracked the value of farm land in each US county from 1870 to 1890, Donaldson & Hornbeck (2013) estimate the relationship between nominal land values and a variable that has been referred to in the literature as market access, a term that can be approximated by an inverse trade cost–weighted sum of the sizes of all other counties from the perspective of the county under study.[35] This relationship is analogous to that in Equation 4, in which market access is then just (under the assumption of symmetric trade costs) a transformation of λ_{oo}, and the variable on the left-hand side is $\ln(V_d)$, where V_d is the value of agricultural land in county d, rather than $\ln(\mathrm{GDP}_d)$. Importantly, a similar relation between land value and market access is true even when all factors other than land are mobile. The results of this estimation procedure imply that the removal of railroads would have reduced land values by 64%, or 3.4% of GNP, somewhat larger than Fogel's estimate.

A more recent setting in which we have seen a dramatic improvement in the transportation technology within a country is China's remarkable new system of expressways, which began to open in 1989. Faber (2014) studies the most aggressive wave of new expressway construction, from 1992 to 2003. He isolates exogenous variation in expressway placement across Chinese counties by solving for the minimum-cost spanning tree network that would connect a set of important cities mentioned in a key planning document and then by working with a sample of rural counties that excludes the counties contained in these cities. Because of this, Faber's estimates identify the causal effect of transport access on a rural county relative to other rural counties that did not gain access. Surprisingly, this relative effect is actually negative: Rural counties crossed by the expressway network did relatively worse than those rural counties that were bypassed. Faber (2014) argues convincingly that this is a natural consequence of the displacement of economic output and employment from rural counties to the urban counties[36] that were deliberately left out of the exercise (because quasi-experimental variation in expressway placement among urban

[34]Other recent work on the effects of the US railroad network include the following: the effect on land values in Haines & Margo (2008), on urbanization in Atack et al. (2010), on the development of the factory in Atack et al. (2011), and on agricultural improvement in Atack & Margo (2011); the computational general equilibrium model in Herrendorf et al. (2012); and models of the network itself in Cervantes (2013) and Swisher (2014).

[35]That is, the market access of region d can be approximated by $\mathrm{MA}_d = \sum_{o \neq d} \tau_{od}^{-\theta} L_o$, where L_o is the population of region o. For other work on market access, readers are referred to Redding & Venables (2004), Hanson (2005), Head & Mayer (2011), Liu & Meissner (2013), and Bartelme (2015). The concept of market potential in Harris (1954) is also related.

[36]Unfortunately, Faber (2014) could not study such dislocation toward these counties directly, as those locations were deliberately left out of the sample.

locations could not be isolated). These results therefore highlight the relevance of the potential for SUTVA violations in the spatial comparisons, especially those within countries, that have featured prominently in the literature so far.

Numerous other recent studies have examined the impact of within-country transportation infrastructure improvements on welfare.[37] Prominent examples include Fernald (1999), Jacoby (2000), Mu & van de Walle (2007), Banerjee et al. (2012), Duranton & Turner (2012), Alder (2014), Aggarwal (2015), Asturias et al. (2014), Bird & Straub (2014), Gertler et al. (2014), Morten & Oliveira (2014), and Ghani et al. (2015). Space limitations prevent a full discussion of the approaches and findings in these papers, but readers are referred to Gramlich (1994) and Redding & Turner (2014) for authoritative surveys.

4.3. Comparisons Across Regions Within a Country Based on Relative Remoteness

As discussed in Section 4.2, the literature on estimating the gains from market integration has made progress by isolating settings in which regions of the same country have been able to trade at a lower cost with one another. A related identification strategy in the literature has been to use within-country variation in the ability of regions to access foreign markets to study the impact that the removal of international trade liberalization can have.

A leading example of such an approach is provided by Redding & Sturm (2008), who consider the effect on West German cities of the division of Germany in 1949 and the subsequent loss of access to markets in East Germany.[38] Redding & Sturm (2008) estimate that the effects of the German division were substantial and negative but varied across cities within West Germany according to the distance of a city from the East German border. Strikingly, they find almost mirror-image results from the reunification that took place in 1991. Although there are many potentially confounding aspects of division and reunification, besides market access, that may have similarly varied with distance to the border, Redding & Sturm (2008) convincingly rule out a wide range of alternative explanations.

Redding & Sturm (2008) also show that their estimates are very similar, quantitatively, to those expected from a gravity model similar to that in Section 2.2. At first glance, this is surprising given two features of the way that gravity models typically formulate trade costs and the economy: (*a*) Trade costs take an ad valorem form, with separate cost components entering multiplicatively, and (*b*) preferences are CES and the market structure is either perfect or monopolistic competition (such that markups are fixed). Under such assumptions, the effect on destination-region prices of one shifter of trade costs—such as the cost of crossing a border, as in Redding & Sturm (2008), or international tariffs more generally—is invariant to the level of any other shifter of trade costs—such as the distance from the destination region to the border. That is, under these assumptions, regions within a country should not differ in terms of their price exposure to a foreign shock. The way in which they do differ, however, is in terms of the importance (given the availability of domestic substitutes) of foreign markets. In this sense, the empirical approach that is inspired by

[37]An active recent literature (surveyed in Redding & Turner 2014) discusses other impacts of transportation infrastructure investments. Readers are referred to, for example, Chandra & Thompson (2000) on the location of economic activity, Michaels (2008) on wage inequality, Duranton & Turner (2011) on urban traffic, Rothenberg (2013) on firm location choices, Duranton et al. (2014) on US cities' trade patterns, Coşar & Demir (2014) on international exporting behavior, and Fajgelbaum & Redding (2014) on spatial specialization into agriculture and nonagriculture.

[38]For work that uses a related estimation strategy, readers are referred to Brülhart et al. (2013), Coşar & Fajgelbaum (2013), and Fajgelbaum & Redding (2014).

standard modeling assumptions in the literature isolates intranational variation in the costs of external market integration not by exploiting variation in the intensity (as measured by prices) of the foreign shock across domestic regions, but by using variation in the exposure of each local economy (in terms of consumption and production shares) to the foreign economy. This has parallels to the relative exposure approach based on industrial composition described briefly in Section 1.

There is natural appeal of an empirical design that would first seek to estimate the extent of foreign price pass-through—due to, for example, a liberalization of tariffs on international imports—separately for each domestic region, and then use the revealed variation in the intensity of the foreign shock across regions to study the impact of economic integration. I am not aware of any work that has been able to pursue this strategy. But recent work by Atkin & Donaldson (2014) takes steps in this direction by estimating how (in Ethiopia, Nigeria, and the United States) a cost shock at a port city passes through into the prices that remote consumers pay for a good shipped from that city. In contrast to what, as argued above, one would expect to find under standard modeling assumptions, Atkin & Donaldson (2014) find that remote locations indeed see significantly lower price pass-through. This is to be expected if, for example, remote locations source goods from a relatively less competitive trading sector. In principle, future work could use this approach to develop a research design for the study of market integration that is based on differential pass-through of foreign price shocks over space.

5. CONCLUSIONS AND FUTURE WORK

This article reviews recent work on estimates of the gains from market integration, both between and within countries. This body of work has grappled with the challenges of simultaneously studying necessarily general equilibrium issues and trying to isolate variation that credibly answers the counterfactual questions of interest. Two main strategies are deployed in this regard. The first relies on the structure of gravity models to simplify the number of parameters required to perform counterfactuals; in particular, any question that involves a proportional change in fundamentals from an observed starting point requires only knowledge of the elasticities of substitution. The second strategy draws causal inferences from empirical comparisons across regions that experienced differential changes in the trade barriers they face vis-à-vis other regions. Some of the intranational work discussed in Section 4.2 has pursued both strategies in parallel, and the results have been reassuringly similar. But, strangely, in the international settings of Section 4.1, it is still the case that many of the most credible estimates from empirical comparisons are larger than those expected in a simple gravity model. More work is needed to understand this puzzle.

Several additional directions for future work are apparent. First, within the setting of the gravity model as applied to multiple sectors (as is typical), there currently exists an uncomfortable fork in the road: Does one model the economy with perfect competition (as in Armington 1969 or Eaton & Kortum 2002) or with monopolistic competition (as in Krugman 1980, Melitz 2003, or Arkolakis et al. 2012a)? The simulations in Costinot & Rodríguez-Clare (2014) suggest that this modeling choice matters, yet there is typically nothing in the data to guide the choice.[39]

A second priority for future work involves relaxing some of the strong parametric assumptions that underpin much of the work in this area. Recent advances in demand and supply estimation are allowing researchers in other fields to study the welfare consequences of policy changes in

[39]The literature on testing for the home market effect, which is a feature of many multisector monopolistic competition models, is related, however (see Davis & Weinstein 2003 and Hanson & Xiang 2004).

a host of partial and industry-level equilibrium contexts (Ackerberg et al. 2007, Nevo 2011, Matzkin 2013). There are no barriers, in principle, to applying these advances to the more general equilibrium questions discussed in this article.

Third, the work reviewed here has ignored many of the potential effects of trade openness that exist in theory but have yet to have a large impact on empirical work. This includes external economies of scale (as in, e.g., Kucheryavyy et al. 2014), the organization of global supply chains (as in, e.g., Krugman & Venables 1995), and the spread of ideas (as in, e.g., Alvarez et al. 2013, Perla et al. 2014, and Sampson 2014).

Finally, the increasing availability of remotely sensed data, with high spatial resolution, is making it possible for researchers to begin to study the economic interactions among increasingly fine spatial units (Allen & Arkolakis 2014, Costinot & Donaldson 2014, Costinot et al. 2015). This is a natural progression given the often arbitrary nature (at least as it concerns the movement of goods) of the intranational jurisdictions that often delineate statistical units of analysis, as well as the ambition to study economic exchange, and the gains from reductions in the barriers to such exchange, at increasingly primitive levels.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am grateful to Pol Antràs and Arnaud Costinot for comments that greatly improved this work.

LITERATURE CITED

Acemoglu D, Johnson S, Robinson JA. 2001. The colonial origins of comparative development: an empirical investigation. *Am. Econ. Rev.* 91:1369–401

Ackerberg D, Berry S, Benkard L, Pakes A. 2007. Econometric tools for analyzing market outcomes. In *Handbook of Econometrics*, Vol. 6A, ed. J Heckman, E Leamer, pp. 4171–76. Amsterdam: North-Holland

Aggarwal S. 2015. *Do rural roads create pathways out of poverty? Evidence from India*. Work. Pap., Indian Sch. Bus., Hyderabad

Ahlfeldt GM, Redding SJ, Sturm DM, Wolf N. 2014. *The economics of density: evidence from the Berlin Wall*. NBER Work. Pap. 20354

Alder S. 2014. *Chinese roads in India: the effect of transport infrastructure on economic development*. Work. Pap., Univ. North Carolina, Chapel Hill

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

Allen T, Arkolakis C. 2014. Trade and the topography of the spatial economy. *Q. J. Econ.* 129:1085–140

Alvarez FE, Buera FJ, Lucas RE Jr. 2013. *Idea flows, economic growth, and trade*. NBER Work. Pap. 19667

Anderson JE. 2011. The gravity model. *Annu. Rev. Econ.* 3:133–60

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

Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables. *J. Am. Stat. Assoc.* 91:444–55

Arkolakis C, Costinot A, Donaldson D, Rodríguez-Clare A. 2012a. *The elusive pro-competitive effects of trade*. Work. Pap., Yale Univ., New Haven, CT

Arkolakis C, Costinot A, Rodríguez-Clare A. 2009. *New trade models, same old gains?* NBER Work. Pap. 15628

Arkolakis C, Costinot A, Rodríguez-Clare A. 2012b. 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, Garcia-Santana M, Ramos R. 2014. *Competition and the welfare gains from transportation infrastructure: evidence from the Golden Quadrilateral of India*. Work. Pap., Georgetown Univ., Washington, DC

Atack J, Bateman F, Haines M, Margo R. 2010. Did railroads induce or follow economic growth? Urbanization and population growth in the American Midwest, 1850–1860. *Soc. Sci. Hist.* 34:171–97

Atack J, Haines M, Margo R. 2011. Railroads and the rise of the factory: evidence from the United States, 1850–1870. In *Economic Evolution and Revolution in Historical Time*, ed. P Rhode, J Rosenbloom, D Weiman, pp. 162–79. Stanford, CA: Stanford Univ. Press

Atack J, Margo RA. 2011. The impact of access to rail transportation on agricultural improvement: the American Midwest as a test case. *J. Transp. Land Use* 4:5–18

Atkin D, Donaldson D. 2014. *Who's getting globalized? The size and implications of intranational trade costs*. Work. Pap., Univ. Calif., Los Angeles

Autor DH, Dorn D, Hanson GH. 2013a. The China syndrome: local labor market effects of import competition in the United States. *Am. Econ. Rev.* 103:2121–68

Balistreri EJ, Hillberry RH, Rutherford TF. 2011. Structural estimation and solution of international trade models with heterogeneous firms. *J. Int. Econ.* 83:95–108

Banerjee A, Duflo E, Qian N. 2012. *On the road: access to transportation infrastructure and economic growth in China*. NBER Work. Pap. 17897

Bartelme D. 2015. *Trade costs and economic geography: evidence from the US*. Work. Pap., Univ. Calif., Berkeley

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

Bernhofen DM, Brown JC. 2005. An empirical assessment of the comparative advantage gains from trade: evidence from Japan. *Am. Econ. Rev.* 95:208–25

Bernhofen DM, El-Sahli Z, Kneller R. 2013. *Estimating the effects of the container revolution on world trade*. Work. Pap. 2013:4, Dep. Econ., Lund Univ., Swed.

Besley T, Fetzer T, Mueller H. 2015. The welfare cost of lawlessness: evidence from Somali piracy. *J. Eur. Econ. Assoc.* 13:203–39

Bird J, Straub S. 2014. *The Brasilia experiment: road access and the spatial pattern of long-term local development in Brazil*. Work. Pap. 14-495, Toulouse Sch. Econ.

Broda C, Weinstein DE. 2006. Globalization and the gains from variety. *Q. J. Econ.* 121:541–85

Brülhart M, Carrère C, Robert-Nicoud F. 2013. *Trade and towns: on the uneven effects of trade liberalization*. Work. Pap., HEC Lausanne

Burstein A, Cravino J. 2015. Measured aggregate gains from international trade. *Am. Econ. J. Macroecon.* 7(2):181–218

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

Cervantes FP. 2013. *Railroads and economic growth: a trade policy approach*. Work. Pap., Bank of Mexico, Mexico City

Chandra A, Thompson E. 2000. Does public infrastructure affect economic activity? Evidence from the rural interstate highway system. *Reg. Sci. Urban Econ.* 30:457–90

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

Coşar AK, Demir B. 2014. *Domestic road infrastructure and international trade: evidence from Turkey*. Work. Pap., Booth Sch. Bus., Univ. Chicago

Coşar AK, Fajgelbaum PD. 2013. *Internal geography, international trade, and regional specialization*. NBER Work. Pap. 19697

Costinot A, Donaldson D. 2014. *How large are the gains from economic integration? Theory and evidence from U.S. agriculture, 1880–1997*. Work. Pap., Mass. Inst. Technol., Cambridge, MA

Costinot A, Donaldson D, Smith CB. 2015. Evolving comparative advantage and the impact of climate change in agricultural markets: evidence from 1.7 million fields around the world. *J. Polit. Econ.* In press

Costinot A, Rodríguez-Clare A. 2014. Trade theory with numbers: quantifying the consequences of globalization. In *Handbook of International Economics*, Vol. 4, ed. E Helpman, K Rogoff, G Gopinath, pp. 197–261. Amsterdam: North-Holland

Costinot A, Vogel J. 2015. Beyond Ricardo: assignment models in international trade. *Annu. Rev. Econ.* 7:31–62

Davis DR, Weinstein DE. 2003. Market access, economic geography and comparative advantage: an empirical test. *J. Int. Econ.* 59:1–23

de Loecker J, Goldberg PK. 2014. Firm performance in a global market. *Annu. Rev. Econ.* 6:201–27

Deaton A, Heston A. 2010. Understanding PPPs and PPP-based national accounts. *Am. Econ. J. Macroecon.* 2(4):1–35

Deaton AS, Muellbauer J. 1980. An almost ideal demand system. *Am. Econ. Rev.* 70:312–26

Dekle R, Eaton J, Kortum S. 2007. Unbalanced trade. *Am. Econ. Rev.* 97:351–55

Desmet K, Nagy DK, Rossi-Hansberg E. 2014. *The geography of development: evaluating migration restrictions and coastal flooding*. Work. Pap., South. Methodist Univ., Dallas

Dixit A, Norman V. 1980. *Theory of International Trade*. Cambridge, UK: Cambridge Univ. Press

Donaldson D. 2015. Railroads of the Raj: estimating the impact of transportation infrastructure. *Am. Econ. Rev.* In press

Donaldson D, Hornbeck R. 2013. *Railroads and American economic growth: a market access approach*. NBER Work. Pap. 19213

Duranton G, Morrow PM, Turner MA. 2014. Roads and trade: evidence from the U.S. *Rev. Econ. Stud.* 81:681–724

Duranton G, Turner MA. 2011. The fundamental law of road congestion: evidence from the U.S. *Am. Econ. Rev.* 101:2616–52

Duranton G, Turner MA. 2012. Urban growth and transportation. *Rev. Econ. Stud.* 79:1407–40

Eaton J, Kortum S. 1999. International technology diffusion: theory and measurement. *Int. Econ. Rev.* 40:537–70

Eaton J, Kortum S. 2001. *Trade in capital goods*. NBER Work. Pap. 8070

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

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

Eaton J, Kortum S, Neiman B, Romalis J. 2013. *Trade and the global recession*. Work. Pap. 2013-21, Sch. Econ., Univ. Sydney

Eaton J, Kortum SS, Sotelo S. 2012. *International trade: linking micro and macro*. NBER Work. Pap. 17864

Faber B. 2014. Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System. *Rev. Econ. Stud.* 81:1046–70

Fajgelbaum PD, Khandelwal AK. 2014. *Measuring the unequal gains from trade*. NBER Work. Pap. 20331

Fajgelbaum PD, Redding SJ. 2014. *External integration, structural transformation and economic development: evidence from Argentina 1870–1914*. NBER Work. Pap. 20217

Feenstra RC. 1994. New product varieties and the measurement of international prices. *Am. Econ. Rev.* 84:157–77

Feenstra RC. 2003. A homothetic utility function for monopolistic competition models, without constant price elasticity. *Econ. Lett.* 78:79–86

Fernald JG. 1999. Roads to prosperity? Assessing the link between public capital and productivity. *Am. Econ. Rev.* 89:619–38

Feyrer J. 2009a. *Distance, trade, and income: the 1967 to 1975 closing of the Suez Canal as a natural experiment*. NBER Work. Pap. 15557

Feyrer J. 2009b. *Trade and income: exploiting time series in geography*. NBER Work. Pap. 14910

Fogel RW. 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins Univ. Press

Fogel RW. 1979. Notes on the social saving controversy. *J. Econ. Hist.* 39:1–54

Frankel JA, Romer DH. 1999. Does trade cause growth? *Am. Econ. Rev.* 89:379–99

Freund C, Ornelas E. 2010. Regional trade agreements. *Annu. Rev. Econ.* 2:139–66

Gertler PJ, Gonzalez-Navarro M, Gracner T, Rothenberg AD. 2014. *Road quality and local economic activity: evidence from Indonesia's highways*. Work. Pap., Univ. Calif., Berkeley

Ghani E, Goswami AG, Kerr WR. 2015. Highway to success: the impact of the Golden Quadrilateral project for the location and performance of Indian manufacturing. *Econ. J.* In press

Goldberg PK, Pavcnik N. 2007. Distributional effects of globalization in developing countries. *J. Econ. Lit.* 45:39–82

Gramlich EM. 1994. Infrastructure investment: a review essay. *J. Econ. Lit.* 32:1176–96

Haines M, Margo R. 2008. Railroads and local economic development: the United States in the 1850s. In *Quantitative Economic History: The Good of Counting*, ed. J Rosenbloom, pp. 78–99. London: Routledge

Hanson GH. 2005. Market potential, increasing returns and geographic concentration. *J. Int. Econ.* 67:1–24

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

Harrigan J. 2010. Airplanes and comparative advantage. *J. Int. Econ.* 82:181–94

Harris CD. 1954. The market as a factor in the localization of industry in the United States. *Ann. Assoc. Am. Geogr.* 44:315–48

Harrison A, McLaren J, McMillan M. 2011. Recent perspectives on trade and inequality. *Annu. Rev. Econ.* 3:261–89

Harrison A, Rodríguez-Clare A. 2010. Trade, foreign investment, and industrial policy for developing countries. In *Handbook of Development Economics*, Vol. 5, ed. D Rodrik, M Rosenzweig, pp. 4039–214. Amsterdam: North-Holland

Head K, Mayer T. 2011. Gravity, market potential and economic development. *J. Econ. Geogr.* 11:281–94

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

Head K, Mayer T, Thoenig M. 2014. Welfare and trade without Pareto. *Am. Econ. Rev.* 104:310–16

Heckman JJ, Honoré BE. 1990. The empirical content of the Roy model. *Econometrica* 58:1121–49

Helpman E, Krugman P. 1985. *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*. Cambridge, MA: MIT Press

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

Herrendorf B, Schmitz JA Jr., Teixeira A. 2012. The role of transportation in U.S. economic development: 1840–1860. *Int. Econ. Rev.* 53:693–716

Hsieh C-T, Ossa R. 2011. *A global view of productivity growth in China*. NBER Work. Pap. 16778

Hummels D. 2007. Transportation costs and international trade in the second era of globalization. *J. Econ. Perspect.* 21:131–54

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

Imbens GW, Angrist JD. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62:467–75

Imbens GW, Rubin DB. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge, UK: Cambridge Univ. Press

Jacoby HG. 2000. Access to markets and the benefits of rural roads. *Econ. J.* 110:713–37

Jensen R. 2007. The digital provide: information (technology), market performance, and welfare in the South Indian fisheries sector. *Q. J. Econ.* 122:879–924

Johnson RC, Moxnes A. 2013. *Technology, trade costs, and the pattern of trade with multi-stage production*. Work. Pap., Dartmouth Univ., Hanover, NH

Kehoe T, Gibson MJ, Rhul KJ, Bajona C. 2008. *Trade liberalization, growth, and productivity*. Work. Pap., Univ. Minnesota, Minneapolis

Khandelwal AK, Schott PK, Wei S-J. 2013. Trade liberalization and embedded institutional reform: evidence from Chinese exporters. *Am. Econ. Rev.* 103:2169–95

Kovak BK. 2013. Regional effects of trade reform: What is the correct measure of liberalization? *Am. Econ. Rev.* 103:1960–76

Krugman PR. 1979. Increasing returns, monopolistic competition, and international trade. *J. Int. Econ.* 9:469–79

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

Krugman PR, Venables AJ. 1995. Globalization and the inequality of nations. *Q. J. Econ.* 110:857–80

Kucheryavyy K, Lyn G, Rodríguez-Clare A. 2014. *External economies and international trade: a quantitative framework*. Work. Pap., Univ. Calif., Berkeley

Liu D, Meissner CM. 2013. *Market potential and the rise of US productivity leadership*. NBER Work. Pap. 18819

Matzkin R. 2013. Nonparametric identification in structural economic models. *Annu. Rev. Econ.* 5:457–86

Melitz MJ. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica* 71:1695–725

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

Michaels G. 2008. The effect of trade on the demand for skill: evidence from the interstate highway system. *Rev. Econ. Stat.* 90:683–701

Mitra S, Mookherjee D, Torero M, Visaria S. 2013. *Asymmetric information and middleman margins: an experiment with West Bengal potato farmers*. Work. Pap., Boston Univ.

Morten M, Oliveira J. 2014. *Migration, roads and labor market integration: evidence from a planned capital city*. Work. Pap., Stanford Univ., Stanford, CA

Mu R, van de Walle D. 2007. *Rural roads and poor area development in Vietnam*. Policy Res. Work. Pap. 4340, World Bank, Washington, DC

Nevo A. 2011. Empirical models of consumer behavior. *Annu. Rev. Econ.* 3:51–75

Novy D. 2013. International trade without CES: estimating translog gravity. *J. Int. Econ.* 89:271–82

Ortega F, Peri G. 2015. Migration, trade and income. *J. Int. Econ.* In press

Pascali L. 2014. *The wind of change: maritime technology, trade and economic development*. Work. Pap., Warwick Univ., Coventry, UK

Perla J, Tonetti C, Waugh M. 2014. *Equilibrium technology diffusion, trade, and growth*. Work. Pap., Univ. British Columbia, Vancouver

Ramondo N, Rodríguez-Clare A. 2013. Trade, multinational production, and the gains from openness. *J. Polit. Econ.* 121:273–322

Redding SJ. 2013. *Goods trade, factor mobility and welfare*. Work. Pap., Princeton Univ., Princeton, NJ

Redding SJ, Sturm DM. 2008. The costs of remoteness: evidence from German division and reunification. *Am. Econ. Rev.* 98:1766–97

Redding SJ, Turner MA. 2014. *Transportation costs and the spatial organization of economic activity*. NBER Work. Pap. 20235

Redding SJ, Venables AJ. 2004. Economic geography and international inequality. *J. Int. Econ.* 62:53–82

Rodriguez F, Rodrik D. 2001. Trade policy and economic growth: a skeptic's guide to the cross-national evidence. In *NBER Macroeconomics Annual 2000*, ed. BS Bernanke, K Rogoff, pp. 261–338. Cambridge, MA: MIT Press

Rothenberg A. 2013. *Transport infrastructure and firm location choice in equilibrium: evidence from Indonesia's highways*. Work. Pap., Rand Corp., Santa Monica, CA

Rubin DB. 1980. Discussion of "Randomization analysis of experimental data in the Fisher randomization test." *J. Am. Stat. Assoc.* 75:591–93

Rubin DB. 1990. Formal models of statistical inference for causal effects. *J. Stat. Plan. Inference* 25:279–92

Sampson T. 2014. *Dynamic selection: an idea flows theory of entry, trade and growth*. Work. Pap., Lond. Sch. Econ.

Schmitz J. 2012. *New and larger costs of monopoly and tariffs*. Work. Pap., Fed. Reserve Bank Minneapolis

Sequeira S. 2014. *Tariffs and corruption: evidence from a tariff liberalization program*. Work. Pap., Lond. Sch. Econ.

Simonovska I, Waugh ME. 2014a. The elasticity of trade: estimates and evidence. *J. Int. Econ.* 92:34–50

Simonovska I, Waugh ME. 2014b. *Trade models, trade elasticities, and the gains from trade*. NBER Work. Pap. 20495

Steinwender C. 2013. *Information frictions and the law of one price: "When the States and the Kingdom became United."* Work. Pap., Harvard Bus. Sch., Boston, MA

Swisher SN IV. 2014. *Reassessing railroads and growth: accounting for transport network endogeneity*. Work. Pap., Cambridge Univ., Cambridge, UK

Tintelnot F. 2013. *Global production with export platforms*. Work. Pap., Univ. Chicago

Topalova P. 2010. Factor immobility and regional impacts of trade liberalization: evidence on poverty from India. *Am. Econ. J. Appl. Econ.* 2(4):1–41

Wacziarg R, Welch KH. 2008. Trade liberalization and growth: new evidence. *World Bank Econ. Rev.* 22:187–231

Winters LA, McCulloch N, McKay A. 2004. Trade liberalization and poverty: the evidence so far. *J. Econ. Lit.* 42:72–115

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

Young A. 2000. The razor's edge: distortions and incremental reform in the People's Republic of China. *Q. J. Econ.* 115:1091–135

Annual Review of Economics

Volume 7, 2015

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Economics* articles may be found at http://www.annualreviews.org/errata/economics