The Gains from Market Integration*

Dave Donaldson

Stanford University

February 2015

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 intra-national trade.

*Forthcoming in the *Annual Review of Economics*. Author's e-mail address: ddonald@stanford.edu. I am grateful to Pol Antràs and Arnaud Costinot for comments that greatly improved this work.

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 U.S. Postal Service will even deliver packages on a Sunday. Equally, waves of post-War 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 intra-national trade more than international trade. Much recent work exploits a fundamental symmetry between intra- and inter-national 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 intra-national 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 multi-regional 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 inter-regional economics. Recent decades have seen a sharpened understanding of

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

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 comparison 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 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 necessarily affecting all of these regions, though perhaps differentially so, at the same time. As a result, the prospects for finding a truly unaffected group, that can act as a pure control, seem distant. The literature that 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 and Rodríguez-Clare, 2010; Anderson, 2011; Costinot and Rodríguez-Clare, 2014; Head and Mayer, 2014), the distributional consequences of trade openness (Winters et al., 2004; Goldberg and Pavcnik, 2007; Harrison et al., 2011), the consequences of market integration for firm and industry performance (de Loecker and Goldberg, 2014), as well as wider consequences of improvements in transportation infrastructure (Redding and Turner, 2014). An additional related area concerns the estimation of the magnitude of

[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 differential industrial composition of these regions before the shock. See, for example, Topalova (2010), Autor et al. (2013a) and Kovak (2013).

barriers to trade, without particular attention to the effects of those barriers, as surveyed in Anderson and 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 different 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 towards more tightly specified environments—so-called 'gravity model' settings—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 and Norman (1985). A central result is that in an economy that admits a representative agent, when a region leaves autarky and allows any amount of trade with other regions the representative agent is made weakly better off (and strictly better off if the agent has different preferences, technology or relative endowments than agents elsewhere).[5] This result 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 (because, for example, these barriers are due to a feature of the region's transportation infrastructure) then any reduction in these barriers enlarges

[3]Since the publication of Anderson and van Wincoop (2004), empirical progress has been made in understanding—among other areas—how spatial price gaps can be used to estimate trade costs (Donaldson, 2014; Simonovska and Waugh, 2014a,b), 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 and Schaur, 2013), how imperfect competition in the trading sector affects barriers to trade (Hummels et al., 2009; Mitra et al., 2013; Atkin and Donaldson, 2014), and the importance of novel aspects of the global trading environment such as containerization (Bernhofen et al., 2013) and piracy (Besley et al., 2012).

[4]There is also a rich tradition of models—synthesized in Helpman and Krugman (1987)—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 Hicks-Kaldor criterion (such that, within each region, those helped by integration compensate those harmed); see, for example, Dixit and Norman (1985). Due to space constraints I abstract from distributional considerations here, but see the surveys described in the Introduction for a discussion of work in this area.

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 while 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 so far has focused on essentially bilateral considerations, where 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 (or PTAs, where a region's trade policy discriminates between outside regions depending on whether they are a member of a common PTA or not). See, for example, the recent survey in Freund and Ornelas (2010).

Two general lessons prevail from this literature, again with a distinction between trade barriers that generate revenue and those that don't. 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]

[6] The extreme setting where 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 to 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 in order to attempt to capture those rents. Some recent work has explored this idea empirically—see, for example, Schmitz (2012); Khandelwal et al. (2013); Sequeira (2014).

[8] For example, if region A and B are identical economies apart from the fact that $\kappa_{AC} < \kappa_{BC}$ then A benefits from trading with C more than B does; but then if τ_{BC} were to fall, region A would lose some of its

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, say, t_{AB} would make region A better off even in settings where region A's ability to manipulate its terms-of-trade with a non-discriminatory reduction (in both t_{AB} and t_{AC}) is weak and hence such a non-discriminatory reduction would be welfare-enhancing for region A. 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, where 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 $\frac{dW_i}{d\kappa_{jk}}$ nor $\frac{dW_i}{dt_{jk}}$ is known 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 extreme, 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 and their quantitative applications in the Sections 3.1 and 3.2 below. 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 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

advantage in market C relative to region B and be made worse off as a result.

costs and market structure so as to arrive at a model of a multi-region 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

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

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 o to d[9], the terms $A_o(\boldsymbol{X}, \boldsymbol{\tau})$ and $B_d(\boldsymbol{X}, \boldsymbol{\tau})$ refer to all determinants of o-to-d trade that are specific to the exporter o and importer d, respectively. The terms $\boldsymbol{X}$ and $\boldsymbol{\tau}$ indicate that the terms $A_o(\boldsymbol{X}, \boldsymbol{\tau})$ and $B_d(\boldsymbol{X}, \boldsymbol{\tau})$ could depend in an unrestricted manner on the endogenously-determined total vector of trade flows and 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 above, 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 $A_o(\boldsymbol{X}, \boldsymbol{\tau})$ and $B_d(\boldsymbol{X}, \boldsymbol{\tau})$. This implies that trade flows take an 'irrelevance of irrelevant alternatives' form. A further simplifying restriction is that the partial effect (that is, the effect holding the terms $A_o(\boldsymbol{X}, \boldsymbol{\tau})$ and $B_d(\boldsymbol{X}, \boldsymbol{\tau})$ 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. While 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 and Rodríguez-Clare (2014) summarize this work by providing a set of sufficient conditions for a multi-region model to be a gravity model. On the demand-side, almost all known gravity model microfoundations begin with CES utility across a continuum of horizontally-differentiated varieties (where the trade flow X_{od} above is then the total expenditure on all such varieties

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

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 and Rodríguez-Clare (2014) include: 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 and Kortum (2002) in which each region's ability to produce a given variety is drawn independently from a Fréchet distribution with common shape parameter; under monopolistic competition, models with free entry and fixed costs of production where 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).

Having defined the gravity model and delineated 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 a counterfactual reductions in trade costs? To answer this question Arkolakis et al. (2012b) showed 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 multi-sector 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

[10] An exception is 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 non-separable utility function as in Melitz and 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 and Ossa (2011). An analogous remark applies to changes in preferences.

change in the extent to which a region trades with itself.

One limitation of the result in equation (2) is that as it stands it can only be used for *ex post* analysis—that is, only if one knows the resulting change in λ_{dd} that was due to a particular event. However, one counterfactual question for which this limitation doesn't bind concerns the welfare costs of a move from current levels of trade costs $\boldsymbol{\tau}$ to autarky (where $\tau'_{od} = \infty$ for all $o \neq d$). While this is clearly an unrealistic and extreme scenario it offers the advantage of delivering a very simple expression for the change in welfare. Since $\lambda_{dd} = 1$ 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 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 (such as 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 given a 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 (such as 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 $\boldsymbol{X}$.[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]

[13] In practice it is important to bear in mind that this full vector of trade flows between all regions, $\boldsymbol{X}$, 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] And 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 in Simonovska and 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 and Redding (2015) provide quantitative examples of such cases, such as a comparison between

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 multi-sector versions of the gravity model specify an environment in which production is separable across sectors (just as it is across 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 and Rodríguez-Clare (2014) survey this approach and describe how much of the above logic extends to the multi-sector 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 (such as the Armington (1969) and Eaton and Kortum (2002) members of the gravity model class) and models featuring love of variety preferences and free entry (such as the Krugman (1980) and Melitz (2003) settings). Intuitively, the reason for this non-equivalence 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 and Krugman (1987)), 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 and Rodríguez-Clare (2014) conduct various calculations concerning the losses of a move to autarky, in models with multiple sectors and IO linkages, under the two alternative assumptions that (i) the set of varieties is fixed within each sector and (ii) 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 doesn't 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.

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).

Second, the recent literature on multi-sector gravity models has been extended to the case where sectors interact through input-output (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 and Venables (1995)—is to imagine that production requires a Cobb-Douglas aggregator of input bundles from other sectors, where those input bundles are themselves a CES aggregator (with the same elasticity of substitution as in consumer preferences) of the varieties from each sector.[16] Caliendo and Parro (2015) and Costinot and 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 intra-national 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 non-goods mobility includes Allen and Arkolakis (2014) and Redding (2013) on the mobility of workers, Eaton and Kortum (2001) on the mobility of capital goods, Eaton and Kortum (1999) on the diffusion of technology, and Ramondo and Rodríguez-Clare (2013) and Tintelnot (2013) on the mobility of technology embodied in multi-regional firms.

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. While 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 to data in this area, as I now discuss.

[16]Other approaches, such as those in Yi (2003) and Johnson and Moxnes (2013), yield new insights about how the presence of global supply chains affects what can be learned from trade flows (such as 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.

3 Estimates from quantitative trade models

The gravity model approach, outlined in Section 2.2 above, 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 and 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 using other members of the gravity model class. Particularly prominent examples include Eaton et al. (2011) and Balistreri et al. (2011).

A key contribution of Costinot and Rodríguez-Clare (2014) was to not only survey the literature on quantitative gravity models but also to perform *ex ante* counterfactual calculations on a consistent dataset 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 (that is, with multiple factors of production, multiple sectors with Cobb-Douglas upper-tier preferences, and input-output linkages between sectors used in Cobb-Douglas production functions) these authors calculate that, for the average country in their dataset, a move to autarky would reduce welfare substantially—by somewhere between 27 and 40 percent 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. While this idea has a long tradition in the context of Canadian provinces and U.S. states (McCallum, 1995; Anderson and 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 of Allen and Arkolakis (2014). These authors begin by characterizing a one-sector gravity model with perfect factor mobility.[18]

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

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

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 instantaneous cost of traveling over such a region is presumably relatively small. Allen and Arkolakis (2014) then apply their model to the United States, using counties as the regions in the model. Because (publicly available) inter-regional trade data are not available within the United States at such a fine level their quantitative estimates follow a procedure that differs somewhat from that described in Section 2.2 above, 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 U.S. interstate highway system (IHS) would lead to a welfare loss of 1.1-1.4%, a figure that substantially exceeds the annual costs of maintaining the IHS. As spatially disaggregated data become more and more available the tools developed in Allen and Arkolakis (2014) are sure to see many applications.[19]

A second recent example of a high spatial-resolution approach to *ex ante* counterfactual simulations of the effects of trade barriers comes from Costinot et al. (2015), who 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 FAO's GAEZ model, which aims to simulate how the production of any major crop will fare given local conditions (climate, soil, topography, etc.) at any of approximately 1.7 million grid cells on the Earth. Importantly, GAEZ provides these simulations both for current climatic conditions and those predicted by climatologists to occur due to 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.

productivity depends directly on the number of agents living there.

[19] One example is 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 restriction).

3.2 Additional gravity-based counterfactuals

The discussion above focused 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 extending settings. I give three examples here.

First, a common challenge 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 (such as political changes, or other market reforms as discussed in Goldberg and Pavcnik (2007)). Mexico's entry to NAFTA in 1994 was no exception. For this reason Caliendo and Parro (2015) use a multi-sector gravity model, with rich inter-sectoral input-output linkages, to quantify the impact of NAFTA on trade flows and welfare. These authors feed the actual time path of Mexico-U.S. bilateral tariffs into their gravity model in order to isolate the effect, according to the model, that NAFTA would have had in the absence of all of the other confounding factors. The estimated 'pure NAFTA' effect on Mexico was to raise welfare by 1.3%.

Second, Eaton et al. (2013) aim to decompose the decline in world trade flows and welfare during the global recession of 2008-2009—where the ratio of trade to GDP fell by 20 percent—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 multi-period 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 Broda and Weinstein (2006). These authors aim to isolate a component of the change in U.S. consumer welfare from 1972 to 2001 that is due to the importation of new product varieties (where a variety is defined as a unique narrowly-defined product from a unique country). Drawing on the methodologies suggested in Feenstra (1994), Broda and 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 U.S. consumers have been steadily importing more products from more countries, holding other margins of consumer and producer adjustment (such as the number of varieties produced

[20]The estimation procedure in Broda and 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).

domestically) constant. Accounting for new imported varieties reduces a continuing-variety U.S. import price index by an additional 28% between 1972 and 2001.

3.3 Non-gravity model approaches

The gravity equation (1) explains inter-regional 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 below). 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 and Khandelwal (2014) allows for non-homothetic preferences (of the AIDS sort, following Deaton and Muellbauer (1980)) and non-degenerate 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 micro-data to model the supply-side more flexibly. Recent examples of this include Melitz and Redding (2015) and Head et al. (2014), who consider monopolistic competition models of heterogeneous firms (facing CES preferences) but where those firms' productivity levels are drawn from a log-normal distribution.

Relatedly, Costinot and 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 doesn't 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 and Donaldson (2014) argue that, because of this selection problem, conventional data cannot hope to non-parametrically estimate the gains from trade in a Ricardian model.[21] But they propose a setting, agriculture, where scientific knowledge about the production function—how any crop would be likely to

[21] This non-identification result relates closely to the literature on the Roy model (Heckman and Honoré, 1990); see Costinot and Vogel (2015) in this volume for further discussion of the similarities.

grow under a whole ranged 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 U.S. agricultural production from 1880-1997, where the availability of agricultural output data for a panel of counties from the Census is particularly rich. As one might expect, U.S. counties appear to have been facing agricultural prices that became increasingly aligned across space throughout the 118 years under study. But surprisingly, the estimated U.S.-wide efficiency gains that resulted from this market integration process are substantial: approximately 0.5-1.5 percent per year compounding (depending on the assumptions made about whether inter-county 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 and Donaldson (2014) lack the inter-county 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 (non-parametrically estimated) gains from market integration can be substantial.

4 Evidence based on empirical comparisons

Section 3 above has discussed the many challenges of speaking with confidence about the effects of market integration using theory alone. While reductions in trade barriers are guaranteed to raise welfare for at least one region, in general neoclassical models, the study of any given region in particular is complicated by the reality of multi-region 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 made result in a model that can make successful predictions about changes in trade barriers. And, as reported in Section 3.3, non-gravity 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 barriers τ_{jk} (at home or elsewhere) change. In doing so, researchers have searched for settings where measured welfare in regions exposed to low trade barriers can be compared, empirically, to that in other regions that are exposed to higher trade barriers. In the program evaluation literature (see Angrist and Pischke (2009) for a textbook treatment) these would be referred to as 'treatment' and 'control' regions, respectively.

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 uncounfoundedness requirement (Rubin, 1990).[22] And 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 in Rubin (1980).

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

By contrast, the SUTVA requirement for 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

[22] Many of the studies reviewed below apply instrumental variable (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).

size and direction of such spillovers, in the context of this review, is in short supply and more work is needed to understand the potential for bias due to SUTVA violations here.

4.1 Comparisons across countries

While the discussion so far has been largely agnostic about whether the trade barriers under study exist between regions that represent nations or regions that represent sub-national 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 therefore 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 that by Frankel and 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:

$$\ln(GDP_d) = \beta \, OPENNESS_d + \varepsilon_d \tag{3}$$

where $OPENNESS_d \doteq \frac{EXPORTS_d + IMPORTS_d}{GDP_d}$ for region d. The goal is then to estimate the parameter β, the causal impact of one additional unit of $OPENNESS_d$ on the log of GDP_d, real (PPP-adjusted) per-capita GDP.[23] It is important to note that (as stressed by Frankel and Romer (1999)) the relationship in equation (3) has nothing direct to say about the impact of trade *barriers* on economic welfare. And while it is plausible that the trade barriers a researcher is interested in evaluating (such as 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 a concrete understanding of it. This is true both in gravity model and more general settings (as discussed in Sections 2.2 and 2.1 above).

While the Frankel and 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, it is closely related to that seen in equation (1). To see this, note that if trade is balanced (that is, the value of total exports equals that of total imports) we have $OPENNESS_d \doteq 2\frac{IMPORTS_d}{GDP_d}$. Equation (2) can be written as (after integrating and

[23] In practice, there are substantial concerns about the ability to construct appropriate measures of 'real' GDP in a multilateral cross-country context (Deaton and 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 and Cravino, 2015).

adding an error term, ε_d, that subsumes, among other exogenous characteristics, the log of autarky-level welfare in location d)

$$\ln(GDP_d) = -\frac{1}{\theta}\ln(\lambda_{dd}) + \varepsilon_d = -\frac{1}{\theta}\ln\left(1 - \frac{IMPORTS_d}{GDP_d}\right) + \varepsilon_d \approx \frac{1}{\theta}\frac{IMPORTS_d}{GDP_d} + \varepsilon_d, \tag{4}$$

and 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 and Romer (1999) is not that different from the relationship one would expect to obtain in a simple, one-sector gravity model.

Frankel and Romer (1999) studied 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 uncounfoundedness requirement discussed above. For this reason the key innovation in Frankel and 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 instrumental variable. The IV used was based on a country's geographic position on Earth. Many empirical applications of the gravity equation (1) above 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 and Romer (1999) therefore constructed an instrument for $OPENNESS_d$ that is 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 and 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, that the comparisons induced by their IV satisfy the uncounfoundedness requirement.

What about the SUTVA requirement discussed above? This question hinges on whether the relationship in equation (3) is stable. Reassuringly, and while this point is not discussed in Frankel and Romer (1999) or the ensuing literature described below, as shown in equation (4) there do exist plausible microfoundations in which equation (3) is (approximately equal to) a stable structural relationship between the two endogenous variables,

[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 $OPENNESS_d$ which is essentially just the sum of trade across all trading partners.

$\ln(GDP)$ and $OPENNESS$. Those microfoundations comprise the assumptions inherent to the gravity model class discussed in Section 2.2 above. 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 and Romer (1999) are truly striking. Their preferred estimate of the coefficient in equation (3) above is $\hat{\beta} = 1.97$. While many regard this estimate as surprisingly large based on intuition, we can see from equation (4) that they are also shocking from the perspective of the gravity models described in Section 2.2 above. A common estimate of the trade elasticity is $\hat{\theta} = 5$ (Head and Mayer, 2014). Hence equation (4) suggests that a one-sector gravity model, without input-output linkages, would predict $\beta \cong \frac{1}{2\theta}$, or (using $\hat{\theta} = 5$) a $\hat{\beta}$ that is approximately 20 times smaller than the estimate in Frankel and 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 a puzzle.

There are a number of potential responses to this puzzle. The most common in the literature by a considerable margin has been that the Frankel and Romer (1999) estimates are biased upwards due to omitted variable bias 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 (LATE), in the language of Imbens and Angrist (1994), that is particularly high. This would require that the countries who are induced to be more open to trade by their proximity to large countries (that is, by the IV) have a higher marginal effect of openness. Yet, as we have seen in equation (4) above, most tractable models for thinking about such issues belong to the gravity model class where (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 and Romer (1999) suffer from a violation of the SUTVA requirement. This would be the case if countries who enjoy their openness because of their relative proximity to other countries are made better off by this proximity, but countries who are relatively distant other countries are actually *harmed* by their remoteness; in such a case the IV estimate being identified by the empirical comparisons in Frankel and Romer (1999) is really the relative effect, or the difference between a positive effect on proximate countries and a negative effect on distant countries. While 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, as I discuss in Section 4.2 below. Third, it should be stressed that the comparison here to gravity models is a to one-sector gravity models without input-output linkages. As described in Section 2.2 above, 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 omitted variable bias (OVB) in the Frankel and 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 and Peri (2014), which considers 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 and Rodrik (2001) showed that the coefficient β in Frankel and 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 and Romer (1999) has recently been provided in two influential papers by James Feyrer. Both found ways to identify time variation in the effective distance, between certain pairs of countries, that changes over time. Armed with such variation, Feyrer (2009a) and Feyrer (2009b) have been able to control, through the use of time-differencing procedures, for any time-invariant (and additively separable) unobservable determinants of a nation's GDP.

Feyrer (2009a) starts by considering the fact that the share of international trade that is conducted by air, rather than by sea or land, has grown rapidly over the past fifty years. Some inter-city (and hence inter-country) distances, such as that between Los Angeles and Tokyo, are the same whether the journey is by air or sea; but others, like that between London and Tokyo, are significantly longer by sea than by air. Feyrer (2009a) 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 both 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 fifty 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 (2009b) 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 (since air shipping was uncommon in the time period, 1967-1975), since the shortest distance by which one could travel by sea between many pairs of cities (such as 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 off of 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 in Feyrer (2009a) or the temporary closure of the Suez Canal in Feyrer (2009b)—as an instrumental variable for $OPENNESS$ in a second-stage equation like equation (3) above. Crucially, this IV regression can now be performed with country fixed-effects. In addition, the OVB concern in Frankel and Romer (1999) is also plausibly weaker here. One source of bias might be due to 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 (2009a) and the Suez Canal IV in Feyrer (2009b) is potentially useful because of the unique features of the latter case. In particular, while 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 (2009b) 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) and Feyrer (2009b) are striking. Feyrer's estimates of the (equivalent of) parameter β in equation (3) range from 1.15 to 5.65 in Feyrer (2009a) (depending on the exact procedure used, such as long differences, fixed-effects or first differences), and from 0.31 to 0.53 in Feyrer (2009b) (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 (2009b) are considerably lower than

[25] Feyrer (2009a) and Feyrer (2009b) actually estimate a slight variant of equation (3): $\ln(GDP_d) = \gamma \ln(IMPORTS_d) + \varepsilon_d$. But with approximately balanced trade $IMPORTS_d \simeq \frac{OPENNESS_d}{2GDP_d}$, this amounts

those in Frankel and Romer (1999), but still somewhat higher than would be predicted by a simple gravity model (i.e. approximately $\frac{1}{2\hat{\theta}} \simeq 0.1$).

Two additional features of the results in the Feyrer (2009b) paper 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 (such as within five years) of the event in question. This is in contrast to the results in Frankel and Romer (1999), which were cross-sectional and hence plausibly long-run in nature, and the results in Feyrer (2009a), which were estimated from the slow rise of air shipping relative to sea shipping. The estimates in Feyrer (2009b) therefore suggest that market integration can affect a nation's living standards significantly and quickly. Second, Feyrer (2009b) 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 above, where the effect of trade barriers on GDP is symmetric with regards to whether these barriers rise or fall. But in models where trade contributes new capital equipment (Eaton and 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, while a reduction in trade barriers would raise GDP, an increase in trade barriers would not lower GDP.

Overall, the results in Frankel and Romer (1999), Feyrer (2009a) and Feyrer (2009b) 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 reduces GDP or falling trade barriers raises GDP? Potential explanations include, as discussed above: the possibility that the IV techniques here are estimating a LATE that is particularly large, the possibility that the empirical comparisons being made here do not satisfy the SUTVA requirement, that possibility that even these results suffer from residual forms of omitted variable bias, or simply the possibility that the stylized and parametric quantitative gravity models—especially those with just one sector and no input-output linkages—against which the empirical results here are being compared are too pessimistic about the size of

to estimating $\ln(GDP_d) = \delta \ln(OPENNESS_d) + \varepsilon_d$, with $\delta \equiv \frac{\gamma}{1-\gamma}$, which is a log-linear version of Frankel and Romer's semi-log specification in equation (3). Feyrer (2009a) estimates $\hat{\gamma}$ ranging from 0.42-0.78 (i.e. $\hat{\delta}$ from 0.72 to 3.55) and the equivalent in Feyrer (2009b) 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 and Romer's semi-log coefficient β by using the level of *OPENNESS* for the median country in Table A1 of Frankel and Romer (1999).

the gains from trade. More work is needed to shed light on these potential explanations.

However, an intriguing counterpoint to the results discussed so far has been recently provided by Pascali (2014). Like Feyrer (2009a), Pascali (2014) seeks to identify the effect of *OPENNESS* on GDP by using variation in shipping technologies over time. To do so, Pascali (2014) 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 maximize sail speeds, among certain pairs of ports. Pascali (2014) 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) where 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 in Feyrer (2009a). 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 (2009a), to use these two variables (interacted with time dummies) as instruments for *OPENNESS* in a panel data regression version of equation (3)—where 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 (2014) also reports that the coefficient on an interaction term between *OPENNESS* and a measure of the quality of a country's institutions, where the variation in this measure is instrumented for with the "settler mortality" instrument from Acemoglu et al. (2001), is positive and statistically significant. And while the sum of these coefficients is positive, it does not appear to be the case that the effect of *OPENNESS* on economic development is statistically significantly greater than zero even for the countries with high-quality institutions.[27] These results therefore stand in stark contrast to those in Frankel and Romer (1999), Feyrer (2009a) and Feyrer (2009b). That is, while

[26] This is done separately by season and also separately for the pre- and post-1860 periods since 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.

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 of 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 that are 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 and Rodrik (2001) have argued 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 where a country's external tariffs changed considerably in order to study market integration from this angle. However, as discussed in Wacziarg and Welch (2008), "countries carrying out trade reforms often simultaneously adopt policies favoring domestic deregulation, privatization, and other microeconomic reforms and macroeconomic adjustments" (p. 206), which then means that empirical comparisons of trade liberalizing countries to control group countries may violate the unconfoundedness requirement.[28]

An impressive study in this regard is due to Bernhofen and 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 post-autarky 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 free trade, 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 above is that it is especially challenging to estimate meaningful treatment effects when using units of observation that are as heterogenous as countries, and that isolating comparisons that hold all equal across countries apart the treatment in question is hopeless. For this reason a large body of empirical work, across many fields,

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

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 for this is that countries rarely permit policy barriers that prevent the intra-national 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 non-policy trade barriers, such as transportation infrastructure, to estimate the gains from intra-national economic integration.

Before discussing that literature further, it is important to keep in mind two factors that may make intra-national settings different from international ones (from the perspective that concerns us here, the use of inter-regional 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 (due to 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 of this 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. To take just one 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 smaller and smaller. For example, the gravity approaches discussed in Section 2.2 above 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

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

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

[31] This can be seen from equation 1 with finite trade costs τ_{od}. That said, there have been recent attempts, such as Helpman et al. (2008) and Eaton et al. (2012), to build frameworks that explain the presence of many

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 (2014) examined 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 is available, Donaldson (2014) pursues two complementary avenues of empirical analysis. In the first, Donaldson (2014) compares the evolution of districts before and after they gained access to the railroad network 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 for 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 (2014) 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 above. Armed with such a model, Donaldson (2014) 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; that is, 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. Further, the coefficient on λ_{oo}, which would be equal to $\frac{-1}{\theta}$ in equation (4),

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).

is approximately equal to the predicted value in this context. Put together, Donaldson (2014) 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 (1870-90 in particular) was the focus of Robert Fogel's landmark analysis of the impact that railroads could have had on the U.S. economy Fogel (1964). Using a so-called 'social savings' calculation, which argues that the benefits from railroads can by bounded above (in a competitive economy) by the additional amount it would cost to ship the post-railroad quantities at the pre-railroad transport costs, Fogel (1964) finds that the railroads had a modest impact on agriculture (where he argued the largest effects would be seen), an upper-bound of about 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 and Hornbeck (2013) take a different approach to the evaluation of the gains from market integration brought about by U.S. railroads, an approach that builds on the gravity model literature discussed above.[34] Given that the setting is one in which movement and mobility of population was relatively high, and that the U.S. railroads enabled entire regions of the country to see substantial settlement for the first time, Donaldson and 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 Agricultural Census data, which tracked the value of farm land in each U.S. county from 1870-1890, Donaldson and 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

[33] In order to estimate how railroads reduced trade costs, Donaldson (2014) employs a spatial price gap approach like that discussed in footnote 3. Armed with estimates of trade costs he then estimates the trade elasticity, θ, from a gravity equation like 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).

[34] Other recent work on the effects of the U.S. railroad network include: the effect on land values in Haines and 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 and Margo (2011); the computational general equilibrium model in Herrendorf et al. (2012); and models of the network itself in Cervantes (2013) and Swisher IV (2014).

[35] That is, market access of region d can be approximated by $MA_d = \sum_{o \neq d} \tau_{od}^{-\theta} L_o$, where L_o is the population of region o. For other work on market access see Redding and Venables (2004), Head and Mayer (2011) and Liu and Meissner (2013). The concept of "market potential" in Harris (1954) is also related.

that in equation (4), where 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 $\ln(GDP_d)$. 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 that saw 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 (2015) 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 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 (2015) argues convincingly that this is a natural consequence of displacement of economic 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 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 Jacoby (1998), Fernald (1999), Mu and van de Walle (2007), Banerjee et al. (2012), Duranton and Turner (2012) Ghani et al. (2012), Alder (2014), Asturias et al. (2014), Bird and Straub (2014), Gertler et al. (2014), and Morten and Oliveira (2014). Space limitations prevent a

[36]Unfortunately, Faber (2015) could not study such dislocation towards these counties directly, as those locations were deliberately left out of the sample (for fear that comparisons between urban and rural counties would violate the unconfoundedness requirement).

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

full discussion of the approaches and findings in these papers but see Gramlich (1994) and Redding and Turner (2014) for authoritative surveys.

4.3 Comparisons across regions within a country based on relative remoteness

As discussed in the previous subsection, the literature on estimating the gains from market integration has made progress by isolating settings where regions of the same country have been able to trade at 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 and 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 and Sturm (2008) estimate that the effects of German division was 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. While there are many potentially aspects of division and reunification, besides market access, that may have similarly varied with distance to the border, Redding and Sturm (2008) convincingly rule out a wide ranged of alternative explanations.

Redding and Sturm (2008) also show that their estimates are very similar, quantitatively, to those one would expect from a gravity model like that in Section 2.2. At first glance this is a surprise given two features of the way that gravity models typically formulate trade costs and the economy: (i) trade costs take an *ad valorem* form, with separate cost components entering multiplicatively; and (ii) preferences are CES and the market structure is either perfect or monopolistic competition (such that mark-ups 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 and Sturm (2008), or, say, 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

[38] For work that uses a related estimation strategy see Brulhart et al. (2013), Cosar and Fajgelbaum (2013), and Fajgelbaum and Redding (2014).

empirical approach that is inspired by standard modeling assumptions in the literature isolates intra-national 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 the use of 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 the Introduction.

There is a natural appeal of an empirical design that would first seek to estimate the extent of foreign price pass-through—due to, say, 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 and Donaldson (2014) take steps in this direction by estimating how (in Ethiopia, Nigeria and the U.S.A.) a cost shock at a port city passes through into the prices that remote consumers pay for a good that is shipped from that city. In contrast to what, as argued above, one would expect to find under standard modeling assumptions, Atkin and 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 paper has reviewed 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 have been deployed in this regard. The first has been to rely on the structure of gravity models to simplify the number of parameters required to perform counterfactuals; in particular, because of the CES-like structure of the gravity equation (1), 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 has been to draw causal inferences from empirical comparisons across regions that experienced differential changes in the trade barriers they face vis-a-vis other regions. Some of the intra-national work discussed in Section 4.2 above 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 one would expect to see in a simple gravity model. More work is needed to understand this puzzle.

A number of additional directions for future work are apparent. First, within the setting of the gravity model as applied to multiple sectors (as is typical), there exists currently an uncomfortable fork in the road: does one model the economy with perfect competition (as in Armington (1969) or Eaton and Kortum (2002)) or with monopolistic competition (as in Krugman (1980); Melitz (2003), or Arkolakis et al. (2012a))? The simulations in Costinot and Rodríguez-Clare (2014) suggest that this modeling choice matters and 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 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, for example, Kucheryavyy et al. (2014)), the organization of global supply chains (as in, for example, Krugman and Venables (1995)) or the spread of ideas (as in, for example, 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 (Costinot and Donaldson, 2014; Costinot et al., 2015; Allen and Arkolakis, 2014). This is a natural progression given the often arbitrary nature (at least as concerns the movement of goods) of the intra-national 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.

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

References

ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2001): "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91, 1369–1401.

ACKERBERG, D., S. BERRY, L. BENKARD, AND A. PAKES (2007): "Econometric Tools for Analyzing Market Outcomes," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Amsterdam: North-Holland.

AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2014): "The Economics of Density: Evidence from the Berlin Wall," NBER Working Papers 20354, National Bureau of Economic Research, Inc.

ALDER, S. (2014): "Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development," Working paper.

ALLEN, T. (2014): "Information Frictions in Trade," *Econometrica*, 82, 2041–2083.

ALLEN, T. AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129, 1085–1140.

ALVAREZ, F. E., F. J. BUERA, AND J. ROBERT E. LUCAS (2013): "Idea Flows, Economic Growth, and Trade," NBER Working Papers 19667, National Bureau of Economic Research, Inc.

ANDERSON, J. E. (2011): "The Gravity Model," *Annual Review of Economics*, 3, 133–160.

ANDERSON, J. E. AND E. VAN WINCOOP (2003): "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93, 170–192.

——— (2004): "Trade Costs," *Journal of Economic Literature*, 42, 691–751.

ANGRIST, J. AND J. PISCHKE (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455.

ARKOLAKIS, C., A. COSTINOT, D. DONALDSON, AND A. RODRÍGUEZ-CLARE (2012a): "The Elusive Pro-Competitive Effects of Trade," Working papers.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2009): "New Trade Models, Same Old Gains?" Nber working paper.

——— (2012b): "New Trade Models, Same Old Gains?" *American Economic Review*, 102, 94–130.

Armington, P. S. (1969): "A Theory of Demand for Products Distinguished by Place of Production," *IMF Staff Papers*, 16, 159–178.

Asturias, J., M. Garcia-Santana, and R. Ramos (2014): "Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India," Working paper.

Atack, J., F. Bateman, M. Haines, and R. Margo (2010): "Did Railroads Induce or Follow Economic Growth? Urbanization and Population Growth in the American Midwest, 1850-1860," *Social Science History*, 34, 171–197.

Atack, J., M. Haines, and R. Margo (2011): "Railroads and the Rise of the Factory: Evidence from the United States, 1850-1870," in *Economic Evolution and Revolution in Historical Time*, ed. by P. Rhode, J. Rosenbloom, and D. Weiman, Palo Alto, CA: Stanford University Press, 162–179.

Atack, J. and R. A. Margo (2011): "The Impact of Access to Rail Transportation on Agricultural Improvement: The American Midwest as a Test Case," *The Journal of Transport and Land Use*, 4, 5–18.

Atkin, D. and D. Donaldson (2014): "Who's Getting Globalized? The Size and Implications of Intranational Trade Costs," Working paper.

Autor, D. H., D. Dorn, and G. H. Hanson (2013a): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–68.

Balistreri, E. J., R. H. Hillberry, and T. F. Rutherford (2011): "Structural Estimation and Solution of International Trade Models with Heterogeneous Firms," *Journal of International Economics*, 83, 95–108.

Banerjee, A., E. Duflo, and N. Qian (2012): "On the Road: Access to Transportation Infrastructure and Economic Growth in China," Working Papers 4826, eSocialSciences.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003): "Plants and Productivity in International Trade," *American Economic Review*, 93, 1268–1290.

BERNHOFEN, D. M. AND J. C. BROWN (2005): "An Empirical Assessment of the Comparative Advantage Gains from Trade: Evidence from Japan," *American Economic Review*, 95, 208–225.

BERNHOFEN, D. M., Z. EL-SAHLI, AND R. KNELLER (2013): "Estimating the Effects of the Container Revolution on World Trade," Working Papers 2013:4, Lund University, Department of Economics.

BESLEY, T., T. FETZER, AND H. MUELLER (2012): "One Kind of Lawlessness: Estimating the Welfare Cost of Somali Piracy," Working Papers 626, Barcelona Graduate School of Economics.

BIRD, J. AND S. STRAUB (2014): "The Brasilia Experiment: Road Access and the Spatial Pattern of Long-term Local Development in Brazil," TSE Working Papers 14-495, Toulouse School of Economics (TSE).

BRODA, C. AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 121, 541–585.

BRULHART, M., C. CARRERE, AND F. ROBERT-NICOUD (2013): "Trade and Towns: On the Uneven Effects of Trade Liberalization," Working paper.

BURSTEIN, A. AND J. CRAVINO (2015): "Measured Aggregate Gains from International Trade," *American Economic Journal: Macro*, In press.

CALIENDO, L. AND F. PARRO (2015): "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, In press.

CERVANTES, F. P. (2013): "Railroads and Economic Growth: A Trade Policy Approach," Working paper.

CHANDRA, A. AND E. THOMPSON (2000): "Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system," *Regional Science and Urban Economics*, 30, 457–490.

CHANEY, T. (2008): "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707–21.

COSAR, A. K. AND B. DEMIR (2014): "Domestic Road Infrastructure and International Trade: Evidence from Turkey," Working paper.

COSAR, A. K. AND P. D. FAJGELBAUM (2013): "Internal Geography, International Trade, and Regional Specialization," NBER Working Paper 19697.

COSTINOT, A. AND D. DONALDSON (2014): "How Large are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-1997," Working paper.

COSTINOT, A., D. DONALDSON, AND C. B. SMITH (2015): "Evolving Comparative Advantage and the Impact of Climate Change in Agricultural Markets: Evidence from 1.7 Million Fields around the World," *Journal of Political Economy*, In press.

COSTINOT, A. AND A. RODRÍGUEZ-CLARE (2014): "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in *Handbook of International Economics*, ed. by K. R. Elhanan Helpman and G. Gopinath, Elsevier, vol. 4, 197 – 261.

COSTINOT, A. AND J. VOGEL (2015): "Beyond Ricardo: Assignment Models in International Trade," *Annual Review of Economics*, In press.

DAVIS, D. R. AND D. E. WEINSTEIN (2003): "Market access, economic geography and comparative advantage: an empirical test," *Journal of International Economics*, 59, 1–23.

DE LOECKER, J. AND P. K. GOLDBERG (2014): "Firm Performance in a Global Market," *Annual Review of Economics*, 6, 201–227.

DEATON, A. AND A. HESTON (2010): "Understanding PPPs and PPP-based National Accounts," *American Economic Journal: Macroeconomics*, 2, 1–35.

DEATON, A. S. AND J. MUELLBAUER (1980): "An Almost Ideal Demand System," *American Economic Review*, 70, 312–26.

DEKLE, R., J. EATON, AND S. KORTUM (2007): "Unbalanced Trade," *American Economic Review*, 97, 351–355.

DESMET, K., D. K. NAGY, AND E. ROSSI-HANSBERG (2014): "The Geography of Development: Evaluating Migration Restrictions and Coastal Flooding," Working paper.

DIXIT, A. AND V. NORMAN (1985): *Theory of International Trade*, Cambridge Books, Cambridge University Press.

DONALDSON, D. (2014): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *American Economic Review*, In press.

DONALDSON, D. AND R. HORNBECK (2013): "Railroads and American Economic Growth: A Market Access Approach," NBER Working Paper 19213.

DURANTON, G., P. M. MORROW, AND M. A. TURNER (2014): "Roads and Trade: Evidence from the U.S." *Review of Economic Studies*, 81, 681–724.

DURANTON, G. AND M. A. TURNER (2011): "The Fundamental Law of Road Congestion: Evidence from the U.S." *American Economic Review*, 101, 2616–2652.

——— (2012): "Urban Growth and Transportation," *Review of Economic Studies*, 79, 1407–1440.

EATON, J. AND S. KORTUM (1999): "International Technology Diffusion: Theory and Measurement," *International Economic Review*, 40, 537–70.

——— (2001): "Trade in Capital Goods," NBER Working Paper 8070.

——— (2002): "Technology, Geography, and Trade," *Econometrica*, 70, 1741–1779.

EATON, J., S. KORTUM, AND F. KRAMARZ (2011): "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 79, 1453–1498.

EATON, J., S. KORTUM, B. NEIMAN, AND J. ROMALIS (2013): "Trade and the Global Recession," Working Papers 2013-21, University of Sydney, School of Economics.

EATON, J., S. S. KORTUM, AND S. SOTELO (2012): "International Trade: Linking Micro and Macro," NBER Working Paper 17864.

FABER, B. (2015): "Trade Integration, Market Size and Industrialization: Evidence from China's National Trunk Highway System," *Review of Economic Studies*, In press.

FAJGELBAUM, P. AND S. J. REDDING (2014): "External Integration, Structural Transformation and Economic Development: Evidence from Argentina 1870-1914," NBER Working Papers 20217, National Bureau of Economic Research, Inc.

FAJGELBAUM, P. D. AND A. K. KHANDELWAL (2014): "Measuring the Unequal Gains from Trade," NBER Working Paper 20331.

FEENSTRA, R. C. (1994): "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84, 157–77.

——— (2003): "A homothetic utility function for monopolistic competition models, without constant price elasticity," *Economics Letters*, 78, 79–86.

FERNALD, J. G. (1999): "Roads to Prosperity? Assessing the Link between Public Capital and Productivity," *American Economic Review*, 89, 619–638.

FEYRER, J. (2009a): "Trade and Income – Exploiting Time Series in Geography," NBER Working Paper 14910.

——— (2009b): "Distance, Trade, and Income – The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment," NBER Working Paper 15557.

FOGEL, R. W. (1964): *Railroads and American Economic Growth: Essays in Econometric History*, Baltimore: Johns Hopkins University Press.

——— (1979): "Notes on the Social Saving Controversy," *Journal of Economic History*, 39, 1–54.

FRANKEL, J. A. AND D. H. ROMER (1999): "Does Trade Cause Growth?" *American Economic Review*, 89, 379–399.

FREUND, C. AND E. ORNELAS (2010): "Regional Trade Agreements," *Annual Review of Economics*, 2, 139–166.

GERTLER, P. J., M. GONZALEZ-NAVARRO, T. GRACNER, AND A. D. ROTHENBERG (2014): "Road Quality and Local Economic Activity: Evidence from Indonesia's Highways," Working paper.

GHANI, E., A. G. GOSWAMI, AND W. R. KERR (2012): "Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," NBER Working Paper 18524.

GOLDBERG, P. K. AND N. PAVCNIK (2007): "Distributional Effects of Globalization in Developing Countries," *Journal of Economic Literature*, 45, 39–82.

GRAMLICH, E. M. (1994): "Infrastructure Investment: A Review Essay," *Journal of Economic Literature*, 32, 1176–1196.

HAINES, M. AND R. MARGO (2008): "Railroads and Local Economic Development: The United States in the 1850s," in *Quantitative Economic History: The Good of Counting*, ed. by J. Rosenbloom, London: Routledge, 78–99.

HANSON, G. H. AND C. XIANG (2004): "The Home-Market Effect and Bilateral Trade Patterns," *American Economic Review*, 94, 1108–1129.

HARRIGAN, J. (2010): "Airplanes and comparative advantage," *Journal of International Economics*, 82, 181–194.

HARRIS, C. D. (1954): "The market as a factor in the localization of industry in the United States," *Annals of the Association of American Geographers*, 44, 315–348.

HARRISON, A., J. MCLAREN, AND M. MCMILLAN (2011): "Recent Perspectives on Trade and Inequality," *Annual Review of Economics*, 3, 261–289.

HARRISON, A. AND A. RODRÍGUEZ-CLARE (2010): *Trade, Foreign Investment, and Industrial Policy for Developing Countries*, Elsevier, vol. 5 of *Handbook of Development Economics*, chap. 63, 4039–4214.

HEAD, K. AND T. MAYER (2011): "Gravity, market potential and economic development," *Journal of Economic Geography*, 11, 281–294.

——— (2014): "Gravity Equations: Workhorse, Toolkit, and Cookbook," in *Handbook of International Economics*, ed. by K. R. Elhanan Helpman and G. Gopinath, Elsevier, vol. 4.

HEAD, K., T. MAYER, AND M. THOENIG (2014): "Welfare and Trade without Pareto," *American Economic Review Papers and Proceedings*, 104, 310–16.

HECKMAN, J. J. AND B. E. HONORÉ (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121–49.

HELPMAN, E. AND P. KRUGMAN (1987): *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, vol. 1 of *MIT Press Books*, The MIT Press.

HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 123, 441–487.

HERRENDORF, B., J. A. SCHMITZ JR., AND A. TEIXEIRA (2012): "The Role Of Transportation In U.S. Economic Development: 1840-1860," *International Economic Review*, 53, 693–716.

HSIEH, C.-T. AND R. OSSA (2011): "A Global View of Productivity Growth in China," NBER Working Paper 16778.

HUMMELS, D. (2007): "Transportation Costs and International Trade in the Second Era of Globalization," *Journal of Economic Perspectives*, 21, 131–154.

HUMMELS, D., V. LUGOVSKYY, AND A. SKIBA (2009): "The trade reducing effects of market power in international shipping," *Journal of Development Economics*, 89, 84–97.

HUMMELS, D. L. AND G. SCHAUR (2013): "Time as a Trade Barrier," *American Economic Review*, 103, 2935–59.

IMBENS, G. W. AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–75.

JACOBY, H. G. (1998): "Access to markets and the benefits of rural roads," Policy Research Working Paper Series 2028.

JENSEN, R. (2007): "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *The Quarterly Journal of Economics*, 122, 879–924.

JOHNSON, R. C. AND A. MOXNES (2013): "Technology, Trade Costs, and the Pattern of Trade with Multi-Stage Production," Working paper.

KEHOE, T., M. J. GIBSON, K. J. RHUL, AND C. BAJONA (2008): "Trade Liberalization, Growth, and Productivity," Working paper.

KHANDELWAL, A. K., P. K. SCHOTT, AND S.-J. WEI (2013): "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters," *American Economic Review*, 103, 2169–95.

KOVAK, B. K. (2013): "Regional Effects of Trade Reform: What Is the Correct Measure of Liberalization?" *American Economic Review*, 103, 1960–76.

KRUGMAN, P. (1980): "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 70, 950–59.

KRUGMAN, P. R. (1979): "Increasing returns, monopolistic competition, and international trade," *Journal of International Economics*, 9, 469–479.

KRUGMAN, P. R. AND A. J. VENABLES (1995): "Globalization and the Inequality of Nations," *The Quarterly Journal of Economics*, 110, 857–80.

KUCHERYAVYY, K., G. LYN, AND A. RODRÍGUEZ-CLARE (2014): "External Economies and International Trade: A Quantitative Framework," Working paper.

LIU, D. AND C. M. MEISSNER (2013): "Market Potential and the Rise of US Productivity Leadership," NBER Working Paper 18819.

MATZKIN, R. (2013): "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 5, 457–486.

McCallum, J. (1995): "National Borders Matter: Canada-U.S. Regional Trade Patterns," *American Economic Review*, 85, 615–23.

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

Melitz, M. J. and G. I. P. Ottaviano (2008): "Market Size, Trade, and Productivity," *Review of Economic Studies*, 75, 295–316.

Melitz, M. J. and S. J. Redding (2015): "New Trade Models, New Welfare Implications," *American Economic Review*, In press.

Michaels, G. (2008): "The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System," *The Review of Economics and Statistics*, 90, 683–701.

Mitra, S., D. Mookherjee, M. Torero, and S. Visaria (2013): "Asymmetric Information and Middleman Margins: An Experiment with West Bengal Potato Farmers," Working paper.

Morten, M. and J. Oliveira (2014): "Migration, Roads and Labor Market Integration: Evidence from a Planned Capital City," Working paper.

Mu, R. and D. van de Walle (2007): "Rural roads and poor area development in Vietnam," Policy Research Working Paper Series 4340, The World Bank.

Nevo, A. (2011): "Empirical Models of Consumer Behavior," *Annual Review of Economics*, 3, 51–75.

Novy, D. (2013): "International trade without CES: Estimating translog gravity," *Journal of International Economics*, 89, 271–282.

Ortega, F. and G. Peri (2014): "Migration, Trade and Income," *Journal of International Economics*, In press.

Pascali, L. (2014): "The wind of change: Maritime technology, trade and economic development," Working paper.

Perla, J., C. Tonetti, and M. Waugh (2014): "Equilibrium Technology Diffusion, Trade, and Growth," Working paper.

Ramondo, N. and A. Rodríguez-Clare (2013): "Trade, Multinational Production, and the Gains from Openness," *Journal of Political Economy*, 121, 273 – 322.

REDDING, S. AND A. J. VENABLES (2004): “Economic geography and international inequality,” *Journal of International Economics*, 62, 53–82.

REDDING, S. J. (2013): “Goods Trade, Factor Mobility and Welfare,” Working paper.

REDDING, S. J. AND D. M. STURM (2008): “The Costs of Remoteness: Evidence from German Division and Reunification,” *American Economic Review*, 98, 1766–97.

REDDING, S. J. AND M. A. TURNER (2014): “Transportation Costs and the Spatial Organization of Economic Activity,” NBER Working Paper 20235.

RODRIGUEZ, F. AND D. RODRIK (2001): “Trade Policy and Economic Growth: A Skeptic’s Guide to the Cross-National Evidence,” in *NBER Macroeconomics Annual 2000, Volume 15*, NBER, 261–338.

ROTHENBERG, A. (2013): “Transport Infrastructure and Firm Location Choice in Equilibrium: Evidence from Indonesia’s Highways,” Working paper.

RUBIN, D. B. (1980): “Discussion of “Randomization Analysis of Experimental Data in the Fisher Randomization Test”,” *Journal of the American Statistical Association*, 75, 591–593.

——— (1990): “Formal Models of Statistical Inference for Causal Effects,” *Journal of Statistical Planning and Inference*, 25, 279–292.

SAMPSON, T. (2014): “Dynamic Selection: An Idea Flows Theory of Entry, Trade and Growth,” Working paper.

SCHMITZ, J. (2012): “New and larger costs of monopoly and tariffs,” Working paper.

SEQUEIRA, S. (2014): “Tariffs and Corruption: Evidence from a Tariff Liberalization Program,” Working paper.

SIMONOVSKA, I. AND M. E. WAUGH (2014a): “The elasticity of trade: Estimates and evidence,” *Journal of International Economics*, 92, 34–50.

——— (2014b): “Trade Models, Trade Elasticities, and the Gains from Trade,” NBER working paper 20495.

STEINWENDER, C. (2013): “Information Frictions and the Law of One Price: “When the States and the Kingdom became United”,” Tech. rep.

SWISHER IV, S. N. (2014): "Reassessing Railroads and Growth: Accounting for Transport Network Endogeneity," Working paper.

TINTELNOT, F. (2013): "Global Production with Export Platforms," Tech. rep.

TOPALOVA, P. (2010): "Factor Immobility and Regional Impacts of Trade Liberalization: Evidence on Poverty from India," *American Economic Journal: Applied Economics*, 2, 1–41.

WACZIARG, R. AND K. H. WELCH (2008): "Trade Liberalization and Growth: New Evidence," *World Bank Economic Review*, 22, 187–231.

WINTERS, L. A., N. MCCULLOCH, AND A. MCKAY (2004): "Trade Liberalization and Poverty: The Evidence So Far," *Journal of Economic Literature*, 42, 72–115.

YI, K.-M. (2003): "Can Vertical Specialization Explain the Growth of World Trade?" *Journal of Political Economy*, 111, 52–102.

YOUNG, A. (2000): "The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China," *The Quarterly Journal of Economics*, 115, 1091–1135.