Determinants of Trade Margins: Insights Using State Export Data

Authors	Subhayu Bandyopadhyay, and Cletus C. Coughlin
Working Paper Number	2014-006C
Revision Date	November 2015
Citable Link	https://doi.org/10.20955/wp.2014.006
Suggested Citation	Bandyopadhyay, S., Coughlin, C.C., 2015; Determinants of Trade Margins: Insights Using State Export Data, Federal Reserve Bank of St. Louis Working Paper 2014-006. URL https://doi.org/10.20955/wp.2014.006

Determinants of Trade Margins: Insights Using State Export Data*

Cletus C. Coughlin#

and

Subhayu Bandyopadhyay+

November 18, 2015

Abstract

We adapt the heterogeneous firm trade models of Helpman, Melitz, and Rubinstein (2008) and Lawless (2010) to analyze extensive and intensive trade margins using state-level exports to foreign nations. Our theoretical analysis provides definitive predictions for the effects of changes in fixed costs, variable costs, and foreign income on the extensive margin, while for the intensive margin the predictions regarding changes in fixed costs are definitive, but the effects of changes in variable costs and foreign income are not. The number of exporting firms of a state is used to measure the extensive margin, while the intensive margin is approximated by the average firm exports of a state. Various count-data models, such as the standard negative binomial and its hurdle extension, are used to address non-trading pairs and overdispersion in the extensive trade estimations, while a Heckman correction is examined to handle sample selection issues in the intensive margin estimations. As the theory predicts, we find more consistent and statistically significant effects of changes in cost-related variables on the extensive than on the intensive margin of trade. Unlike Lawless (2010), but consistent with a truncated Pareto distribution, empirical findings suggest that variable costs reduce average exports. A noteworthy finding is that U.S. foreign direct investment has a positive effect on both margins.

JEL Codes: F10, R10

Keywords: state exports, extensive margin, intensive margin, negative binomial, hurdle model, Heckman correction

*The views expressed are those of the authors and do not necessarily represent official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System. The authors thank Diana A. Cooke, Jonas Crews, Rodrigo J. Guerrero, and David A. Lopez, for excellent research assistance, and Andrew Cassey, John Conlon, John Gardner, Martina Lawless, Andrew McCallum, and Nida Çakir Melek, for their comments.

#Senior Vice President, Federal Reserve Bank of St. Louis and Policy Associate, University of Nottingham Centre on Globalisation and Economic Policy. E-mail: Coughlin@stls.frb.org

+Research Officer, Federal Reserve Bank of St. Louis, Research Division, PO Box 442, St. Louis, MO 63166-0442, U.S.A.; and Research Fellow, IZA, Bonn, Germany. E-mail: Bandyopadhyay@stls.frb.org

1. Introduction

Relatively few firms export and these firms usually export to a small number of markets.[1] Not surprisingly, these exporting firms possess productivity advantages over non-exporting firms. The theoretical foundations for these results can be found in models of firm behavior, most notably in Melitz (2003), based on firm heterogeneity in productivity and on the costs of exporting. One implication is that a productivity threshold must be exceeded for a firm to export to a particular country. Moreover, this productivity threshold affects extensive (i.e., number of firms) and intensive (i.e., average exports per firm) trade margins. Interest in these margins, especially in the context of trade liberalization, is growing rapidly.[2]

Most of the literature on exports of heterogeneous firms has focused on the national level. The departure point of this paper is to recognize that states within a nation offer different environments to firms, which can have differential effects on exports from that state. First, we adapt the models of Helpman, Melitz, and Rubinstein (2008, we henceforth refer to this paper as HMR), and Lawless (2010), to the context of individual states' exporting to foreign nations to derive predictions for changes in parameters on exports. Then, we use state-level data to examine the potential determinants of how many firms in a state export to a specific country (i.e., the extensive margin of state exports) and the average sales of those firms to that country (an approximation for the intensive margin of a firm's exports).[3] Among other results, as the theory predicts, we find relatively more statistically significant effects of changes in cost-related variables on the extensive margin compared to the effects on the intensive margin of state exports. A notable theoretical departure of this paper is to extend our analysis to consider the truncated Pareto distribution of firm productivity levels as a special case of truncated distributions, which, following HMR, we analyze first. The truncated Pareto distribution leads to a dependence of average firm exports on variable costs - a result that contrasts with Lawless (2010). This result is supported empirically, when we find that fees associated with crossing the border have a negative effect on the intensive margin.

By examining trade margins at the level of individual states, we add to existing knowledge in various ways. First, state-level data allows for more accurate measurement of the

[1] Bernard, Jensen, Redding, and Schott (2007) found that 4 percent of 5.5 million U.S. firms in 2000 were exporters.
[2] See Lawless (2010), Markusen (2010), Kehoe and Ruhl (2009), and Berthou and Fontagné (2008).
[3] See Cassey (2011) for a summary of the state export literature.

impact of distance. For a large geographic area, such as the United States, the location from which to measure distance is unclear. State-level data shrinks the range of possibilities and the likely extent of measurement error. Second, we show the heterogeneity across states in terms of determinants and their quantitative effects. These results can be useful for state policymakers and likely produce additional questions about comparative advantage across states and the role of location. Third, we provide additional insights on Lawless' (2010) findings, especially whether trade costs have a systematic impact on the intensive margin for some states. It is possible that trade costs have little effect on average exports per firm at the national level, but could be important for a subset of states. Fourth, by examining numerous variables, we take an admittedly small step advocated by Head and Thierry (2013) toward measuring variables other than size, distance, and borders that potentially affect trade flows.

An important estimation issue in analyzing trade margins arises because frequently no firms in a state export to a specific destination. As stressed by Santos Silva and Tenreyro (2006) and HMR, disregarding and misusing the information associated with non-trading pairs can cause biased estimates.[4] For the extensive margin, non-trading pairs produce zeros, with the number of zeros being larger for smaller states.[5] Thus, count data models, such as the hurdle negative binomial model, are used to account for the multitude of zeros. The extensive-margin zeros result in missing values for their counterparts on the intensive margin. As a result, a Heckman correction is examined for the intensive trade estimations to handle the possible sample selection bias.

This paper is complementary to recent papers by Hillberry and Hummels (2008), Santos Silva and Tenreyro (2006), HMR, Lawless (2010), and Coughlin (2014). The first paper examines firm-level trade flows within the United States, while we examine U.S. international trade flows at the state level. Hillberry and Hummels (2008) find that the number of unique establishment/destination pairs declines sharply for distances up to 200 miles, with little decline thereafter. They also find that the average value of a shipment declines with distance, but the

[4] Armenter and Koren's (2014) "balls-and-bins" model is another of a recent analysis that focuses on zeros.
[5] Zeros are of particular interest because the percentage of zeros is often too high to be consistent with standard count data models and zeros often reflect corner solutions. The process generating zeros might depend on other driving forces (or respond differently to the same forces) than the process for strictly positive outcomes.

decline over (ever-longer) short distances is minimal and then the decline becomes only slightly more pronounced over longer distances.

Lawless (2010) examines the number of U.S. firms exporting to 156 countries in 2006 as a function of country size and distance. Following Eaton, Kortum, and Kramarz (2004), she decomposes exports to each country into the number of firms exporting (i.e., the extensive margin) and average export sales per firm (i.e., the intensive margin). With respect to country size, she finds a positive, statistically significant effect on both margins, with the magnitude larger for the extensive margin. For distance, she finds a negative, statistically significant impact on both margins, with the (absolute) magnitude larger for the extensive margin. She also finds that most proxies for trade costs affect only the extensive margin. Similar to Lawless (2010), we estimate basic and extended gravity models using data from 2006. In addition to estimating pooled models, we estimate separate regressions for each state and for each trade margin using foreign destination size, the distance from state to the destination, and other variables capturing trade costs. These other variables include language, infrastructure variables, geographic variables, ease of trading across borders, the destination's legal environment, and the existence of both formal and informal networks. The impact of foreign direct investment on the extensive and intensive trade margins is examined as well. The pooled estimation also generates insights concerning the role of state size.

Coughlin (2014) explores changes in state-level trade during the Great Trade Collapse and the subsequent rebound. He found that adjustments on the intensive margin occurred to a much greater extent than on the extensive margin during both the collapse and the rebound. In other words, for the majority of states, relatively larger changes in exports per firm than in the number of exporting firms were observed.

The remainder of the paper is arranged as follows. Section 2 presents the theoretical model underlying the empirical analysis. Section 3 describes the data used in the empirical analysis. Section 4 discusses the estimated models and the results. Section 5 completes the paper with a summary of the key contributions of the analysis.

2. Theory: Heterogeneous Firms, Trade Costs, and State Exports

We adapt the model of HMR to the context of state exports, and also follow Lawless (2010) in terms of the theoretical developments.[6] The origins of these models can be found in Melitz (2003) and in Chaney (2008). Key factors driving the results are that firms differ in their productivity and, to export, must incur fixed as well as variable export costs. As a result, not all (empirically few) domestic producers are also exporters and, in fact, trade flows between potential origins and destinations are often zero.

In the tradition of the recent trade literature, we consider heterogeneous monopolistically competitive firms producing different varieties within an industry. US states are indexed by i (i=1, 2…50), and foreign nations by j (j=1,2,…M). Following HMR, we assume that consumers in nation j consume a continuum of products indexed by k, where the set of products available for consumption in that nation is B_j. The standard utility function characterizing consumer preferences in the foreign nation is

$$U_j = \left[\int_{k \in B_j} x_j\left(k\right)^{\frac{\varepsilon-1}{\varepsilon}} dk \right]^{\frac{\varepsilon}{\varepsilon-1}}, \ \varepsilon > 1, \tag{1}$$

where ε is a constant elasticity of substitution between products. Utility maximization subject to budget constraint yields the following demand function for product k in nation j

$$x_j\left(k\right) = \frac{p_j\left(k\right)^{-\varepsilon} Y_j}{P_j^{1-\varepsilon}}, \tag{2}$$

where Y_j is nation j's total expenditure (given exogenously), and P_j is its Dixit-Stiglitz aggregate price index, such that:

[6] The key difference of the following analysis from Lawless (2010) is that along the lines of HMR, we consider a truncated probability distribution characterizing firm productivities. Accordingly, although the qualitative results of the analysis are generally similar to Lawless (2010), some important differences emerge in the special case of the truncated Pareto distribution that we consider in the last part of this section.

$$P_j = \left[\int_{k \in B_j} p_j(k)^{1-\varepsilon} dk \right]^{\frac{1}{1-\varepsilon}}. \tag{3}$$

We assume that the mass of firms in the industry in state *i* is given exogenously at $\bar{N}_i$, and that there is no fixed cost in producing for the domestic market.[7] We assume that the marginal input cost of any product produced by a state *i* firm is a constant c_i, which is not necessarily equal across different U.S. states. Also, assume that a state *i* producer of product *k* has a productivity level $a(k)$, such that its marginal cost of production is $c_i / a(k)$.

For this firm to export its product to nation *j*, in addition to the standard variable cost, it has to incur two additional costs. First, there is an iceberg type tariff/transportation cost, such that for each unit reaching the foreign market, the firm needs to produce $t_{ij}(>1)$ units, where $t_{ij}-1$ units melt away between production and delivery in the export market. There is a fixed cost F_{ij} faced by all firms of state *i* in order to export their product to nation *j*. Limiting our focus to exporting firms, the profit from exports to nation *j* of firm *k* is:

$$\pi_{ij}(k) = p_{ij}(k)x_{ij}(k) - \frac{c_i}{a(k)} t_{ij} x_{ij}(k) - F_{ij}, \tag{4}$$

where $x_{ij}(k)$ is the level of exports to nation *j* of firm *k* of state *i*.

The demand function in Equation (2) implies that this firm perceives its price elasticity of demand in the export market as ε, which follows from the fact that for a sufficiently large set of consumption goods in the foreign nation, this firm's effect on the aggregate price level of the foreign nation is negligible. Using this fact, marginal revenue - marginal cost equalization yields the profit maximizing export and price levels for the firm as:

$$p_{ij}(k)\left(1 - \frac{1}{\varepsilon}\right) = \frac{c_i}{a(k)} t_{ij} \Rightarrow p_{ij}(k) = \frac{\varepsilon \tau_{ij}}{(\varepsilon - 1)a(k)}, \tau_{ij} = t_{ij} c_i. \tag{5}$$

[7] This follows HMR, and, as in their paper, this immediately implies that all domestic firms produce for the domestic market because the monopolistically competitive domestic price includes a markup over marginal cost, and, in addition, there is no fixed cost for domestic sales.

Plugging the price level obtained from Equation (5) into Equation (4), and using Equation (2), we get the firm's (optimized) profit from exports as:

$$\pi_{ij}\left(k\right)=\left(\frac{\tau_{ij}}{a\left(k\right)}\right)^{1-\varepsilon}\frac{\mu Y_j}{P_j^{1-\varepsilon}}-F_{ij}\text{, where }\mu=\varepsilon^{-\varepsilon}\left(\varepsilon-1\right)^{\varepsilon-1}. \tag{6}$$

Using Equation (6), we find that positive (or zero) export profit (i.e., $\pi_{ij}\left(k\right)\geq 0$) can be obtained if and only if:[8]

$$a\left(k\right)\geq\tilde{a}_{ij}=\frac{\tau_{ij}}{P_j}\left(\frac{F_{ij}}{\mu Y_j}\right)^{\frac{1}{\varepsilon-1}}\Rightarrow\frac{\partial\tilde{a}_{ij}}{\partial\tau_{ij}}>0\text{, and, }\frac{\partial\tilde{a}_{ij}}{\partial F_{ij}}>0, \tag{7}$$

where $\tilde{a}_{ij}$ is the minimum (or threshold) productivity level required for a firm from state *i* to profitably export to nation *j*. Now, the export revenue of a firm of state *i* from its exports to nation *j* is:

$$s_{ij}\left(a\left(k\right)\right)=p_{ij}\left(k\right)x_{ij}\left(k\right)=\left[\frac{P_j\left(\varepsilon-1\right)a\left(k\right)}{\varepsilon\tau_{ij}}\right]^{\varepsilon-1}Y_j\Rightarrow\frac{\partial s_{ij}\left(a\left(k\right)\right)}{\partial\tau_{ij}}<0, \tag{8}$$

which means that as the variable costs of exporting from state *i* to nation *j* increase, the export revenue of a state *i* firm from its exports to nation *j* must fall.

We assume that the productivity level a is distributed as a truncated probability distribution, characterized by a probability density function $g\left(a\right)$, with support $\left(a_L, a_H\right)$, where $a_L < a_H$.[9] Let E_{ij} be the sum of export revenues from exports to nation *j* by firms of state *i*. Noting that only firms above the threshold productivity level $\tilde{a}_{ij}$ export to nation *j*, and also that the mass of firms from state *i* is $\bar{N}_i$, state *i*'s aggregate export revenues are:

[8] We assume that the foreign nation's price index includes prices of a large basket of goods from its own firms as well as from firms from nations other than the US, such that we can take this price level as given with respect to changes in τ_{ij}.

[9] This distribution, as noted in HMR, presents a more realistic picture of heterogeneous firms, such that no infinitely productive firm needs to be assumed.

$$E_{ij} = \int_{\tilde{a}_{ij}}^{a_H} s_{ij}\left(a\right)\bar{N}_i g\left(a\right)da \text{, for } \tilde{a}_{ij} < a_H \text{; } E_{ij} = 0 \text{, otherwise.} \tag{9}$$

Analogously, the total number of firms from state *i* exporting to nation *j* is:

$$N_{ij} = \bar{N}_i \int_{\tilde{a}_{ij}}^{a_H} g\left(a\right)da = \bar{N}_i\left[1 - G\left(\tilde{a}_{ij}\right)\right], \tag{10}$$

Where $G\left(a\right)$ is the cumulative density function associated with $g\left(a\right)$. In other words, $1 - G\left(\tilde{a}_{ij}\right)$ represents the fraction of firms of state *i* who export to nation *j*.

Now, consider any parameter θ that can affect the trading equilibrium. Also, let us focus on states that have non-zero exports before and after the change in this parameter. The change in the number of exporting firms can be obtained by differentiating Equation (10), to yield:

$$\frac{\partial N_{ij}}{\partial \theta} = -\bar{N}_i g\left(\tilde{a}_{ij}\right)\frac{\partial \tilde{a}_{ij}}{\partial \theta}. \tag{11}$$

Equation (11) shows that a change in θ potentially alters the threshold productivity level $\tilde{a}_{ij}$, leading to a smaller (larger) equilibrium number of state *i* firms exporting to nation *j*, if the threshold is raised (lowered).

The change in aggregate export revenues of state *i*'s firms from exports to nation *j* is:

$$\frac{\partial E_{ij}}{\partial \theta} = \bar{N}_i \int_{\tilde{a}_{ij}}^{a_H} \frac{\partial s_{ij}}{\partial \theta} g\left(a\right)da - \bar{N}_i s_{ij}\left(\tilde{a}_{ij}\right)g\left(\tilde{a}_{ij}\right)\frac{\partial \tilde{a}_{ij}}{\partial \theta}. \tag{12}$$

The first term on the right-hand-side of Equation (12) is the change in exports of state *i* to nation *j* due to a change in the export revenues of an existing exporting firm, while the second term is the change in these exports due to the entry (or exit) of firms from state *i* from nation *j*'s market, because of a change in the relevant threshold productivity level.

Finally, consider average export revenue of firms from a state from their exports to nation *j*. This is given by E_{ij} / N_{ij}. The effect of a change in θ on this average is:

$$\frac{\partial\left(E_{ij}/N_{ij}\right)}{\partial\theta}=\frac{N_{ij}\dfrac{\partial E_{ij}}{\partial\theta}-E_{ij}\dfrac{\partial N_{ij}}{\partial\theta}}{N_{ij}^2}. \tag{13}$$

Using Equations (11) and (12) in Equation (13), we get:

$$\frac{N_{ij}^2}{\bar{N}_i}\left[\frac{\partial\left(E_{ij}/N_{ij}\right)}{\partial\theta}\right]=N_{ij}\int_{\tilde{a}_{ij}}^{a_H}\frac{\partial s_{ij}}{\partial\theta}g\left(a\right)da+\left[E_{ij}-N_{ij}s_{ij}\left(\tilde{a}_{ij}\right)\right]g\left(\tilde{a}_{ij}\right)\frac{\partial\tilde{a}_{ij}}{\partial\theta}, \tag{14}$$

where $E_{ij}-N_{ij}s_{ij}\left(\tilde{a}_{ij}\right)>0$. This is because $N_{ij}s_{ij}\left(\tilde{a}_{ij}\right)$ is the export revenue of the least productive firm of state *i* scaled by the total number of the state's firms exporting to *j*, and hence must be lower than the aggregate state *i* export revenues from trade with nation *j* $\left(i.e., E_{ij}\right)$. This is because E_{ij} also includes export revenues of firms whose productivities are above the cut-off level $\tilde{a}_{ij}$.

Now, let us consider three specific applications of the comparative statics implied by the analysis above. We first consider a change in the fixed cost F_{ij}, then turn to the effect of a change in the variable cost parameter τ_{ij}, and finally consider the effect of a change in nation *j*'s income level Y_j.

Case 1: Change in fixed exporting cost $F_{ij}\left(i.e., d\theta=dF_{ij}\right)$

From Equation (8), we can see that $\dfrac{\partial s_{ij}}{\partial F_{ij}}=0$ for all firms. Therefore, there is no change in exports due to the scale of operation of existing exporting firms. Effects of the change of the fixed cost on the extensive margin can be analyzed using Equations (11) and (12), which yield:

$$\frac{\partial N_{ij}}{\partial F_{ij}}=-\bar{N}_i g\left(\tilde{a}_{ij}\right)\frac{\partial\tilde{a}_{ij}}{\partial F_{ij}}<0\text{, and, }\frac{\partial E_{ij}}{\partial F_{ij}}=-\bar{N}_i s_{ij}\left(\tilde{a}_{ij}\right)g\left(\tilde{a}_{ij}\right)\frac{\partial\tilde{a}_{ij}}{\partial F_{ij}}<0. \tag{15}$$

Thus, the number of firms exporting from state *i* to nation *j*, as well as the state's aggregate export revenues from sales to nation *j* must fall in response to a rise in the fixed cost F_{ij}. Using Equation (14) we get:

$$\frac{\partial\left(E_{ij}/N_{ij}\right)}{\partial F_{ij}}=\frac{\bar{N}_i\left[E_{ij}-N_{ij}s_{ij}\left(\tilde{a}_{ij}\right)\right]g\left(\tilde{a}_{ij}\right)\frac{\partial\tilde{a}_{ij}}{\partial F_{ij}}}{N_{ij}^2}>0, \qquad (16)$$

We summarize the empirically observable implications of this analysis as the following:

A rise in the fixed exporting cost F_{ij} should reduce state i's aggregate export revenues from sales to nation j, and reduce the number state i firms exporting to nation j, but increase the average export revenues from sales to nation j.

Case 2: Change in variable exporting cost due to change in transportation cost t_{ij} or change in state level variable production cost (i.e., $d\theta=d\tau_{ij}$; where $d\tau_{ij}=t_{ij}dc_i$, or $d\tau_{ij}=c_i dt_{ij}$).

Using Equations (7) and (8) in Equations (11) and (12):

$$\frac{\partial N_{ij}}{\partial\tau_{ij}}=-\bar{N}_i g\left(\tilde{a}_{ij}\right)\frac{\partial\tilde{a}_{ij}}{\partial\tau_{ij}}<0\text{, and }\frac{\partial E_{ij}}{\partial\tau_{ij}}=\bar{N}_i\int_{\tilde{a}_{ij}}^{a_H}\frac{\partial s_{ij}}{\partial\tau_{ij}}g\left(a\right)da+s_{ij}\left(\tilde{a}_{ij}\right)\frac{\partial N_{ij}}{\partial\tau_{ij}}<0. \qquad (17)$$

Thus, a rise in variable cost τ_{ij} reduces both the number of state *i* firms exporting to nation *j*, and also the aggregate export revenues of the state from nation *j*. Using Eq. (14), the effect on average export revenue from sales to nation *j* is:

$$\frac{N_{ij}^2}{\bar{N}_i}\left[\frac{\partial\left(E_{ij}/N_{ij}\right)}{\partial\tau_{ij}}\right]=N_{ij}\int_{\tilde{a}_{ij}}^{a_H}\frac{\partial s_{ij}}{\partial\tau_{ij}}g\left(a\right)da+\left[E_{ij}-N_{ij}s_{ij}\left(\tilde{a}_{ij}\right)\right]g\left(\tilde{a}_{ij}\right)\frac{\partial\tilde{a}_{ij}}{\partial\tau_{ij}}. \qquad (18)$$

The first term on the right-hand side of Equation (18) is negative, reflecting shrinkage of the average scale of firm exports to nation *j* due to a reduction on the intensive margin of each firm. The last term in Equation (18) is positive, reflecting an expansion of the average scale due to exit of the least productive firms - which tends to raise average scale of surviving firms. The net effect of these two terms is ambiguous in general, but it may be possible to determine which effect dominates if we use some specific probability density functions, such as the one associated with the truncated Pareto distribution. We present that analysis at the end of this section. Empirically observable implications of this analysis are that:

An increase of the variable cost parameter τ_{ij} should reduce state i's export revenues from sales to nation j, and reduce the number of the state's firms exporting to nation j, while its effect on average export revenues from sales to nation j is ambiguous, in general.

Case 3: Change in foreign income (i.e., $d\theta = dY_j$)

The analysis of this case is similar to Case 2, and hence we focus our discussion on the underlying intuition and results rather than providing the formal proof, which is available upon request. The first thing to notice is that the profit maximizing export price for each product sold to foreign nation *j* must be independent of foreign income Y_j. This is because, given ε, Eq. (5) ties down the export price of product *k* to a constant markup above the effective marginal cost of the exporting the marginal unit to nation *j*, which is $\tau_{ij} / a(k)$. Furthermore, Equation (2) shows that demand for each product rises proportionally with foreign income. Thus, a rise in Y_j must raise both the scale of exports and the export revenues of each existing domestic firm selling to foreign nation *j* (i.e., an increase on the *intensive margin*). In addition, Eq. (6) reveals that all domestic firms' (including previously non-exporting ones) potential profit from exporting to nation *j* must rise because price, marginal cost, and fixed cost do not change, but the scale of potential export to nation *j* (i.e., x_{ij}) increases. This rise in potential profit brings some previously non-exporting firms of each state into the market for exports to nation *j* (an increase on the *extensive margin*). However, analogous to the previous case explored, average exports E_{ij} / N_{ij} may or may not rise, because relaxation of the extensive margin introduces some less productive firms, which pulls average scale in the negative direction.

A rise in foreign income Y_j should increase each state's aggregate export revenues from sales to nation j, and increase the number of each state's firms who export to nation j, while the effect on a state's average export revenue from sales to nation j is ambiguous.

The general truncated distribution used until this point yields unambiguous results for all cases except for the effect of variable cost parameter τ_{ij} and of foreign income level Y_j on average export revenues E_{ij} / N_{ij}. To shed more light on this issue, we conclude this section by analyzing the special case of a truncated Pareto distribution (see HMR, who also use this distribution). Using this distribution, we show that the result obtained by Lawless (2010) that average export revenues are independent of the variable cost parameter (under Pareto) is a special case of the truncated Pareto distribution. This special case requires a_H (the highest productivity level associated with any probability mass) to tend to infinity. If this condition is not met, average export revenues are indeed dependent on the variable cost parameter as well as on the foreign income level Y_j.

Truncated Pareto Distribution

Consider a truncated Pareto distribution such that:

$$g\left(a\right)=\frac{\phi a_L^{\phi} a_H^{\phi} a^{-\phi-1}}{a_H^{\phi}-a_L^{\phi}}\text{; and } G\left(a\right)=\frac{a_H^{\phi}\left(1-a_L^{\phi}a^{-\phi}\right)}{a_H^{\phi}-a_L^{\phi}};\ \phi>0. \tag{19}$$

Using Equations (9) and (10):

$$\frac{E_{ij}}{N_{ij}}=\frac{\int_{\tilde{a}_{ij}}^{a_H} s_{ij}\left(a\right)g\left(a\right)da}{1-G\left(\tilde{a}_{ij}\right)}\quad . \tag{20}$$

Using Equation (19), Equation (20) can be reduced to:

$$\frac{E_{ij}}{N_{ij}}=\frac{\phi\varepsilon F_{ij}\left[1-\left(\tilde{a}_{ij}/a_H\right)^{1+\phi-\varepsilon}\right]}{\left(1+\phi-\varepsilon\right)\left[1-\left(\tilde{a}_{ij}/a_H\right)^{\phi}\right]}. \tag{21}$$

As $a_H\to\infty$, for a finite $\tilde{a}_{ij}$, $\tilde{a}_{ij}/a_H\to 0$, and we get $\frac{E_{ij}}{N_{ij}}\to\frac{\phi\varepsilon F_{ij}}{1+\phi-\varepsilon}$, which is the expression obtained in Lawless (2010), where the average export revenue is independent of variable cost parameter τ_{ij} and the foreign income level Y_j. However, if a_H is finite, such that $\tilde{a}_{ij} / a_H$ does not approach zero, $\frac{E_{ij}}{N_{ij}}$ is a function of $\tilde{a}_{ij} / a_H$, which is, in turn, a function of τ_{ij} and also of Y_j. This can be established easily by considering an example where $\phi = \varepsilon = 2$. Using these parameters, Equation (21) yields:

$$\frac{E_{ij}}{N_{ij}} = \frac{4F_{ij}}{1+\left(\tilde{a}_{ij} / a_H\right)}. \tag{22}$$

As τ_{ij} rises, $\tilde{a}_{ij}$ rises, and as Y_j rises, $\tilde{a}_{ij}$ falls (see Equation 7). Given F_{ij} and a_H, Equation (22) implies that a rise in τ_{ij} leads to a fall in $\frac{E_{ij}}{N_{ij}}$, while a rise in Y_j must raise $\frac{E_{ij}}{N_{ij}}$. This example proves that the independence result of Lawless (2010) does not carry over to the case of a truncated Pareto distribution where a_H does not tend toward infinity. Our empirical analysis throws more light on the directions of this dependence between variable costs and export revenues on the one hand, and between the foreign income level and export revenues on the other.

3. Data

The following analysis is focused on two dependent variables ─ one for the extensive margin and one for the intensive margin ─ and their relationships with numerous independent variables. Given 50 states and 187 countries in the sample, the maximum number of unique observations for a variable is 9350. However, the number of usable observations is generally less than 9350 because some variables are only available at the state or country level and some values for variables are withheld for confidentiality reasons. The variables, which pertain to 2006 unless otherwise indicated, are listed and defined in Table 1. Summary statistics in levels for these variables are contained in Table 2. All data using dollars are denominated in 2005 U.S. dollars.

For the dependent variables ─ the number of firms and their average export sales by country ─ the paper relies on the U.S. Census Bureau's *Profile of U.S. Exporting Companies*.

Rather than the nation as a whole as the basic geographic unit, we use the number of firms in an individual state *s* exporting to a specific country *c*, $Firms_{s,c}$, and the average exports of those firms to a specific country, $Exportsperfirm_{s,c}$. The state focus also produces numerous state-country pairs (1598 in total) that have a value of zero for both margins. In addition, 2155 observations for total exports, $Exports_{s,c}$, and average exports, $Exportsperfirm_{s,c}$, are withheld for confidentiality reasons when a small number of a state's firms export to a country.

Across states, the variation in *Firms* is much larger than in *Exportsperfirm*. For example, a comparison of the 5 largest states using gross state product with the 5 smallest states shows that the ratio of the average number of firms is nearly 40, while the ratio of average exports per firm is 1.3. Comparing the 25 largest with the 25 smallest states, the ratio of the average number of exports declines to 7.3, while the ratio of the average exports per firm declines to 1.1.

Turning to the independent variables, variables to capture demand and trade costs in export destinations are used. Based on the theoretical model, the impacts of these variables on the extensive margin are usually definitive, but their impacts on the intensive margin are usually ambiguous.

Starting with demand, the gross domestic product of the destination country for a state's exports, GDP_c, is used as a proxy for market size. A second variable possibly affecting the demand for exports of U.S. producers relates to U.S. foreign direct investment in the destination market, FDI_c. Through various demand and cost mechanisms, trade can be affected by the internal networks of multinational firms.[10] Whether this investment will affect trade flows measured at the state level and whether foreign direct investment complements or substitutes for trade flows is uncertain on theoretical grounds. Moreover, this uncertainty extends to the extensive margin as it is possible that increased foreign direct investment might be associated with a smaller number of exporting firms in a state.

All other independent variables can be viewed as proxies for trade costs. These proxies include measures associated with geography (i.e., natural trade frictions) and those that result from government policies (i.e., man-made trade frictions). In the following discussion we identify some independent variables that do not appear in the reported results, either because

[10] See Bernard, Jensen, Redding, and Schott (2010) for a discussion of intra-firm trade and product contractibility.

they were not found to be statistically significant or were highly collinear with measures that performed better.

As is standard in gravity models, we include a distance measure that is the distance from the largest city in an exporting state to the largest city in the importing country, $Distance_{s,c}$. Transportation costs should increase the farther the distance between a state and a country; however, despite the importance of geography for transportation costs, geographic distance is only a first rough approximation.[11]

Another variable thought to affect trade costs, expressed as a 0/1 dummy, is whether the export destination uses English as an official language, $English_c$. If the destination country also uses English as an official language, trade costs should tend to be lower because of increased ease of communication between buyers and sellers. In addition to language, we examine the impact of two communication infrastructure variables. $Phones_c$ is the number of mobile cellular subscriptions per 100 people in country c and $Internet_c$ is the number of internet users per 100 people in country c. Higher values of these variables should make it easier to acquire information and transact business.

Two geographic variables thought to affect trade costs ─ population density, $Popdensity_c$, and area, $Landarea_c$, are examined. These variables are thought to capture the internal geography of the export destination that affects the costs of serving the market. Countries with larger population densities or smaller geographic areas should tend to have lower trade costs because of lower distribution network and internal transportation costs.

Trade costs are also affected by government policies and actions. $TradeFreedom_c$ is an index that measures the extent to which goods moving across international borders are unaffected by tariffs and non-tariff barriers.[12] Higher values of this index correspond to less costly trade impediments.[13]

[11] Despite the common practice of using distance as a proxy for transport costs, it is imperfectly correlated with transport costs. For examples, see Clark (2007) and Giuliano, Spilimbergo, and Tonon (2014). Studies focused on transport costs provide a more complete view that goes beyond distance. See Korinek (2011) for references.
[12] For construction details, go to www.heritage.org/index/Trade-Freedom.
[13] A trade restrictiveness index is the preferred index from a theoretical perspective, but the trade freedom index is available for more countries. See Coughlin (2010) for an elementary discussion of trade restrictiveness indices.

In addition to tariffs and non-tariff barriers, governments can affect the costs of trading across borders through various administrative and bureaucratic measures. Three specific variables are examined.[14] First, documents must be completed as part of processing imports at the port of entry. The number of required documents per imported shipment, $Tradedocuments_c$, includes the documents required by various entities, such as government ministries, custom authorities, port and container terminal authorities, health and technical control agencies, and banks. Second, time is required to move a shipment from arrival through the port. This time, denoted by $Timeprocess_c$, is in terms of the number of calendar days. Third, there are fees associated with the process of crossing the border. $Tradefees_c$ are the official, administrative fees of importing, excluding tariffs and trade taxes. These fees include the costs for documents, administrative fees for customs clearance and technical control, customs brokers' fees, terminal handling charges, and costs for inland transportation. Larger values of each of these measures, which can be viewed as transport costs unrelated to distance, should be related positively to trade costs.[15]

This paper also considers two measures of governance because stronger legal institutions might facilitate international trade. First, the rule of law, $Legal_c$, measures the confidence of agents in the legal system. Included in this measure is the quality of contract enforcement, property rights, the police, and the courts as well as the likelihood of crime and violence. A second measure, $CorruptionControl_c$, focuses on corruption. This measures the perceptions of agents about the extent to which public power is used for private gain and the effective capture of government by elites and private interests. For both measures, higher values indicate better governance and, thus, should reflect lower trade costs.

The final trade cost variable follows directly from the literature examining the effect of information networks on trade. At the state level, empirical analysis has frequently found a positive connection between a state's exports to a specific country and the state's number of

[14] These variables are found in the World Bank's *Doing Business Survey*; see Djankov, Freund, and Pham (2010).
[15] Clark (2007) notes that up to 75 percent of international transport costs can be invariant with respect to distance.

residents born in that country.[16] Trade from a state to a country might be less costly the larger the number of residents in a state that were born in the destination country, $Foreignborn_{s,c}$.

4. Estimation and Results

As discussed previously, separate models of the extensive and intensive trade margins for each state are estimated. Because the underlying distributions of the number of exporting firms per country and the exports per firm are much different, the appropriate estimation technique also differed. This section begins by discussing the extensive margin.

Estimation of the Extensive Margin

For the estimation of the extensive margin, this paper estimates various count-data models. The models, based on Poisson (P) and negative binomial (NB) distributions, address two major issues. One issue is overdispersion (i.e., variance exceeding the mean) and the other is zeros, possibly excess zeros (i.e., the frequency of zeros exceeds the predicted frequency based on single-equation estimations).[17]

In Figure 1 a representative histogram is shown. The horizontal axis shows the number of firms in Arkansas exporting to each of 187 countries. Along the vertical axis, the density of the number of firms is shown. In Arkansas, there are no exporting firms for 51 (or 27 percent) of the 187 potential importing countries. The general shape of the density function is consistent with a Poisson distribution as the density generally declines with rightward movements along the horizontal axis. This general shape is also demonstrated by larger states; however, much less density is associated with the smallest numbers on the horizontal axis.

In the present case, for a given state we examine the number of the state's firms exporting to numerous countries *c* (*c=1...n*). The expected number of a state's exporting firms to these countries is $\lambda_1, \lambda_2, ... \lambda_n$. Almost without exception, histograms, such as the one shown in Figure 1, and statistical tests indicate overdispersion. In other words, the variance greatly exceeds the mean. Thus, as a starting point, we estimate the following negative binomial model:

---

[16] Bandyopadhyay, Coughlin, and Wall (2008) summarize this literature. Coughlin and Wall (2011) find evidence that ethnic networks are associated with increased trade on the intensive margin, but not on the extensive margin.
[17] For a textbook introduction to handling zeros in count data models, see Winkelmann (2008), Chapter 6. For an application using various estimation models to examine excess zeros, see Gurmu and Trivedi (1996).

$$\ln \lambda_c = \beta' x_j + \varepsilon_j . \tag{23}$$

Where β is a parameter vector to be estimated, x_j is a vector of observable characteristics (including a constant) that influences the number of firms exporting to a country, and ε_j is gamma distributed with mean 1.0 and variance alpha. This allows the variance to exceed the mean.

With a standard count model in the current context, a Poisson or a negative binomial model attempts to identify the covariates that statistically explain the number of a state's firms that export to a particular country. While the negative binomial model deals with the issue of overdispersion, the issue of excess zeros remains a possibility.[18] To confront excess zeros, two approaches that utilize dual regime data generating processes are employed. Both approaches have theoretical appeal from an economics perspective.

With a hurdle model the first stage is a binary choice model, such as a logit model, which examines whether or not a state exports to a country.[19] In other words, is there a level of productivity sufficiently high in a state to compensate for the additional costs of exporting to a country? In the second stage a count model, such as a Poisson or negative binomial model, truncated at zero is used to estimate how many of a state's firms export to a country.[20] In other words, how many firms have productivity sufficiently high to overcome the additional trade costs?

The second approach relies on zero-inflated models.[21] In the first stage, a binary choice model is employed to differentiate between structural and sampling zeros. In this context, structural zeros occur when there is no chance of exports from a state to a country, while sampling zeros occur when exports are possible, but there are no exports during the sample period.[22] Those states with no chance of exporting do not have productivity sufficiently high to

[18] If the overdispersion results from a large number of zeros in the data, then the results produced by a negative binomial regression model will be biased.
[19] Hurdle count data models, first discussed in Mullahy (1986), are used frequently in health-care studies. For examples, see Gerdtham (1997), Álvarez and Delgado (2002), and Lahiri and Xing (2004).
[20] See Winkelmann (2008), Chapter 6, for an overview of handling zeros in count data models.
[21] The seminal article on zero-inflated models is Lambert (1992). For more recent examples, see Wang (2003) and Burger, Van Oort, and Linders (2009).
[22] Zero-inflated models are simply adding mass to the zero point relative to standard count models.

overcome the additional costs of exporting, while the remaining zeros reflect randomness. In the second stage, a standard, non-truncated, count model for the number of exporters from a state to a country is estimated.

Thus, in addition to a standard negative binomial (NB) count model, we estimate hurdle models using Poisson (HP) and negative binomial (HNB) count models and zero-inflated models using Poisson (ZIP) and negative binomial (ZINB) count models. Based on goodness-of-fit measures, the results indicate that negative binomial estimation for standard gravity models outperform the Poisson counterparts. In comparisons across three binomial models – standard negative binomial, zero-inflated negative binomial, and hurdle negative binomial – the results indicate that the hurdle negative binomial models generally perform best for pooled estimations, but that standard negative binomial models perform the best for the individual state estimations. Because of these findings, we present the pooled results for the HNB model and the individual states results for the NB model.

Extensive Margin: Results

Three sets of results are presented. The first set of results is generated by pooling the data for all states. Thus, the parameter estimates, after controlling for state fixed effects or state size using gross state product, are assumed to be the same across states. Both a simple, or benchmark, gravity model that includes as independent variables only the gross domestic product of the importing country, GDP_c, and its distance from the exporting state, $Distance_{s,c}$, and an extended gravity model that incorporates additional explanatory variables are presented and examined. The remaining two sets of results, one for a simple gravity and one for an extended gravity model, are based on separate regressions for each state, which allows the parameter estimates to vary across states.

Pooled Estimates. The results of estimating hurdle negative binomial models of the extensive margin using pooled state-level data with state fixed effects (columns 1-3) or gross state product (columns 4-6) for the simple and extended gravity models are listed in Table 3. Heteroskedasticity-robust standard errors underlie the estimates in columns 1, 2, 4, and 5, while cluster-robust standard errors underlie the estimates in columns 3 and 6. The clustering is based on destination country. Several results stand out. First, the statistical significance of the dispersion parameter, *alpha*, indicates that a negative binomial model is appropriate. Second, the likelihood ratio tests comparing the hurdle with the standard negative binomial results shown in the last row of Table 3 indicate that the improvement from using a hurdle model is statistically significant. [23]

Turning to the results in columns 1 and 4 in Table 3 for the simple gravity model, one finds similar results using state fixed effects and gross state product. Gross domestic product of the importing country, GDP_c, is a positive, statistically significant determinant of the number of exporters. The coefficient estimates, 0.74 and 0.73, are roughly identical for the two models. Meanwhile, the distance between the importing country and exporting state, $Distance_{s,c}$, is a negative, statistically significant determinant of the number of exporters. The coefficient estimate varies only slightly between models, -0.90 using state fixed effects and -0.99 using gross state product.

Turning to the extended model, no surprising results appear in columns 2 and 5 of Table 3. All variables exhibit their expected sign and are statistically significant. However, the (absolute) estimates for both gross domestic product and distance decline relative to the results for the simple gravity model. Using state fixed effects, the estimate for gross domestic product declines from 0.74 to 0.44, while, using gross state product, *GSP*, the estimate declines from 0.73 to 0.43. The coefficient estimate for distance declines in absolute terms from -0.90 to -0.71 using state fixed effects and from -0.99 to -0.78 using gross state product.

Turning to the additional independent variables, regardless of the use of state fixed effects or gross state product, U.S. foreign direct investment in a destination country, FDI_c, is positively related to the number of state exporters. This result suggests a complementarity between state exporters and U.S. foreign direct investment. On the cost side, destinations in which English is an official language, $English_c$, seem to contribute to an increased number of state exporters. This finding is consistent with Lawless (2010) and many other studies.

Similar to Lawless (2010), an increased number of state exporters also tends to be associated with a more developed communications infrastructure. The number of mobile cellular

[23] Ferrante and Novelli (2013) illustrates a hurdle negative binomial approach in the context of export activity.

subscriptions per 100 people, *Phones*, is generally associated with more exporters. The positive and statistically significant results for a geographic variable, *Popdensity*, are similar to those in Lawless (2010). Finally, also similar to Lawless (2010), higher levels of official, administrative fees, *Tradefees*, decrease the number of state exporters.

A final set of results in columns 3 and 6 shows the impact of estimations producing cluster-robust standard errors rather than simply heteroskedasticity-robust standard errors. Coefficient estimates are, of course, unaffected. The major change is that the standard errors based on clustering exceed the corresponding heteroskedasticity-robust standard errors. Most variables remain statistically significant; however, two variables that were previously identified as statistically significant, *Phones* and *Popdensity*, are no longer statistically significant.

Turning to some additional results, details on state fixed effects are not reported in Table 3. Wyoming is used as the base. Without exception, the state fixed effects are statistically significant. An additional finding shown in column 4-6 is that larger states tend to have more exporting firms. More noteworthy, however, is the magnitude of the estimate, roughly 1.20 regardless of the model, which suggests that a given percentage increase in state size is associated with a more-than-proportionate increase in exporting firms.

Simple Gravity Model – Estimates by State. The extensive-margin results for the simple gravity model are presented in Table 4. The estimates are based on a standard negative binomial model because this model generally outperformed the hurdle negative binomial model at the state level. For example, hurdle models were significantly better than the standard negative binomial model in only 12 states using a likelihood ratio test and hurdle models had a lower AIC than their counterpart standard model in only 15 states. Using 187 observations for each state, the pseudo R^2 values range from 0.10 in Florida to 0.26 in Montana, North Dakota, and Wyoming. The dispersion parameter (*alpha*), which provides a test of the negative binomial model versus a Poisson model, is statistically significant for all 50 states.

The results for GDP_c and $Distance_{s,c}$ are consistent with expectations and with Lawless (2010). For each state, GDP_c is a positive, statistically significant determinant of the number of exporting firms, while $Distance_{s,c}$ is a negative, statistically significant determinant. The mean estimate for the coefficient on GDP_c is 0.74, with a range from 0.59 for Florida to 0.86 for New

Hampshire and West Virginia. Meanwhile, the mean estimate for the coefficient on $Distance_{s,c}$ is -1.00, with a range from -4.18 for Hawaii to -0.49 for New Hampshire. Other relatively large (in absolute value) coefficients are estimated for Alaska (-2.24) and Florida (-2.01).

Extended Gravity Model – Estimates by State. Turning to the extensive-margin results summarized in Table 5 for the extended gravity model, one sees a richer set of results.[24] Similar to the results for the simple gravity model, the standard negative binomial model generally outperformed the hurdle negative binomial model. Based on the likelihood ratio test, hurdle models in only 12 states were significantly better than the standard model. Moreover, hurdle models in only 11 states had a lower AIC than their counterpart standard model. Also, for every state except Alaska a likelihood ratio test indicated that the increase in explanatory power of the extended relative to the simple gravity model was statistically significant. Once again, the results for the individual state regressions matched expectations and were consistent with Lawless' (2010) findings. The sample size was 136 because some independent variables were unavailable. The pseudo R^2 values ranged from 0.12 in Florida to 0.35 in Montana. The dispersion parameter was statistically significant in every state.

For each state, identical to the results for the simple model, GDP_c is a positive, statistically significant determinant of the number of exporting firms, while $Distance_{s,c}$ is a negative, statistically significant determinant. Relative to the estimates for the simple model, the estimates tend to be absolutely smaller. The mean estimate for the coefficient on GDP_c is 0.46, with a range from 0.29 for Florida to 0.59 in New Mexico. Meanwhile, the mean estimate for the coefficient on $Distance_{s,c}$ is -0.79, with a range from -3.45 for Hawaii to -0.25 in New Hampshire.[25] The relatively large differences between the coefficient for *Distance* in the extended model and the simple model likely reflect some of the problems associated with using *Distance* as a proxy for transport costs.

Turning to the additional independent variables, U.S. foreign direct investment in a destination country, FDI_c, is positively related to the number of state exporters. This variable, whose parameter estimates range from 0.08 for North Dakota and Virginia to 0.27 for Wyoming,

---

[24] State-by-state results are available upon request.

[25] For Lawless' (2010) preferred extended gravity model, the coefficient on GDP_c is 0.53 and on $Distance_{s,c}$ is -1.01.

with a mean of 0.17, is statistically significant for every state. This result suggests a complementarity between state exporters and U.S. foreign direct investment.

On the cost side, destinations in which English is an official language, $English_c$, seem to contribute to an increased number of state exporters. This finding is consistent with Lawless (2010). For 49 of 50 states the estimated parameter is positive, with statistical significance in 37 states. Coefficient estimates for this variable range from -0.09 for Delaware to 0.80 in Montana, with a mean of 0.34.

Similar to Lawless (2010), an increased number of state exporters also tends to be associated with a more developed communications infrastructure. The number of mobile cellular subscriptions per 100 people, *Phones*, is generally associated with more exporters. The mean estimate is 0.05, with estimates ranging from -0.21 in North Dakota to 0.21 in New Hampshire. For 32 of 50 states the estimated parameter is positive; however, in only 13 is this relationship statistically significant. Of the 18 states with a negative estimate, only one is significant.[26]

The results for a geographic variable, *Popdensity*, are similar to those in Lawless (2010) and similar to the results for the communications infrastructure. The mean coefficient estimate is 0.04, with estimates ranging from -0.09 in Montana to 0.19 in Hawaii. For 39 of 50 states the estimated parameter is positive, with 15 cases of statistical significance. Of the 11 states with a negative estimate, two are statistically significant.[27]

Finally, also similar to Lawless (2010), higher levels of official, administrative fees, *Tradefees*, decrease the number of state exporters.[28] All 50 states exhibit a negative sign, with 49 being statistically significant. The mean coefficient estimate is -0.49, ranging from -0.71 for South Dakota to -0.19 for Texas.[29]

Estimation of the Intensive Margin

[26] When included as a replacement for *Phones*, *Internet* produces similar results.
[27] When replacing *Popdensity*, *Landarea* tends to have a negative impact. Overall, however, *Popdensity* yields better results.
[28] A similar conclusion pertains to related measures for the costs of trading across borders.
[29] Attempts to control for the legal environment in the destination country, using $Legal_c$ and $CorruptionControl_c$, failed to produce results that were statistically significant. A similar comment applies to the possible impact of ethnic networks, using $Foreignborn_{s,c}$.

The estimation of the intensive margin uses two estimation methods - ordinary least squares and a full information maximum likelihood method that incorporates a Heckman correction (1979).[30] The absence of trade between many state-country pairs results in many undefined intensive trade margins. The Heckman procedure deals with the potential sample selection bias as a specification error. We implement this procedure via a full information maximum likelihood estimation. Two equations are estimated. First, a probit equation is estimated to produce an estimate of the inverse Mill's ratio. This ratio is then used as a regressor in the equation estimating the intensive margin.

Intensive Margin: Results

Similar to the preceding discussion of results for the extensive margin, three sets of results are presented. The first set of results is generated by pooling the data for all states and estimating both simple and extended gravity models. For comparability with the results for the extensive margin, the same independent variables are used. The remaining two sets of results, one for a simple gravity and one for an extended gravity model, are based on separate regressions for each state.

Pooled Estimates. The results of estimating the intensive margin using state-level data simultaneously for both the simple and extended gravity models are listed in Table 6. All results are based on a full information maximum likelihood estimation that includes a Heckman correction are reported. Similar to the extensive margin estimations, heteroskedasticity-robust standard errors underlie the estimates in columns 1, 2, 4, and 5, while cluster-robust standard errors underlie the estimates in columns 3 and 6. The clustering is based on destination country. For the pooled estimates, the Heckman correction adds value.[31] One indicator is that the correlation coefficient for the probit and main models prior to the inclusion of the Mill's ratio (ρ) is significantly different from zero in the extended gravity models, which signifies that there will be sample selection bias if the inverse Mill's ratio is not included in the main model. While this correlation coefficient is not significant in the simple models, we still use full information maximum likelihood to assist comparability with their extended model counterparts.

[30] Puhani (2000) provides evidence that the full information maximum likelihood estimator is preferable to Heckman's limited-information two-step method.
[31] The variables used in the probit portion are listed in the notes in Table 6.

Columns 1 and 2 of Table 6 contain the results for simple and extended gravity models that include state fixed effects, while columns 4 and 5 show results using gross state product. The coefficient estimates indicate that *GDP* and $Distance_{s,c}$ are statistically significant. The absolute values of the coefficients associated with the extended model are less than those associated with the simple gravity model.

Turning to the results in columns 2 and 5 for the additional variables in the extended model, foreign direct investment is a positive, statistically significant determinant of exports per firm. A similar comment can be made for population density in the foreign country. Meanwhile, English as an official language, mobile phone subscriptions, and trade fees are each a negative, statistically significant determinant of the intensive margin using state-level data.

Similar to the previous results regarding the extensive margin, the estimations based on cluster-robust standard errors produce results that differ slightly from estimations based simply on heteroskedasticity-robust standard errors. The results in columns 3 and 6 are based on cluster-robust standard errors, which exceed the corresponding heteroskedasticity-robust standard errors. Most variables remain statistically significant; however, two variables that were previously identified as statistically significant, *English* and *Tradefees*, are no longer statistically significant.

While not reported in detail, the use of state fixed effects adds explanatory power in a statistical sense as all but 7 fixed effects in the simple model and 6 fixed effects in the extended model are statistically significant. Also, when state fixed effects are replaced by state gross product for the models in Table 6, we find that larger states tend to have larger exports per firm than smaller states. The estimated coefficient is 0.28 in the simple and 0.29 in the extended model.

Simple Gravity Model – Estimates by State. The results for the simple gravity model are listed in Table 7. Contrary to the results for the pooled estimations, OLS estimation was found to be superior for state-by-state estimation. Estimations for 36 states failed to converge, and ρ was only significant two times. Given the small sample issues associated with the Heckman correction, this finding for the estimation by individual state is not surprising.[32] Because the dependent variable is a state's average exports per firm to a country and a given state is unlikely to export to all countries, the number of observations for these regressions varies substantially across states. Even if exports occur, information for a specific state-country pair might not be available for confidentiality reasons. As a result, the sample size ranges from 19 in Wyoming to 183 in California.

Turning to the results, the adjusted R^2 range is from 0.08 in North Dakota to 0.71 in Vermont. Generally speaking, the adjusted R^2 is, at least, 0.40. Only eight states have an adjusted R^2 less than 0.40. Recall that the discussion of the theoretical underpinnings of our estimation did not provide strong expectations concerning the signs of the independent variables. The results based on the simple gravity model indicate that GDP_c is positively related to a state's average exports. For each state the estimated sign is positive, with statistical significance in 49 cases. The estimated coefficient ranges from 0.21 in North Dakota to 0.70 in West Virginia. The mean estimate is 0.46, which is somewhat larger than the estimate of 0.29 produced by Lawless (2010) using national data.[33]

Meanwhile, the estimated sign for $Distance_{s,c}$ is generally negative (e.g., 45 of 50), but this variable is statistically significant in only 29 of the 45 states.[34] These estimates generally fall between -1.00 and zero; however, Alaska exhibits an estimate of -2.19 and Hawaii shows an estimate of -1.76. The mean estimate is -0.39, which is (in absolute value) larger than the estimate of -0.26 produced by Lawless (2010). Part of the difference can be attributed to the estimates for Alaska and Hawaii, which assume a greater weight in the state calculations due to the simple averaging process than they do in Lawless (2010). Potentially, the use of state data should generate more accurate estimates because of more accurate measures of distance.

<u>Extended Gravity Model – Estimates by State</u>. Similar to the estimation for the simple gravity model, OLS estimation was found to be superior for the extended gravity model.

---

[32] See Puhani (2000) and Manning, Duan, and Rogers (1987). Sample selection (Heckman) models do not perform well when censoring is either quite low or relatively high. For the state-by-state estimations, censoring problems are frequent. Meanwhile, the pooled models have "moderate" censoring rates.
[33] See Table 3 in Lawless (2010).
[34] For the five cases of a positive coefficient, not one is statistically significant.

Estimations for 38 states failed to converge, while ρ was statistically significant only one time. Turning to the results for the extended model summarized in Table 8, the inclusion of more variables yields very different degrees of success in explaining the intensive margin relative to the simple gravity model.[35] For most states (45) the extended gravity model generates a larger adjusted R^2 relative to the simple gravity model; however, for 15 of the 50 states an F test indicated that the extended model did not outperform the simple model.

Turning to the number of statistically significant results for GDP_c and $Distance_{s,c}$, both numbers decline relative to the simple model. For GDP_c, the number declines from 49 to 43, while the decline is from 29 to 27 for $Distance_{s,c}$. The mean estimates are close to those generated in Lawless (2010).[36] For example, the mean of the parameter estimates for GDP_c is 0.31, while the estimate in Lawless (2010) is 0.36. For $Distance_{s,c}$, the mean estimate of -0.34 is virtually identical to the one in Lawless (2010). Relative to the coefficient estimates for the simple gravity model, the estimates also tend to be absolutely smaller.

For the additional variables, only FDI_c and $Popdensity_c$ exhibit a consistently positive impact on average exports per firm. *FDI* is significant in 30 of 47 states, which is more frequent than our distance measure, while $Popdensity_c$ is significant in 15 of 39 states. Meanwhile, $English_c$, $Phones_c$, and $Tradefees_c$ are frequently negative, but rarely significant.[37] Overall, the results for these other variables, except FDI_c, add little to understanding the intensive margin.

5. Discussion and Concluding Comments

The novel contribution of this paper is to use state-level export data in the context of heterogeneous exporting firms to investigate the effects of different parameters on extensive and intensive margin of state exports. We first present a theoretical analysis, and then use various relevant econometric methods, to throw light on the determinants of the margins of state exports. Consistent with theory and prior studies at the national level, changes in the extensive margin across states are more closely connected to explanatory variables than are changes in the intensive margin. Our study also provides a rich set of information on the diversity across states with respect to the impacts of the various independent variables on both trade margins. This

---

[35] State-by-state results are available upon request.
[36] See Lawless (2010), Table 7.
[37] Lawless (2010) finds statistical significance for *Phones*.

granular sub-national approach is needed, because, for virtually every independent variable, one observes large differences in both the parameter estimates and statistical significance across states.

Not surprisingly, standard GDP and distance related variables for gravity-based models are also key explanatory variables for both trade margins in this paper. The extensive margin is relatively more responsive than the intensive margin to percentage changes in these variables. In addition, numerous other variables capturing other dimensions of trade and transportation costs, such as the importance of a common language, communications infrastructure, population density, and border costs, affect the extensive trade margin. State size also affects positively both trade margins. What is especially noteworthy is that increases in state size appear to have a more-than-proportionate effect on the number of exporting firms.

Similar to Lawless (2010), we find few consistent results for explaining the intensive trade margin in the state-by-state regressions. However, the single-equation estimation of the intensive trade margin does produce a number of results at odds with her results. Specifically, we find three variables to be statistically significant determinants of the intensive margin. Both the official use of English and fees associated with crossing the border are found to have negative effects, while population density is found to have a positive effect.

A noteworthy finding of this paper is to highlight the empirical importance of U.S. foreign direct investment in affecting the margins of state exports. Both margins are positively related to U.S. foreign direct investment. This complementary relationship needs further theoretical and empirical analysis, which remains in our agenda for future research.

References

Álvarez, Begoña and Delgado, Miguel A. (2002) "Goodness-of-Fit Techniques for Count Data Models: An Application to the Demand for Dental Care in Spain," *Empirical Economics*, 27(3), 543-567.

Armenter, Roc and Koren, Miklós (2014) "A Balls-and-Bins Model of Trade," *American Economic Review*, 104(7), 2127-2151.

Bandyopadhyay, Subhayu; Coughlin, Cletus C.; and Wall, Howard J. (2008) "Ethnic Networks and State Exports," *Review of International Economics*, 16(1), 199-213.

Bernard, Andrew B.; Jensen, J. Bradford; Redding, Stephen J.; and Schott, Peter K. (2007) "Firms in International Trade," *Journal of Economic Perspectives*, 21(3), 105-130.

Bernard, Andrew B.; Jensen, J. Bradford; Redding, Stephen J.; and Schott, Peter K. (2010) "Intrafirm Trade and Product Contractibility," *American Economic Review*, 100(2), 444-448.

Berthou, Antoine and Fontagné, Lionel (2008) "The Euro and the Intensive and Extensive Margins of Trade: Evidence form French Firm Level Data," CEPII, Working Paper No. 06, April.

Burger, Martijn; Van Oort, Frank; and Linders, Gert-Jan (2009) "On the specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation," *Spatial Economic Analysis*, 4(2), 167-190.

Cassey, Andrew J. (2011) "State Foreign Export Patterns," *Southern Economic Journal*, 78(2), 308-329.

Chaney, Thomas (2008) "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98(4), 1707-1721.

Clark, Don (2007) "Distance, Production, and Trade," *Journal of International Trade and Economic Development*, 16(3), 359-371.

Coughlin, Cletus C. (2014) "The Great Trade Collapse and Rebound," Federal Reserve Bank of St. Louis *Review*, 96(1), 13-33.

Coughlin, Cletus C. (2010) "Measuring International Trade Policy: A Primer on Trade Restrictiveness Indices," Federal Reserve Bank of St. Louis *Review*, 92(5), 381-394.

Coughlin, Cletus C. and Wall, Howard J. (2011) "Ethnic Networks and Trade: Intensive vs. Extensive Margins," *Economics Letters*, 113(1), 73-75.

Djankov, Simeon; Freund, Caroline; and Pham, Cong S. (2010) "Trading on Time," *Review of Economics and Statistics*, 92(1), 166-173.

Eaton, Jonathan; Kortum, Samuel; and Kramarz, Francis (2004) "Dissecting Trade: Firms, Industries, and Export Destinations," *American Economic Review*, 94(2), 150-154.

Ferrante, Maria Rosario and Novelli, Marco (2013) "Heterogeneity and Number of Export Destinations of Italian Firms: A Hurdle Negative Binomial Regression Approach," *Global Economy Journal*, 13(3), 391-416.

Gerdtham, Ulf-G (1997) "Equity in Health Care Utilization: Further Evidence Based on Hurdle Models and Swedish Micro Data," *Health Economics*, 6(3), 303-319.

Giuliano, Paola; Spilimbergo, Antonio; and Tonon, Giovanni (2014) "Genetic Distance, Transportation Costs, and Trade," *Journal of Economic Geography*, 14(1), 179-198.

Gurmu, Shiferaw and Trivedi, Pravin K. (1996) "Excess Zeros in Count Models for Recreational Trips," *Journal of Business and Economic Statistics*, 14(4), 469-477.

Head, Keith and Mayer, Thierry (2013) "What Separates Us? Sources of Resistance to Globalization," *Canadian Journal of Economics*, 46(4), 1196-1231.

Heckman, James J. (1979) "Sample Selection Bias as a Specification Error," *Econometrica*, 47(1), 153-161.

Helpman, Elhanan; Melitz, Marc; and Rubinstein, Yona (2008) "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 123(2), 441-487.

Hillberry, Russell and Hummels, David (2008) "Trade Responses to Geographic Frictions: A Decomposition Using Micro-data," *European Economic Review*, 52(3), 527-550.

Kehoe, Timothy J. and Ruhl, Kim J. (2009) "How Important Is the New Goods Margin in International Trade?" Federal Reserve Bank of Minneapolis, Research Department Staff Report 324, June.

Korinek, Jane (2011) "Clarifying Trade Costs in Maritime Transport," *Organization for Economic Cooperation and Development*, TAD/TC/WP(2008)10/FINAL.

Lahiri, Kajal and Xing, Guibo (2004) "An Econometric Analysis of Veterans' Health Care Utilization Using Two-Part Models," *Empirical Economics*, 29(2), 431-449.

Lambert, Diane (1992) "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing," *Technometrics*, 34(1), 1-14.

Lawless, Martina (2010) "Deconstructing Gravity: Trade Costs and Extensive and Intensive Margins," *Canadian Journal of Economics*, 43(4), 1149-1172.

Manning, W. G.; Duan, N.; and Rogers, W. H. (1987) "Monte Carlo Evidence on the Choice Between Sample Selection and Two-Part Models," *Journal of Econometrics*, 35(1), 59-82.

Markusen, James R. (2010) "Expansion of Trade at the Extensive Margin: A General Gains-from-Trade Result and Illustrative Examples," NBER, Working Paper No. 15926, April.

Melitz, Marc J. (2003) "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6), 1695-1725.

Mullahy, John (1986) "Specification and Testing of Some Modified Count Data Models," *Journal of Econometrics*, 33(3), 341-365.

Puhani, Patrick A. (2000) "The Heckman Correction for Sample Selection and its Critique," *Journal of Economic Surveys*, 14(1), 53-68.

Santos Silva, J. M. C. and Tenreyro, Silvana (2006) "The Log of Gravity," *Review of Economics and Statistics*, 88(4), 641-658.

Wang, Peiming (2003) "A Bivariate Zero-Inflated Negative Binomial Regression Model for Count Data with Excess Zeros," *Economics Letters*, 78(3), 373-378.

Winkelmann, Rainer (2008) *Econometric Analysis of Count Data*, 5th edition, Springer, Berlin.

Table 1
Variables for Trade Margin Regressions: Abbreviation and Definition

Name	Definition
$Firms_{s,c}$	number of firms in state *s* exporting to the country *c*
$Exports_{s,c}$ *	exports from state *s* to country *c*
$Exportsperfirm_{s,c}$ *	average exports per firm from state *s* to country *c*
$Distance_{s,c}$ *	distance (kilometers) from largest city in state *s* to largest city in country *c*
GSP_s *	gross state product of state *s*
GDP_c *	gross domestic product of country *c*
FDI_c *	value ($m) of U.S. foreign direct investment in country *c*
$English_c$	if English is an official language of country *c*, then value is 1; otherwise 0
$Phones_c$ *	mobile cellular subscriptions per 100 people in country *c*
$Internet_c$ *	internet users per 100 people in country *c*
$Popdensity_c$ *	population per square kilometer in country *c*
$Landarea_c$*	area in square kilometers in country *c*
$TradeFreedom_c$	index measuring the extent to which traded goods are unaffected by tariff and non-tariff barriers - higher values indicate lower barriers in country *c*
$Tradedocuments_c$	number of documents per a standardized shipment in country *c* required for import clearance
$Timeprocess_c$	days to move a standardized shipment through country *c*'s port
$Tradefees_c$ *	official, administrative fees in dollars per imported container in country *c* including the costs for documents, administrative fees for customs clearance and technical control, customs brokers fees, terminal handling charges, and inland transportation - tariffs and trade taxes are excluded
$Legal_c$	confidence of agents in country *c* that existing rules/laws will be followed and enforced – higher percentile ranks reflect better governance
$CorruptionControl_c$	perceptions of agents in country *c* that public power is exercised for public rather than private interests - higher percentile ranks reflect better governance
$Foreignborn_{s,c}$ *	number born in country *c* living in state *s*

*Indicates variable expressed as a natural logarithm in regressions

Table 2
Summary Statistics

Name	Mean	Range	
		Minimum	Maximum
State-Country Variables			
$Firms_{s,c}$	114.9	0 *(1598 pairs)*	16,505
$Exports_{s,c}$	112,044,961	0 *(1598 pairs)*	48,035,483,648
$Exportsperfirm_{s,c}$	338,263	3,198	43,507,948
$Distance_{s,c}$	9,469	334	19,497
$Foreignborn_{s,c}$	7,990	0 *(56 pairs)*	3,928,701
State Specific Variables			
GSP_s	251.9 b.	22.9 b.	1,673.1 b.
Country Specific Variables			
GDP_c	182.3 b.	0.1 b.	4,649.3 b.
FDI_c	13,946	-17	393,540
$English_c$*	Na	0	1
$Phones_c$	56	1	151
$Internet_c$	23	≈0	88
$Popdensity_c$	301.1	≈0	17,727
$Landarea_c$	635,451	28	16,377,740
$TradeFreedom_c$	68	17	90
$Tradedocuments_c$	10	2	20
$Timeprocess_c$	35	3	139
$Tradefees_c$	1,379	323	4,421
$Legal_c$	49	1	100
$CorruptionControl_c$	49	1	100

** English (dummy): Of 187 countries, 56 (30 percent) use English as an official language while 131 do not.*

Table 3
Extensive Margin: Pooled Results

	Hurdle Negative Binomial Coefficient Estimates					
	State FE			GSP		
	(1)	(2)	(3)	(4)	(5)	(6)
GDP	0.74***	0.44***	0.44***	0.73***	0.43***	0.43***
Distance	-0.90***	-0.71***	-0.71***	-0.99***	-0.78***	-0.78***
FDI		0.17***	0.17***		0.17***	0.17***
English		0.34***	0.34***		0.35***	0.35***
Phones		0.03**	0.03		0.04**	0.04
Popdensity		0.04***	0.04		0.04***	0.04
Tradefees		-0.49***	-0.49***		-0.51***	-0.51***
GSP				1.20***	1.18***	1.18***
Constant	-9.37***	-1.91***	-1.91*	-28.06***	-20.40***	-20.40***
Alpha	0.73***	0.39***	0.39***	0.86***	0.50***	0.50***
Log-likelihood	-34723.8	-26319.0	-26319.0	-35394.1	-27050.1	-27050.1
Sample $Size_1$	9350	6800	6800	9350	6800	6800
Sample $Size_2$	7752	5844	5844	7752	5844	5844
AIC	69655.5	52866.1	52866.1	70806.2	54138.2	54138.2
Likelihood Ratio	191.0***	216.8***	216.8***	104.1***	146.1***	146.1***

Notes: Statistical significance: 1% level -- ***; 5% level -- **; 10% level -- *. Sample $Size_1$ refers to the sample size of the logit model, while Sample $Size_2$ refers to the sample size for the truncated negative binomial model. Only the results from the negative binomial model within the hurdle estimation are reported. Regressions 1, 2, 4, and 5 incorporate heteroskedasticity-robust standard errors, while regressions 3 and 6 incorporate cluster-robust standard errors that consider destination country clusters. However, for regressions 1 and 4, the results are identical to the cluster-robust results. Log-likelihoods are specifically log-pseudolikelihoods because of the use of robust standard errors. The likelihood ratio compares each hurdle model to its standard negative binomial counterpart.

Table 4

Negative Binomial Coefficient Estimates: Simple Gravity Model for Extensive Margin

State	*GDP*	*Distance*	*Constant*	*Alpha*	Pseudo R^2
AL	0.74***	-1.17***	-4.33***	0.72***	0.17
AK	0.79***	-2.24***	1.67	1.02***	0.25
AZ	0.74***	-0.73***	-7.96***	0.52***	0.20
AR	0.77***	-0.97***	-7.74***	0.73***	0.19
CA	0.72***	-1.04***	-2.08	0.73***	0.13
CO	0.76***	-0.76***	-8.18***	0.57***	0.19
CT	0.80***	-0.70***	-9.58***	0.72***	0.17
DE	0.76***	-0.76***	-9.77***	0.80***	0.19
FL	0.59***	-2.01***	9.21***	0.98***	0.10
GA	0.66***	-1.05***	-2.19**	0.71***	0.13
HI	0.72***	-4.18***	21.91***	1.09***	0.23
ID	0.68***	-1.10***	-4.43**	0.76***	0.19
IL	0.77***	-0.90***	-5.87***	0.70***	0.15
IN	0.76***	-0.84***	-7.22***	0.75***	0.16
IA	0.71***	-0.98***	-5.52***	0.61***	0.19
KS	0.69***	-0.76***	-6.95***	0.46***	0.20
KY	0.79***	-1.02***	-6.74***	0.82***	0.17
LA	0.66***	-0.94***	-4.31***	0.73***	0.15
ME	0.69***	-0.81***	-7.57***	0.56***	0.23
MD	0.62***	-0.56***	-6.33***	0.50***	0.17
MA	0.80***	-0.73***	-8.63***	0.65***	0.16
MI	0.77***	-0.74***	-8.06***	0.78***	0.16
MN	0.74***	-0.76***	-7.07***	0.63***	0.17
MS	0.74***	-1.14***	-5.23***	0.71***	0.19
MO	0.77***	-1.01***	-6.17***	0.79***	0.16
MT	0.82***	-1.43***	-6.12**	0.79***	0.26
NE	0.73***	-0.77***	-8.38***	0.50***	0.22
NV	0.78***	-1.20***	-5.66***	0.80***	0.19
NH	0.86***	-0.49***	-13.84***	0.72***	0.19
NJ	0.72***	-0.92***	-4.60***	0.71***	0.14
NM	0.72***	-0.92***	-7.54***	0.61***	0.23
NY	0.75***	-0.82***	-5.65***	0.70***	0.13
NC	0.76***	-0.92***	-5.92***	0.72***	0.15
ND	0.66***	-0.96***	-6.12***	0.50***	0.26
OH	0.79***	-0.89***	-6.59***	0.73***	0.15
OK	0.76***	-0.73***	-9.00***	0.66***	0.18
OR	0.71***	-1.10***	-3.89	0.78***	0.17
PA	0.76***	-0.87***	-6.15***	0.68***	0.15
RI	0.82***	-0.73***	-11.19***	1.02***	0.17
SC	0.75***	-0.95***	-6.06***	0.70***	0.16
SD	0.77***	-1.07***	-7.83***	0.73***	0.23
TN	0.75***	-1.01***	-5.50***	0.76***	0.16
TX	0.72***	-0.77***	-5.19***	0.57***	0.14
UT	0.74***	-0.81***	-7.60***	0.56***	0.19
VT	0.77***	-0.56***	-11.87***	0.47***	0.25
VA	0.66***	-0.65***	-6.26***	0.56***	0.16
WA	0.71***	-1.09***	-3.28	0.67***	0.16
WV	0.86***	-0.53***	-14.65***	0.65***	0.24
WI	0.77***	-1.08***	-4.89***	0.70***	0.16
WY	0.82***	-0.78***	-12.84***	0.79***	0.26

Notes: Statistical significance: 1% level -- ***; 5% level -- **; 10% level -- *.
Sample size is 187 for every state.

Table 5
Summary of Negative Binomial Results: Extended Gravity Model for Extensive Margin

			Sign (Statistical Significance)	
Independent Variable	Mean	Range	Positive	Negative
GDP	0.46	0.29 – 0.59	50 (50)	0 (0)
FDI	0.17	0.08 – 0.27	50 (50)	0 (0)
Distance	-0.79	-3.45 – -0.25	0 (0)	50 (50)
English	0.34	-0.09 – 0.80	49 (37)	1 (0)
Phones	0.05	-0.21 – 0.21	32 (13)	18 (1)
Popdensity	0.04	-0.09 – 0.19	39 (15)	11 (2)
Tradefees	-0.49	-0.71 – -0.19	0 (0)	50 (49)
Alpha	0.33	0.12 – 0.80	50 (50)	0 (0)
Pseudo R^2	0.22	0.12 – 0.35		
Sample Size	136			

Table 6
Intensive Margin: Pooled Results

	FIML (Heckman) Coefficient Estimates					
	State FE			GSP		
	(1)	(2)	(3)	(4)	(5)	(6)
GDP	0.43***	0.30***	0.30***	0.43***	0.30***	0.30***
Distance	-0.38***	-0.35***	-0.35***	-0.38***	-0.35***	-0.35***
FDI		0.12***	0.12***		0.13***	0.13***
English		-0.06**	-0.06		-0.08**	-0.08
Phones		-0.14***	-0.14***		-0.14***	-0.14***
Popdensity		0.06***	0.06**		0.06***	0.06**
Tradefees		-0.07**	-0.07		-0.07**	-0.07
GSP				0.28***	0.29***	0.29***
Constant	3.33***	6.15***	6.15***	-0.76*	2.04***	2.04**
Log-likelihood	-7350.6	-7111.5	-7111.5	-7832.6	-7610.1	-7610.1
Sample $Size_1$	5479	5429	5429	5479	5429	5429
Sample $Size_2$	4529	4479	4479	4529	4479	4479
λ	0.04	0.17***	0.17**	0.05	0.24***	0.24**
σ	0.89	0.86	0.86	0.99	0.96	0.96
ρ	0.05	0.20***	0.20**	0.05	0.25***	0.25**

Notes: Statistical significance: 1% level -- ***; 5% level -- **; 10% level -- *. Sample $Size_1$ refers to the sample size of the first stage, while Sample $Size_2$ refers to the sample size of the second stage. Only the results from the second stage of the Heckman models are provided. The variables used in the first stage are *GDP, Distance, FDI, Legal, Tradefees, Landarea, Popdensity*, and *Internet*. Sample size refers to the number of observations used in the second stage of the model. Regressions 1, 2, 4, and 5 incorporate heteroskedasticity-robust standard errors, while regressions 3 and 6 incorporate cluster-robust standard errors that consider destination country clusters. Log-likelihoods are specifically log-pseudolikelihoods because of the use of robust standard errors.

Table 7

OLS Coefficient Estimates: Simple Gravity Model for Intensive Margin

State	*GDP*	*Distance*	*Constant*	Adjusted R^2	Sample Size
AL	0.42***	-0.73***	8.61***	0.39	109
AK	0.48***	-2.19**	20.34*	0.44	31
AZ	0.55***	0.03	-1.90	0.54	136
AR	0.50***	-0.17	1.04	0.60	80
CA	0.32***	-0.04	4.57***	0.50	183
CO	0.51***	-0.38**	2.11	0.64	136
CT	0.51***	-0.18	0.61	0.66	122
DE	0.53***	0.04	-1.88	0.41	69
FL	0.29***	-0.34***	7.64***	0.49	170
GA	0.34***	-0.44***	7.93***	0.53	152
HI	0.33***	-1.76***	18.65***	0.21	32
ID	0.52***	-0.27	0.90	0.48	82
IL	0.40***	-0.08	2.85***	0.57	160
IN	0.51***	-0.46***	3.23***	0.68	135
IA	0.50***	-0.37***	3.07***	0.69	106
KS	0.58***	-0.44**	1.94	0.58	110
KY	0.62***	-0.60***	1.90	0.65	116
LA	0.40***	-0.99***	12.39***	0.29	116
ME	0.52***	-0.08	-1.15	0.55	81
MD	0.37***	-0.03	2.81**	0.47	142
MA	0.43***	-0.19	2.85**	0.50	149
MI	0.44***	-0.54***	5.90***	0.55	140
MN	0.42***	-0.28**	4.10***	0.50	140
MS	0.50***	-0.58***	4.95***	0.41	91
MO	0.47***	-0.43***	4.09**	0.56	118
MT	0.58***	0.25	-5.88**	0.45	47
NE	0.52***	-0.38**	2.31	0.55	95
NV	0.54***	-0.11	-0.94	0.55	87
NH	0.49***	-0.19	0.77	0.57	97
NJ	0.32***	-0.19*	5.71***	0.43	161
NM	0.40***	-0.24	3.15	0.37	61
NY	0.34***	-0.46***	7.63***	0.46	170
NC	0.38***	-0.62***	8.16***	0.45	145
ND	0.21***	0.06	6.46***	0.08	52
OH	0.50***	-0.47***	3.87***	0.67	148
OK	0.38***	0.01	2.14*	0.35	111
OR	0.43***	-0.23	3.52	0.36	116
PA	0.39***	-0.57***	7.25***	0.57	160
RI	0.43***	-0.46**	4.58***	0.43	87
SC	0.40***	-0.30***	5.18***	0.54	125
SD	0.53***	-0.09	-1.29	0.54	59
TN	0.48***	-0.57***	5.62***	0.51	120
TX	0.31***	-0.46***	9.57***	0.40	168
UT	0.39***	-0.64***	7.41***	0.42	111
VT	0.62***	-0.62***	0.77	0.71	76
VA	0.42***	-0.36***	5.12***	0.56	141
WA	0.51***	-0.36	2.76	0.58	134
WV	0.70***	-0.14	-4.54**	0.60	58
WI	0.45***	-0.26***	3.17***	0.66	143
WY	0.58	-0.76	2.89	0.09	19

Notes: Statistical significance: 1% level -- ***; 5% level -- **; 10% level -- *.

Table 8

Summary of Results: Extended Gravity Model for Intensive Margin

			Sign (Statistical Significance)	
Independent Variable	Mean	Range	Positive	Negative
GDP	0.31	-0.12 – 0.56	49 (43)	1 (0)
FDI	0.13	-0.31 – 0.74	47 (30)	3 (1)
Distance	-0.34	-2.58 – 0.39	8 (0)	42 (27)
English	-0.10	-1.08 – 0.60	16 (4)	34 (7)
Phones	-0.13	-1.00 – 0.27	8 (2)	42 (12)
Popdensity	0.06	-0.27 – 0.27	39 (15)	11 (3)
Tradefees	-0.09	-0.76 – 0.39	18 (2)	32 (5)
Adjusted R^2	0.54	-0.02 – 0.73		
Sample Size	90	19 – 134		

Figure 1
Representative Histogram: Arkansas