NBER WORKING PAPER SERIES

GRAVITY AND EXTENDED GRAVITY: USING MOMENT INEQUALITIES TO ESTIMATE A MODEL OF EXPORT ENTRY

Eduardo Morales
Gloria Sheu
Andrés Zahler

Working Paper 19916
http://www.nber.org/papers/w19916

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2014, Revised June 2017

An earlier draft of this paper circulated under the title "Gravity and Extended Gravity: Estimating a Structural Model of Export Entry." The views in this paper should not be purported to reflect those of the United States Department of Justice. An earlier draft of this paper circulated under the title “Gravity and Extended Gravity: Estimating a Structural Model of Export Entry.” We would like to thank Pol Antràs, Kamran Bilir, Jan De Loecker, Michael Dickstein, Gene Grossman, Ricardo Hausmann, Elhanan Helpman, Guido Imbens, Robert Lawrence, Andrew Levy, Marc Melitz, Julie Mortimer, Ezra Oberfield, Ariel Pakes, Charly Porcher, Stephen Redding, Dani Rodrik, Esteban Rossi-Hansberg, Felix Tintelnot, and Elizabeth Walker for insightful conversations, and seminar participants at BEA, Boston College, Brandeis, Columbia, CREI, Duke, ERWIT, Harvard, LSE, MIT, NBER, Princeton, Sciences-Po, SED, UCLA, University of Chicago-Booth, Wisconsin, and the World Bank for very helpful comments. We also thank Jenny Nuñez, Luis Cerpa, the Chilean Customs Authority, and the Instituto Nacional de Estadísticas for assistance with building the dataset. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. Andres Zahler acknowledges the Nucleo Milenio Initiative NS100017 "Intelis Centre" for partial funding.

Gravity and Extended Gravity: Using Moment Inequalities to Estimate a Model of Export Entry
Eduardo Morales, Gloria Sheu, and Andrés Zahler
NBER Working Paper No. 19916
February 2014, Revised June 2017
JEL No. F14,L65

ABSTRACT

Exporting firms often enter foreign markets that are similar to previous export destinations. We develop a dynamic model in which a firm's exports in a market may depend on how similar the market is to the firm's home country (gravity) and to its previous export destinations (extended gravity). Given the large number of export paths from which forward-looking firms may choose, we use a moment inequality approach to structurally estimate our model. Using data from Chilean exporters, we estimate that having similarities with a prior export destination in geographic location, language, and income per capita jointly reduce the cost of foreign market entry by 69% to 90%. Reductions due to geographic location (25% to 38%) and language (29% to 36%) have the largest effect. Extended gravity thus has a large impact on export entry costs.

Eduardo Morales
Department of Economics
Princeton University
310 Fisher Hall
Princeton, NJ 08544
and NBER
ecmorales@princeton.edu

Gloria Sheu
Department of Justice
gloria.sheu@usdoj.gov

Andrés Zahler
Universidad Diego Portales
no email available

1 Introduction

Exporting firms continuously enter and exit foreign countries, and the decision of which countries to export to is an important determinant of aggregate trade flows.[1] When selecting which countries to enter, firms tend to choose markets similar to their prior export destinations.[2] We establish that this cross-country correlation in firms' export decisions occurs because entry costs in a market are smaller for firms that have previously exported to similar markets. We refer to this path dependence in entry costs as *extended gravity*.

Extended gravity has significant implications for trade policy. It predicts that reducing trade barriers in a country will increase entry not only in its own market but also in other markets that are connected to it through extended gravity. This suggests that import policies in one country generate externalities for other countries. As for export policy, extended gravity means that export promotion measures will have the largest impact when targeted toward destination countries that share characteristics with large export markets.

We estimate a new model of firm export dynamics. In our model, entry costs in a market depend on how similar it is both to the firm's home country and to other countries to which the firm has previously exported. Our model is thus consistent with gravity in that it allows the probability to be higher that firms export to markets that are close to their country of origin. But the model also allows firms export decisions to depend on their previous export history. While gravity reflects proximity between origin and destination markets, extended gravity depends on proximity between past and subsequent destinations. Our model imposes only weak restrictions on how firms both determine the set of countries to which they consider exporting and forecast the profits they would obtain upon entering these markets. We introduce an approach that exploits these weak assumptions on firms' consideration and information sets in order to estimate bounds on the impact of extended gravity on entry costs.

To measure the importance of extended gravity, we use a firm-level dataset for Chile that includes information on exports by year and destination as well as on a broad set of firm characteristics during 1995–2005. We estimate our model for the chemicals sector, which is consistently among the top two manufacturing sectors in Chile by volume of exports during our sample period. Our estimates show that similarity in geographic location, language, and income per capita with a prior export destination jointly reduce the cost of foreign market entry by 69% to 90%. The reductions due to similarity in geographic location (25% to 38%) and language (29% to 36%) have the largest impact. Conversely, we cannot rule out that similarity in income per capita has no extended gravity effects on its own.

Extended gravity is consistent with foreign market entry requiring a costly adaptation process: some firms are better prepared to enter certain countries because they have previously

[1] See Bernard et al. (2007, 2010); Hillberry and Hummels (2008); Head and Mayer (2014).

[2] For evidence on spatial correlation in export flows, see Evenett and Venables (2002); Lawless (2009, 2013); Albornoz et al. (2012); Chaney (2014); Defever et al. (2015); Meinen (2015).

served similar markets and have thus already partly completed this adaptation. This process may entail changes in the branding, labeling, and packaging of the product, as well as product modifications that reflect local tastes or legal requirements imposed by national regulators.[3]

In order to test our hypothesis that extended gravity impacts export entry costs, we rely on a dynamic multi-country model to structurally identify costs that may depend on gravity and extended gravity. We account for gravity by allowing these costs to depend on whether each destination shares a continent, language, or similar income per capita with Chile. By contrast, extended gravity depends on whether the destination shares a border, continent, language, or similar income per capita with a country to which the firm exported in the previous year. If extended gravity is important, firms in our framework decide whether to enter a country taking into account the impact of this decision on future entry costs in other markets.

The standard approach to the estimation of entry models relies on deriving choice probabilities from a theoretical framework and finding the parameter values that maximize the likelihood of the entry choices observed in the data (Das et al., 2007). This approach is not feasible in our setting. Evaluating these probabilities involves examining the dynamic implications of every possible bundle of export destinations. Given the cardinality of the potential choice set (for a given number of countries J, this set includes 2^J elements), computing the value function for each of its elements is infeasible unless very strong simplifying assumptions are imposed on the firm's actual choice set and state vector.[4] The impossibility of computing this value function implies that we cannot solve the model and perform counterfactuals. However, using moment inequalities, we can estimate bounds on the effect of extended gravity on export entry costs. Our estimator requires neither computing the value function of the firm nor artificially reducing the dimensionality of the firm's choice set or state vector.

Our inequalities come from applying an analogue of Euler's perturbation method. Specifically, we impose one-period deviations on the observed export path of each firm. Our moment inequalities are robust to different assumptions on: (a) how forward-looking firms are, as captured by firms' planning horizons; (b) firms' choice sets; and (c) firms' information sets. In addition, our inequalities do not impose any parametric restrictions on the distribution of firms' expectational errors, which may vary flexibly across firms, countries, and time periods.

In identifying extended gravity effects from firms' observed export choices, we face the challenge of separating path dependence from unobserved heterogeneity. Unobservable (to the researcher) determinants of the decision to export that are specific to each firm and correlated both over time and across countries that share geographic location, language or

[3] Adaptation may also involve searching for a local distributor (Chaney, 2014, 2016a,b) or hiring workers with knowledge of specific markets (Labanca et al., 2014).

[4] For example, even if the firm's actual choice set includes only 20 destinations and the expected profits in each of them depend only on a state variable that takes 5 values, the state vector will still include over 10^{13} elements. Furthermore, as we show in a simulation in Appendix D, misspecifications of the firm's consideration or information sets may significantly bias estimates of the impact of extended gravity on entry costs.

similar income per capita will generate export paths similar to those that we would observe if these extended gravity factors were an important determinant of firms' choices. In order to separately identify the effect of these unobservables from the path dependence generated by extended gravity, in our moment inequality estimation we allow for firm-, year-, and group-of-countries-specific fixed effects. Specifically, we allow the firm's export decision to depend on fixed effects that are specific to each firm and year, but common across countries that are located in the same continent, share an official language, or have similar income per capita.

Our estimates show that extended gravity has a significant effect on entry costs. Previously serving a market that shares a border with a destination (e.g. Poland and Germany) reduces entry costs by 25% to 38%.[5] Sharing a continent without sharing border (e.g. Poland and France) reduces entry costs by 19% to 29%, and sharing only language (e.g. Portugal and Brazil) by 29% to 36%. Our estimates for having similar income per capita are less informative; they indicate that the reduction in entry costs may be any number smaller than 29%. The combined effect of all four extended gravity covariates is between 69% and 90%, which implies that, for example, a Chilean firm exporting to Germany will face subsequent entry costs in Austria that are between a tenth and a third of the costs faced by a firm that is exporting only to destinations that do not share any extended gravity covariates with Austria.

Our paper is related to several strands of the literature. First, it relates to papers that estimate export entry costs. Das et al. (2007) estimate fixed and sunk costs of breaking into exporting generally. Dickstein and Morales (2016) estimate fixed and sunk costs by destination, but ignore the presence of extended gravity effects.[6] In contrast to this prior literature, we estimate the impact of extended gravity on export entry costs.

Second, our paper relates to previous work showing that firms tend to export to countries similar to their prior destinations; e.g. Eaton et al. (2008); Lawless (2009, 2013); Albornoz et al. (2012); Chaney (2014); Defever et al. (2015). Except for Chaney (2014), none of these papers structurally estimates a model of forward-looking firms that incorporates a mechanism rationalizing this export behavior. We build such a model, emphasize the importance of entry cost dynamics to explain observed export dynamics, and identify extended gravity effects under weak assumptions on firms' information and consideration sets and planning horizons.

Third, our paper introduces a new moment inequality procedure to deal with multiple discreteness problems (Hendel, 1999). These are decision problems in which agents violate the single-choice assumption inherent to multinomial discrete choice models. These problems feature in the store-network choice literature (Jia, 2008; Holmes, 2011; Ellickson et al., 2013; Zheng, 2016), in the demand estimation literature (Allenby et al., 2002, 2007; Dubé, 2004), and in the work on multinational companies (Tintelnot, 2016) and on the sourcing decisions of

[5] We report here projections of a 95% confidence set. See Section 5 for details on estimation.

[6] Additional export cost estimates are provided in Roberts and Tybout (1997); Arkolakis (2010); Moxnes (2010); Aw et al. (2011); Eaton et al. (2014); Arkolakis et al. (2015a); Irarrazabal et al. (2015); Eaton et al. (2016); Fitzgerald et al. (2016); Ruhl and Willis (2016); Bai et al. (2017).

importers (Antràs et al., 2017). In these papers, the set including all bundles of alternatives an agent may choose is very large. The literature contains three approaches to dealing with these large-dimensional discrete choice sets: first, exploiting the increasing or decreasing differences property of the agent's objective function (Jia, 2008; Antràs et al., 2017; Arkolakis and Eckert, 2017); second, modeling multiple discreteness problems as an aggregation of simple discrete choices happening at different points in time (Hendel, 1999; Sieg and Zhang, 2012; Arcidiacono et al., 2016); and third, using moment inequalities (Holmes, 2011).

Fourth, our paper contributes to an empirical literature on moment inequalities (Katz, 2007; Ishii, 2008; Ho, 2009; Pakes, 2010; Ho and Pakes, 2014; Eizenberg, 2014; Pakes et al., 2015; Pakes and Porter, 2015; Wollmann, 2016; Illanes, 2016; Dickstein and Morales, 2016). Our approach is closest to that in Holmes (2011), but differs from it in that we do not form inequalities by changing the order in which we observe firms entering markets. We implement instead an analogue of Euler's perturbation method (Hansen and Singleton, 1982; Luttmer, 1999), building thus inequalities that are valid under weak restrictions on firms' expectations.

Fifth, extended gravity has implications for the interpretation of gravity parameters. The standard gravity equation (Tinbergen, 1962) predicts that trade flows between two countries depend only on their size and measures of trade resistance between them. Anderson and van Wincoop (2003) take into account third country effects through multilateral resistance terms: given two countries, higher barriers between one of them and the rest of the world raises imports from the other one. Extended gravity works in the opposite direction: it creates benefits for firms from directing their exports towards markets that share characteristics with a large number of countries, especially if those markets require high adaptation costs to enter.

The rest of the paper is organized as follows. Section 2 describes our data and presents stylized facts that motivate the rest of the paper. Section 3 introduces a model of entry into export markets, and Section 4 derives moment inequalities from it. Section 5 describes our estimation approach, and sections 6 and 7 present the results. Section 8 concludes.

2 Data

In this section, we describe our sources of data and provide evidence suggestive of the importance of extended gravity in determining firms' export destinations.

2.1 Data Sources

Our data covers the period 1995–2005 and comes from two separate sources. The first is the Chilean customs database, which covers the universe of exports of Chilean firms. The second is the Chilean Annual Industrial Survey (*Encuesta Nacional Industrial Anual*, or ENIA), which includes all manufacturing plants with at least 10 workers. We aggregate the plant-level information in ENIA to obtain firm-level information, and merge it with customs using firm identifiers. We thus observe both the export and domestic activity of each firm.

Our dataset includes all firms operating in the chemicals sector (sector 24, ISIC rev. 3.1), which is among the top two Chilean manufacturing sectors by volume of exports in every sample year. We observe both exporters and non-exporters, and use for estimation an unbalanced panel that includes all firms active for at least two consecutive years between 1995 and 2005. An observation is a firm-country-year combination. The per-year average number of firm-country pairs with positive exports is approximately 650, out of which around 150 correspond to firms that were not exporting to the same country in the previous year, and around 125 correspond to firms that did not continue exporting to the same country in the following year. These export events are generated by a per-year average of 110 firms exporting to around 70 countries in total. For each firm-year-country combination we have information on the value of goods sold in US dollars, and we transform them into year 2000 values using the US CPI.

We complement our customs-ENIA data with a database of country characteristics. We obtain information on the primary official language, continent, and names of bordering countries of each possible destination market from CEPII (Mayer and Zignago, 2011). We collect data on real GDP per capita from the World Bank World Development Indicators. We construct our gravity and extended gravity variables from these country characteristics.

The gravity variables relate Chile to each destination. We create four individual dummy variables that equal one if these destinations do *not* share border, continent, language, or similar income per capita with Chile.[7] We denote them as "Grav. Border", "Grav. Cont.", "Grav. Lang.", and "Grav. GDPpc".[8] The extended gravity variables relate each potential destination to a firm's prior export bundle. We define separate dummies for sharing border, continent, language, or similar income per capita with at least one country the firm exported to in the previous year, and not with Chile itself. We denote them as "Ext. Grav. Border", "Ext. Grav. Cont.", "Ext. Grav. Lang.", and "Ext. Grav. GDPpc". Thus, an extended gravity dummy equals one for a given firm-country-year observation if the country does not share the corresponding characteristic with Chile but shares it with some other country to which the firm exported in the previous year.[9] For example, in the case of Austria, all four extended gravity variables equal one for a firm that exported to Germany in the previous year.

[7]Using 2002 data, the World Bank classifies countries into four groups according to their GDP per capita. Low income is 735 USD or less, lower middle income is 736 to 2,935 USD, upper middle income is 2,936 to 9,075 USD, and high income is 9,076 USD or more. Chile belongs to the upper middle income group. Whenever two countries belong to the same group, we refer to them as "sharing similar income per capita".

[8]Formally, $(\text{Grav. Cont.})_j = 1 - \text{continent}(h, j)$, where $\text{continent}(h, j)$ is a dummy variable that equals one if countries h and j share continent. The other two gravity variables are defined analogously.

[9]Formally, $(\text{Ext. Grav. Lang.})_{ijt} = (1-d_{ijt-1}) \times (1-\text{language}(h, j')) \times \mathbb{1}\{\sum_{j'} d_{ij't-1} \times \text{language}(j, j') \geq 0\}$, where d_{ijt} is a dummy variable that equals one if firm i exports to destination j in year t, $\text{language}(j, j')$ is a dummy variable that equals one if countries j and j' share language, and $\mathbb{1}\{A\}$ is an indicator function that equals 1 if A is true. The other three extended gravity variables are defined analogously.

2.2 Motivating Evidence

We provide here descriptive evidence suggestive of the relevance of extended gravity. We also discuss alternative economic mechanisms that may generate similar export behavior.

As an illustration of the export patterns suggestive of the presence of extended gravity, Figure 4 shows the 2000 to 2005 export path of a firm in our sample.[10] Prior to 2000, this firm's single export destination outside South America was the United States. Its export destinations in this time period were thus consistent with gravity: they are either large (the United States), or close to Chile (Bolivia and Peru), or both (Argentina). However, from the year 2000 onwards, this firm entered markets that are small and far away from Chile, but related to its prior destinations. The firm expanded through Central America by consecutively entering countries bordering prior destinations: it entered Mexico (which borders the United States) in 2000; Guatemala (which borders Mexico) in 2001; Belize, Honduras, and El Salvador (which all border Guatemala) in 2003; and, finally, Nicaragua (which borders Honduras) in 2004. Simultaneously, the firm also expanded through Europe. While Mexico and Guatemala (the Central American countries geographically close to the United States) appear to be the firm's gateway into Central America, the United Kingdom (the European country linguistically close to the United States) seems to be its gateway into Europe. From the United Kingdom, the firm jumped successively into France (in 2003), the Netherlands (in 2004) and Spain (in 2005). Given the short export spells in each of these countries, it is hard to square this export behavior with the hypothetically large sunk entry costs that gravity forces would predict.

By systematically entering markets similar to its prior destinations, the firm in Figure 4 exemplifies patterns present in our data. Table 1 shows export entry probabilities computed using all observations in our sample.[11] The overall entry probability is 0.66%. If our extended gravity story holds, the entry probability in a potential destination will be larger among firms that exported in the previous year to markets that share some extended gravity covariate with it. This prediction matches the evidence in Table 1. The probability of entering a destination conditional on previously exporting to a connected market is always larger than the unconditional one. This probability increase depends on the characteristic shared between both markets: it is more than twofold if both markets share income per capita or language, more than fourfold if they share continent, and approximately tenfold if they share a border.

The evidence in Table 1 is only suggestive of the presence of extended gravity. Other economic forces can also explain these findings. The discussion of these other forces informs the specification of the model introduced in Section 3, written with the aim of guiding us towards an identification approach that controls for these alternative explanations when measuring extended gravity effects.

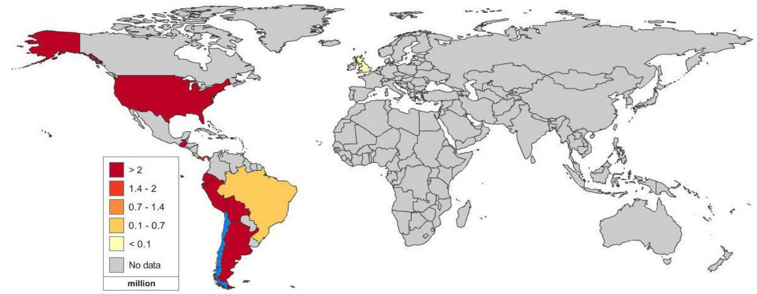
[10] Table A.1 in Appendix A.1 lists all destinations of this firm for every year between 1995 and 2005.

[11] The entry probability in a country equals the number of firms exporting to it in year t and not in $t-1$, divided by the number of non-exporters in $t-1$. Table 1 presents averages across countries of these probabilities.

Figure 1: Export Path of an Individual Firm

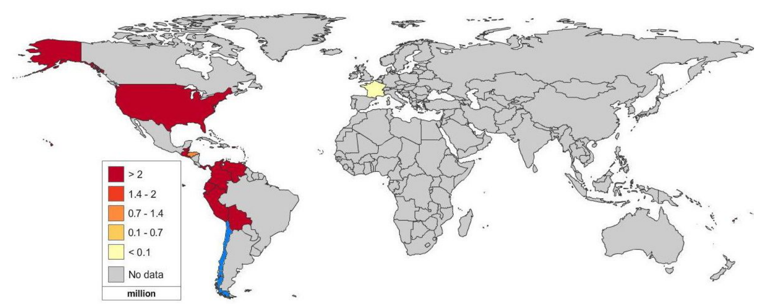
(a) Year 2000

(b) Year 2001



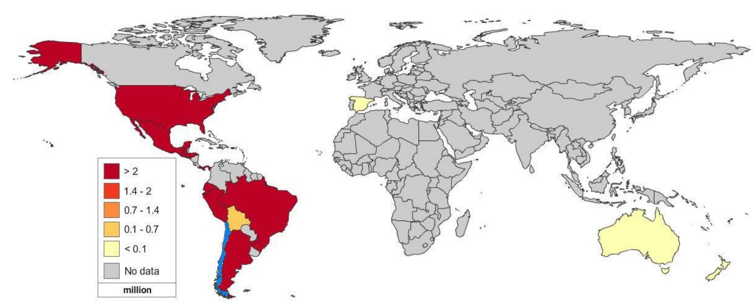
(c) Year 2002

(d) Year 2003



(e) Year 2004

(f) Year 2005



Note: Export destinations of an illustrative firm between years 2000 and 2005.

Table 1: Entry Probabilities

	Probability of Entry	Number of Entries
Overall:	0.66%	1638
Extended Gravity:		
If Ext. Grav. Border = 1	6.74%	397
If Ext. Grav. Cont. = 1	2.79%	525
If Ext. Grav. Lang. = 1	1.59%	205
If Ext. Grav. GDPpc = 1	1.53%	588
If All Ext. Grav. = 0	0.31%	770

First, suppose that firms rank countries by proximity to Chile and spread out gradually to more distant markets (i.e. exports are only determined by standard gravity). Two countries ranked consecutively are likely to belong to the same continent and, thus, in this case, a firm already exporting to a continent will be more likely to subsequently enter other countries in the same continent. This correlation in export entry, however, would be driven by distance between Chile and each destination, not by distance between destinations. It is thus key to account for gravity in order to correctly identify extended gravity. Consequently, in our model, we allow firms' export decisions to depend flexibly on gravity forces.

Second, the higher probability of exporting to a country among firms previously exporting to related markets could reflect similarity in firm-specific demand or supply across these markets. Under this interpretation, for example, the higher probability of exporting to a market for a firm previously exporting to a bordering country would not be due to a reduction in entry costs caused by this export experience (state dependence), but due to similarity in preferences for this firm's output among customers living in these two countries (unobserved heterogeneity). While our benchmark estimates account only for firm-year unobserved heterogeneity, we also show that these are largely robust to allowing firms' export decisions to depend on serially correlated firm- and year-specific unobserved covariates that are common across groups of countries that share any of the extended gravity variables we consider.[12]

Third, the patterns documented in Table 1 could also be due to the combined effect of two forces: (a) most firms do not export at all or export only to a small number of destinations, while a few firms export widely; (b) the more countries a firm exports to, the more likely it is that every new destination shares some characteristic with one of its previous destinations. Thus, correctly identifying extended gravity requires accounting for factors that make some firms, *ceteris paribus*, more likely to export to any country, as well as for the set of potential new export destinations that each firm has in each time period. We exploit our data to account

[12]For example, some firms may be more likely to export to European countries and others more likely to export to Asian countries, and these patterns may be entirely due to factors unobserved to us (e.g. firms sell different varieties that are differentially demanded in different continents) and correlated over time.

for as many determinants of firms' exports as possible. Additionally, we impose only minimal assumptions on the sets of countries that firms consider exporting to in each period.

Finally, even if one can conclude that exporting to a country does indeed cause an increase in the probability of subsequently exporting to other destinations that share some characteristic with it, there are still different mechanisms that may generate this path dependence in export decisions. In our structural estimation, we show that one mechanism present in the data is the reduction in export entry costs due to extended gravity. When identifying this channel, we control for two alternative sources of path dependence. First, we allow our extended gravity variables to also impact exports through other channels. Second, we allow firms' information sets to flexibly evolve as they enter foreign markets. This second mechanism accounts for the possible presence of learning by exporting: as firms export to a country, they gain information about the demand for their products both in that country and in countries similar to it. Our estimates of the impact of extended gravity on export entry costs thus control for the possible presence of learning and should not be attributed to it.

As an intermediate step between the motivating evidence presented in this section and the structural estimates presented in sections 6 and 7, Appendix A presents additional reduced-form evidence on the relevance of extended gravity in determining firms' export destinations.[13]

3 Empirical Model

In this section, we present the model that guides our identification of the impact of extended gravity on the costs that firms face when entering a new destination. Time is discrete and indexed by t. All firms are located in a country h and choose which export markets to export to.[14] We index all firms active in year t by $i = 1, \ldots, N_t$, and the potential destination markets by $j = 1, \ldots, J$. The creation and destruction of firms is treated as exogenous.

3.1 Demand, Variable Trade Costs, and Market Structure

Firms face an isoelastic demand in every country: $q_{ijt} = p_{ijt}^{-\eta} P_{jt}^{\eta-1} Y_{jt}$. The quantity demanded, q_{ijt}, thus depends on the price the firm sets, p_{ijt}, the total expenditure in the market, Y_{jt}, and the price index, P_{jt}, which captures the competition that the firm faces in this market.

A firm produces one unit of output with a_{it} bundles of inputs. The cost of each bundle is w_t, and the marginal production cost is thus $a_{it} w_t$. When firms sell abroad, they also pay

[13] In Appendix A.2, we present entry probabilities similar to those in Table 1, separately for firms of different size and firms with different number of prior export destinations; in Appendix A.3, we present estimates from logit models of export participation that control for gravity; in Appendix A.4, we additionally control for firm-specific unobservables that are common to countries that share a continent, language or similar income per capita; and, in Appendix A.5, we show that the direct impact of extended gravity on export participation is unaffected by whether we allow it to also impact firms' export revenues.

[14] In our empirical application, the market h is Chile. For ease of notation, we eliminate the subindex h.

"iceberg" trade costs: they must ship τ_{ijt} units of output for one unit to reach market j. These costs account for transport costs and for *ad valorem* tariffs charged by country j on goods originating from h. The marginal cost to the firm of selling in market j is thus $\tau_{ijt}a_{it}w_t$.

Conditional on entering a foreign market, exporters behave as monopolistically competitive firms and, thus, the revenue firm i obtains if it exports to market j in period t is

$$r_{ijt} \equiv p_{ijt}q_{ijt} = \left[\frac{\eta}{\eta-1}\frac{\tau_{ijt}a_{it}w_t}{P_{jt}}\right]^{1-\eta} Y_{jt}. \tag{1}$$

We model the impact of variable trade costs τ_{ijt} on export revenues as

$$\tau_{ijt}^{1-\eta} = \exp(\xi_{jt} + \xi_i + X^{\tau}_{ijt}\xi^{\tau}) + \varepsilon^{\tau}_{ijt}, \tag{2}$$

where ξ_{jt} is a country-year term, ξ_i is a firm-specific term, and $X^{\tau}_{ijt} = (d_{ijt-1}, X^e_{ijt}, \ln(a_{it}))$. The variable d_{ijt-1} is a dummy that equals one if firm i exported to country j in year $t-1$, and X^e_{ijt} is a vector that accounts for the extended gravity variables introduced in Section 2.1: "Ext. Grav. Border", "Ext. Grav. Cont.", "Ext. Grav. Lang.", and "Ext. Grav. GDPpc". Finally, ε^{τ}_{ijt} is unobserved and we assume that

$$\mathbb{E}_{jt}[\varepsilon^{\tau}_{ijt}|X^{\tau}_{ijt}, d_{ijt}, \mathcal{J}_{it}] = 0, \tag{3}$$

where $\mathbb{E}_{jt}[\cdot]$ denotes an expectation conditional on a destination-year pair jt, and $\mathcal{J}_{it}$ denotes the information set of the firm when deciding where to export to. Equation (3) thus implies that the unobserved trade costs, ε^{τ}_{ijt}, do not affect the decision of whether to export to market j in year t, as captured by the dummy variable d_{ijt}.

As we show in Appendix B.1, equations (1) to (3) imply that

$$r_{ijt} = \exp(\alpha_{jt} + \alpha_i + X^r_{ijt}\alpha^r) + \varepsilon^R_{ijt}, \tag{4}$$

where α_{jt} is a country-year component common to all firms, α_i is a firm-specific term, and $X^r_{ijt} = (d_{ijt-1}, X^e_{ijt}, \ln(r_{iht}))$. The variable ε^R_{ijt} is unobserved and satisfies

$$\mathbb{E}_{jt}[\varepsilon^R_{ijt}|X^r_{ijt}, d_{ijt}, \mathcal{J}_{it}] = 0. \tag{5}$$

As described in Section 5, our estimation procedure uses equations (4) and (5) to generate a proxy for the potential export revenues r_{ijt} for every firm, country, and year.[15] Equation (4) allows export revenues to depend on gravity (through α_{jt}), extended gravity (through X^e_{ijt}),

[15]While our data includes information on export sales for firm-country-year triplets with positive exports, $r_{ijt}d_{ijt}$, estimating extended gravity effects requires a proxy for r_{ijt} in all remaining cases. Equations (1) to (3) play no role in our analysis except by providing a microfoundation for the relationship between X^r_{ijt} and r_{ijt} in equations (4) and (5). We attach no structural interpretation to the parameters entering equation (4).

firms' domestic sales, r_{iht}, and firms' lagged export status, d_{ijt-1}, in a way that captures prevalent features of our data.[16,17] The term ε^R_{ijt} makes our model compatible with any residual variation in export revenues across firms, countries and years (Eaton et al., 2011).

As discussed in Section 2.2, flexibly modeling the determinants of export revenues is desirable for the identification of extended gravity: any variation in r_{ijt} that affects firms' export participation decisions, is not controlled for explicitly in the export revenue equation (4), and is correlated with our extended gravity covariates would be confounded in our estimation with the true impact of these covariates on export entry costs. However, in practice, the predicted export revenues in our setting are similar across many different specifications of the vector of observed covariates X^r_{ijt} and fixed effects included in the revenue equation (4).[18]

3.2 Fixed and Sunk Export Costs

Exporters also face fixed and sunk export costs. Fixed costs are independent of both the firm's export history and how much it sells to a destination. They account for the cost of advertising, updating information on a market, and participating in trade fairs. We assume

$$f_{ijt} = f^o_j + u_{ic_jt} + \varepsilon^F_{ijt}, \tag{6}$$

where index c_j represents a group of countries to which j belongs, and u_{ic_jt} thus denotes factors that are firm- and year-specific but common to a group c_j (e.g. common to all countries sharing both continent and language). The observable part of fixed costs, f^o_j, is modeled as

$$f^o_j = \gamma^F_0 + \gamma^F_c(\text{Grav. Cont.})_j + \gamma^F_l(\text{Grav. Lang.})_j + \gamma^F_g(\text{Grav. GDPpc})_j, \tag{7}$$

where each gravity term is defined in footnote 8. Both u_{ic_jt} and ε^F_{ijt} are unobserved to the researcher. While we impose no assumption on the distribution of u_{ic_jt}, we assume that

$$\mathbb{E}[\varepsilon^F_{ijt}|d_{ijt}, \mathcal{J}_{it}] = 0. \tag{8}$$

Sunk costs are independent of the quantity exported to a destination, and a firm only has to pay them if it was not exporting to this destination in the previous year. They account for

[16] As shown in Table B.1 (Appendix B.2), the elasticity of r_{ijt} with respect to r_{iht} is 0.4. Our model rationalizes this estimate by allowing τ_{ijt} to depend on a_{it} in equation (2). A model in which τ_{ijt} is a demand shifter is isomorphic to ours. Thus, one can interpret the relationship between τ_{ijt} and a_{it} as a relationship between productivity and "quality" (Verhoogen, 2008; Kugler and Verhoogen, 2012; Fieler et al., 2016).

[17] The dependency of r_{ijt} on d_{ijt-1} is consistent with Ruhl and Willis (2016). This dependency may be due to firms' learning (Albornoz et al., 2012; Berman et al., 2015), partial-year effects (Bernard et al., 2015; Gumpert et al., 2016), or customer capital accumulation (Fitzgerald et al., 2016; Piveteau, 2016).

[18] As we show in Appendix B.2, the key predictors of export revenues are: (a) the firm fixed effect α_i; (b) the firm's domestic sales, r_{iht}; (c) the distance between home and destination markets; and (d) the aggregate sectoral imports of country j in year t. Covariates (c) and (d) are accounted for by the country-year effect α_{jt} in equation (4). All other covariates have little additional explanatory power for export revenues in our data.

expenses in building distribution networks, hiring workers with specific skills (e.g. knowledge of foreign languages), and adapting the exported products to destination-specific preferences and legal requirements. We model them as

$$s_{ijt} = s_j^o - e_{ijt}^o + \varepsilon_{ijt}^S. \tag{9}$$

The observable part of sunk costs depends both on gravity, s_j^o, and extended gravity, e_{ijt}^o. We model the gravity term as

$$s_j^o = \gamma_0^S + \gamma_c^S(\text{Grav. Cont.})_j + \gamma_l^S(\text{Grav. Lang.})_j + \gamma_g^S(\text{Grav. GDPpc})_j, \tag{10}$$

and the extended gravity term as

$$\begin{aligned} e_{ijt}^o = \gamma_b^E(\text{Ext. Grav. Border})_{ijt} + \gamma_c^E(\text{Ext. Grav. Cont.})_{ijt} + \gamma_l^E(\text{Ext. Grav. Lang.})_{ijt} \\ + \gamma_g^E(\text{Ext. Grav. GDPpc})_{ijt}, \end{aligned} \tag{11}$$

with each extended gravity variable defined in footnote 9. The term e_{ijt}^o thus depends on all export destinations of firm i in year $t-1$: it accounts for the possibility that entry costs in a market are smaller for those firms that have previously exported to countries similar to it. We assume that ε_{ijt}^S is unobserved to the firm when it is deciding on its set of export destinations,

$$\mathbb{E}[\varepsilon_{ijt}^S | d_{ijt}, \mathcal{J}_{it}] = 0. \tag{12}$$

As discussed in Section 2.2, correctly identifying extended gravity requires controlling for the impact on firms' export decisions of firm-specific unobserved (to the researcher) covariates that are correlated over time and across countries that share characteristics giving rise to extended gravity relationships. We allow for such unobserved covariates through the term u_{ic_jt} in equation (6). We do not restrict the correlation in u_{ic_jt} across firms, groups of countries, and time periods, nor its relationship to firms' export decisions. Thus, our model is consistent with firms conditioning on u_{ic_jt} when deciding where to export to. Our baseline results in Section 6 assume that $c_j = c$ for all j. In Section 7, we present additional results in which we clasify countries into groups c_j according to their continent, language, or income per capita.

3.3 Export Profits

The assumptions on demand, variable production costs and market structure in Section 3.1, and the assumptions on fixed and sunk costs in Section 3.2, imply that the potential static profits of exporting to a destination j are

$$\pi_{ijt} = r_{ijt} - \tau_{ijt} a_{it} w_t q_{ijt} - f_{ijt} - (1 - d_{ijt-1}) s_{ijt} = \eta^{-1} r_{ijt} - f_{ijt} - (1 - d_{ijt-1}) s_{ijt}, \tag{13}$$

and the total potential static profits of exporting to a bundle b of destinations are

$$\pi_{ibt} = \sum_{j \in b} \pi_{ijt}. \tag{14}$$

Through the dependency of export revenues and sunk costs on last period's exports, the static export profits in period t, π_{ibt}, will depend on the export bundle chosen in $t-1$. However, *conditional on the* $t-1$ *export bundle*, all prior years' export destinations have no impact on the static profits in period t.[19] The export bundle in $t-2$ will thus directly affect the static profits in $t-1$, but it will affect those in t only indirectly through the optimal set of export destinations in $t-1$. Thus, the dynamic problem of the firm exhibits one-period dependence.

3.4 Optimal Export Destinations

While we use b to denote a generic bundle of countries that a firm may choose to export to, we use o_{it} to denote the export bundle actually chosen by firm i in period t. Formally, o_{it} is a vector that indicates the export status in each of the J export markets: $o_{it} = (d_{i1t}, \ldots, d_{ijt}, \ldots, d_{iJt})$, where, as a reminder, d_{ijt} equals one if firm i exports to market j in year t, and zero otherwise. Assumption 1 indicates how firms choose the vector o_{it} in every time period.

Assumption 1 *For every firm i and period t, let o_{it} denote the observed bundle of export destinations, $\mathcal{J}_{it}$ denote the information set, and $\mathcal{B}_{it}$ denote the consideration set. Then*

$$o_{it} = \underset{b \in \mathcal{B}_{it}}{\operatorname{argmax}} \, \mathbb{E}\left[\Pi_{ibt,L_{it}} | \mathcal{J}_{it}\right], \tag{15}$$

where $\mathbb{E}[\cdot]$ denotes the expectation consistent with the data generating process and

$$\Pi_{ibt,L_{it}} = \pi_{ibt} + \sum_{l=1}^{L_{it}} \delta^l \pi_{io_{it+l}(b)t+l}, \tag{16}$$

where δ is the discount factor and $o_{it+l}(b)$ denotes the optimal export bundle that firm i would choose at period $t+l$ if it had exported to the bundle b in period t.

Assumption 1 characterizes the firm's observed export bundle as the outcome of an optimization problem defined by three elements: (1) the L_{it} periods ahead discounted sum of profits $\Pi_{ibt,L_{it}}$; (2) the consideration set $\mathcal{B}_{it}$ or set of export bundles among which the firm selects its preferred one; (3) the information set $\mathcal{J}_{it}$ or set of variables the firm uses to predict its

[19] This would not be true if sunk costs in a country for firms that had exported two periods ago to it were lower than those for firms that had never exported to it. Extending our methodology to allow for such entry costs is straightforward. However, Roberts and Tybout (1997) provide evidence consistent with the assumption that export entry decisions exhibit only one-period dependence.

potential export profits in each of the bundles included in $\mathcal{B}_{it}$. Identifying the impact of extended gravity on sunk costs requires imposing some restrictions on these three elements of the optimization problem. Assumptions 2 to 4 indicate the restrictions we impose.

Assumption 2 $L_{it} \geq 1$.

We impose only weak restrictions on how forward-looking firms are when deciding their optimal export bundle. Our model is compatible with firms that take into account the effect of their current choices on future profits in any of the three following ways: (a) only one period ahead, $L_{it} = 1$; (b) any finite number $p > 2$ of periods ahead, $L_{it} = p$; or, (c) an infinite number of periods ahead, $L_{it} = \infty$ (i.e. perfectly forward looking firms). Furthermore, different firms may have different planning horizons, and the planning horizon of a firm may change over time; i.e. L_{it} may be different from $L_{i't'}$ for $i \neq i'$ or $t \neq t'$. This heterogeneity in planning horizons accommodates differences across managers in their investment preferences.[20]

Equation (15) imposes that firms' expectations are rational but leaves their information sets unrestricted. Assumption 3 indicates the restriction we impose on them.

Assumption 3 *$Z_{it} \subseteq \mathcal{J}_{it}$, where Z_{it} is a vector of observed covariates.*

We thus impose that the researcher observes a vector Z_{it} that is included in the firm's information set $\mathcal{J}_{it}$.[21] Specifically, to compute our estimates, we specify the vector Z_{it} as

$$Z_{it} = (Z_{ijt}, j = 1, \ldots, J), \tag{17a}$$

$$Z_{ijt} = (f_j^o, s_j^o, e_{ijt}^o, d_{ijt-1}), \tag{17b}$$

where, as a reminder, f_j^o, s_j^o and e_{ijt}^o are components of fixed and sunk costs that depend exclusively on gravity and extended gravity variables, and d_{ijt-1} captures the lagged export status of the firm in country j. Equation (17) thus only requires firms to know whether each foreign country shares continent, language or similar income per capita either with Chile or with at least one country to which they exported to in the previous year. It is reasonable to assume that all potential exporters have this information. Beyond this minimal content, we do not impose any assumption on firms' information sets. The variables not in Z_{it} that complete the information set $\mathcal{J}_{it}$ may thus vary flexibly across firms and years.

Assumption 3 and the definition of Z_{it} in equation (17) allow for a large degree of unobserved heterogeneity in the uncertainty firms face when deciding which countries to export to. They are compatible with firms having different information both on the export revenue they

[20] Bandiera et al. (2015) find that managers have heterogeneous utility functions. Pennings and Garcia (2008) show that this heterogeneity matters for their investment decisions, and Cheng and Steinwender (2016) show that different managers react differently to trade shocks.

[21] Whenever we indicate that a random vector Z_{it} is included in the true information set $\mathcal{J}_{it}$, $Z_{it} \subseteq \mathcal{J}_{it}$, we formally mean that the distribution of Z_{it} conditional on $\mathcal{J}_{it}$ is degenerate.

would obtain in each market, r_{ijt}, and on the fixed cost component u_{ic_jt}. These differences may be due to firms' investing differentially in acquiring information or having different prior export experience across markets (consistent with within-firm learning).[22]

Equation (15) also leaves unrestricted the consideration set $\mathcal{B}_{it}$. The *potential* choice set among which firms may choose their optimal export bundle includes all combinations of foreign countries. Given that the number of countries J to which at least one firm in our data exports is over 100, it is unrealistic to assume that firms evaluate the trade-offs, as captured by the function $\mathbb{E}[\Pi_{ibt,L_{it}}|\mathcal{J}_{it}]$, of exporting to each of these 2^J bundles of countries. The firm's consideration set $\mathcal{B}_{it}$ is thus likely smaller than the potential choice set. However, the lack of data on firms' consideration sets makes it hard to correctly specify the set $\mathcal{B}_{it}$ of every firm and period in our sample. For this reason, we do not specify a consideration set $\mathcal{B}_{it}$ for every firm and year, but just impose a minimal content requirement on it.

Assumption 4 *$\mathcal{A}_{it} \subset \mathcal{B}_{it}$, where $\mathcal{A}_{it}$ is known to the researcher.*

This assumption imposes that the researcher must list the elements of a subset $\mathcal{A}_{it}$ of the true consideration set $\mathcal{B}_{it}$. Specifically, to compute our estimates, we specify the set $\mathcal{A}_{it}$ as

$$\mathcal{A}_{it} = \{o_{it}\} \cup \{o_{it}^{j \to j'}, \forall j = 1, \dots, J, \text{ and } j' = 1, \dots, J \text{ such that } j' \in \mathcal{A}_{ijt}\}, \tag{18a}$$

$$\mathcal{A}_{ijt} = \{j' = 1, \dots, J \text{ such that } f_j^o = f_{j'}^o \text{ and } u_{ic_jt} = u_{ic_{j'}t}\}, \tag{18b}$$

where $o_{it}^{j \to j'}$ is the bundle constructed by swapping the observed destination j for the alternative one j'. Formally, for a bundle o_{it} with $d_{ijt} = 1$ and $d_{ij't} = 0$, the bundle $o_{it}^{j \to j'} = (d'_{i1t}, \dots, d'_{iJt})$ is constructed as: (a) $d_{ij''t} = d'_{ij''t}$ if $j'' \neq j$ and $j'' \neq j'$; (b) $d_{ijt} - 1 = d'_{ijt}$; and (c) $d_{ij't} + 1 = d'_{ij't}$. The set $\mathcal{A}_{it}$ includes the observed bundle, o_{it}, plus all other ones built by swapping an observed destination j for an alternative one j' that belongs to the set $\mathcal{A}_{ijt}$. According to the definition of $\mathcal{A}_{ijt}$, destinations j and j' must share: (a) the component of fixed costs f_j^o, defined in equation (7); and (b) the unobserved fixed costs term u_{ic_jt}, defined in equation (6). Requirement (a) implies that both j and j' must have the same gravity relationship to the home country of the firm: either both or none of them share continent, language or similar income per capita with country h. Depending on how we define the groups of countries assumed to share the value of u_{ic_jt}, requirement (b) may additionally require j and j' to share a continent, language or similar income per capita with each other.[23] Con-

[22] Dickstein and Morales (2016) find that large firms have more information relevant to predict r_{ijt} than small firms, and their evidence suggests that this informational advantage is due to larger investments in acquiring information. Multiple papers provide evidence consistent with within-firm learning (Albornoz et al., 2012; Berman et al., 2015; Arkolakis et al., 2015b; Timoshenko, 2015a,b; Bastos et al., 2016; Fitzgerald et al., 2016).

[23] For example, if we assume that u_{ic_jt} is common to all countries located in the same continent ($c_j = c_{j'}$ if j and j' belong to the same continent), requirement (a) implies that either both or none of j and j' have Spanish as official language (Chile's official language) and share similar income per capita with Chile; and requirement (b) requires both j and j' to be located in the same continent. In this case, we could thus hypothetically swap the United Kingdom for Germany, but not for the United States, as they are located in different continents.

sequently, Assumption 4 and equation (18) imply consideration sets that include *at least* the observed choice, o_{it}, plus small perturbations around it. The set of bundles not in $\mathcal{A}_{it}$ that complete the consideration set $\mathcal{B}_{it}$ may vary flexibly across firms and years.

3.5 Parameters to Identify, Identification Approach and Prior Literature

The unknown model parameters are the demand elasticity η; the discount factor δ; the export revenue parameters entering equation (4),

$$\alpha \equiv (\{\alpha_{jt}\}_{j,t}, \{\alpha_i\}_i, \alpha^r); \tag{19}$$

the fixed and sunk costs parameters entering equations (7), (10), and (11),

$$\gamma \equiv (\gamma_0^F, \gamma_c^F, \gamma_l^F, \gamma_g^F, \gamma_0^S, \gamma_c^S, \gamma_l^S, \gamma_g^S, \gamma_b^E, \gamma_c^E, \gamma_l^E, \gamma_g^E); \tag{20}$$

the planning horizon L_{it}, information set $\mathcal{J}_{it}$, and consideration set $\mathcal{B}_{it}$ of every firm i and year t in the sample; and the joint distribution of the unobserved determinants of export revenues, fixed and sunk costs, defined respectively in equations (4), (6), and (9).

Prior literature that has estimated single-agent export entry models has done so by assuming away extended gravity effects, $\gamma_b^E = \gamma_c^E = \gamma_l^E = \gamma_g^E = 0$; fixing the value of δ to a number close to 1; specifying the exact planning horizon, L_{it}, and the precise content of both the information and the consideration sets, $\mathcal{J}_{it}$ and $\mathcal{B}_{it}$, of every firm and year in the sample; and imposing parametric restrictions on the distributions of the unobserved determinants of export profits. Given these assumptions, the remaining parameters are point identified. This approach cannot be applied in our setting. Once we allow the extended gravity parameters to differ from zero, computational feasibility forces us to impose strong assumptions on planning horizons and information and consideration sets so that we can estimate the remaining parameters through maximum likelihood or a method of moments approach.[24]

Even if computational feasibility was not a constraint, computing extended gravity estimates that depend on precise definitions of the firm's planning horizon, information and consideration sets would be undesirable. As the simulation in Appendix D.2 illustrates, extended gravity estimates are biased if these model elements are misspecified. Given the lack of data on firms' planning horizons, information and consideration sets and our aim of correctly identifying the extended gravity parameters, we opt for imposing only the relatively weak restrictions indicated in assumptions 2 to 4 and equations (17) and (18). Imposing only these weak restrictions is not without costs: they are not strong enough to point identify the extended gravity parameters and, furthermore, the resulting model is not suitable to analyze

[24] A method of moments approach is feasible if $L_{it} = 0$ for all it pairs and $\mathcal{J}_{it}$ is such that firms have perfect foresight; see Jia (2008), Tintelnot (2016), Antràs et al. (2017), and Arkolakis and Eckert (2017).

how firms' export decisions change in response to counterfactual changes in the environment.

We quantify the importance of extended gravity in reducing export entry costs by identifying bounds on the following vector of *relative* extended gravity parameters

$$\kappa \equiv (\kappa_b, \kappa_c, \kappa_l, \kappa_g) \equiv \left(\frac{\gamma_b^E}{\gamma_{all}^S}, \frac{\gamma_c^E}{\gamma_{all}^S}, \frac{\gamma_l^E}{\gamma_{all}^S}, \frac{\gamma_g^E}{\gamma_{all}^S} \right), \qquad \gamma_{all}^S \equiv \gamma_0^S + \gamma_c^S + \gamma_l^S + \gamma_g^S. \tag{21}$$

The parameter κ_b captures the relative reduction due to extended gravity in border in the sunk costs of entering a country that differs from Chile in all three gravity variables included in our analysis. The parameters κ_c, κ_l and κ_g capture analogous relative reductions due to extended gravity in continent, language and similarity in income per capita. For example, for a firm entering Germany, κ_g indicates the relative reduction in sunk costs if previously exporting to the United States, κ_c indicates the corresponding reduction if previously exporting to Romania, $\kappa_c + \kappa_g$ if previously exporting to Spain, $\kappa_b + \kappa_c + \kappa_g$ if exporting to France, and $\kappa_b + \kappa_c + \kappa_l + \kappa_g$ if exporting to Austria. Focusing on identifying the parameter vector κ, instead of the parameter vector γ, has several advantages. First, the value of κ is independent of the units in which export sales are measured and, thus, is easier to interpret. Second, identifying bounds on κ does not require fixing any parameter to a normalizing constant and, thus, these bounds are scale-invariant. Third, the assumptions required to identify κ are weaker than those needed to identify all elements of γ.[25] Fourth, it is computationally much simpler.[26]

4 Deriving Moment Inequalities

In Section 4.1, we derive conditional moment inequalities from the model described in Section 3. In Section 4.2, we transform these conditional moment inequalities into unconditional ones. In Section 4.3, we illustrate how these unconditional moment inequalities may be used to compute bounds on the elements of the extended gravity parameter vector κ.

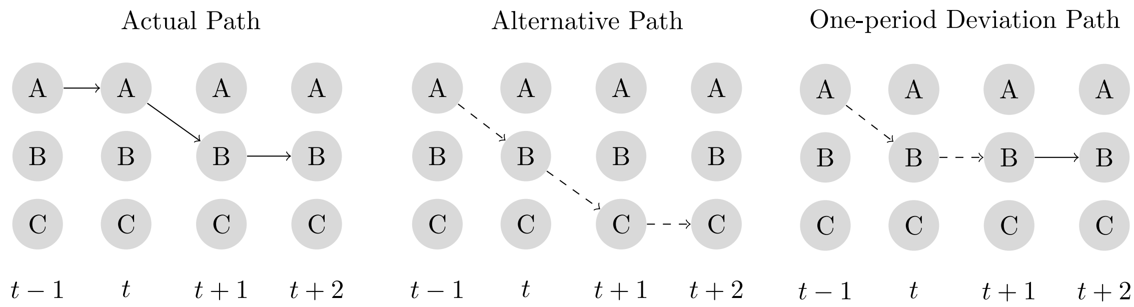
4.1 One-period Deviations

We apply an analogue of Euler's perturbation method to derive inequalities: we compare the stream of profits along a firm's observed sequence of bundles with the stream along alternative sequences that differ from the observed one in just one period. Denoting the observed sequence

[25] E.g. identifying γ_0^F requires parametric restrictions on the distribution of u_{ic_jt}, and identifying γ_0^F, γ_c^F, γ_l^F and γ_g^F requires expanding the set $\mathcal{A}_{it}$ to include alternative export bundles that differ from the observed one both in the number of export destinations and in the gravity characteristics of the countries included in them.

[26] Ho and Rosen (2016) discuss how the computational cost of standard moment inequality inference procedures increases with the dimensionality of the parameter vector to estimate: these procedures require evaluating whether each point in the parameter space verifies a condition determining its inclusion in the confidence set. As γ includes 12 parameters, computing a confidence set for it is computationally expensive: a grid covering a 12-dimensional space requires a very large number of points in order to keep the discretization error small.

Figure 2: Actual and Counterfactual Path: Example



The left panel describes the actual export path: the firm chooses destination A in periods $t-1$ and t, and destination B in periods $t+1$ and $t+2$. The middle panel describes the path that would have been optimal conditional on choosing destination B in year t. The right panel describes a one-period deviation path from the actual one: it is identical to the actual path except for swapping destination A for destination B in period t. The solid arrows denote transitions observed in the data; the dotted arrows denote counterfactual transitions.

as $o_{i1}^T = \{\ldots, o_{it-1}, o_{it}, o_{it+1}, \ldots\}$, we form inequalities by comparing the expected discounted sum of profits generated by it to that generated by an alternative sequence that differs from o_{i1}^T in the bundle chosen in t, $\{\ldots, o_{it-1}, o_{it}^{j\to j'}, o_{it+1}, \ldots\}$, where, as a reminder, $o_{it}^{j\to j'}$ is the bundle that results from swapping destination j by j' in o_{it}. Given the one-period dependence in export profits imposed in our model (see Section 3.3), the difference in the discounted sum of profits generated by the observed and the alternative paths depends only on the difference in static profits in periods t, $\pi_{ijj't}$, and $t+1$, $\pi_{ijj't+1}$,

$$\underbrace{\pi_{ijt} - \pi_{ij't}}_{\pi_{ijj't}} + \underbrace{\delta \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi_{ij''t+1}^{j\to j'})}_{\pi_{ijj't+1}}, \tag{22}$$

where $\pi_{ij''t+1}^{j\to j'}$ are the potential profits of i in country j'' and year $t+1$ if it chooses $o_{it}^{j\to j'}$ in t.

To provide intuition on equation (22), we present in Figure 2 the example of a firm that must choose which of three possible markets to export to. The left panel describes the firm's observed path: it chooses market A in periods $t-1$ and t, and market B in periods $t+1$ and $t+2$. From Assumption 1, these are the firm's optimal export destinations in each period. The middle panel describes the choice that would have been optimal in $t+1$ and $t+2$ if the firm had deviated from the optimal path and chosen destination B in period t. The right panel describes an alternative path built as a one-period deviation from the observed one: it deviates from it only in that market B is chosen in period t. Using the model's notation: $o_{it} = (d_{iAt}, d_{iBt}, d_{iCt}) = (1,0,0)$ and $o_{it}^{A\to B} = (0,1,0)$. The static profits of exporting to market B in period $t+2$ are independent of the choice made by the firm in period t: profits exhibit one-period dependence and, thus, they depend on export history only through the export bundle in the previous period. This is why period $t+2$ static profits do not enter equation (22). Conversely, the difference in static profits in periods t and $t+1$ between the actual and the period-t deviating path will generally be different from zero.

Proposition 1 shows how we use one-period deviations to derive moment inequalities.

Proposition 1 *Suppose assumptions 1, 2, and 3 hold; then, for any pair of countries j and j' such that $o_{it}^{j \to j'} \in \mathcal{B}_{it}$, and any Z_{it},*

$$\mathbb{E}\left[\pi_{ijj't} + \delta \pi_{ijj't+1} \middle| d_{ijt}(1-d_{ij't}) = 1, Z_{it}\right] \geq 0. \tag{23}$$

The proof of Proposition 1 is in Appendix C.1. Equation (23) indicates that, conditional on any vector of covariates included in the firm's information set, Z_{it}, and firm i exporting to country j and not to j' in period t ($d_{ijt}(1-d_{ij't}) = 1$), the expected discounted sum of profits along the observed export path is weakly larger than the expected discounted sum along an alternative path that swaps the observed destination j for the alternative j' in period t. Equation (23) is a revealed preference inequality and, thus, for it to hold, the bundle $o_{it}^{j \to j'}$ must belong to the consideration set of the firm in period t, $\mathcal{B}_{it}$.

To gain intuition on the inequality in equation (23), we use again the example in Figure 2. Since, according to the left panel, destination A was chosen in period t, Assumption 1 implies that, conditional on the information available to the firm in period t, the firm weakly prefers the optimal export path that includes destination A in t (in the left panel) to the optimal path that includes destination B in t (in the middle panel). However, once country B is selected in period t, choices in subsequent periods that are optimal conditional on this choice (in the middle panel) must be preferred over choices that would have been optimal only if destination A had been selected in t (in the left panel). Transitivity of preferences thus ensures that the optimal path described in the left panel is, given the information available to the firm in t, weakly preferred over the one-year deviation path described in the right panel.[27]

The inequality in equation (23) illustrates how the general approach in Pakes (2010) and Pakes et al. (2015) can be applied to single-agent dynamic discrete choice models. Our strategy of using one-period deviations to build estimating equations follows the methodology in Hansen and Singleton (1982) and Luttmer (1999), but is adapted to our moment inequality setting.[28]

[27] According to Proposition 1, it need not be the case that the realized or ex post difference in profits is positive. Formally, our model does not imply that $d_{ijt}(1-d_{ij't})(\pi_{ijj't} + \delta \pi_{ijj't+1}) \geq 0$ for every i, j, j' and t.

[28] Arcidiacono and Miller (2011); Scott (2013); Aguirregabiria and Magesan (2013, 2016); and Traiberman (2016) also use one-period deviations to estimate dynamic discrete choice models. These papers, however, fully specify the agents' planning horizon, and information and consideration sets. Also, for every realization of these information sets, these procedures require estimating nonparametrically the probability that any alternative in the consideration set is chosen. Given the dimensionality of any reasonable specification of the information and consideration sets in our setting, performing this nonparametric estimation is infeasible in our case.

Thus, our inequalities do not condition on choices that the firm makes in periods later than the deviating period t. This differentiates our inequalities from those in Holmes (2011) and Illanes (2016), and it is important for our purposes because conditioning on the firm's subsequent choices would rule out the definition of $(\varepsilon^R_{ijt}, \varepsilon^F_{ijt}, \varepsilon^S_{ijt})$ as expectational errors.[29]

4.2 From Conditional to Unconditional Moment Inequalities

The inequalities in equation (23) have two properties that complicate their applicability in estimation. First, they condition on a particular pair of destinations j and j', implying that the number of inequalities that one can construct is larger than the sample size. Second, they condition on the vector Z_{it}, which may take many values. These two characteristics imply that the sample analogue of most of the moments in equation (23) will average over very few observations. To facilitate estimation, we exploit the many conditional moment inequalities in equation (23) to derive a finite number of unconditional inequalities that aggregate across pairs of actual and counterfactual destinations and across observations with different values of Z_{it}.[30] Conditioning on a finite set of moments, while convenient, may entail a loss of information relative to the many conditional moment inequalities in equation (23). However, as Section 6 shows, the inequalities we employ nonetheless generate economically meaningful bounds on the parameters of interest. Proposition 2 characterizes our unconditional inequalities.

Proposition 2 *Suppose equation* (23) *and Assumption 4 hold; then, for any function* $\Psi(\cdot)$ *such that*

$$\Psi(Z_{ijt}, Z_{ij't}) \geq 0 \tag{24}$$

for all values of $(Z_{ijt}, Z_{ij't})$ *in their support, it holds that*

$$\mathbb{E}\left[\sum_{j=1}^{J} \sum_{j' \in \mathcal{A}_{ijt}} \Psi(Z_{ijt}, Z_{ij't}) d_{ijt}(1 - d_{ij't})(\pi_{ijj't} + \delta\pi_{ijj't+1})\right] \geq 0. \tag{25}$$

The proof of Proposition 2 is in Appendix C.2. The inequality in equation (25) sums over all pairs (j, j') such that j is an actual export destination, $d_{ijt} = 1$, and j' is a market to which the

[29] Equations (4), (6) and (9) assume that the vector $(\varepsilon^R_{ijt}, \varepsilon^F_{ijt}, \varepsilon^S_{ijt})$ is mean independent of $\mathcal{J}_{it}$, but do not restrict its relationship to any information set $\mathcal{J}_{it'}$ such that $t' > t$. This is consistent with the interpretation of $(\varepsilon^R_{ijt}, \varepsilon^F_{ijt}, \varepsilon^S_{ijt})$ as expectational errors. If we were to apply the inequalities in Holmes (2011) to our setting, we would need to assume that the vector $(\varepsilon^R_{ijt}, \varepsilon^F_{ijt}, \varepsilon^S_{ijt})$ is mean independent of the information sets $\mathcal{J}_{it'}$ in every period t'. If were to apply the inequalities in Illanes (2016) to our setting, we would need to assume that the vector $(\varepsilon^R_{ijt}, \varepsilon^F_{ijt}, \varepsilon^S_{ijt})$ belongs to the information set $\mathcal{J}_{it}$. We opt for our approach because, as discussed in Dickstein and Morales (2016), allowing for expectational errors is key to the estimation of export entry costs.

[30] Menzel (2014); Chernozhukov et al. (2014); Bugni et al. (2016b) introduce inference procedures in settings with many moment inequalities. Andrews and Shi (2013); Chernozhukov et al. (2013); Armstrong (2014, 2015); Armstrong and Chan (2016) study conditional moment inequality models.

firm does not export, $d_{ij't}=0$, but that is included in the set $\mathcal{A}_{ijt}$ specified by the researcher, $j' \in \mathcal{A}_{ijt}$. Assumption 4 requires all bundles formed by swapping country j for an alternative j' included in $\mathcal{A}_{ijt}$ to belong to the consideration set $\mathcal{B}_{it}$. This motivates our choice of $\mathcal{A}_{ijt}$ in equation (18b) as including only countries that are similar to the observed destination j; i.e. countries that the firm is likely to have considered when selecting j as destination.

When summing over pairs of destinations j and j', the inequality in equation (25) weights them according to a function $\Psi(\cdot)$ that must satisfy two restrictions: (a) it is weakly positive; (b) it is a function only of variables observed to the researcher and assumed to belong to the firm's information set $\mathcal{J}_{it}$. This motivates our choice of Z_{ijt}, described in equation (17b), as including only variables that every firm is likely to know.

Without additional restrictions on the set $\mathcal{A}_{ijt}$ and the instrument function $\Psi(\cdot)$, the profit differences $\pi_{ijj't}$ and $\pi_{ijj't+1}$ in equation (25) will depend on all observed and unobserved determinants of export revenues, fixed and sunk costs, and all parameters included the vectors α and γ defined in equations (19) and (20). For our aim of computing bounds on the extended gravity parameters in the vector κ defined in equation (21), inequalities that depend on all elements of γ or on the unobserved determinant of fixed costs u_{ic_jt} are problematic. As discussed in footnote 26, computing a confidence set for all elements of γ is very costly. Additionally, if the moments we use for estimation depend on the unobserved term u_{ic_jt}, then we would need to either assume that it is mean independent of the firm's information set, $\mathcal{J}_{it}$, or impose parametric restrictions on its distribution. Either of these distributional assumptions on u_{ic_jt}, if inaccurate, will bias our estimates of κ. The following proposition indicates how we solve these two problems by restricting the set $\mathcal{A}_{ijt}$ and the function $\Psi(\cdot)$.

Proposition 3 *Suppose equations* (5), (8), (12), *and* (18b) *hold, and that*

$$\Psi(Z_{ijt}, Z_{ij't}) = 0 \quad \text{if} \quad s^o_j \neq \gamma^S_{all} \text{ or } s^o_{j'} \neq \gamma^S_{all}; \tag{26}$$

then the moment

$$\mathbb{E}\left[\sum_{j=1}^{J}\sum_{j'\in\mathcal{A}_{ijt}} \Psi(Z_{ijt}, Z_{ij't}) d_{ijt}(1-d_{ij't})(\pi_{ijj't}+\delta\pi_{ijj't+1})\right] \tag{27}$$

depends only on the distribution of a vector of observed covariates and the parameter vector $\theta^* \equiv (\alpha, \kappa, \eta, \gamma^S_{all})$.

The proof of Proposition 3 is in Appendix C.3. Importantly, this proposition indicates that, if we restrict $\mathcal{A}_{ijt}$ according to equation (18b), $\Psi(\cdot)$ according to equation (26), and the distribution of the vector $(\varepsilon^R_{ijt}, \varepsilon^F_{ijt}, \varepsilon^S_{ijt})$ according to equations (5), (8), and (12), then the moment in equation (27) does not depend on unobserved determinants of export profits, and depends on the vector γ only through the parameters γ^S_{all} and κ defined in equation (21).[31]

The following corollary indicates how we derive the finite set of unconditional moment inequalities we use to estimate bounds on κ.

Corollary 1 *Given propositions 2 and 3 and any K functions $\{\Psi_k(Z_{ijt}, Z_{ij't})\}_{k=1}^K$ satisfying equations* (24) *and* (26)*; then, for every $k = 1, \ldots, K$, it holds:*

$$\mathbb{m}_k(\theta^*) \equiv \mathbb{E}\left[\sum_{j=1}^{J} \sum_{j' \in \mathcal{A}_{ijt}} \Psi_k(Z_{ijt}, Z_{ij't}) d_{ijt}(1 - d_{ij't})(\pi_{ijj't} + \delta\pi_{ijj't+1})\right] \geq 0. \tag{28}$$

Appendix C.4 lists the $K = 10$ instrument functions we use in our estimation. Let's define a vector θ of unknown parameters whose true value is θ^*. Then, denoting by Θ the set of all values of θ such that $\mathbb{m}_k(\theta) \geq 0$, for all $k = 1, \ldots, K$, Corollary 1 implies that $\theta^* \in \Theta$.

4.3 Using Inequalities to Bound Extended Gravity Parameters: Intuition

We illustrate here how, by properly selecting alternative destinations to compare to the observed ones, one can construct inequalities that bound the parameters of interest.

In the example in Table 2, a firm enters the United Kingdom in year 8, and we consider an alternative path in which it enters Germany instead. Both countries are in Europe and have similar income per capita, but differ in that the former is English-speaking and the latter is German-speaking. Assume that the firm exported only to the United States in year 7 and does not export anywhere in year 9. Therefore, in terms of extended gravity effects, the actual and counterfactual paths differ only in that the firm benefits from extended gravity in language in the former but not in the latter. Indexing the observed destination with j and the alternative one with j', the difference in year 8 static profits between actual and counterfactual paths is:

$$\begin{aligned}
\pi_{ijj'8} &= \eta^{-1} r_{ij8} - f_{ij8} - s_{ij8} - (\eta^{-1} r_{ij'8} - f_{ij'8} - s_{ij'8}) \qquad (29)\\
&= \eta^{-1}(r_{ij8} - r_{ij'8}) - (u_{ic_j8} - u_{ic_{j'}8}) - (\varepsilon^F_{ij8} - \varepsilon^F_{ij'8}) + (e^o_{ij8} - e^o_{ij'8}) - (\varepsilon^S_{ij8} - \varepsilon^S_{ij'8})\\
&= \eta^{-1}(r_{ij8} - r_{ij'8}) - (u_{ic_j8} - u_{ic_{j'}8}) - (\varepsilon^F_{ij8} - \varepsilon^F_{ij'8}) + \gamma^E_l - (\varepsilon^S_{ij8} - \varepsilon^S_{ij'8})\\
&= \eta^{-1}(r^o_{ij8} - r^o_{ij'8} + \varepsilon^R_{ij8} - \varepsilon^R_{ij'8}) - (u_{ic_j8} - u_{ic_{j'}8}) - (\varepsilon^F_{ij8} - \varepsilon^F_{ij'8}) + \gamma^E_l - (\varepsilon^S_{ij8} - \varepsilon^S_{ij'8}),
\end{aligned}$$

where the first line uses equation (13); the second line applies equations (6) and (9), and takes into account that Germany and the United Kingdom share all gravity variables affecting the observable components of fixed costs, $f^o_j = f^o_{j'}$, and sunk costs, $s^o_j = s^o_{j'}$; the third line exploits

[31]The intuition behind Proposition 3 is the following. First, according to equation (18b), destinations j and j' must satisfy that $f^o_j = f^o_{j'}$; thus, the moment in equation (27) differences out all terms that depend on $(\gamma^F_0, \gamma^F_c, \gamma^F_l, \gamma^F_g)$. Second, also according to equation (18b), j and j' must also satisfy that $u_{ic_jt} = u_{ic_{j'}t}$, which are thus also differenced out. Finally, according to equation (26), j and j' must share no gravity characteristic with Chile; thus, equation (27) depends on $(\gamma^S_0, \gamma^S_c, \gamma^S_l, \gamma^S_g)$ only through the scalar γ^S_{all}.

Table 2: Example of a 1-period Export Event

		$t=7$	$t=8$	$t=9$
Observed	United Kingdom	0	1	0
	Germany	0	0	0
Alternative	United Kingdom	0	0	0
	Germany	0	1	0

that the firm's single export destination in year 7 shares language with j but not with j'; and the fourth line uses the expression for export revenues in equation (4), with the notational simplification $r^o_{ijt} \equiv \exp(\alpha_{jt} + \alpha_i + X^r_{ijt}\alpha^r)$. As the firm does not export in year 9, its export profits in this year are zero both in the actual and counterfactual paths, and then

$$d_{ij''9}(\pi_{ij''9} - \pi^{j\rightarrow j'}_{ij''9}) = 0, \qquad \text{for all } j'' \in J. \tag{30}$$

Therefore, the equivalent of the difference in profits in equation (22) is

$$\begin{gathered}\pi_{ijj'8} + \delta\pi_{ijj'9} = \\ \eta^{-1}(r^o_{ij8} - r^o_{ij'8} + \varepsilon^R_{ij8} - \varepsilon^R_{ij'8}) - (u_{ic_j8} - u_{ic_{j'}8}) - (\varepsilon^F_{ij8} - \varepsilon^F_{ij'8}) + \gamma^E_l - (\varepsilon^S_{ij8} - \varepsilon^S_{ij'8}).\end{gathered} \tag{31}$$

However, not every possible pair of observed and alternative destinations may be used to build our inequalities. As propositions 2 and 3 show, our inequalities compare an observed destination j only to those alternative ones included in the set $\mathcal{A}_{ijt}$ defined in equation (18b). By restricting in this way the set of alternative destinations, our inequalities include only profit differences between destinations j and j' such that $u_{ic_jt} = u_{ic_{j'}t}$ for every firm i and year t. Therefore, Germany is a valid alternative to the United Kingdom only if we assume that u_{ic_jt} is common across countries that share continent and similar income per capita. Imposing this assumption, the difference in profits in equation (31) becomes

$$\begin{aligned}\pi_{ijj'8} + \delta\pi_{ijj'9} &= \eta^{-1}(r^o_{ij8} - r^o_{ij'8} + \varepsilon^R_{ij8} - \varepsilon^R_{ij'8}) - (\varepsilon^F_{ij8} - \varepsilon^F_{ij'8}) + \gamma^E_l - (\varepsilon^S_{ij8} - \varepsilon^S_{ij'8}) \\ &= \gamma^S_{all}(\tilde{\eta}^{-1}(r^o_{ij8} - r^o_{ij'8} + \varepsilon^R_{ij8} - \varepsilon^R_{ij'8}) + \kappa_l) - \varepsilon^F_{ij8} + \varepsilon^F_{ij'8} - \varepsilon^S_{ij8} + \varepsilon^S_{ij'8},\end{aligned} \tag{32}$$

where the second line rewrites the difference in profits as a function of the relative extended gravity parameter of interest κ_l and $\tilde{\eta} = \eta\gamma^S_{all}$. If we additionally assume that $\varepsilon^R_{ijt} = \varepsilon^F_{ijt} = \varepsilon^S_{ijt} = 0$ for every i, j, and t, and $\gamma^S_{all} > 0$, then equation (32) defines a lower bound on κ_l as a function of observed determinants of export revenue and the parameter vector $(\tilde{\eta}, \alpha)$:

$$\gamma^S_{all}(\tilde{\eta}^{-1}(r^o_{ij8} - r^o_{ij'8}) + \kappa_l) \geq 0 \quad \longrightarrow \quad \kappa_l \geq \tilde{\eta}^{-1}(r^o_{ij'8} - r^o_{ij8}). \tag{33}$$

Our model however allows ε^R_{ijt}, ε^F_{ijt} and ε^S_{ijt} to differ from zero: equations (5), (8) and (12) impose only that these terms are mean zero conditional on the firm's information set. Therefore, deriving bounds on κ_l that depend only on observed covariates requires averaging profit differences such as those in equation (32) across sets of firms, observed and alternative destinations, and years, selected on the basis of variables that belong to the firms' information sets. For each inequality in equation (28), the instrument function $\Psi_k(Z_{ijt}, Z_{ij't})$ selects the observations that the corresponding moment averages over. How can we define a function $\Psi_k(Z_{ijt}, Z_{ij't})$ such that the corresponding inequality identifies a lower bound on κ_l?

A moment inequality will help identify a lower bound on κ_l if it averages across paths such that, as in Table 2, the firm benefits from extended gravity in language more in the observed than in the alternative path. An instrument function that satisfies the requirements in equations (24) and (26) and that selects observations likely to verify this condition is:

$$\Psi_k(Z_{ijt}, Z_{ij't}) = \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, e^o_{ijt} - e^o_{ij't} = \gamma_l\}. \tag{34}$$

This function takes value one if three conditions are satisfied, and zero otherwise. The first condition requires that neither j nor j' share continent, language or similar income per capita with Chile. The second one requires that firm i is exporting to neither j nor j' in year $t-1$; and the third one requires that countries j and j' are identical in every extended gravity covariate other than language, which benefits only the observed destination j.

The function in equation (34) guarantees that, for all observations entering the moment inequality defined by it, the difference in static profits in period t, $\pi_{ijj't}$, is analogous to that in equation (29). However, it does not guarantee that the difference in period $t+1$ static profits, $\pi_{ijj't+1}$, will equal zero as in equation (30). Imagine that, instead of the path described in Table 2, we observe the one in Table 3, in which the firm exports to the United Kingdom also in year 9. In this case, the difference in export profits in year 9 in the United Kingdom is

$$d_{ij9}(\pi_{ij9} - \pi^{j\to j'}_{ij9}) = \gamma^S_{all} - \gamma^E_c - \gamma^E_g + \varepsilon^S_{ij9}, \tag{35}$$

or, in words, the sunk costs of entering the United Kingdom for a firm that only exports to Germany in year 8. Therefore, the equivalent of equation (32) for the example in Table 3 is

$$\begin{aligned} &\pi_{ijj'8} + \delta\pi_{ijj'9} = \\ &\gamma^S_{all}(\tilde{\eta}^{-1}(r^o_{ij8} - r^o_{ij'8} + \varepsilon^R_{ij8} - \varepsilon^R_{ij'8}) + \kappa_l + \delta(1 - \kappa_c - \kappa_g)) - \varepsilon^F_{ij8} + \varepsilon^F_{ij'8} - \varepsilon^S_{ij8} + \varepsilon^S_{ij'8} + \delta\varepsilon^S_{ij9}. \end{aligned} \tag{36}$$

The examples in tables 2 and 3 assume that the firm does not export to any country other than the United Kingdom in year 9. More generally, when swapping an observed destination by an alternative one in a year t, one needs to keep track of how this change affects, through extended gravity effects, the sunk costs in any other country to which the firm starts exporting in year $t+1$. We illustrate this case through an example in Appendix C.5.

Table 3: Example of a 1-period Export Event

		$t=7$	$t=8$	$t=9$
Observed	United Kingdom	0	1	1
	Germany	0	0	0
Alternative	United Kingdom	0	0	1
	Germany	0	1	0

As equations (32) and (36) illustrate, the instrument function $\Psi_k(\cdot)$ in equation (34) does not fully determine the shape of the difference in profits $\pi_{ijj't} + \delta\pi_{ijj't+1}$ for the observations that enter the corresponding moment $\mathbb{m}_k(\cdot)$. This shape depends on the observed export destinations in years t and $t+1$. However, all profit differences corresponding to observations that make the instrument function in equation (34) equal to one will share the feature that their derivative with respect to the parameter κ_l is likely positive. Consequently, the resulting moment inequality is increasing κ_l and, thus, identifies a lower bound on κ_l.[32]

The examples above also show that expectational errors always enter additively in the function $\pi_{ijj't} + \delta\pi_{ijj't+1}$. As $\Psi_k(\cdot)$ is a function of variables included in the firms' information sets, the average of these expectational errors across the observations that make $\Psi_k(\cdot)$ equal to one will equal zero asymptotically. Consequently, the resulting moment inequalities will depend exclusively on observed covariates and a finite number of parameters to estimate.

5 Estimation

Our parameter vector of interest is κ, defined in equation (21). We treat all other parameters as nuisance parameters. For estimation, we use the inequalities $\{\mathbb{m}_k(\theta) \geq 0, k = 1, \dots, 10\}$ described in equation (28) and Appendix C.4. As the examples in equations (32) and (36) illustrate, and Appendix C.3 proves, we can rewrite these moments as a function of κ, the vector of revenue parameters α, the rescaled elasticity of demand $\tilde{\eta}$, and the sunk cost γ^S_{all}. Furthermore, these moments are homogeneous of degree one in γ^S_{all} and, thus, we can set this parameter to any arbitrary positive constant without affecting the bounds on $(\alpha, \kappa, \tilde{\eta})$.

We estimate bounds on $(\alpha, \kappa, \tilde{\eta})$ in two steps. First, we use data on export revenues and moment equalities to obtain point estimates of α. Second, we use moment inequalities and these estimates of α to obtain confidence sets for $(\kappa, \tilde{\eta})$. This two-step estimator is preferred over an alternative approach that uses only the inequalities $\{\mathbb{m}_k(\theta) \geq 0, k = 1, \dots, 10\}$ to estimate bounds on $(\alpha, \kappa, \tilde{\eta})$. First, the two-step approach uses different sources of variation to identify α and κ: we use data on export revenues conditional on export participation to identify α, and data on foreign market entry and exit to identify κ. If we had estimated both α

[32] If the partial derivative of $\mathbb{m}_k(\cdot)$ with respect to κ_l, is positive, then, holding all other parameters constant, $\mathbb{m}_k(\cdot)$ increases with κ_l and, thus, the inequality $\mathbb{m}_k(\theta) \geq 0$ will be violated at very low values of κ_l.

and κ using only our inequalities, then separate identification of α and κ would be exclusively due to the functional form restrictions we impose on revenue, fixed and sunk cost in equations (4), (7), (10) and (11). Second, if we were to identify α using only inequalities, then it would also be set identified (instead of point identified). Finally, due to computational constraints, identifying α using inequalities would force us to limit its dimensionality.[33]

5.1 First Step Estimation

Following equations (4) and (5), we estimate α using data on revenues, r_{ijt}, and its determinants, X^r_{ijt}, for the subsample of firms, countries, and years with positive exports, $d_{ijt} = 1$. We use nonlinear least squares to compute the estimates $\hat{\alpha}$; we provide additional details in Appendix B.2. As X^r_{ijt} is observed independently of the value of d_{ijt}, we use $\hat{\alpha}$ to define a proxy for the potential revenues of every firm, country and year: $\hat{r}_{ijt} \equiv \exp(\hat{\alpha}_{jt} + \hat{\alpha}_i + X^r_{ijt}\hat{\alpha}^r)$.

5.2 Second Step Estimation

Given $\hat{r}_{ijt}$ for every firm, country, and year, we use the sample analogue of the ten moment inequalities described in equation (28) and Appendix C.4 to compute a 95% confidence set for the vector $(\kappa, \tilde{\eta})$. When computing this set, we apply the procedure described in Section 10.2 of Andrews and Soares (2010) to account for the fact that our moments depend on the preliminary estimate $\hat{\alpha}$. Specifically, we apply the Generalized Moment Selection procedure and adjust the variance matrix of the moments to account for their dependency on $\hat{\alpha}$.[34] In sections 6 and 7, we report confidence sets for different linear functions of the elements of κ; we compute these as projections of the 95% confidence set for $(\kappa, \tilde{\eta})$.[35]

6 Results

We present estimates of bounds on linear combinations of the four extended gravity parameters: the relative reduction in sunk costs due to extended gravity in border, κ_b, continent,

[33] See the discussion in footnote 26. A third possible approach is to combine in one step the equalities we use to identify α and the inequalities that identify $(\kappa, \tilde{\eta})$. This approach has the same computational limitations as using only inequalities to estimate $(\alpha, \kappa, \tilde{\eta})$: we still need to limit the dimensionality of α so that a grid covering all feasible values of $(\alpha, \kappa, \tilde{\eta})$ neither is too coarse nor incorporates an infeasibly large number of points.

[34] Section 8 in Andrews and Shi (2013) contains a similar adjustment for the case of a conditional moment inequality estimator. We thank the authors of Andrews et al. (2016) for sharing with us an unpublished version of their work that includes details on the implementation of the estimator in Section 8 of Andrews and Shi (2013); we adapt their implementation procedure to our combination of a nonlinear least squares first step estimator and an unconditional moment inequality second step estimator. All details are in Appendix C.6.

[35] Our projected confidence sets are conservative. Instead of applying the procedure introduced in Bugni et al. (2016a) (which has better power properties), we report these projected sets for two reasons. First, they are sufficiently tight to provide economically meaningful information on the effect of extended gravity on sunk costs. Second, as we report confidence sets for many functions of κ, it is computationally more convenient to compute a single confidence set for $(\kappa, \tilde{\eta})$ and then project it multiple times.

Table 4: Bounds on Individual Extended Gravity Parameters

Border	Continent	Language	GDPpc
[5.71%, 13.33%]	[19.05%, 28.57%]	[28.57%, 36.19%]	[0%, 28.57%]

Notes: This table reports bounds on the vector κ defined in equation (21). It uses the regression results described in column I of Table B.1. The confidence intervals are projections of a confidence set for $(\kappa, \tilde{\eta})$ computed following the procedure in Section 10.2 of Andrews and Soares (2010).

κ_c, language, κ_l, and income per capita, κ_g. To compute these estimates, we assume here that the unobserved heterogeneity in expected fixed export costs is common across countries: $u_{ic_jt} = u_{it}$, for all j. We impose no other restriction on the distribution of u_{it}.[36]

As described in Section 5, our extended gravity estimates depend on prior estimates of the revenue parameters α, which we use to compute a proxy of the potential export revenue of every firm, country and year, $\hat{r}_{ijt}$. Our estimates of α, reported in Table B.1 in Appendix B.2, reveal that: new exporters sell small amounts; firms' exports increase in the size of the destination market and generally decrease in any measure of distance between home and foreign markets; and more productive firms, as proxied either by value added per worker or by domestic sales, export larger amounts. For the purpose of computing confidence sets on extended gravity parameters, the main characteristic of the nine revenue regressions reported in Table B.1 is that, as we show in Table B.2 in Appendix B.2, they all generate very similar predicted export revenues $\hat{r}_{ijt}$. Our moment inequality estimates are thus robust to several different specifications of the export revenue regression. To confirm this, we report here and in Appendix C.7 two sets of extended gravity estimates that differ in the number of covariates used to compute the corresponding predicted export revenues.[37] No matter which specification we use to construct $\hat{r}_{ijt}$, the moment inequality estimates are very similar.

As shown in Table 4, we estimate the extended gravity effect due to border to be between approximately 6% and 13%, the effect due to continent to be between 19% and 29%, the effect due to language to be between 29% and 36%, and the effect due to similarity in income per capita to be lower than 29%. Panel A in Figure 3 represents these estimates graphically: except for the case of similarity in income per capita, our estimated bounds are tight and reject the null that extended gravity effects are zero. One should not conclude from these

[36] In Section 7, we relax this assumption and allow u_{ic_jt} to vary across countries that differ in the continent of location, in official language, or in income per capita.

[37] We report here results that rely on the "long" revenue regression described in column I of Table B.1. It includes: firm and year fixed effects; the firm's value added per worker, share and average wages of skilled and unskilled workers; a large set of distance measures between foreign and home countries; this same set interacted with a dummy for first year of exports to a country; extended gravity covariates; and a measure of the foreign market's size. We report in Appendix C.7 results that use the "short" revenue regression described in column VI of Table B.1. This one includes only: firm and year fixed effects; the firm's domestic sales; the physical distance between foreign and home countries; and the aggregate imports in the foreign market. Besides fixed effects, the "long" regression includes 28 regressors, while the "short" one only 3.

Table 5: Bounds on Combinations of Extended Gravity Parameters

Border + Continent	Language + GDPpc	Continent + GDPpc	Continent + GDPpc + Border	Continent + Language	Continent + Language + Border	Continent + Language + GDPpc	All
[24.76%, 38.10%]	[32.38%, 60.95%]	[24.76%, 49.52%]	[34.29%, 57.14%]	[47.62%, 62.86%]	[55.24%, 74.29%]	[57.14%, 81.90%]	[68.57%] 89.52%]

Notes: This table reports bounds on sums of elements of the vector κ defined in equation (21). It uses the regression results described in column I of Table B.1. The confidence intervals reported in this table are projections of a 5-dimensional confidence set for $(\kappa, \tilde{\eta})$ computed following the procedure in Section 10.2 of Andrews and Soares (2010).

estimates that linguistic factors matter more than geographic ones. Contiguous countries will also generally share continent and, as shown in Table 5 and Panel A in Figure 3, their joint effect is between 25% and 38%, comparable to the effect of sharing common language.

Table 5 and Panel B in Figure 3 present bounds on combinations of extended gravity parameters. Noticeably, firms that have export experience in a country that shares border, continent, language and similar income per capita with a new export destination will pay sunk export costs 69% to 90% smaller than those paid by a firm that has no prior export experience (or experience only in countries unrelated to that new destination). This result puts into question the standard assumption that firms consider different foreign countries as independent export markets. Our results suggest that this equivalence between markets and countries is, in some cases, inaccurate. When two countries are very similar to each other, firms that jump between them face barriers only slightly larger than those they would face when jumping across regions of the same country.

Panel C in Figure 3 illustrates our estimates through the example of a firm that enters the United States. If this firm was previously exporting to Canada, a country that shares border, continent, language and similar income per capita with the United States, it will pay only between one tenth and one third of the entry costs paid by a firm that was not exporting in the prior year. Similarly, Panel D of Figure 3 shows that the cost of starting to export to Germany for a firm whose only export destinations in the previous year are, for example, China or Argentina will be between four and ten times larger than the corresponding cost for a firm that exported to Austria instead. These results reflect that, once a firm has adapted its products or workforce to successfully export to Austria, the additional cost required to enter Germany is relatively small. Furthermore, firms are forward-looking and, thus, when considering whether to enter a country like Austria, they take into account that the investment required to break into this market will also eventually allow them to enter other similar countries such as Germany. Therefore, the decision of whether to make the investment needed to enter a foreign country will depend on how similar it is to other markets.

The difference in the entry costs in Germany of previous exporters to Austria and of previous exporters to France reflects the role of language in generating extended gravity effects.

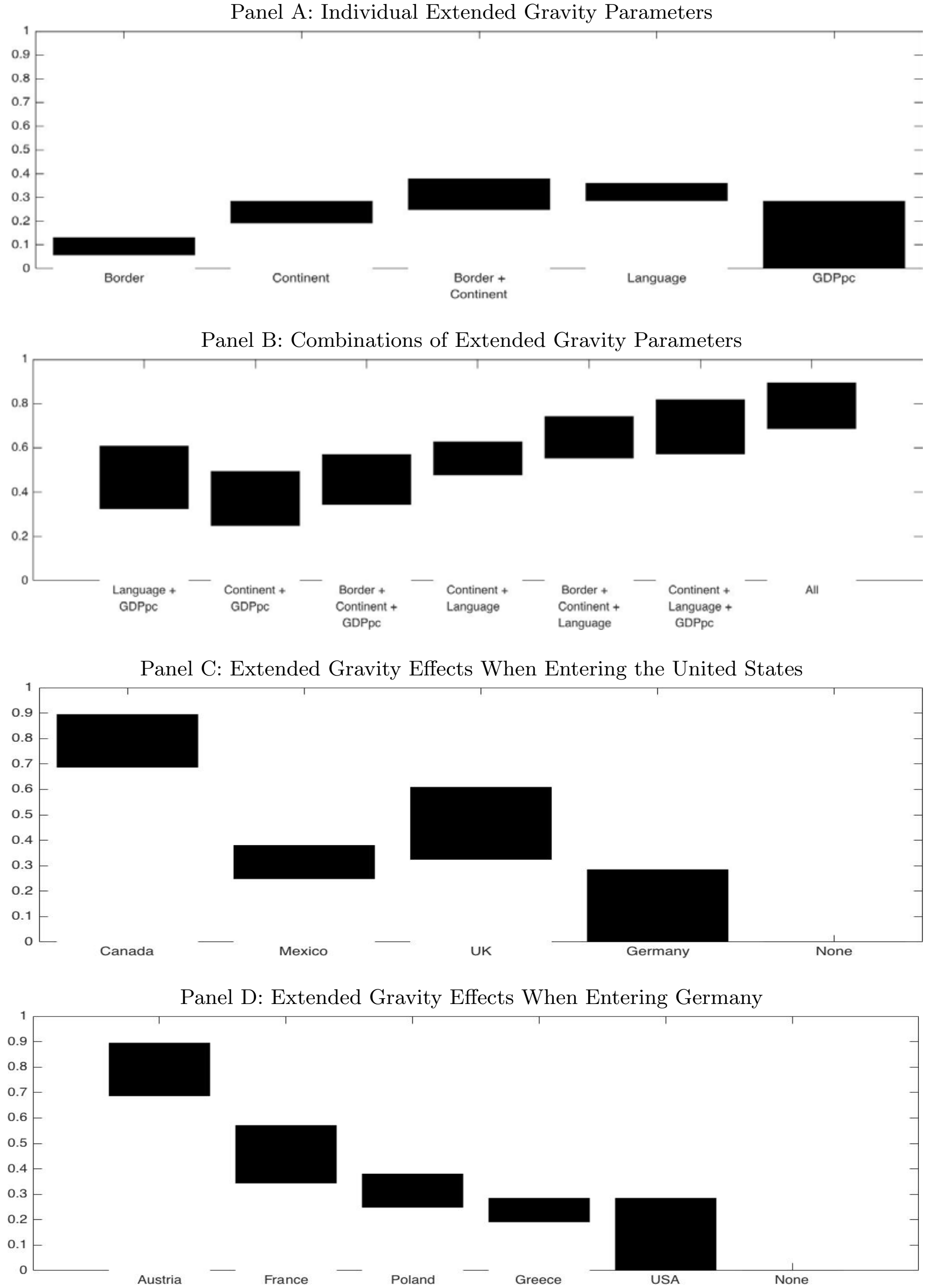
Once the linguistic connection is not present, the total reduction in entry costs due to extended gravity effects is at most 57%. If we further omit the income per capita connection and consider the sunk costs in Germany of previous exporters to Poland, then the advantage is between 25% and 38%. Finally, sharing only continent (as is the case of Greece and Germany) implies, as discussed above, an approximate reduction in entry costs of 19% to 29%.

Is it easier to enter the United States from Mexico or from the United Kingdom? Mexico shares all geographic factors (border and continent) with the United States, while the United Kingdom shares all non-geographic factors (language and similar income per capita). While we cannot reject that the extended gravity effects enjoyed by a firm previously exporting to Mexico are the same as those experienced by a firm previously exporting to the United Kingdom, most parameter values in our confidence set indicate that the combined effect of language and income per capita is larger than the combined effect of border and continent.

Two features of our estimates are worth noting. First, although we identify κ from firms' discrete export decisions, our estimates do not depend on a normalization by scale. The reason is that κ captures not the absolute but the relative reduction in sunk costs. Second, our estimates show that, even though the confidence interval for a parameter may be large, confidence intervals for linear combinations of this and other parameters can be smaller. For example, the confidence interval for κ_g is nearly 30 percentage points wide; however, the confidence interval on the sum of all four extended gravity parameters, $\kappa_b + \kappa_c + \kappa_l + \kappa_g$, is 21 percentage points wide. Combinations of parameters may thus be better identified than each of them by itself; estimates will have this property whenever the covariates that multiply these parameters are positively correlated. Panel (e) in Figure 4 illustrates this: the projection of the confidence set for $(\kappa, \tilde{\eta})$ on the space (κ_c, κ_g) slopes negatively and, thus, includes both parameter values that combine high values of κ_c with low values of κ_g and values that combine low values of κ_c with high values of κ_g. The bounds on $\kappa_c + \kappa_g$ are thus narrower than those on κ_g. The reason why the projection of the confidence set for $(\kappa, \tilde{\eta})$ on the space (κ_c, κ_g) slopes negatively is that countries that belong to the same continent tend to have similar income per capita: if a firm enters a country that shares a continent with a prior export destination, this country will also likely share similar income per capita with that destination. In these cases, our estimator cannot determine whether the firm's entry decision is due to extended gravity in continent or in income per capita, but it can determine that the sum of both effects must be large enough to explain the observed export decision.[38]

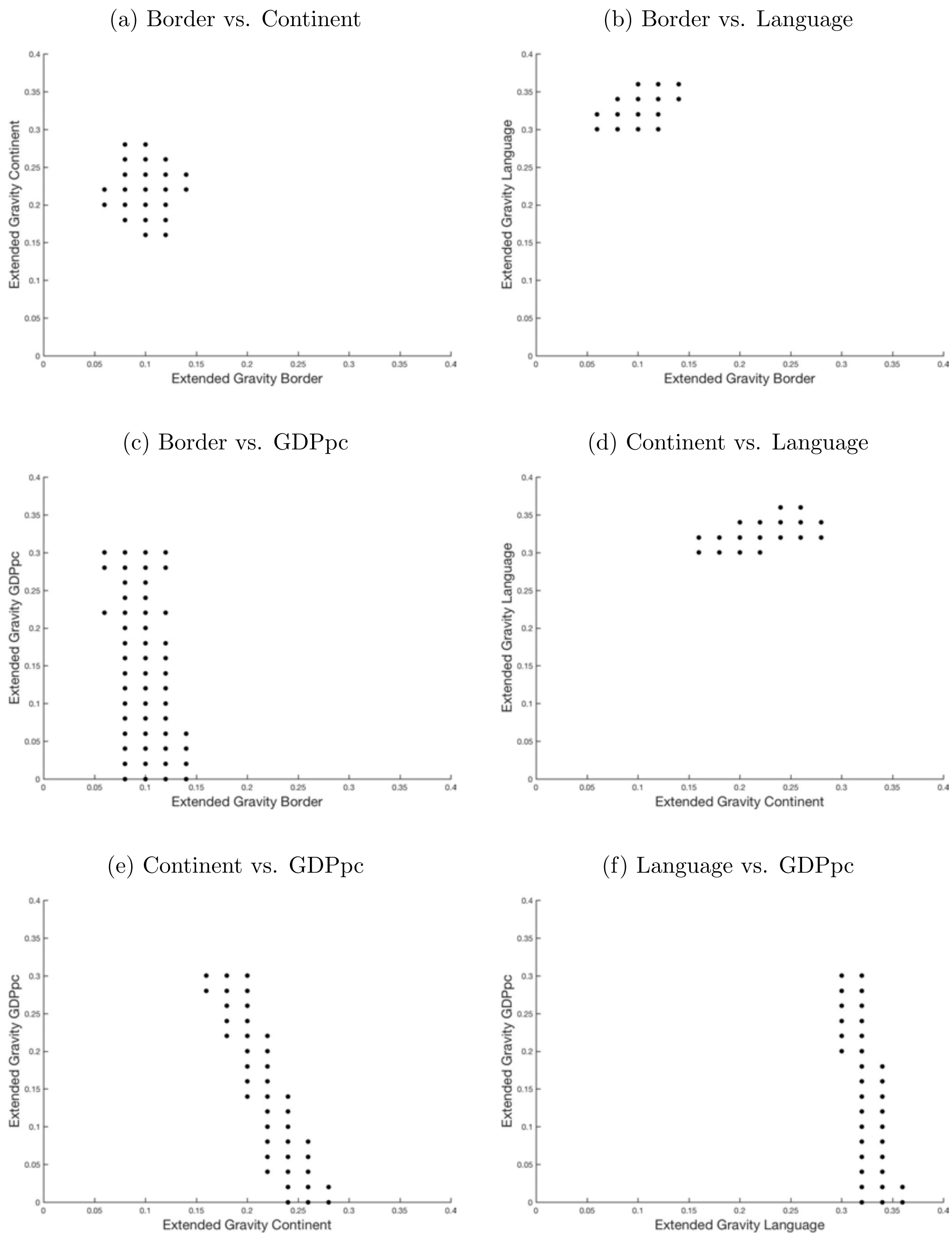
[38] This is one of the reasons why we do not aim to identify the effect of each gravity variable on sunk costs (the other one being an increase in the dimensionality of the parameter vector to estimate). In the case of Chile, gravity variables are very correlated: most Spanish-speaking countries are located in South America and have similar levels of income per capita. For example, if we observe firms entering Argentina instead of the larger US market, our model will indicate that export costs in Argentina are lower; however, we cannot discern whether this is due to Argentina sharing continent, language or similar income per capita with Chile. For this reason, we estimate extended gravity effects relative to the sunk costs of exporting to a country that differs from Chile in *all* gravity variables: we identify bounds on $\kappa_b = \gamma_b^E / \gamma_{all}^S$, but not, for example, on γ_b^E / γ_c^S.

Figure 3: Bounds on Extended Gravity Parameters



Notes: These figures represent the projected confidence intervals reported in Panel A of tables 4 and 5.

Figure 4: Projected Confidence Set



Notes: These confidence sets are two-dimensional projections of a 5-dimensional confidence set for $(\kappa, \tilde{\eta})$ computed following the procedure in Section 10.2 of Andrews and Soares (2010).

7 Robustness

In this section, we present estimates consistent with the unobserved heterogeneity in export profits, u_{ic_jt}, varying across groups of countries. We treat u_{ic_jt} as a firm i, year t, and group-of-countries c_j fixed effect, and define the groups c_j by the countries' continent, language, or income per capita. We thus impose no restrictions on the distribution of u_{ic_jt} across country groups, firms, and time periods. This implies that only those inequalities in which we swap an observed export destination for an alternative one belonging to the same country group will be useful to identify κ: any difference in the firm's export decisions in two countries j and j' that do not belong to the same country group, $c_j \neq c_{j'}$, can be explained by specific realizations of u_{ic_jt} and $u_{ic_{j'}t}$. Our approach to build inequalities is thus based on differencing out the unobserved heterogeneity that affects firms' export decisions, and we do so by choosing actual and counterfactual destinations j and j' such that $c_j = c_{j'}$ and, therefore, $u_{ic_jt} - u_{ic_{j'}t} = 0$.

Inequalities that compare export profits in countries that share continent, language, or similar income per capita create challenges for the identification of extended gravity parameters. For example, consider the case in which we treat u_{ic_jt} as a firm-year-continent fixed effect and, thus, build inequalities that compare destinations that belong to the same continent. This complicates the identification of the continent extended gravity parameter, κ_c. A firm that benefits from extended gravity in continent when entering country j in year t would have equally benefited from it if it had entered an alternative destination j' located in the same continent: if the firm is exporting in $t-1$ to a country sharing continent with j, it is also then exporting to a country sharing continent with j'. The difference in static profits in t, $\pi_{ijj't}$, will thus not depend κ_c. While the difference in static profits in year $t+1$, $\pi_{ijj't+1}$, may still depend on κ_c, inequalities that difference out a firm-year-continent fixed effect will have low identification power for κ_c. However, as long as there is variation in language and income per capita across countries sharing a continent, these inequalities may still be useful to identify extended gravity effects due to language, κ_l, and income per capita, κ_g. Similarly, when we allow for firm-year-language or firm-year-income-per-capita-group fixed effects, our inequalities lose identification power for the corresponding extended gravity parameters.

The results we obtain when we allow u_{ic_jt} to vary across groups of countries, reported in Table 6, are consistent with those we obtain when we assume that this term is common to all countries, reported in Table 4. When we allow u_{ic_jt} to differ across continents, the border extended gravity effect is estimated to be between 10% and 23%, the one due to language to be between 12% and 27%, and the one due to similarity in income per capita to be between 9% and 21%. As expected based on the discussion above, the continent extended gravity effect is not identified in this case. Allowing u_{ic_jt} to vary across countries that differ in their official language yields a border extended gravity effect between 3% and 9%, a continent extended gravity effect between 12% and 24%, and an extended gravity effect due to similarity income

Table 6: Bounds on Individual Extended Gravity Parameters

Panel A: Firm-Year-Continent Fixed Effects			
Border	Continent	Language	GDPpc
[10.37%, 22.70%]	[-, -]	[11.85%, 26.67%]	[8.89%, 20.74%]
Panel B: Firm-Year-Language Fixed Effects			
Border	Continent	Language	GDPpc
[2.96%, 8.89%]	[11.85%, 23.70%]	[-, -]	[14.81%, 34.07%]
Panel C: Firm-Year-GDP Per Capita Fixed Effects			
Border	Continent	Language	GDPpc
[4.10%, 7.18%]	[18.46%, 28.72%]	[18.46%, 28.72%]	[-, -]

Notes: This table reports bounds on the vector κ defined in equation (21). It relies on the regression estimates in column I of Table B.1. The bounds in panels A, B, and C are computed under the assumption that u_{ic_jt} is, respectively, a firm-year-continent, a firm-year-language, and a firm-year-income-per-capita-group fixed effect. The confidence intervals are projections of a confidence set for $(\kappa, \tilde{\eta})$ computed following the procedure in Section 10.2 of Andrews and Soares (2010).

per capita between 15% and 34%. Finally, when allowing for firm-year-income-per-capita-group fixed effects, we obtain a border extended gravity effect between 4% and 7%, and extended gravity effects due to continent and language between 19% and 29%.

Our inequalities do not allow for firm-, year-, and country-specific unobserved heterogeneity in export profits known to the firm when deciding on its set of export destinations. However, as the simulation in Appendix D.3 shows, if substantial unobserved heterogeneity impacts the firm's export decisions, the identified set defined by inequalities that ignore such heterogeneity is likely to be empty. We use the model specification tests in Andrews and Soares (2010) and Bugni et al. (2015) to test the null hypothesis that the identified sets defined by the inequalities employed in tables 4 and 6 are nonempty: the p-values for both the "test BP" and the "test RS" are always larger than 10% (see Appendix C.6.2 for details). One should not interpret our results as suggesting that export profits do not differ across firms for unobserved reasons. Such unobserved determinants are accounted for in our model through the terms ε^R_{ijt}, ε^F_{ijt} and ε^S_{ijt}, defined in equations (4), (6), and (9). However, consistent with equations (5), (8), and (12), our estimates do not contradict the assumption that these terms are also unobserved to the firm when deciding which new destinations to enter.

8 Concluding Remarks

We use moment inequalities to estimate a dynamic model of firm entry into spatially related export markets. The traditional approach assumes that the firm's export decision in a market is independent of the decision taken in any other market. Conversely, our model allows for dynamic complementarities across markets in the firm's export decisions. These decisions are thus potentially very complex, and this complexity makes moment inequalities ideal.

Our results show that extended gravity is an important determinant of firms' entry costs. A firm that exports to countries that share border, continent, language and similar income per capita with a particular market will face sunk costs in it that are between 69% and 90% smaller than those faced by a firm whose export destinations do not share any of these four characteristics with this market. Among the four extended gravity factors we consider, border and language are to be the most important. Exporting to a country located in the continent of a subsequent new destination reduces sunk costs by 19% to 29% if both countries do not share border, and by 25% to 38% if they do. This effect is similar to the impact of sharing language, which reduces entry costs by 29% to 36%. The impact of sharing similar income per capita is imprecisely estimated to be below 29%.

Although our estimator has the advantage of being consistent with a flexible specification of the firm's information and consideration sets, and planning horizon, it also has limitations. In particular, our benchmark estimates assume away the existence of firm-year-country specific factors that influence firms' entry decisions but are not in our data. If these unobserved factors are correlated across countries connected through some extended gravity variable, then our extended gravity estimates will not be capturing only state dependence in trade costs but also the effect of unobserved heterogeneity in potential export profits. To address this concern, we also present estimates that allow for firm-year-group-of-countries specific fixed effects. No matter whether we define country groups by their continent, language or income per capita, the resulting estimates are generally consistent with the benchmark ones.

Given that sunk costs are important determinants of firms' export decisions (Das et al., 2007) and that the extensive margin of firms drives much of the variation in aggregate trade across destinations (Bernard et al., 2010), our findings suggest that shocks to the profitability of exporting to a market will have important effects on neighboring countries and in countries that share a language with them. For example, our analysis suggests that an increase in trade barriers between the United States and China will impact Chinese exports to Mexico and Canada. Quantifying this impact, however, requires correctly specifying firms' information and consideration sets and planning horizons, as well as solving the resulting combinatorial dynamic discrete choice problem. This requires dealing with computational challenges still unsolved in the literature, and, thus, we leave this quantification for future work.

References

Aguirregabiria, Victor and Arvind Magesan, "Euler Equations for the Estimation of

Dynamic Discrete Choice Structural Models," *Advances in Econometrics*, 2013, *31*, 3–44. [19]

— **and** — , "Solution and Estimation of Dynamic Discrete Choice Structural Models Using Euler Equations," *mimeo*, October 2016. [19]

Albornoz, Facundo, Héctor F. Calvo Pardo, Gregory Corcos, and Emanuel Ornelas, "Sequential Exporting," *Journal of International Economics*, 2012, *88* (1), 1–24. [1, 3, 11, 15]

Allenby, Greg M., Jaehwan Kim, and Peter E. Rossi, "Modeling Consumer Demand for Variety," *Marketing Science*, 2002, *21* (Summer), 229–250. [3]

— **,** — **, and** — , "Product Attributes and Models of Multiple Discreteness," *Journal of Econometrics*, 2007, *138*, 208–230. [3]

Anderson, James E. and Eric van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 2003, *93* (1), 170–192. [4]

Andrews, Donald W.K. and Gustavo Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78*, 119–157. [26, 27, 28, 31, 33, 20, 24, 25]

— **and Patrik Guggenberger**, "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 2009, *25*, 669–709. [24]

— **and Xiaoxia Shi**, "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, *81* (2), 609–666. [20, 26]

— **, Wooyoung Kim, and Xiaoxia Shi**, "Stata Commands for Testing Conditional Moment Inequalities/Equalities," *Stata Journal*, 2016, *forthcoming.* [26]

Antràs, Pol, Teresa Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S.," *mimeo*, February 2017. [4, 16]

Arcidiacono, Peter and Robert A. Miller, "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, 2011, *79*, 1823–1867. [19]

— **, Patrick Bayer, Jason R. Blevins, and Paul B. Ellickson**, "Estimation of Dynamic Discrete Choice Models in Continuous Time with an Application to Retail Competition," *Review of Economic Studies*, 2016, *83* (3), 889–931. [4]

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151–1199. [3]

— **and Fabian Eckert**, "Combinatorial Discrete Choice," *mimeo*, January 2017. [4, 16]

— **, Sharat Ganapati, and Marc-Andreas Muendler**, "The Extensive Margin of Exporting Goods: A Firm-level Analysis," *mimeo*, November 2015. [3]

— , **Theodore Papageorgiou, and Olga A. Timoshenko**, "Firm Learning and Growth," *mimeo*, September 2015. [15]

Armstrong, Tim B., "Weighted KS Statistic for Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 2014, *182* (2), 92–116. [20]

— , "Asymptotically Exact Inference in Conditional Moment Inequality Models," *Journal of Econometrics*, 2015, *186* (1), 51–65. [20]

— **and Hock Peng Chan**, "Multiscale Adaptive Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 2016, *194* (1), 24–43. [20]

Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu, "R&D Investment, Exporting, and Productivity Dynamics," *American Economic Review*, 2011, *101* (4), 1312–1344. [3]

Bai, Xue, Kala Krishna, and Hong Ma, "How you Export Matters: Export Mode, Learning and Productivity in China," *Journal of International Economics*, 2017, *104* (1), 122–137. [3]

Bandiera, Oriana, Luigi Guiso, Andrea Prat, and Raffaella Sadun, "Matching Firms, Managers, and Incentives," *Journal of Labor Economics*, 2015, *33* (3), 623–681. [14]

Bastos, Paulo, Daniel A. Dias, and Olga A. Timoshenko, "Learning, Prices, and Firm Dynamics," *mimeo*, 2016. [15]

Berman, Nicolas, Vincent Rebeyrol, and Vincent Vicard, "Demand Learning and Firm Dynamics: Evidence from Exporters," *mimeo*, November 2015. [11, 15]

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott, "Firms in International Trade," *Journal of Economic Perspectives*, 2007, *21* (3), 105–130. [1]

— , — , — , **and** — , "The Margins of US Trade," *American Economic Review*, 2010, *99*, 487–493. [1, 34]

— , **Renzo Massari, Jose-Daniel Reyes, and Daria Taglioni**, "Exporter-Dynamics and Partial-Year Effects," *mimeo*, November 2015. [11]

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi, "Alternative Specification Tests for Partially Identified Models Defined by Moment Inequalities," *Journal of Econometrics*, 2015, *185* (1), 259–282. [33, 23, 24]

— , — , **and** — , "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 2016, *forthcoming.* [26]

— , **Mehmet Caner, Anders Bredahl Kock, and Soumedra Lahiri**, "Inference in Partially Identified Models with Many Moment Inequalities Using Lasso," *mimeo*, 2016. [20]

Butler, J. S. and Robert Moffitt, "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica*, 1982, *50* (3), 761–764. [7]

Cameron, A. Colin and Pravin K. Trivedi, *Microeconometrics: Methods and Applications*, New York: Cambridge University Press, 2005. [23]

Card, David and Daniel Sullivan, "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 1998, *56* (3), 497–530. [6]

Chaney, Thomas, "The Network Structure of International Trade," *American Economic Review*, 2014, *104* (11), 3600–3634. [1, 2, 3]

_ , "The Gravity Equation in International Trade: An Explanation," *Journal of Political Economy*, 2016, *forthcoming.* [2]

_ , "Networks in International Trade," in Yann Bramoulle, Andrea Galleoti, and Brian Rogers, eds., *Oxford Handbook of the Economics of Networks*, Oxford: Oxford University Press, 2016. [2]

Cheng, Chen and Claudia Steinwender, "Import Competition, Heterogeneous Preferences of Managers, and Productivity," *mimeo*, December 2016. [14]

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato, "Testing Many Moment Inequalities," *mimeo*, November 2014. [20]

_ , Sokbae Lee, and Adam M. Rosen, "Intersection Bounds: Estimation and Inference," *Econometrica*, 2013, *81* (2), 667–737. [20]

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout, "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 2007, *75* (3), 837–873. [2, 3, 34]

Defever, Fabrice, Bendikt Heid, and Mario Larch, "Spatial Exporters," *Journal of International Economics*, 2015, *95* (1), 145–156. [1, 3]

Dickstein, Michael J. and Eduardo Morales, "What Do Exporters Know?," *mimeo*, December 2016. [3, 4, 15, 20, 30]

Dubé, Jean-Pierre, "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science*, 2004, *23* (1), 66–81. [3]

Eaton, Jonathan, David Jinkins, James R. Tybout, and Daniel Yi Xu, "Two Sided Search in International Markets," *mimeo*, June 2016. [3]

_ , Marcela Eslava, David Jinkins, C. J. Krizan, and James R. Tybout, "A Search and Learning Model of Export Dynamics," *mimeo*, February 2014. [3]

_ , _ , Maurice Kugler, and James R. Tybout, "The Margins of Entry into Export Markets: Evidence from Colombia," in Elhanan Helpman, Dalia Marin, and Thierry Verdier, eds., *The Organization of Firms in a Global Economy*, Cambridge: Harvard University Press, 2008. [3]

_ , Samuel Kortum, and Francis Kramraz, "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 2011, *79* (5), 1453–1498. [11]

Eizenberg, Alon, "Upstream Innovation and Product Variety in the U.S. Home PC Market," *Review of Economic Studies*, 2014, *81*, 1003–1045. [4]

Ellickson, Paul, Stephanie Houghton, and Christopher Timmins, "Estimating Network Economies in Retail Chains: A Revealed Preference Approach," *Rand Journal of Economics*, 2013, *44* (2), 169–193. [3]

Evenett, Simon J. and Anthony J. Venables, "Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows," *mimeo*, June 2002. [1]

Fieler, Ana Cecilia, Marcela Eslava, and Daniel Yi Xu, "Trade, Skills and Quality-Upgrading: A Theory with Evidence from Colombia," *mimeo*, November 2016. [11]

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi, "How Exporters Grow," *mimeo*, January 2016. [3, 11, 15]

Gumpert, Anna, Andreas Moxnes, Natalia Ramondo, and Felix Tintelnot, "Exporters' and Multinational Firms' Life-cycle Dynamics," *mimeo*, October 2016. [11]

Hansen, Lars Peter and Kenneth J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 1982, *50* (5), 1269–1286. [4, 19]

Head, Keith and Thierry Mayer, "Gravity Equations: Workhorse,Toolkit, and Cookbook," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, eds., *Vol.4 of Handbook of International Economics*, Elsevier, 2014, pp. 131–195. [1]

Heckman, James J., "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process and Some Monte Carlos Evidence," in Charles F Manski and Daniel McFadden, eds., *Structural Analysis of Discrete Choice Data and Econometric Applications*, Cambridge: MIT Press, 1981, pp. 179–195. [6]

Hendel, Igal, "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns," *Review of Economic Studies*, 1999, *199* (April), 423–446. [3, 4]

Hillberry, Russell and David Hummels, "Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data," *European Economic Review*, 2008, *52* (3), 527–550. [1]

Ho, Katherine, "Insurer-Provider Networks in the Medical Care Market," *American Economic Review*, 2009, *99* (1), 393–430. [4]

_ **and Adam Rosen**, "Partial Identification in Applied Research: Benefits and Challenges," *mimeo*, August 2016. [17]

_ **and Ariel Pakes**, "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 2014, *104* (12), 3841–3884. [4]

Holmes, Thomas J., "The Diffusion of Wal-Mart and Economies of Density," *Econometrica*, 2011, *79* (1), 253–302. [3, 4, 20]

Illanes, Gaston, "Switching Costs in Pension Plan Choice," *mimeo*, 2016. [4, 20]

Irarrazabal, Alfonso A., Andreas Moxnes, and Luca David Opromolla, "The Tip of the Iceberg: A Quantitative Framework for Estimating Trade Costs," *Review of Economics and Statistics*, 2015, *97* (4), 777–792. [3]

Ishii, Joy, "Compatibility, Competition, and Investment in Network Industries: ATM Networks in the Banking Industry," *mimeo*, 2008. [4]

Jia, Panle, "What Happens When Wal-Mart Comes to Town: An Empirical Analysis of the Discount Retailing Industry," *Econometrica*, 2008, *76* (6), 1263–1316. [3, 4, 16]

Judd, Kenneth L., *Numerical Methods in Economics*, Cambridge, Massachusetts: The MIT Press, 1998. [7]

Katz, Michael, "Supermarkets and Zoning Laws," *Unpublished PhD Dissertation*, 2007, *Harvard University.* [4]

Kugler, Maurice and Eric Verhoogen, "Prices, Plant Size, and Product Quality," *Review of Economic Studies*, 2012, *79* (1), 307–339. [11]

Labanca, Claudio, Danielken Molina, and Marc-Andreas Muendler, "Preparing to Export," *mimeo*, December 2014. [2]

Lawless, Martina, "Firm Export Dynamics and the Geography of Trade," *Journal of International Economics*, 2009, *77* (2), 1149–1172. [1, 3]

_ , "Marginal Distance: Does Export Experience Reduce Firm Trade Costs?," *Open Economies Review*, 2013, *24* (5), 819–841. [1, 3]

Luttmer, Erzo C. J., "What Level of Fixed Costs can Reconcile Consumption and Stock Market Returns?," *Journal of Political Economy*, 1999, *107* (5), 969–997. [4, 19]

Mayer, Thierry and Soledad Zignago, "Notes on CEPII's Distances Measures," *mimeo*, December 2011. [5]

Meinen, Phillipp, "Sunk Costs of Exporting and the Role of Experience in International Trade," *Canadian Journal of Economics*, 2015, *48* (1), 335–367. [1]

Menzel, Konrad, "Consistent Estimation with Many Moment Inequalities," *Journal of Econometrics*, 2014, *182*, 329–350. [20]

Moxnes, Andreas, "Are Sunk Costs in Exporting Country-Specific?," *Canadian Journal of Economics*, 2010, *43* (2), 467–493. [3]

Pakes, Ariel, "Alternative Models for Moment Inequalities," *Econometrica*, 2010, *78* (6), 1783–1822. [4, 19]

_ **and Jack R. Porter**, "Moment Inequalities for Multinomial Choice with Fixed Effects," *mimeo*, December 2015. [4]

_ , _ , **Katherine Ho, and Joy Ishii**, "Moment Inequalities and their Application," *Econometrica*, 2015, *83* (1), 315–334. [4, 19]

Pennings, Joost M. E. and Philip Garcia, "The Informational Content of the Shape of Utility Functions: Financial Strategic Behavior," *Managerial and Decision Economics*, 2008, *30* (2), 93–90. [14]

Piveteau, Paul, "An Empirical Dynamic Model of Trade with Consumer Accumulation," *mimeo*, January 2016. [11]

Roberts, Mark J. and James R. Tybout, "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 1997, *87* (4), 545–564. [3, 13]

Romano, Joseph P. and Azeem M. Shaikh, "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 2008, *138* (9), 2786–2807. [24]

Ruhl, Kim J. and Jonathan L. Willis, "New Exporter Dynamics," *International Economic Review*, forthcoming 2016. [3, 11]

Scott, Paul T., "Dynamic Discrete Choice Estimation of Agricultural Land Use," *mimeo*, December 2013. [19]

Sieg, Holger and Jipeng Zhang, "The Effectiveness of Private Benefits in Fundraising of Local Charities," *International Economic Review*, 2012, *53* (2), 349–374. [4]

Timoshenko, Olga A., "Learning versus Sunk Costs Explanations of Export Persistence," *European Economic Review*, 2015, *79*, 113–128. [15]

_ , "Product Switching in a Model of Learning," *Journal of International Economics*, 2015, *95* (2), 233–249. [15]

Tinbergen, Jan, *Shaping the World Economy*, New York: The Twentieth Century Fund, 1962. [4]

Tintelnot, Felix, "Global Production with Export Platforms," *Quarterly Journal of Economics*, 2016, *forthcoming.* [3, 16]

Traiberman, Sharon, "Occupations and Import Competition," *mimeo*, December 2016. [19]

Verhoogen, Eric A., "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 2008, *123* (2), 489–530. [11]

Wollmann, Thomas G., "Trucks Without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles," *mimeo*, May 2016. [4]

Zheng, Fanyin, "Spatial Competition and Preemptive Entry in the Discount Retail Industry," *mimeo*, November 2016. [3]

Online Appendix for “Extended Gravity’

Eduardo Morales
Princeton University

Gloria Sheu
U.S. Department of Justice

Andrés Zahler
Universidad Diego Portales

June 24, 2017

A Reduced-Form Evidence

In this section, we present reduced-form evidence on the relevance of extended gravity in determining firms' export destinations. Specifically, we complement the motivating evidence presented in Table 1 in Section 2.2 by addressing some alternative explanations for the transition probabilities documented in that table.

A.1 Export Path of an Illustrative Firm

For the same firm illustrated in Figure 1 in Section 2.2, this table lists all countries to which the firm exported in each year in our sample. Country names listed in italics have English as their primary official language. Countries are listed in increasing order according to their distance to Chile. We refer the reader to Section 2.2 for a description of the export behavior of this firm that illustrates how extended gravity forces may be an important explanation of its choice of which foreign countries to start to export in each year.

Table A.1: Export Path of an Individual Firm

Countries	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Argentina	X	X	X	X	X	X	X				X
Bolivia							X	X	X	X	X
Peru	X		X	X		X	X	X	X	X	X
Ecuador			X	X	X		X	X	X	X	X
Brazil							X				X
Colombia									X		
Venezuela									X		
Panama						X	X		X	X	X
Nicaragua										X	
El Salvador									X	X	X
Honduras									x	X	X
Belize									X		
Guatemala							X	X	X	X	X
Mexico						X					X
United States		X	X	X	X	X	X	X	X	X	X
Spain											X
France									X		
The Netherlands										X	
United Kingdom							X	X			
Australia											X

Notes: For the same firm illustrated in Figure 1, this table lists all countries to which the firm exported in each year in our sample. Country names listed in italics have English as their primary official language.

A.2 Transition Probabilities by Firm Size

As discussed in Section 2.2, extended gravity effects are not the only economic mechanism that can rationalize the transition probabilities reported in Table 1. An alternative explanation is the joint effect of the following two forces: (a) most firms either do not export at all or export to a very few destinations, while a small set of firms export widely; (b) the more countries a firm exports to, the more likely it is that every new destination shares some characteristic with a previous destination of the firm.

The model reported in Section 3 and the results presented in sections 6 and 7 account for this alternative explanation both by allowing observed and unobserved firm characteristics to determine the number of countries

Table A.2: Transition Probabilities By Firm Size

	Overall		Large Firms		Small Firms	
	Prob. of Entry	Num. of Entries	Prob. of Entry	Num. of Entries	Prob. of Entry	Num. of Entries
Overall:	0.53%	1638	1.03%	1264	0.30%	374
Extended Gravity:						
If Ext. Grav. Border = 1	6.74%	397	6.78%	321	5.94%	76
If Ext. Grav. Cont. = 1	2.79%	525	2.86%	391	2.88%	134
If Ext. Grav. Lang. = 1	1.59%	205	1.46%	145	1.83%	60
If Ext. Grav. GDPpc = 1	1.53%	588	1.59%	453	1.52%	135
If All Ext. Grav. = 0	0.31%	770	0.69%	585	0.16%	185

Notes: *Large Firms* have domestic sales above median; *Small Firms* have domestic sales below median. Median domestic sales in our sample equal approximately 1.8 million dollars.

Table A.3: Transition Probabilities By Number of Exporters

	Overall		Many Destinations		Few Destinations	
	Prob. of Entry	Num. of Entries	Prob. of Entry	Num. of Entries	Prob. of Entry	Num. of Entries
Overall:	0.53%	1638	1.97%	871	0.38%	767
Extended Gravity:						
If Ext. Grav. Border = 1	6.74%	397	7.22%	361	1.21%	36
If Ext. Grav. Cont. = 1	2.79%	525	3.23%	497	0.95%	28
If Ext. Grav. Lang. = 1	1.59%	205	1.87%	196	0.34%	9
If Ext. Grav. GDPpc = 1	1.53%	588	2.31%	498	0.57%	90
If All Ext. Grav. = 0	0.31%	770	0.99%	134	0.33%	636

Notes: *Many Destinations* probabilities correspond to firms with above median number of destinations; *Few Destinations* probabilities correspond to firms with number of destinations below or equal to the median. The median number of destinations in our sample is equal to 4.

a firm exports to in every time period and by flexibly modeling the set of export destinations to which each firm considers exporting (i.e. its consideration set).

As preliminary evidence that the transition probabilities reported in Table 1 are not entirely due to this possible alternative explanation, tables A.2 and A.3 report transition probabilities analogous to those in Table 1 for subgroups of firms that differ both in their domestic sales (interpreted here as a proxy for the firm's productivity) and in their total number of export destinations in the previous year. To facilitate the comparison, the first two columns in both of these tables include the overall numbers reported in Table 1.

The first row in tables A.2 and A.3 shows that the overall probability of entry is larger for large firms and for firms already exporting to a large number of destinations. The remaining rows in these tables show that, for all the four subgroups of firms considered in them, the general finding that firms are more likely to enter a country when previously exporting to similar destinations survives. For example, the relative increase in the entry probability associated to previously exporting to a bordering country is: approximately thirteen when considering all firms; seven when considering firms with large domestic sales; close to twenty when considering firms with low domestic sales; and slightly below four no matter whether we look at firms with above or below median number of export destinations in the year prior to entry. As the remaining rows in tables A.2 and A.3 illustrate, similar patterns hold for the relative increase in export entry probabilities associated to previously exporting to countries located in the same continent, sharing the same official language or having similar income per capita as the potential new export destination.

A.3 Accounting for Gravity Forces

As discussed in Section 2.2, an alternative explanation for the transition probabilities reported in Table 1 is the importance of gravity variables in determining the entry behavior of potential exporters. If firms rank countries by proximity to Chile and spread out gradually to more distant markets, then a firm already exporting to a continent would be more likely to start exporting to other countries in the same continent. Similarly, if Chilean firms enter destinations with languages more similar to Spanish before entering those with more distinct languages (e.g. French speaking countries are prioritized over German speaking ones), the fact that we observe firms successively entering countries that share language would be exclusively due to their ranking in terms of linguistic distance to the country of origin (gravity) and not due to their linguistic distance to prior destinations of the firm (extended gravity).

The model described in Section 3 accounts for this alternative explanation by allowing the export profits that a firm would make in any given destination upon entry to flexibly depend on several measures of distance between that destination and the firm's home market, Chile. Specifically, as equation (2) in Section 3 shows, we allow marginal trade costs in our model to vary flexibly by destination country. Additionally, as equations (7) and (10) show, we also allow fixed and sunk costs to depend on whether the destination shares continent, official language or similar income per capita with Chile. Therefore, we allow any measure of similarity between countries that we use to define extended gravity effects (similar geographic location, language or income per capita) to also impact export decisions through standard gravity effects. Based on this model, sections 6 and 7 present the moment inequality estimates of those model parameters that determine the importance of extended gravity effects on export entry costs.

As a preliminary step, we show in Table A.4 how reduced-form export entry probabilities depend on extended gravity variables while simultaneously controlling for standard gravity covariates. Specifically, we present in Table A.4 estimates of several binary logit models of the probability that a firm exports a positive amount to a destination in a given year.

Besides the extended gravity covariates of interest (described in Section 2.1), we allow in the baseline specification (column I) the probability that a firm exports to a potential destination to depend on: (a) measures of distance between the potential destination and Chile (gravity); (b) measures of previous export experience of the firm in this potential destination; and (c) interactions of the measures in (a) and (b). The covariates in (a) account for gravity; those in (b) account for persistence in export status; and those in (c) account for heterogeneity in this persistence across potential destinations.

As measures of gravity, we include dummy variables that equal one for those potential destinations that do not share any border ("Grav. Border"), continent ("Grav. Cont."), language ("Grav. Lang."), similar income per capita ("Grav. GDPpc"), or any free trade agreement ("Grav. FTA") with Chile. As measures of previous export experience of a firm in a potential destination, we include a dummy variable that equals one for those firm-destination-year triplets for which the firm was not exporting to the corresponding destination in the previous year ("Entry").

The results of the baseline specification in column I of Table A.4 show that: (a) the most important determinant of export participation is prior export participation (as indicated by the large negative coefficient on "Entry"); (b) all our measures of distance between the destination and the origin countries have a negative impact on the probability that firm exports to that destination (as indicated by the negative coefficients on all gravity variables); (c) the persistence in export status is generally stronger in destinations that are further away (in the gravity sense) from the country of origin of the firms (as indicated by the negative coefficients on the interactions of "Entry" with both "Grav. Border" and "Grav. Lang."); (d) all extended gravity variables have a large and statistically significant positive impact on export participation.

A shortcoming of the specification in column I of Table A.4 is that it does not control directly for the heterogeneity across firms that differ in productivity (or domestic sales) in their probability of exporting to any given destination. This heterogeneity is documented in Table A.2 in Appendix A.2. The productivity of a firm is likely to impact its export probability by affecting its potential export revenue; i.e. the sales revenue that it would obtain if it were to export. In columns II to IV of Table A.4, we account for this impact by expanding the baseline specification in column I with an additional proxy for this potential export revenue. Specifically, we include either domestic sales as an additional control (in column II) or a predicted measure of potential export revenues that we construct by projecting the observed export revenues (of those firms, countries and years with positive exports) on a wide set of firm and country characteristics that we observe for every firm,

Table A.4: Logit

Variables: (β)	I	II	III	IV	V	VI	VII	VIII
Domestic Sales		0.215^a				0.207^a		
		(42.367)				(41.071)		
Revenue			3.075^a	2.625^a			3.508^a	2.749^a
			(13.597)	(13.636)			(15.588)	(14.306)
Grav. Border	-0.578^a	-0.525^a	-0.478^a	-0.544^a				
	(-5.795)	(-5.682)	(-4.774)	(-5.477)				
Grav. Cont.	-1.131^a	-1.049^a	-1.075^a	-1.135^a	-1.718^a	-1.662^a	-1.601^a	-1.681^a
	(-10.559)	(-9.545)	(-9.945)	(-10.658)	(-19.228)	(-17.711)	(-17.501)	(-18.666)
Grav. Lang.	-0.576^a	-0.488^a	-0.747^a	-0.683^a	-0.667^a	-0.639^a	-0.889^a	-0.826^a
	(-6.639)	(-5.237)	(-7.611)	(-7.939)	(-7.928)	(-6.921)	(-10.245)	(-9.698)
Grav. GDPpc	-0.792^a	-0.636^a	-0.721^a	-0.728^a	-0.773^a	-0.641^a	-0.709^a	0.723^a
	(-9.871)	(-7.254)	(-8.912)	(-9.313)	(-9.840)	(-7.364)	(-8.838)	(-9.112)
Grav. FTA	-0.383^a	-0.504^a	-0.441^a	-0.415^a				
	(-6.455)	(-8.264)	(-7.303)	(-6.939)				
Entry	-3.532^a	-3.087^a	-3.975^a	-3.401^a	-4.643^a	-4.398^a	-4.569^a	-4.614^a
	(-6.877)	(-5.923)	(-7.583)	(-6.605)	(-67.538)	(-62.513)	(-65.600)	(-66.992)
Entry $\times$	-0.104	-0.128^c	-0.042	-0.120^c				
Grav. Dist.	(-1.535)	(-1.844)	(-0.609)	(-1.755)				
Entry $\times$	-0.303^a	-0.343^a	-0.304^a	-0.300^a				
Grav. Border	(-2.379)	(-2.662)	(-2.339)	(-2.357)				
Entry $\times$	0.156	0.053	0.018	0.166	0.059	-0.069	-0.022	0.055
Grav. Cont.	(1.178)	(0.395)	(0.134)	(1.248)	(0.554)	(0.636)	(-0.207)	(0.519)
Entry $\times$	-0.544^a	-0.603^a	-0.448^a	-0.445^a	-0.606^a	-0.695^a	-0.482^a	-0.518^a
Grav. Lang.	(-4.828)	(-5.025)	(-3.900)	(-3.940)	(-5.385)	(-6.223)	(-4.527)	(-4.945)
Entry $\times$	0.217^a	0.129	0.205^b	0.195^b	0.184^b	0.095	0.158^c	0.154^c
Grav. GDPpc	(2.352)	(1.350)	(2.177)	(2.107)	(2.018)	(1.012)	(1.702)	(1.676)
Entry $\times$	0.923^a	0.873^a	0.845^a	0.829^a	0.746^a	0.697^a	0.695^a	0.667^a
Ext. Grav. Border	(15.705)	(14.944)	(14.298)	(14.307)	(13.807)	(12.779)	(12.509)	(12.231)
Entry $\times$	1.293^a	1.211^a	1.307^a	1.280^a	1.403^a	1.314^a	1.390^a	1.371^a
Ext. Grav. Cont.	(19.703)	(17.814)	(19.694)	(19.743)	(22.423)	(19.744)	(21.379)	(21.585)
Entry $\times$	0.411^a	0.353^a	0.387^a	0.363^a	0.427^a	0.391^a	0.417^a	0.402^a
Ext. Grav. Lang.	(5.598)	(4.802)	(5.233)	(5.068)	(6.096)	(5.373)	(5.663)	(5.661)
Entry $\times$	1.027^a	0.870^a	1.005^a	0.989^a	0.949^a	0.794^a	0.941^a	0.923^a
Ext. Grav. GDPpc	(16.778)	(13.093)	(16.243)	(16.288)	(16.042)	(12.109)	(15.437)	(15.421)
Num. Obs.	234,896	234,896	234,896	234,896	234,896	234,896	234,896	234,896

Notes: *a* denotes 1% significance, *b* denotes 5% significance, *c* denotes 10% significance. Estimates are obtained by MLE, and t-statistics are in parentheses. The dependent variable is a dummy variable for positive exports. All specifications include year fixed effects. "Domestic Sales", "Revenue", and "Grav. Dist" are in logs. All other covariates are dummy variables.

country and time period (in columns III and IV).[39] The results in columns II to IV in Table A.4 show that

[39] Using different firm and country characteristics as covariates in this projection yields different revenue proxies. Table B.1 presents projection estimates for six different sets of covariates. In spite of the differences across these sets, Table B.2 shows that the pairwise correlation coefficients between the export revenues generated by these six specifications are close to one. Therefore, for the sake of interpreting the results in Table A.4, it is not very important what the exact vector of observed covariates we use to construct our measure of predicted revenues is. Specifically, column III in Table A.4 uses a proxy for revenue consistent with the specification in column VI of Table B.1 (which projects export revenue on a

adding either domestic sales or a measure of potential export revenues as an additional covariate in our logit model does not significantly affect the estimates of the different gravity and extended gravity measures. All the resulting estimates are very similar to those in column I, which we have already described above.

The specifications in columns V to VII differ from those in columns I to IV only in that they exclude those gravity measures that we will also be excluding from the specification of the fixed and sunk export costs in our structural model (see equations (7) and (10) in Section 3): "Grav. Border", "Grav. FTA", and "Grav. Dist". As the results in Table A.4 show, eliminating these covariates from the logit model reduces the coefficient on "Entry" and "Grav. Cont" but has nearly no impact on the four extended gravity coefficients. Given that our moment inequality estimates focus exclusively on estimating bounds on these four extended gravity coefficients, we find their robustness to the specific set of gravity variables we control for reassuring.

A.4 Accounting for Unobserved Heterogeneity Across Country Groups

As discussed in Section 2.2, an alternative explanation for the transition probabilities reported in Table 1 is the importance of unobserved (to the researcher) demand and supply conditions that are correlated across destinations with similar geographic location, language, and/or income per capita. Under this alternative explanation, for example, the higher probability of exporting to a market among those firms previously exporting to a different market located in the same continent would not be due to a reduction in entry costs generated by this previous export experience. Instead, it would just be the consequence of export profits being similar across these two markets.

The model described in Section 3 accounts for this alternative explanation by allowing export profits to flexibly depend on unobserved covariates that are firm- and year-specific and correlated across groups of countries that share continent, language and/or similar income per capita. We thus allow firms to have some unobserved features that, for example, make some more likely to export to English-speaking countries and others more likely to export to French-speaking countries. For example, firms may specialize in different varieties of a product that are differently demanded in different destinations, or firms' workforce may include workers with different skills (e.g. linguistic skills) that specifically valuable to break into certain markets, or firms may have access to distributors that specialize in different sets of countries. Specifically, as equation (6) shows, we allow the export fixed costs of firm i in country j and year t, f_{ijt}, to depend on an unobserved component u_{ic_jt} common to all countries j in the same group or "cluster" c_j. In Section 7, we present results in which we define clusters of countries by their continent, language or income per capita group. When computing these results, we treat these unobserved terms u_{ic_jt} as firm-, year-, group-of-countries-specific fixed effects and, thus, we impose no assumption on their distribution.

As a preliminary step, we present here results in which we allow for an unobserved term u_{ic_jt} to affect the probability that a firm i exports to country j in year t but, relative to our moment inequality estimation, restrict its distribution in two ways: (a) we assume that it is constant over time, $u_{ic_jt} = u_{ic_j}$; (b) we assume that it is distributed normally and independently across firms and clusters of countries. Specifically, we assume here that

$$\begin{aligned} d_{ijt} = \mathbb{1}\{&\beta_1 \, \text{revenue}_{ijt} + \beta_2 \, \text{gravity}_j + \beta_3(1 - d_{ijt-1}) + \beta_4[(1 - d_{ijt-1}) \times \text{gravity}_j] \\ &+ \beta_5[(1 - d_{ijt-1}) \times \text{ext.gravity}_{ijt}] + u_{ic_j} + \varepsilon_{ijt} > 0\}, \end{aligned} \tag{A.1}$$

where d_{ijt} is a dummy variable equal to 1 if firm i exports to country j during year t, "revenue" is a proxy for the potential export revenue, "gravity" and "ext.gravity" are vectors of observed of gravity and extended gravity covariates, and both u_{ic_j} and ε_{ijt} are unobserved to the researcher. We assume that ε_{ijt} is independent across firms, countries and time periods, and follows a logistic distribution with location parameter equal to zero and scale parameter equal to one. Concerning the term u_{ic_j}, we assume that it is independent across firms and country clusters $c = 1, \ldots, C$, and follows a normal distribution with mean equal to zero and variance equal to an unknown parameter σ_c^2. Jointly with these distributional assumptions, the definition of d_{ijt} in equation (A.1) implies a mixed logit model for d_{ijt}. The correlation structure in this mixed logit is such that

$$cov(u_{ic_j} + \varepsilon_{ijt}, u_{ic_{j'}} + \varepsilon_{ij't}) = \begin{cases} \sigma_c^2 + \sigma_\varepsilon^2 & \text{if } i = i',\ j = j', \text{ and } t = t', \\ \sigma_c^2 & \text{if } i = i',\ c_j = c_{j'}, \text{ and } j \neq j' \text{ or } t' \neq t, \\ 0 & \text{otherwise,} \end{cases}$$

large set of firm and country characteristics) while column IV in Table A.4 uses a proxy for revenue consistent with the specification in column I of Table B.1 (which projects export revenue on a minimal set of characteristics). Consistently with the high correlations in Table B.2, the estimates in columns III and VI of Table B.1 are very similar to each other.

where $\sigma_\varepsilon^2 = var(\varepsilon_{ijt})$ is the variance of the standard logistic distribution; i.e. $\sigma_\varepsilon^2 = \pi^2/3$.

We present estimates of the vector $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$ and σ_c in Table A.5. In all columns, we use the same measures of "revenue", "gravity", and "ext.gravity" as in column III in Table A.4. The only difference across the different columns in Table A.5 is the definition of the country clusters $c = 1, \ldots, C$. These different definitions of country clusters imply different cross-country correlation patterns in the unobserved determinants of the firm's export status. As an example, the estimates in column I in Table A.5 are based on the assumptions that two different countries j and j' belong to the same group or cluster, $c_j = c_{j'}$, if they share continent, language, and similar income per capita. Column II defines broader groups: $c_j = c_{j'}$ if countries j and j' share continent and language, independent of their income per capita. Columns III to VII allow for other partitions of countries. Specifically, column III includes firm-, continent-, income-per-capita-specific random effects; column IV includes firm-, language-, income-per-capita-specific random effects; column V includes firm-, continent-specific random effects; column VI includes firm- language-specific random effects; and column VII includes firm-, income-per-capita-specific random effects.

Given a definition of the clusters of countries $c = 1, \ldots, C$, the log-likelihood function corresponding to each of the models in Table A.5 is

$$(NC)^{-1} \sum_{i=1}^{N} \sum_{c=1}^{C} \log \Big[\int_{u_{ic}} L_{ic}(u_{ic})(1/\sigma_c)\phi(u_{ic}/\sigma_c) du_{ic} \Big], \tag{A.2}$$

with

$$L_{ic}(u_{ic}) \equiv \prod_{\substack{j=1, \\ c=c_j}}^{J} \Big\{ \prod_{t=1}^{T} \{ \mathcal{L}(d_{ijt} | \{d_{ij't-1}; j' = 1, \ldots, J\}, x_{ijt}, u_{ic}; \beta)\} \mathcal{L}_0(d_{ij0} | x_{ij0}, u_{ic}; \gamma) \Big\}, \tag{A.3}$$

where $\prod_{j=1,c=c_j}^{J}$ denotes the product over all countries j included in the cluster c; the vector x_{ijt} includes the observed measures of "revenue", "gravity", and "ext.gravity"; $\beta \equiv (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$; γ is a vector of reduced-form parameters; and $\phi(\cdot)$ denotes the standard normal density. The variable N denotes the total number of firms that appear at least one year in the sample, and the variable C denotes the total number of clusters of countries that we allow for. The individual likelihood function $\mathcal{L}(\cdot)$ equals

$$\begin{gathered} \mathcal{L}(d_{ijt} | \{d_{ij't-1}; j' = 1, \ldots, J\}, x_{ijt}, u_{ic}; \beta) = \\ \big(P(d_{ijt} = 1 | \{d_{ij't-1}; j' = 1, \ldots, J\}, x_{ijt}, u_{ic}; \beta)\big)^{d_{ijt}} \big(1 - P(d_{ijt} = 1 | \{d_{ij't-1}; j' = 1, \ldots, J\}, x_{ijt}, u_{ic}; \beta)\big)^{1-d_{ijt}}, \end{gathered} \tag{A.4}$$

with

$$\begin{gathered} P(d_{ijt} = 1 | \{d_{ij't-1}; j' = 1, \ldots, J\}, x_{ijt}, u_{ic}; \beta) = \\ \frac{\exp(\beta_1 \, \text{revenue}_{ijt} + \beta_2 \, \text{gravity}_j + \beta_3(1 - d_{ijt-1}) + \beta_4(1 - d_{ijt-1})\text{gravity}_j + \beta_5(1 - d_{ijt-1})\text{ext.gravity}_{ijt} + u_{ic_j})}{1 + \exp(\beta_1 \, \text{revenue}_{ijt} + \beta_2 \, \text{gravity}_j + \beta_3(1 - d_{ijt-1}) + \beta_4(1 - d_{ijt-1})\text{gravity}_j + \beta_5(1 - d_{ijt-1})\text{ext.gravity}_{ijt} + u_{ic_j})}. \end{gathered} \tag{A.5}$$

The individual likelihood function $\mathcal{L}_0(\cdot)$ equals

$$\mathcal{L}_0(d_{ij0} | x_{ij0}, u_{ic}; \gamma) = \big(P_0(d_{ij0} = 1 | x_{ij0}, u_{ic}; \gamma)\big)^{d_{ij0}} \big(1 - P_0(d_{ij0} = 1 | x_{ij0}, u_{ic}; \gamma)\big)^{1-d_{ij0}}, \tag{A.6}$$

for some unknown probability function $P_0(\cdot)$. Our model does not have a prediction for the functional form of the function $P_0(\cdot)$. This function determines the probability of the initial conditions $\{d_{ij0}; i = 1, \ldots, I, \; j = 1, \ldots, J\}$ as a function of the exogenous covariates included in the vector $\{x_{ij0}; i = 1, \ldots, I, \; j = 1, \ldots, J\}$ and the firm- and group-of-countries specific effects $\{u_{ic_j}; i = 1, \ldots, I, \; j = 1, \ldots, J\}$. Following Heckman (1981), we approximate $P_0(\cdot)$ through a reduced-form function of the exogenous variables included in the vector x_{ijt} and the unobserved effect u_{ic_j}; specifically, we assume that[40]

$$P_0(d_{ij0} = 1 | x_{ij0}, u_{ic}; \gamma) = \frac{\exp(\gamma_1 \, \text{revenue}_{ij0} + \gamma_2 \, \text{gravity}_j + u_{ic_j})}{1 + \exp(\gamma_1 \, \text{revenue}_{ij0} + \gamma_2 \, \text{gravity}_j + u_{ic_j})}. \tag{A.7}$$

[40] An alternative approach to model $P_0(\cdot)$ is to assume that all firms are in steady state in the first year of our sample (e.g. Card and Sullivan, 1998). Not only is this assumption unrealistic in our setting but, additionally, computing the steady state export probability of each firm in each destination country is particularly complicated in our case due to the presence of the extended gravity effects.

For any given value of the parameter vector (β, σ_c), computing the log-likelihood function in equations (A.2) to (A.7) requires numerically computing the integral

$$\int_{u_{ic}} L_{ic}(u_{ic})(1/\sigma_c)\phi(u_{ic}/\sigma_c)du_{ic}, \tag{A.8}$$

for every firm i in the sample and every cluster c in which we partition the J possible export destinations. To compute this integral, we use the Gauss-Hermite quadrature (e.g. Butler and Moffitt, 1982) and, thus, approximate it as

$$\int_{u_{ic}} f_{ic}(u_{ic})(1/\sigma_g)\phi(u_{ic}/\sigma_g)du_{ic} \approx \pi^{-\frac{1}{2}} \sum_{l=1}^{n} \tilde{w}_l f_{ic}(\sqrt{2}\sigma_g \tilde{u}_l), \tag{A.9}$$

where $\tilde{w}_l$ and $\tilde{u}_l$ are the Gauss-Hermite quadrature weights and nodes. In our approximation, we fix $n = 10$ and use the weights and nodes reported in Table 7.4 in Judd (1998).

A comparison of the estimates of the different models presented in Table A.5 to those of the model presented in column III in Table A.4, shows that, while the estimates of the coefficients on gravity measures are sensitive to the introduction of firm- and group-of-countries-specific random effects, the estimates of the extended gravity effects are generally robust to the introduction of these random effects u_{ic} in the statistical model. As in Table A.4, the coefficients on the four extended gravity covariates are always positive and generally statistically significant at 1%. The only exception happens in those cases in which the random effect is defined exactly at the same level as the extended gravity covariate. In columns V to VII, we allow for random effects at the continent-, language- and income-per-capita groups and the corresponding extended gravity coefficients become either significant only at the 5% level (in the case of column V) or not statistically significant at any commonly used statistical level (in the case of columns VI and VII). Conversely, all other extended gravity covariates remain significant at the 1% significance level.

A.5 Accounting for Extended Gravity in Marginal Costs

As discussed in Section 2.2, even if one can conclude that exporting to a country does indeed increase the probability of subsequently exporting to other destinations that share some characteristic with it, there may be different economic mechanisms that can generate this path dependence in export destinations. Specifically, the transition probabilities reported in Table 1 could be due to extended gravity variables impacting not the costs that a firm must pay upon entering a new market (sunk export costs) but the per-unit transport cost that firms must pay to sell abroad.

As we discuss in Section 3, the per-unit transport costs that a firm faces in a destination affect the probability that the firm exports to such destination by impacting the potential export revenue that such firm would make in such destination upon entry. Therefore, to show that our measures of extended gravity effects, understood as changes in sunk export costs, are not sensitive to whether we account for the possible impact of the same extended gravity variables on per-unit transport costs, we compute predicted potential export revenues allowing different vectors of covariates to impact such revenues. The estimates on these different specifications of export revenues are in Table B.1. Among the six revenue projections in Table B.1, those in columns I, II and IV account for the possible impact of extended gravity effects on per-unit transport costs (and, thus, on export revenues), while those in columns III, V and VI do not. The six columns in Table A.6 use the measure of predicted revenues generated by the specification described in the corresponding column of Table B.1. As the results in Table A.6 show, the differences in the maximum likelihood estimates of the parameters on the four extended gravity are very small. These estimates are thus robust to whether we account for the potential impact of the same extended gravity covariates on potential export revenues.[41]

[41] Table A.6 shows the robustness of the estimates in column I of Table A.5 to different specifications of the predicted export revenues. For each of the specifications in columns II to VII of Table A.5, tables analogous to Table A.6 are available upon request. They all show that the mixed logit estimates are not affected by the particular statistical model used to predict the potential revenue from exporting, revenue$_{ijt}$, among those in Table B.1.

Table A.5: Mixed Logit

Variables: (β)	I	II	III	IV	V	VI	VII
Revenue	0.014^{a}	0.013^{a}	0.015^{a}	0.013^{a}	0.015^{a}	0.017^{a}	0.016^{a}
	(5.415)	(5.126)	(6.100)	(5.088)	(5.897)	(6.343)	(6.444)
Grav. Border	-0.315^{a}	-0.268^{a}	-0.123	-0.183^{b}	-0.193^{b}	-0.057	-0.054
	(-3.188)	(-2.757)	(-1.278)	(-1.918)	(1.936)	(-0.597)	(0.559)
Grav. Cont.	-0.037	-0.050	-0.246^{b}	-0.263^{a}	-0.055	-0.334^{a}	-0.327^{a}
	(-0.356)	(-0.481)	(-2.513)	(-2.600)	(-0.035)	(-3.488)	(-3.399)
Grav. Lang.	-0.724^{a}	-0.669^{a}	-0.651^{a}	-0.619^{a}	-0.623^{a}	-0.578^{a}	-0.586^{a}
	(-7.659)	(-7.255)	(-7.611)	(-6.769)	(-7.028)	(-6.851)	(-6.901)
Grav. GDPpc	0.179^{b}	0.141^{c}	0.153^{b}	0.197^{a}	0.077	0.055	0.052
	(2.338)	(1.933)	(2.145)	(2.719)	(1.055)	(0.773)	(0.727)
Grav. FTA	-0.338^{a}	-0.259^{a}	-0.306^{a}	-0.304^{a}	-0.319^{a}	-0.505^{a}	-0.376^{a}
	(-4.978)	(-4.288)	(-5.162)	(-5.146)	(-5.349)	(-5.770)	(-6.404)
Entry	-2.180^{a}	-2.356^{a}	-2.652^{a}	-2.940^{a}	-3.354^{a}	-2.906^{a}	-2.608^{a}
	(-3.877)	(-4.232)	(-4.913)	(-5.416)	(-5.905)	(-5.446)	(-4.962)
Entry ×	-0.124^{c}	-0.128^{c}	-0.075	0.036	0.014	-0.068	-0.091
Grav. Dist.	(-1.655)	(-1.726)	(-1.049)	(0.497)	(0.187)	(-0.595)	(-1.291)
Entry ×	-0.919^{a}	-0.910^{a}	-1.097^{a}	-1.086^{a}	-1.168^{a}	-1.153^{a}	-1.182^{a}
Grav. Border	(-6.791)	(-6.793)	(-8.243)	(-8.191)	(-8.512)	(-8.722)	(-8.873)
Entry ×	-1.032^{a}	-0.864^{a}	-0.703^{a}	-1.032^{a}	0.278^{c}	-0.734^{a}	-0.915^{a}
Grav. Cont.	(-6.930)	(-5.854)	(-4.898)	(-7.223)	(1.921)	(-5.374)	(-6.629)
Entry ×	-0.145	-0.165	-0.395^{a}	-0.173	-0.380^{a}	0.001	-0.428^{a}
Grav. Lang.	(-1.149)	(-1.319)	(-3.313)	(1.404)	(-3.169)	(0.006)	(-3.656)
Entry ×	-0.727^{a}	-0.623^{a}	-0.567^{a}	-0.694^{b}	-0.273^{b}	-0.388^{a}	0.019
Grav. GDPpc	(-6.504)	(-5.902)	(-5.158)	(-6.285)	(-2.588)	(-3.766)	(0.171)
Entry ×	0.761^{a}	0.830^{a}	0.793^{a}	0.872^{a}	0.699^{a}	0.942^{a}	0.889^{a}
Ext. Grav. Border	(9.814)	(11.140)	(10.822)	(11.931)	(9.552)	(13.334)	(12.576)
Entry ×	1.552^{a}	1.263^{a}	1.307^{a}	1.465^{a}	0.201^{b}	1.180^{a}	1.467^{a}
Ext. Grav. Cont.	(17.001)	(14.224)	(14.570)	(17.029)	(2.271)	(14.468)	(18.380)
Entry ×	0.277^{a}	0.293^{a}	0.251^{a}	0.224^{b}	-0.130	-0.111	0.236^{a}
Ext. Grav. Lang.	(2.843)	(3.117)	(2.735)	(2.409)	(-1.353)	(-1.244)	(2.668)
Entry ×	0.708^{a}	0.657^{a}	0.573^{a}	0.576^{a}	0.229^{a}	0.429^{a}	-0.056
Ext. Grav. GDPpc	(7.810)	(7.719)	(6.287)	(6.495)	(2.767)	(5.301)	(-0.602)
Firm RE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Continent RE	Yes	Yes	Yes	No	Yes	No	No
Language RE	Yes	Yes	No	Yes	No	Yes	No
GDPpc RE	Yes	No	Yes	Yes	No	No	Yes
Std. Dev. RE: (σ_c)	9.374^{a}	8.908^{a}	9.169^{a}	9.272^{a}	5.714^{a}	8.199^{a}	8.826^{a}
	(9.227)	(8.762)	(9.022)	(9.133)	(5.620)	(8.033)	(8.665)
Num. Obs.	234,896	234,896	234,896	234,896	234,896	234,896	234,896
Num. Obs. per (i, c_j)	23	40	46	40	133	93	232

Notes: a denotes 1% significance, b denotes 5% significance, c denotes 10% significance. Estimates are obtained by MLE, and t-statistics are in parentheses. The dependent variable is a dummy variable for positive exports. The explanatory variable Revenue denotes predicted revenue (in logs), as generated by the estimates in Column I of Table B.1. Column I includes firm-continent-language-GDPpc specific random effects. Column II includes firm-continent-language specific random effects. Column III includes firm-continent-GDPpc specific random effects. Column IV includes firm-language-GDPpc specific random effects. Column V includes firm-continent specific random effects. Column VI includes firm-language specific random effects. Column VII includes firm-GDPpc specific effects.

Table A.6: Mixed Logit With Different Predicted Export Revenues

Variables: (β)	I	II	III	IV	V	VI
Revenue	0.014^{a}	0.014^{a}	0.012^{a}	0.018^{a}	0.027^{a}	0.022^{a}
	(5.415)	(4.439)	(4.970)	(6.021)	(11.701)	(9.001)
Grav. Border	-0.315^{a}	-0.317^{a}	-0.310^{a}	-0.318^{a}	-0.313^{a}	-0.245^{a}
	(-3.184)	(-3.207)	(-3.136)	(-3.219)	(-3.169)	(-2.479)
Grav. Cont.	-0.037	-0.057	-0.033	-0.036	-0.050	-0.029
	(-0.356)	(-0.535)	(-0.317)	(-0.342)	(-0.479)	(-0.272)
Grav. Lang.	-0.724^{a}	-0.713^{a}	-0.719^{a}	-0.749^{a}	-0.762^{a}	-0.748^{a}
	(-7.660)	(-7.477)	(-7.567)	(-7.931)	(-8.237)	(-7.948)
Grav. GDPpc	0.179^{a}	0.148^{b}	0.174^{b}	0.181^{a}	0.177^{a}	0.179^{a}
	(2.334)	(1.910)	(2.284)	(2.375)	(2.325)	(2.320)
Grav. FTA	-0.302^{a}	-0.299^{a}	-0.308^{a}	-0.302^{a}	-0.284^{a}	-0.321^{a}
	(-4.978)	(-4.801)	(-5.075)	(-4.982)	(-4.665)	(-5.200)
Entry	-2.179^{a}	-2.205^{a}	-2.165^{a}	-2.160^{a}	-2.164^{a}	-2.780^{a}
	(-3.875)	(-3.901)	(-3.847)	(-3.844)	(-3.858)	(-4.883)
Entry ×	-0.124^{b}	-0.121^{c}	-0.127^{b}	-0.127^{b}	-0.128^{b}	-0.049
Grav. Dist.	(-1.657)	(-1.603)	(-1.691)	(-1.695)	(-1.713)	(-0.639)
Entry ×	-0.919^{a}	-0.934^{a}	-0.917^{a}	-0.909^{a}	-0.924^{a}	-0.925^{a}
Grav. Border	(-6.794)	(-6.883)	(-6.776)	(-6.715)	(-6.828)	(-6.785)
Entry ×	-1.032^{a}	-1.036^{a}	-1.034^{a}	-1.029^{a}	-0.975^{a}	-1.080^{a}
Grav. Cont.	(-6.930)	(-6.913)	(-4.898)	(-6.912)	(-6.557)	(-7.213)
Entry ×	-0.145	-0.160	-0.149	-0.121	-0.177^{c}	-0.231^{b}
Grav. Lang.	(-1.149)	(-1.244)	(-1.174)	(-0.951)	(-1.428)	(-1.832)
Entry ×	-0.726^{a}	-0.717^{a}	-0.722^{a}	-0.727^{a}	-0.715^{a}	-0.707^{a}
Grav. GDPpc	(-6.498)	(-6.341)	(-6.458)	(-6.504)	(-6.188)	(-6.246)
Entry ×	0.761^{a}	0.791^{a}	0.749^{a}	0.765^{a}	0.696^{a}	0.739^{a}
Ext. Grav. Border	(9.814)	(10.001)	(9.657)	(9.865)	(8.950)	(9.328)
Entry ×	1.552^{a}	1.517^{a}	1.560^{a}	1.495^{a}	1.523^{b}	1.508^{a}
Ext. Grav. Cont.	(17.003)	(16.345)	(17.091)	(16.969)	(16.621)	(16.159)
Entry ×	0.277^{a}	0.269^{a}	0.284^{a}	0.275^{a}	0.224^{b}	0.226^{b}
Ext. Grav. Lang.	(2.847)	(2.669)	(2.914)	(2.828)	(2.266)	(2.218)
Entry ×	0.708^{a}	0.751^{a}	0.709^{a}	0.706^{a}	0.701^{a}	0.732^{a}
Ext. Grav. GDPpc	(7.810)	(8.177)	(7.826)	(7.782)	(7.704)	(7.932)
Firm RE	Yes	Yes	Yes	Yes	Yes	Yes
Continent RE	Yes	Yes	Yes	Yes	Yes	Yes
Language RE	Yes	Yes	Yes	Yes	Yes	Yes
GDPpc RE	Yes	Yes	Yes	Yes	Yes	Yes
Std. Dev. RE: (σ_g)	9.374^{a}	9.380^{a}	9.372^{a}	9.365^{a}	9.346^{a}	9.375^{a}
	(9.227)	(9.236)	(9.224)	(9.218)	(9.202)	(9.232)
Num. Obs.	234,896	234,896	234,896	234,896	234,896	234,896

Notes: a denotes 1% significance, b denotes 5% significance, c denotes 10% significance. Estimates are obtained by MLE, and t-statistics are in parentheses. The dependent variable is a dummy variable for positive exports. The explanatory variable For each column, "Revenue" denotes predicted revenue (in logs), as generated by the estimates in the corresponding column of Table B.1. All columns include firm-continent-language-GDPpc specific random effects.

B Export Revenue: Details

B.1 Export Revenue Equation: Details

In this section, we show that the assumptions on demand, variable trade and production costs, and market structure introduced in Section 3.1 imply the expressions for r_{ijt} in equations (4) and (5).

Given the constant elasticity of substitution demand function $q_{ijt} = p_{ijt}^{-\eta} P_{jt}^{\eta-1} Y_{jt}$ in every market, the constant marginal production cost $a_{it}w_t$, the constant variable trade costs τ_{ijt}, and the monopolistic competition assumptions, we can write the potential revenue that firm i would obtain in market j at period t as indicated in equation (1); i.e.

$$r_{ijt} = \left[\frac{\eta}{\eta - 1} \frac{\tau_{ijt} a_{it} w_t}{P_{jt}} \right]^{1-\eta} Y_{jt}.$$

Assuming that trade costs in h are common across firms, $\tau_{iht} = \tau_{ht}$, we can similarly write the potential revenue that firm i will obtain in the home market at period t as

$$r_{iht} = \left[\frac{\eta}{\eta - 1} \frac{\tau_{ht} a_{it} w_t}{P_{ht}} \right]^{1-\eta} Y_{ht}. \tag{B.1}$$

Taking the ratio of these two expressions, we can express the potential export revenues of firm i in market j and period t relative to its domestic sales in the same time period as

$$\frac{r_{ijt}}{r_{iht}} = \left[\frac{\tau_{ijt}}{\tau_{ht}} \frac{P_{ht}}{P_{jt}} \right]^{1-\eta} \frac{Y_{jt}}{Y_{ht}}, \tag{B.2}$$

and, multiplying by r_{iht} on both sides of the equality, we obtain

$$r_{ijt} = \left[\frac{\tau_{ijt}}{\tau_{ht}} \frac{P_{ht}}{P_{jt}} \right]^{1-\eta} \frac{Y_{jt}}{Y_{ht}} r_{iht}. \tag{B.3}$$

Given the expression for variable trade costs τ_{ijt} in equation (2), we can rewrite potential export revenues as in equation (4),

$$r_{ijt} = r_{ijt}^o + \varepsilon_{ijt}^R,$$

with the observed component of revenue being equal to,

$$r_{ijt}^o = \exp(\xi_{jt} + \xi_i + X_{ijt}^\tau \xi^\tau) \left[\frac{1}{\tau_{ht}} \frac{P_{ht}}{P_{jt}} \right]^{1-\eta} \frac{Y_{jt}}{Y_{ht}} r_{iht}, \tag{B.4}$$

and the unobserved component being equal to

$$\varepsilon_{ijt}^R = \varepsilon_{ijt}^\tau \left[\frac{1}{\tau_{ht}} \frac{P_{ht}}{P_{jt}} \right]^{1-\eta} \frac{Y_{jt}}{Y_{ht}} r_{iht}. \tag{B.5}$$

Using the expression for r_{iht} in equation (B.1), the definition $X_{ijt}^\tau \equiv (d_{ijt-1}, X_{ijt}^e, \ln(a_{it}))$ and additionally defining $\xi^\tau \equiv (\xi_d, \xi_e, \xi_a)$, we can further rewrite r_{ijt}^o as

$$r_{ijt}^o = \exp(\xi_{jt} + \xi_i) \left[r_{iht}^{\frac{1}{1-\eta}} Y_{ht}^{-\frac{1}{1-\eta}} \frac{(\eta - 1) P_{ht}}{\eta \tau_{ht} w_t} \right]^{\xi_a} \exp(\xi_d d_{ijt-1} + \xi_e X_{ijt}^e) \left[\frac{1}{\tau_{ht}} \frac{P_{ht}}{P_{jt}} \right]^{1-\eta} \frac{Y_{jt}}{Y_{ht}} r_{iht}.$$

Defining the country-year and firm unobserved effects entering the export sales equation,

$$\begin{aligned} \exp(\alpha_{jt}) &= \exp(\tau_{jt}) \left[Y_{ht}^{-\frac{1}{1-\eta}} \frac{(\eta - 1) P_{ht}}{\eta \tau_{ht} w_t} \right]^{\xi_a} \left[\frac{1}{\tau_{ht}} \frac{P_{ht}}{P_{jt}} \right]^{1-\eta} \frac{Y_{jt}}{Y_{ht}}, \\ \exp(\alpha_i) &= \exp(\tau_i), \end{aligned}$$

and the parameter vector

$$\alpha^r = (\xi_d, \xi_e, \xi_r), \qquad \xi_r = \frac{\xi_a + 1 - \eta}{1 - \eta},$$

we can rewrite r^o_{ijt} as in equation (4); i.e.

$$r^o_{ijt} = \exp(\alpha_{jt} + \alpha_i + X^r_{ijt}\alpha_r), \qquad \text{with} \qquad X^r_{ijt} = (d_{ijt-1}, X^e_{ijt}, \ln(r_{iht})).$$

Concerning the unobservable component of export revenues, ε^R_{ijt}, the following derivation shows that, given its definition in equation (B.5) and the mean independence condition in equation (3), the variable ε^R_{ijt} must satisfy the mean independence condition in equation (5):

$$\begin{aligned}
\mathbb{E}_{jt}[\varepsilon^R_{ijt}|X^r_{ijt}, d_{ijt}, \mathcal{J}_{it}] &= \mathbb{E}_{jt}[\varepsilon^R_{ijt}|d_{ijt-1}, X^e_{ijt}, r_{iht}, d_{ijt}, \mathcal{J}_{it}] \\
&= \mathbb{E}_{jt}\left[\varepsilon^\tau_{ijt}\left[\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right]^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht}\middle|d_{ijt-1}, X^e_{ijt}, r_{iht}, d_{ijt}, \mathcal{J}_{it}\right] \\
&= \mathbb{E}_{jt}[\varepsilon^\tau_{ijt}|d_{ijt-1}, X^e_{ijt}, r_{iht}, d_{ijt}, \mathcal{J}_{it}] \times \left[\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right]^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= \mathbb{E}_{jt}[\varepsilon^\tau_{ijt}|d_{ijt-1}, X^e_{ijt}, a_{it}, d_{ijt}, \mathcal{J}_{it}] \times \left[\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right]^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= \mathbb{E}_{jt}[\varepsilon^\tau_{ijt}|X^\tau_{ijt}, d_{ijt}, \mathcal{J}_{it}] \times \left[\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right]^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= 0 \times \left[\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right]^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} = 0,
\end{aligned}$$

where the first equality uses the definition of X^r_{ijt}, the second equality uses equation (B.5), the third equality takes into account that price indices and market sizes are constant conditional on a country-year pair, the fourth equality takes into account that r_{iht} is a deterministic function of a_{it} and variables that vary only at the country-year pair level, the fifth equality uses the definition of X^τ_{ijt}, and the last equality applies the mean independence restriction in equation (3).

Summing up, this section shows that the description of potential export revenues in equations (4) to (5) is a consequence of the assumptions imposed in equations (1) to (3).

B.2 Revenue Regression

Table B.1 contains the Nonlinear Least Squares (NLS) estimates of the parameter vector entering the expression for potential export revenues in equation (4). Let's denote the observed measure of export revenues for a firm i in market j and year t as r^{obs}_{ijt}. We assume that observed export revenues equal potential export revenues for those firm-market-year triplets with positive exports. Therefore,

$$r^{obs}_{ijt} = d_{ijt}r_{ijt} = d_{ijt}(\exp(\alpha_{jt} + \alpha_i + X^r_{ijt}\alpha^r) + \varepsilon^R_{ijt}),$$

where the second equality uses equation (4). Combining this expression with equation (5), we can derive the following moment condition

$$\mathbb{E}_{jt}[r^{obs}_{ijt} - \exp(\alpha_{jt} + \alpha_i + X^r_{ijt}\alpha^r)|X^r_{ijt}, d_{ijt} = 1] = 0. \tag{B.6}$$

Notice that, if present in the data, measurement error in observed export revenues r^{obs}_{ijt} is also likely to be mean zero conditional on the vector of observed covariates X^r_{ijt} and the export dummy d_{ijt}. Therefore, the mean independence condition in equation (B.6) is also likely to hold in the presence of additive measurement error in export revenues.

Table B.1 presents the estimates of the parameters entering the moment condition in equation (B.6) under several different specifications of the vector of observed covariates X^r_{ijt} and the vector of fixed effects entering the export revenue equation. Columns I to VI include firm and year fixed effects, while columns VII to IX include firm and country-year pair specific fixed effects. Specifically, the regression equation whose estimates are shown in column IX corresponds exactly to that in equation (4) in Section 3.1 in the main text: it includes firm and country-year fixed effects and, as observed covariates, it includes the firm's domestic sales, an indicator of lagged export participation in each country ("Entry" equals 1 the firm did not export to the corresponding market in the previous year), and the vector of extended gravity covariates X^e_{ijt}.

While the assumptions on demand, variable trade costs and market structure in Section 3.1 suggest that country-year specific fixed effects should be included in our revenue regression, its inclusion is problematic when

Table B.1: Revenue Regressions

Variables: (β)	I	II	III	IV	V	VI	VII	VIII	IX
log(Dom. Rev.)		0.386^a				0.328^a	0.403^a	0.388^a	0.384^a
		(29.186)				(20.892)	(32.016)	(32.241)	(32.767)
Emp. $\times$ Chile	0.444^a	0.142^a	0.380^a		0.184^a				
	(13.476)	(7.201)	(11.504)		(5.301)				
Sk. Emp.	0.057^b		0.103^a	0.258^a	0.236^a				
	(2.184)		(4.413)	(14.207)	(9.775)				
Unsk. Emp.	-0.002		0.015	0.078^a	0.091^a				
	(-0.145)		(1.339)	(8.128)	(8.182)				
Avg. Sk. Wage	0.643^a		0.816^a	0.069^a	0.063^a				
	(4.736)		(5.591)	(5.146)	(4.386)				
Avg. Unsk. Wage	0.009^a		0.004^c	-0.003	-0.008^a				
	(2.751)		(1.286)	(-0.978)	(-2.411)				
VA/Emp.	0.369^a		0.341^a	0.333^a	0.307^a				
	(24.400)		(22.632)	(22.244)	(19.589)				
Grav. Dist.	-0.638^a	-0.526^a	-0.538^a	-0.617^a	-0.596^a	-1.337^a			
	(-8.100)	(-6.461)	(-8.096)	(-7.768)	(-8.285)	(-46.453)			
Grav. Border	1.206^a	1.157^a	0.839^a	1.195^a	0.978^a				
	(11.504)	(10.722)	(9.221)	(11.219)	(10.044)				
Grav. Cont.	-0.079	-0.089	-0.421^a	-0.208^b	-0.531^a				
	(-0.706)	(-0.795)	(-4.047)	(-1.827)	(-4.788)				
Grav. Lang.	-0.084^c	-0.311^a	0.122^a	-0.059	-0.030				
	(-1.614)	(-6.188)	(2.493)	(-1.118)	(-0.567)				
Grav. GDPpc	-1.013^a	-1.388^a	-0.143^b	-0.999^a	-0.170^b				
	(-3.352)	(-2.968)	(-1.839)	(-2.925)	(-2.110)				
Grav. FTA	0.202^a	-0.182^a	0.237^a	0.200^a	0.199^a				
	(4.439)	(-4.101)	(5.241)	(4.332)	(4.274)				
Landlocked	-1.234^c	-1.289^b	-0.991^c	-1.127^c	-1.267				
	(-1.447)	(-1.856)	(-1.334)	(-1.494)	(-1.247)				
Agg. Imports	0.857^a	0.871^a	-0.782^a	0.849^a	0.833^b	1.107^a			
	(30.755)	(31.649)	(29.503)	(30.241)	(29.788)	(41.867)			
GDPpc	-0.067^a	-0.109^a	-0.055^a	-0.067^a	-0.523^a				
	(-3.440)	(-5.616)	(-2.899)	(-3.449)	(-2.488)				
Ext. Grav. Border	-0.205^a	-0.282^a		-0.202^a					
	(-6.259)	(-8.125)		(-6.111)					
Ext. Grav. Cont.	-0.449^a	-0.501^a		-0.346^a					
	(-9.590)	(-8.671)		(-7.136)					
Ext. Grav. Lang.	0.029	0.395^a		0.033					
	(0.883)	(10.863)		(0.967)					
Ext. Grav. GDPpc	0.966^a	1.617^a		0.954^a					
	(3.289)	(3.492)		(2.853)					
Entry	-11.028^a	-5.299	-15.916^a	-10.237^a				-1.693^a	-1.899^a
	(-3.309)	(-0.708)	(-6.956)	(-3.095)				(-10.493)	(-10.067)
Entry $\times$ Grav. Dist.	1.464^a	0.593	2.147^a	1.387^a					
	(3.116)	(0.566)	(6.817)	(2.997)					

Notes: continues in the next page.

Table B.1: Revenue Regressions (cont.)

Variables: (β)	I	II	III	IV	V	VI	VII	VIII	IX
Entry $\times$	-1.695^c	-1.028	-1.451	-1.251					
Grav. Cont.	(-1.371)	(-0.793)	(-1.153)	(-0.993)					
Entry $\times$	-1.055^a	-1.452^a	-1.097^a	-1.075^a					
Grav. Lang.	(-4.053)	(-3.146)	(-4.083)	(-4.030)					
Entry $\times$	0.770	2.296^b	-0.045	0.984^c					
Grav. GDPpc	(1.159)	(2.242)	(-0.083)	(1.368)					
Entry $\times$	-1.000^a	-0.409		-0.954^a					-0.177^a
Ext. Grav. Border	(-2.997)	(-0.769)		(-2.752)					(-5.282)
Entry $\times$	1.264^a	1.267^c		0.782^a					0.478^a
Ext. Grav. Cont.	(4.172)	(1.507)		(2.635)					(14.918)
Entry $\times$	0.394^c	-0.975		0.357					-0.103^a
Ext. Grav. Lang.	(1.307)	(-0.884)		(1.144)					(-2.534)
Entry $\times$	-0.734^b	-2.385^a		-0.824^b					0.534^a
Ext. Grav. GDPpc	(-2.018)	(-2.633)		(-2.073)					(2.503)
Adj. R^2	0.8715	0.8876	0.8632	0.8633	0.8496	0.8216	0.9027	0.9075	0.9099
Num. Obs.	8219	8219	8219	8219	8219	8219	8219	8219	8219

Notes: *a* denotes 1% significance, *b* denotes 5% significance, *c* denotes 10% significance. Heteroskedasticity-robust t-statistics are in parentheses. The dependent variable is firm-country-year level observed revenue (conditional on being positive). Firm dummies are included in all specifications. Additionally, specifications I to VI account for year fixed effects, and specifications VII to IX account for country-year pair specific fixed effects. Interaction terms between year dummies and a dummy for selling in Chile are also included.

the aim of this regression is to generate predicted revenues for a large number of firm-country-year triplets. The reason is that we cannot identify the value of the country-year fixed effect α_{jt} for those destination-year pairs in which we observe no exporters in our sample. This is inconvenient, as it implies that, if we were to construct our revenue predictions using an specification that includes such fixed effects, we would not be able to construct a proxy for the potential revenue from exporting r_{ijt} for a large number of observations in our sample.

The specifications in columns I to VI account only firm and year fixed effects. These fixed effects can be precisely estimated for all firms and years in our sample and, therefore, its inclusion does not restrict the set of observations from whom we can construct predicted potential export revenues. To compensate for the lack of country-year fixed effects in these specifications, we include instead a large set of observable country characteristics: both time-invariant gravity covariates (e.g. "Grav. Dist" or "Grav. Border") and time-varying measures of market size ("Agg. Imports").

Specifically, the regressions in columns I to VI differ in the exact set of firm and country characteristics they include. While regressions II and VI include only domestic sales as a summary statistic of all firm characteristics that matter to predict export revenues, regressions I, III, IV and V include instead information on the number of employees, average wages for skilled and unskilled workers, and value added per worker. Similarly, while specifications I to V include a large set of gravity and extended gravity covariates, specification VI only accounts for distance to Chile and aggregate sectoral imports in the destination market.

As discussed in Section 3.1, we are not intrinsically interested in the estimates presented in Table B.1. They are useful only as an intermediate step to generate an observed proxy of potential export revenues for every firm-country-year triplet in the sample. To measure how sensitive our predicted potential export revenues are to the specification of the revenue equation, we present in Table B.2 pairwise correlation coefficients across the nine vectors of predicted revenues generated using the estimates reported in the nine columns in Table B.1.

We can extract several conclusions from the correlation matrix in Table B.2. First, there is no evidence that regressions with country-year fixed effects yield different revenue predictions than those without them. Therefore, we can approximate well the observed export revenues even if we do not account for these fixed effects. Second, as long as we include in the revenue regression the firm's domestic sales, the destination market's distance to Chile and aggregate sectoral imports, and firm and year fixed effects, the resulting predicted revenues are highly correlated with those that we would obtain if we add other observed firm and country characteristics.

Given that our focus is on estimating the impact of extended gravity on export sunk costs, the main implication of the high correlation coefficients shown in Table B.2 is that we will obtain very similar estimates of our parameters of interest independently of the exact model we use to generate predicted export revenues. In fact, both the mixed logit estimates in Table A.6 in Appendix A.5 and the moment inequality estimates in Table 4 in Section 6 show that this is the case.

Table B.2: Revenue - Correlation Matrix

Specification	I	II	III	IV	V	VI	VII	VIII	IX
I	1.000								
II	0.938	1.000							
III	0.994	0.956	1.000						
IV	0.908	0.951	0.943	1.000					
V	0.982	0.948	0.986	0.934	1.000				
VI	0.787	0.900	0.832	0.930	0.847	1.000			
VII	0.973	0.983	0.972	0.974	0.974	1.000	1.000		
VIII	0.975	0.985	0.974	0.976	0.972	0.954	0.953	1.000	
IX	0.974	0.985	0.973	0.976	0.970	0.953	0.995	0.998	1.000

Notes: Correlation coefficients between the different predicted revenues generated by the corresponding regressions in Table B.1. Pairwise correlations between specifications that do not include country-year fixed effects (i.e. all except VII to IX) are computed across the 234,896 firm-destination-year triplets included in our sample. Pairwise correlations involving one of the specifications that includes country-year fixed effects are computed across the 7,937 firm-destination-year triplets in our sample that correspond to destination-year pairs in which at least one exporting firm.

C Moment Inequalities: Details

C.1 Proof of Proposition 1

Equation (15) implies

$$\mathbb{E}\Big[\Pi_{io_{it}t,L_{it}}\Big|\mathcal{J}_{it}\Big] \geq \mathbb{E}\Big[\Pi_{io_{it}^{j\to j'}t,L_{it}}\Big|\mathcal{J}_{it}\Big].$$

Thus, given the definition of $\Pi_{ibt,L_{it}}$ for any bundle b in equation (16), we know that

$$\mathbb{E}\Big[\pi_{io_{it}t} + \sum_{l=1}^{L_{it}} \delta^l \pi_{io_{it+l}(o_{it})t+l}\Big|\mathcal{J}_{it}\Big] \geq \mathbb{E}\Big[\pi_{io_{it}^{j\to j'}t} + \sum_{l=1}^{L_{it}} \delta^l \pi^{j\to j'}_{io_{it+l}(o_{it}^{j\to j'})t+l}\Big|\mathcal{J}_{it}\Big], \tag{C.1}$$

where

$$\pi^{j\to j'}_{io_{it+l}(o_{it}^{j\to j'})t+l}$$

denotes the static profits in year $t+l$ of a firm that chose $o_{it}^{j\to j'}$ in period t but that selected optimally its set of export destinations in every subsequent period; specifically, $o_{it+l}(o_{it}^{j\to j'})$ denotes the optimal export bundle in period $t+l$ conditional on having exported to bundle $o_{it}^{j\to j'}$ in period t. Similarly,

$$\mathbb{E}\Big[\pi_{io_{it}^{j\to j'}t} + \sum_{l=1}^{L_{it}} \delta^l \pi^{j\to j'}_{io_{it+l}(o_{it}^{j\to j'})t+l}\Big|\mathcal{J}_{it}\Big] \geq \mathbb{E}\Big[\pi_{io_{it}^{j\to j'}t} + \sum_{l=1}^{L_{it}} \delta^l \pi^{j\to j'}_{io_{it+l}(o_{it})t+l}\Big|\mathcal{J}_{it}\Big], \tag{C.2}$$

where

$$\pi^{j\to j'}_{io_{it+l}(o_{it})t+l},$$

denotes the static profits in year $t+l$ of a firm that chose $o_{it}^{j\to j'}$ in period t but that, in every subsequent period, selected the export bundle that would have been optimal if it had exported to o_{it} in period t instead; specifically, $o_{it+l}(o_{it})$ denotes the optimal export bundle in period $t+l$ conditional on having exported to bundle o_{it} in period t.

Combining inequalities (C.1) and (C.2), we can derive the inequality

$$\mathbb{E}\Big[\pi_{io_{it}t} + \sum_{l=1}^{L_{it}} \delta^l \pi_{io_{it+l}(o_{it})t+l}\Big|\mathcal{J}_{it}\Big] \geq \mathbb{E}\Big[\pi_{io_{it}^{j\to j'}t} + \sum_{l=1}^{L_{it}} \delta^l \pi^{j\to j'}_{io_{it+l}(o_{it})t+l}\Big|\mathcal{J}_{it}\Big].$$

As long as $L_{it} \geq 1$, the one-period state dependence of our model (see Section 3.3 for a discussion of this property) implies that we can rewrite this inequality as a sum over static profits in periods t and $t+1$ only,

$$\mathbb{E}\Big[\pi_{io_{it}t} + \delta\pi_{io_{it+1}(o_{it})t+1}\Big|\mathcal{J}_{it}\Big] \geq \mathbb{E}\Big[\pi_{io_{it}^{j\to j'}t} + \delta\pi^{j\to j'}_{io_{it+1}(o_{it})t+1}\Big|\mathcal{J}_{it}\Big].$$

According to the definition of the static profits π_{ibt} in equation (14), we can rewrite this expression as

$$\mathbb{E}\Big[\sum_{j''\in o_{it}} \pi_{ij''t} + \delta\sum_{j''=1}^{J} d_{ij''t+1}\pi_{ij''t+1}\Big|\mathcal{J}_{it}\Big] \geq \mathbb{E}\Big[\sum_{j''\in o_{it}^{j\to j'}} \pi_{ij''t} + \delta\sum_{j''=1}^{J} d_{ij''t+1}\pi^{j\to j'}_{ij''t+1}\Big|\mathcal{J}_{it}\Big],$$

where, as indicated in Section 4.1,

$$\pi^{j\to j'}_{ij''t+1}$$

denotes the static profits of exporting to destination j'' in period $t+1$ for a firm that exported to the bundle $o_{it}^{j\to j'}$ in period t. Reorganizing terms in the prior inequality, we obtain

$$\mathbb{E}\Big[\sum_{j''\in o_{it}} \pi_{ij''t} - \sum_{j''\in o_{it}^{j\to j'}} \pi_{ij''t} + \delta\sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi^{j\to j'}_{ij''t+1})\Big|\mathcal{J}_{it}\Big] \geq 0,$$

and, taking into account that the bundles o_{it} and $o_{it}^{j\to j'}$ differ only in that destination j is swapped by destination j', this prior inequality simplifies to

$$\mathbb{E}\Big[\pi_{ijt} - \pi_{ij't} + \delta \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi_{ij''t+1}^{j\to j'})\Big| d_{ijt}(1-d_{ijt}) = 1, \mathcal{J}_{it}\Big] \geq 0.$$

According to Assumption 3, $Z_{it} \subseteq \mathcal{J}_{it}$, and, thus, applying the Law of Iterated Expectations, we can derive the following inequality

$$\mathbb{E}\Big[\pi_{ijt} - \pi_{ij't} + \delta \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi_{ij''t+1}^{j\to j'})\Big| d_{ijt}(1-d_{ijt}) = 1, Z_{it}\Big] \geq 0,$$

which is identical to equation (23) in Proposition 1 with

$$\pi_{ijj't} \equiv \pi_{ijt} - \pi_{ij't},$$
$$\pi_{ijj't+1} \equiv \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi_{ij''t+1}^{j\to j'}). \quad \blacksquare$$

C.2 Proof of Proposition 2

From the conditional moment inequality in equation (23) and given that the vector $(Z_{ijt}, Z_{ij't})$ is a subvector of the vector Z_{it} (see equation (17)), we can derive the unconditional moment

$$\mathbb{E}\big[\Psi(Z_{ijt}, Z_{ij't})d_{ijt}(1-d_{ij't})(\pi_{ijj't} + \delta\pi_{ijj't+1})\big] \geq 0,$$

for any specific pair of countries j and j' such that $o_{it}^{j\to j'}$ belongs to the consideration set of the firm $\mathcal{B}_{it}$ and any positively valued function $\Psi(\cdot)$.

Given the definition of the sets $\mathcal{A}_{it}$ and $\mathcal{A}_{ijt}$ in equation (18), the condition that destination j' belongs to the set $\mathcal{A}_{ijt}$ is sufficient for the bundle

$$o_{it}^{j\to j'}$$

to belong to $\mathcal{B}_{it}$. Therefore, we can conclude that the conditional moment inequality in equation (23) implies that

$$\mathbb{E}\big[\Psi(Z_{ijt}, Z_{ij't})d_{ijt}(1-d_{ij't})(\pi_{ijj't} + \delta\pi_{ijj't+1})\big] \geq 0$$

holds for every pair of countries j and j' such that $j' \in \mathcal{A}_{ijt}$. Summing these inequalities across all pairs of countries j and $j' \in \mathcal{A}_{ijt}$, we obtain

$$\sum_{j=1}^{J} \sum_{j'\in\mathcal{A}_{ijt}} \mathbb{E}\big[\Psi(Z_{ijt}, Z_{ij't})d_{ijt}(1-d_{ij't})(\pi_{ijj't} + \delta\pi_{ijj't+1})\big] \geq 0,$$

and, finally, switching the order of the summatories and the expectation operator yields the inequality in equation (25). $\blacksquare$

C.3 Proof of Proposition 3

From the expression for the static export profits in equation (13), we can rewrite the difference in periods t and $t+1$ static profits as:

$$\pi_{ijt} - \pi_{ij't} + \delta \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi_{ij''t+1}^{j\to j'}) =$$
$$\eta^{-1}r_{ijt} - f_{ijt} - (1-d_{ijt-1})s_{ijt} - \eta^{-1}r_{ij't} + f_{ij't} + (1-d_{ij't-1})s_{ij't}+$$

$$\delta d_{ijt+1} s^{j\to j'}_{ijt+1} - \delta d_{ij't+1} s_{ij't+1} - \delta \sum_{\substack{j''\neq j,\\ j''\neq j'}} d_{ij''t+1}(1-d_{ij''t})(s_{ij''t+1} - s^{j\to j'}_{ij''t+1}),$$

where $s^{j\to j'}_{ij''t+1}$ denotes the potential sunk costs of exporting to country j'' in period $t+1$ for firm i if this one were to swap destination j by destination j' in year t. Given the expressions for export revenues and fixed and sunk export costs in equations (4), (6) and (9), we can rewrite the difference in periods t and $t+1$ static profits as:

$$\begin{gathered}
\pi_{ijt} - \pi_{ij't} + \delta \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi^{j\to j'}_{ij''t+1}) = \\
\eta^{-1} r^o_{ijt} + \eta^{-1}\varepsilon^R_{ijt} - f^o_j - u_{ic_jt} - \varepsilon^F_{ijt} - (1-d_{ijt-1})(s^o_j - e_{ijt} + \varepsilon^S_{ijt}) - \\
\eta^{-1} r^o_{ij't} - \eta^{-1}\varepsilon^R_{ij't} + f^o_{j'} + u_{ic_{j'}t} + \varepsilon^F_{ij't} + (1-d_{ij't-1})(s^o_{j'} - e_{ij't} + \varepsilon^S_{ij't}) + \\
\delta d_{ijt+1}(s^o_j - e^{o,j\to j'}_{ijt+1} + \varepsilon^S_{ijt+1}) - \delta d_{ij't+1}(s^o_{j'} - e^o_{ij't+1} + \varepsilon^S_{ij't+1}) - \\
\delta \sum_{\substack{j''\neq j,\\ j''\neq j'}} d_{ij''t+1}(1-d_{ij''t})(s^o_{j''} - e^o_{ij''t+1} + \varepsilon^S_{ij''t+1} - s^o_{j''} + e^{o,j\to j'}_{ij''t+1} - \varepsilon^S_{ij''t+1}),
\end{gathered}$$

and, canceling terms, we obtain

$$\begin{gathered}
\pi_{ijt} - \pi_{ij't} + \delta \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi^{j\to j'}_{ij''t+1}) = \\
\eta^{-1} r^o_{ijt} + \eta^{-1}\varepsilon^R_{ijt} - f^o_j - u_{ic_jt} - \varepsilon^F_{ijt} - (1-d_{ijt-1})(s^o_j - e^o_{ijt} + \varepsilon^S_{ijt}) - \\
\eta^{-1} r^o_{ij't} - \eta^{-1}\varepsilon^R_{ij't} + f^o_{j'} + u_{ic_{j'}t} + \varepsilon^F_{ij't} + (1-d_{ij't-1})(s^o_{j'} - e^o_{ij't} + \varepsilon^S_{ij't}) + \\
\delta\Big(d_{ijt+1}(s^o_j - e^{o,j\to j'}_{ijt+1} + \varepsilon^S_{ijt+1}) - d_{ij't+1}(s^o_{j'} - e^o_{ij't+1} + \varepsilon^S_{ij't+1}) + \sum_{\substack{j''\neq j,\\ j''\neq j'}} d_{ij''t+1}(1-d_{ij''t})(e^o_{ij''t+1} - e^{o,j\to j'}_{ij''t+1})\Big).
\end{gathered}$$

In order to derive this expression, we have not imposed yet any of the restrictions in Proposition 3. We impose first the restriction that the alternative destination j' must belong to the set $\mathcal{A}_{ijt}$ defined in equation (18b). If we select the counterfactual destination j' in this way, the difference in periods t and $t+1$ static profits simplifies to

$$\begin{gathered}
\pi_{ijt} - \pi_{ij't} + \delta \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi^{j\to j'}_{ij''t+1}) = \\
\eta^{-1} r^o_{ijt} + \eta^{-1}\varepsilon^R_{ijt} - \varepsilon^F_{ijt} - (1-d_{ijt-1})(s^o_j - e^o_{ijt} + \varepsilon^S_{ijt}) - \\
\eta^{-1} r^o_{ij't} - \eta^{-1}\varepsilon^R_{ij't} + \varepsilon^F_{ij't} + (1-d_{ij't-1})(s^o_{j'} - e^o_{ij't} + \varepsilon^S_{ij't}) + \\
\delta\Big(d_{ijt+1}(s^o_j - e^{o,j\to j'}_{ijt+1} + \varepsilon^S_{ijt+1}) - d_{ij't+1}(s^o_{j'} - e^o_{ij't+1} + \varepsilon^S_{ij't+1}) + \sum_{\substack{j''\neq j,\\ j''\neq j'}} d_{ij''t+1}(1-d_{ij''t})(e^o_{ij''t+1} - e^{o,j\to j'}_{ij''t+1})\Big).
\end{gathered}$$

Therefore, the restriction that destination j' must belong to the set $\mathcal{A}_{ijt}$ defined in equation (18b) implies that the resulting difference in static profits will depend neither on the observable parts of fixed export costs, f^o_j and $f^o_{j'}$, nor on the unobserved fixed cost components u_{ic_jt} and $u_{ic_{j'}t}$. If we additionally impose the restriction that both destinations j and j' must share neither continent, nor language nor similar income per capita with Chile (i.e. $s^o_j = s^o_{j'} = \gamma^S_{all}$, as imposed in equation (26)), the expression for the difference in in periods t and $t+1$ static profits simplifies to

$$\begin{gathered}
\pi_{ijt} - \pi_{ij't} + \delta \sum_{j''=1}^{J} d_{ij''t+1}(\pi_{ij''t+1} - \pi^{j\to j'}_{ij''t+1}) = \\
\eta^{-1} r^o_{ijt} + \eta^{-1}\varepsilon^R_{ijt} - \varepsilon^F_{ijt} - (1-d_{ijt-1})(\gamma^S_{all} - e^o_{ijt} + \varepsilon^S_{ijt}) -
\end{gathered}$$

$$\eta^{-1} r^o_{ij't} - \eta^{-1}\varepsilon^R_{ij't} + \varepsilon^F_{ij't} + (1-d_{ij't-1})(\gamma^S_{all} - e^o_{ij't} + \varepsilon^S_{ij't}) +$$
$$\delta\Big(d_{ijt+1}(\gamma^S_{all} - e^{o,j\to j'}_{ijt+1} + \varepsilon^S_{ijt+1}) - d_{ij't+1}(\gamma^S_{all} - e^o_{ij't+1} - \varepsilon^S_{ij't+1}) + \sum_{\substack{j''\neq j,\\ j''\neq j'}} d_{ij''t+1}(1-d_{ij''t})(e^o_{ij''t+1} - e^{o,j\to j'}_{ij''t+1})\Big).$$

Therefore, the restriction in equation (26) that destinations j and j' must verify $s^o_j = s^o_{j'} = \gamma^S_{all}$ implies that the resulting difference in static profits depends on the parameter vector $(\gamma^S_0, \gamma^S_c, \gamma^S_l, \gamma^S_g)$ introduced in equation (10) only through the function $\gamma^S_{all} \equiv \gamma^S_0 + \gamma^S_c + \gamma^S_l + \gamma^S_g$ introduced in equation (21).

Given the last expression for the difference in periods t and $t+1$ static profits and the following mean independence restriction

$$\mathbb{E}\left[\Psi(Z_{ijt}, Z_{ij't}) d_{ijt}(1-d_{ij't}) \quad \times \quad \begin{pmatrix} \eta^{-1}\varepsilon^R_{ijt} \\ \eta^{-1}\varepsilon^R_{ij't} \\ \varepsilon^F_{ijt} \\ \varepsilon^F_{ij't} \\ (1-d_{ijt-1})\varepsilon^S_{ijt} \\ (1-d_{ij't-1})\varepsilon^S_{ij't} \\ d_{ijt+1}\varepsilon^S_{ijt+1} \\ d_{ij't+1}\varepsilon^S_{ij't+1} \end{pmatrix}\right] = 0 \tag{C.3}$$

we can rewrite the moment in equation (27) as

$$\mathbb{E}\Bigg[\sum_{j=1}^{J}\sum_{j'\in\mathcal{A}_{ijt}} \Psi(Z_{ijt}, Z_{ij't}) d_{ijt}(1-d_{ij't})\Big(\eta^{-1}(r^o_{ijt} - r^o_{ij't}) - (1-d_{ijt-1})(\gamma^S_{all} - e^o_{ijt}) + (1-d_{ij't-1})(\gamma^S_{all} - e^o_{ij't})$$
$$+\delta(d_{ijt+1}(\gamma^S_{all} - e^{o,j\to j'}_{ijt+1}) - d_{ij't+1}(\gamma^S_{all} - e^o_{ij't+1})) + \delta\Big(\sum_{\substack{j''\neq j,\\ j''\neq j'}} d_{ij''t+1}(1-d_{ij''t})(e^o_{ij''t+1} - e^{o,j\to j'}_{ij''t+1})\Big)\Big)\Bigg]. \tag{C.4}$$

The mean independence conditions in equation (C.3) are implied by assumptions 1 and 3, the definition of Z_{ijt} in equation (17b), and the mean independence conditions in equations (5), (8), and (12). Equation (C.4) is still not expressed as a function of the parameter vector $(\alpha, \kappa, \tilde{\eta}, \gamma^S_{all})$. To do so, one has to realize that, according to equation (11), for any firm, country and time period, the function e^o_{ijt} is linear in the parameter vector $(\gamma^E_b, \gamma^E_c, \gamma^E_l, \gamma^E_g)$ and, therefore, we can define a function

$$\tilde{e}^o_{ijt} \equiv \frac{1}{\gamma^S_{all}} e^o_{ijt}, \tag{C.5}$$

that is linear in the parameter vector κ defined in equation (21). In words, $\tilde{e}^o_{ijt}$ denotes the extended gravity component e^o_{ijt} normalized by the sunk cost of exporting to a country that shares neither continent, nor language, nor similar income per capita with the firm's home market, γ^S_{all}. Using the equality in equation (C.5), we can rewrite the moment in equation (C.4) as

$$\mathbb{E}\Bigg[\sum_{j=1}^{J}\sum_{j'\in\mathcal{A}_{ijt}} \Psi(Z_{ijt}, Z_{ij't}) d_{ijt}(1-d_{ij't})\gamma^S_{all}\Big(\tilde{\eta}^{-1}(r^o_{ijt} - r^o_{ij't}) - (1-d_{ijt-1})(1 - \tilde{e}^o_{ijt}) + (1-d_{ij't-1})(1 - \tilde{e}^o_{ij't})$$
$$+\delta(d_{ijt+1}(1 - \tilde{e}^{o,j\to j'}_{ijt+1}) - d_{ij't+1}(1 - \tilde{e}^o_{ij't+1})) + \delta\Big(\sum_{\substack{j''\neq j,\\ j''\neq j'}} d_{ij''t+1}(1-d_{ij''t})(\tilde{e}^o_{ij''t+1} - \tilde{e}^{o,j\to j'}_{ij''t+1})\Big)\Big)\Bigg], \tag{C.6}$$

where $\tilde{\eta} \equiv \eta\gamma^S_{all}$. As indicated in Proposition 3, this moment is a function of the parameter vector α (through r^o_{ijt} and $r^o_{ij't}$), of the parameter vector κ (through $\tilde{e}^o_{ijt}$, $\tilde{e}^o_{ij't}$, $\tilde{e}^{o,j\to j'}_{ijt+1}$, $\tilde{e}^o_{ij't+1}$ and, for every j'' distinct from j and j', $\tilde{e}^o_{ij''t+1}$ and $\tilde{e}^{o,j\to j'}_{ij''t+1}$) and of the scalar parameters $\tilde{\eta}$ and γ^S_{all}. Furthermore, it is multiplicative in γ^S_{all}. ■

C.4 Specifying Moments: Details

The instrument functions we use to compute the results in sections 6 and 7 are:

$$
\begin{aligned}
\Psi_1(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. Border})_j = 1, (\text{Ext. Grav. Border})_{j'} = 0\},\\
\Psi_2(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. Border})_j = 0, (\text{Ext. Grav. Border})_{j'} = 1\},\\
\Psi_3(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. Cont.})_j = 1, (\text{Ext. Grav. Cont.})_{j'} = 0\},\\
\Psi_4(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. Cont.})_j = 0, (\text{Ext. Grav. Cont.})_{j'} = 1\},\\
\Psi_5(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. Lang.})_j = 1, (\text{Ext. Grav. Lang.})_{j'} = 0\},\\
\Psi_6(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. Lang.})_j = 0, (\text{Ext. Grav. Lang.})_{j'} = 1\},\\
\Psi_7(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. GDPpc})_j = 1, (\text{Ext. Grav. GDPpc})_{j'} = 0\},\\
\Psi_8(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. GDPpc})_j = 0, (\text{Ext. Grav. GDPpc})_{j'} = 1\},\\
\Psi_9(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = 1, d_{ij't-1} = 0\},\\
\Psi_{10}(Z_{ijt}, Z_{ij't}) &= \mathbb{1}\{s^o_j = s^o_{j'} = \gamma^S_{all}, d_{ijt-1} = 0, d_{ij't-1} = 1\}.
\end{aligned}
$$

Our instrument functions are thus indicator functions that take value one only when a set of restrictions on the characteristics of the actual and counterfactual export destinations of a firm are met.

All ten instrument functions include the restriction that $s^o_j = s^o_{j'} = \gamma^S_{all}$: neither the actual nor the counterfactual destinations are in South America, have Spanish as official language, or have similar income per capita to Chile. In other terms, neither country j nor country j' share any gravity characteristic with the country of origin of the firms. This condition implies that the resulting moment inequalities depend on the vector of sunk costs parameters $(\gamma^S_0, \gamma^S_c, \gamma^S_l, \gamma^S_g)$ only through the parameter γ^S_{all}, defined in equation (21) as $\gamma^S_{all} \equiv \gamma^S_0 + \gamma^S_c + \gamma^S_l + \gamma^S_g$. Furthermore, as equation (7) shows, the condition that $s^o_j = s^o_{j'}$ implies that $f^o_j = f^o_{j'}$ and, thus, the resulting moment inequalities will not depend on the fixed costs parameters $(\gamma^F_0, \gamma^F_c, \gamma^F_l, \gamma^F_g)$.

All other restrictions included in the instrument functions $\Psi_1(\cdot)$ to $\Psi_{10}(\cdot)$ have the purpose of creating moments that identify upper and lower bounds for the parameter vector $(\kappa, \tilde{\eta})$. For example, the function $\Psi_1(Z_{ijt}, Z_{ij't})$ takes value one only for pairs of actual and counterfactual destinations such that firm i benefits in period t from extended gravity due to border in the actual destination j but not in the counterfactual destination j'. The resulting difference in profits $\pi_{ijj't} + \delta\pi_{ijj't+1}$ is thus likely to depend negatively on the parameter that determines the size of the reduction in sunk costs due to extended gravity effects in border, κ_b. Consequently, the moment inequality defined by the instrument function $\Psi_1(Z_{ijt}, Z_{ij't})$ is likely to depend negatively on κ_b and, thus, this inequality will help identify an upper bound on κ_b. Conversely, the moment inequality defined by the instrument function $\Psi_2(Z_{ijt}, Z_{ij't})$ is likely to depend positively on κ_b and, thus, this inequality will help identify a lower bound on κ_b.

The functions $\Psi_3(Z_{ijt}, Z_{ij't})$ and $\Psi_4(Z_{ijt}, Z_{ij't})$ are analogous to $\Psi_1(Z_{ijt}, Z_{ij't})$ and $\Psi_2(Z_{ijt}, Z_{ij't})$ for the case of the extended gravity parameter due to continent, κ_c. The functions $\Psi_5(Z_{ijt}, Z_{ij't})$ and $\Psi_6(Z_{ijt}, Z_{ij't})$ play the same role for the case of extended gravity in language, κ_l, and the functions $\Psi_7(Z_{ijt}, Z_{ij't})$ and $\Psi_8(Z_{ijt}, Z_{ij't})$ do so for extended gravity effects due to sharing income per capita, κ_g.

The instrument functions $\Psi_1(Z_{ijt}, Z_{ij't})$ to $\Psi_8(Z_{ijt}, Z_{ij't})$ select observations for which actual and counterfactual export paths imply the same number of export entries; i.e. the same number of countries to which the firm i export in period t without exporting to them in period $t-1$. The instrument functions $\Psi_9(Z_{ijt}, Z_{ij't})$ to $\Psi_{10}(Z_{ijt}, Z_{ij't})$ alter the total number of foreign market entries. Specifically, in all cases in which $\Psi_9(Z_{ijt}, Z_{ij't})$ equals one, the firm enters the counterfactual destination j' (because $d_{ij't-1} = 0$) while, on the actual path, it was not entering the actual destination j (because $d_{ijt-1} = 1$). The opposite is true in those cases in which $\Psi_{10}(Z_{ijt}, Z_{ij't})$ equals one.

C.5 Using Inequalities to Bound Parameters: Extra Example

As mentioned in Section 4.3, when swapping an observed destination by an alternative one in a year t, one needs to keep track of how this change affects, through extended gravity effects, the sunk costs in any other country to which the firm starts exporting in year $t+1$.

To illustrate, we consider in Table C.1 an example similar to that in Table 3, but in which the firm starts exporting to Australia in year 9. In this example, the difference in static profits $\pi_{ijj'8}$ is identical to that in equation (29), the difference in export profits to the United Kingdom in year 9 equals that in equation (35), but the difference in export profits to Australia, indexed here by j'', in year 9 now equals

$$d_{ij''9}(\pi_{ij''9} - \pi_{ij''9}^{j \to j'}) = -\gamma_l^E, \tag{C.7}$$

reflecting the fact that Australia shares language with the firm's observed destination in year 8, the United Kingdom, but not with its alternative one, Germany. Thus, the difference in profits between the actual and counterfactual paths described in Table C.1 is

$$\begin{aligned} &\pi_{ijj'8} + \delta\pi_{ijj'9} = \\ &\gamma_{all}^S(\tilde{\eta}^{-1}(r_{ij8}^o - r_{ij'8}^o + \varepsilon_{ij8}^R - \varepsilon_{ij'8}^R) + \kappa_l + \delta(1 - \kappa_c - \kappa_g - \kappa_l)) - \varepsilon_{ij8}^F + \varepsilon_{ij'8}^F - \varepsilon_{ij8}^S + \varepsilon_{ij'8}^S + \delta\varepsilon_{ij9}^S. \end{aligned} \tag{C.8}$$

Table C.1: Example of a 1-period Export Event

		$t=7$	$t=8$	$t=9$
Observed	United Kingdom	0	1	1
	Germany	0	0	0
	Australia	0	0	1
Alternative	United Kingdom	0	0	1
	Germany	0	1	0
	Australia	0	0	1

C.6 Inference: Details

C.6.1 Confidence Set for the True Parameter

We describe here the procedure we follow to compute the confidence set for the true parameter vector $(\kappa, \tilde{\eta})$. This procedure implements the asymptotic version of the Generalized Moment Selection (GMS) test described on page 135 of Andrews and Soares (2010), corrected following the procedure described in Section 10.2 of the same paper to account for the fact that our moments depend on a consistent and asymptotically normal preliminary estimator $\hat{\alpha}$.

We base our confidence set on the modified method of moments (MMM) statistic. Specifically, we index the finite set of inequalities that we use for estimation by $k = 1, \ldots, K$ and denote them as

$$\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) \geq 0, \qquad k = 1, \ldots, K,$$

where θ_κ denotes the unknown parameter vector whose true value is κ and $\theta_{\tilde{\eta}}$ denotes the unknown scalar parameter whose true value is $\tilde{\eta}$. For each $k = 1, \ldots, K$, the moment function $\bar{m}_k(\cdot)$ is the sample analogue of the moment in equation (28) after normalizing it by γ_{all}^S:

$$\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) = \frac{1}{N}\left\{\sum_{t=1}^{T}\sum_{i=1}^{N_t}\left\{\sum_{j=1}^{J}\sum_{j' \in \mathcal{A}_{ijt}} \Psi_k(Z_{ijt}, Z_{ij't}) d_{ijt}(1 - d_{ij't})(1/\gamma_{all}^S)(\pi_{ijj't} + \delta\pi_{ijj't+1})\right\}\right\},$$

where, conditional on equation $\Psi_k(Z_{ijt}, Z_{ij't})$ satisfying the restrictions in equation (24) and (26) and $\mathcal{A}_{ijt}$ being defined as in equation (18b), the true difference in static profits normalized by γ_{all}^S, $(1/\gamma_{all}^S)(\pi_{ijj't} + \delta\pi_{ijj't+1})$, is exclusively a function of the parameter vector $(\alpha, \kappa, \tilde{\eta})$. The number $N = \sum_{t=1}^{T} N_t$ denotes the total number of observations. To simplify the notation for the rest of this section, we define

$$m_{k,it}(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) \equiv \sum_{j=1}^{J}\sum_{j' \in \mathcal{A}_{ijt}} \Psi_k(Z_{ijt}, Z_{ij't}) d_{ijt}(1 - d_{ij't})(1/\gamma_{all}^S)(\pi_{ijj't} + \delta\pi_{ijj't+1}), \tag{C.9}$$

and, thus, for every $k = 1, \ldots, K$, we rewrite $\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}})$ as

$$\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) = \frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N_t} m_{k,it}(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}), \tag{C.10}$$

The MMM statistic is therefore defined as

$$Q(\theta) = \sum_{k=1}^{K} \Big(\min \Big\{ \frac{\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}})}{\hat{\sigma}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}})}, 0 \Big\} \Big)^2, \tag{C.11}$$

where

$$\hat{\sigma}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) = \Big(\frac{1}{N} \sum_{t=1}^{T} \sum_{i=1}^{N_t} \big(m_{k,it}(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) - \bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) \big)^2 \Big)^{\frac{1}{2}}. \tag{C.12}$$

The total number of moment inequalities employed for identification, K, will depend on the finite number of unconditional moment inequalities that we derive from the conditional revealed-preference moment inequalities described in equation (23); Appendix C.4 contains additional details on the unconditional moments that we employ. Given the set of unconditional moment inequalities $k = 1, \ldots, K$ and the test statistic in equation (C.11), we compute confidence sets for the true parameter value $(\kappa, \tilde{\eta})$ using the following steps:

Step 1: define a grid Θ_g that will contain the confidence set. We define this grid as an orthotope with as many dimensions as there are scalars in the parameter vector $(\theta_\kappa, \theta_{\tilde{\eta}})$. In the case of the confidence set for the parameter vector $(\kappa_b, \kappa_c, \kappa_l, \kappa_g, \tilde{\eta})$, $(\theta_\kappa, \theta_{\tilde{\eta}})$ is a 5-dimensional orthotope. To define the limits of this 5-dimensional orthotope, we solve the following nonlinear optimization

$$min_{(\theta_\kappa, \theta_{\tilde{\eta}})}\ \mathbb{d} \cdot (\theta_\kappa, \theta_{\tilde{\eta}}) \qquad \text{subject to} \qquad \bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) + \ln N \geq 0, \tag{C.13}$$

where $\mathbb{d}$ is one of the elements of the matrix

$$\mathbb{D} = (\mathbb{d}_{1+}, \mathbb{d}_{1-}, \mathbb{d}_{2+}, \mathbb{d}_{2-}, \mathbb{d}_{3+}, \mathbb{d}_{3-}, \mathbb{d}_{4+}, \mathbb{d}_{4-}, \mathbb{d}_{5+}, \mathbb{d}_{5-})' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Given that $\mathbb{D}$ has ten elements, we will therefore solve ten nonlinear optimizations like that in equation (C.13). Denote the six 10-dimensional vectors $(\theta_\kappa, \theta_{\tilde{\eta}})$ that solve each of these optimizations as

$$((\theta_\kappa, \theta_{\tilde{\eta}})_{1+}, (\theta_\kappa, \theta_{\tilde{\eta}})_{1-}, (\theta_\kappa, \theta_{\tilde{\eta}})_{2+}, (\theta_\kappa, \theta_{\tilde{\eta}})_{2-}, (\theta_\kappa, \theta_{\tilde{\eta}})_{3+}, (\theta_\kappa, \theta_{\tilde{\eta}})_{3-}, (\theta_\kappa, \theta_{\tilde{\eta}})_{4+}, (\theta_\kappa, \theta_{\tilde{\eta}})_{4-}, (\theta_\kappa, \theta_{\tilde{\eta}})_{5+}, (\theta_\kappa, \theta_{\tilde{\eta}})_{5-})',$$

and compute the ten boundaries of the 5-dimensional orthotope Θ_g as

$$\begin{pmatrix} \mathbb{d}_{1+} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{1+} & \mathbb{d}_{1-} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{1-} \\ \mathbb{d}_{2+} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{2+} & \mathbb{d}_{2-} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{2-} \\ \mathbb{d}_{3+} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{3+} & \mathbb{d}_{3-} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{3-} \\ \mathbb{d}_{4+} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{4+} & \mathbb{d}_{4-} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{4-} \\ \mathbb{d}_{5+} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{5+} & \mathbb{d}_{5-} \cdot (\theta_\kappa, \theta_{\tilde{\eta}})_{5-} \end{pmatrix},$$

where the first column contains the minimum value of the element of $(\theta_\kappa, \theta_{\tilde{\eta}})$ indicated by the corresponding row and the second column contains the corresponding maximum. Once we have these ten limits of the 5-dimensional orthotope Θ_g we fill it up with 400,000 equidistant points.

Step 2: choose a point $\theta_p \in \Theta_g$. The following steps will test the null hypothesis that the vector θ_p is identical to the true value of θ:

$$H_0 : (\kappa, \tilde{\eta}) = \theta_p \qquad \text{vs.} \qquad H_1 : (\kappa, \tilde{\eta}) \neq \theta_p.$$

Step 3: evaluate the MMM test statistic at θ_p:

$$Q(\hat{\alpha}, \theta_p) = \sum_{k=1}^{K} \Big(\min \Big\{ \frac{\bar{m}_k(\hat{\alpha}, \theta_p)}{\hat{\sigma}_k(\hat{\alpha}, \theta_p)}, 0 \Big\} \Big)^2, \tag{C.14}$$

where $\bar{m}_k(\hat{\alpha}, \theta_p)$ is equal to the moment $\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}})$ evaluated at $(\theta_\kappa, \theta_{\tilde{\eta}}) = \theta_p$; and analogously for $\hat{\sigma}_k(\hat{\alpha}, \theta_p)$.

Step 4: compute the correlation matrix of moments evaluated at θ_p:

$$\hat{\Omega}(\hat{\alpha}, \theta_p) = Diag^{-\frac{1}{2}}(\hat{\Sigma}(\hat{\alpha}, \theta_p))\hat{\Sigma}(\hat{\alpha}, \theta_p) Diag^{-\frac{1}{2}}(\hat{\Sigma}(\hat{\alpha}, \theta_p)), \tag{C.15}$$

where $Diag(\hat{\Sigma}(\hat{\alpha}, \theta_p))$ is the $K \times K$ diagonal matrix whose diagonal elements are equal to those of $\hat{\Sigma}(\hat{\alpha}, \theta_p)$, $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\hat{\alpha}, \theta_p))$ is a matrix such that $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\hat{\alpha}, \theta_p)) Diag^{-\frac{1}{2}}(\hat{\Sigma}(\hat{\alpha}, \theta_p)) = Diag^{-1}(\hat{\Sigma}(\hat{\alpha}, \theta_p))$ and $\hat{\Sigma}(\hat{\alpha}, \theta_p)$ is an asymptotic variance matrix computed to reflect the fact that $\hat{\alpha}$ has been previously estimated. Specifically, the variance matrix $\hat{\Sigma}(\hat{\alpha}, \theta_p)$ is the sum of two terms:

$$\hat{\Sigma}(\hat{\alpha}, \theta_p) = \hat{\Sigma}_1(\hat{\alpha}, \theta_p) + \hat{\Sigma}_2(\hat{\alpha}, \theta_p), \tag{C.16}$$

where the first term takes into account the variance of the moments given $\hat{\alpha}$ and the second term takes into account how the variance of $\hat{\alpha}$ affects the variance of our moments. Generally, the variance matrix $\hat{\Sigma}(\hat{\alpha}, \theta_p)$ should include an extra term that takes into account the covariance between both sources of randomness; in our case, the observations being used to estimate $\hat{\alpha}$ do not coincide with the observations entering the moments $\{\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}), k = 1, \ldots, K\}$ and, thus, we set this third term to zero for simplicity. The first term in equation (C.16) equals:

$$\hat{\Sigma}_1(\hat{\alpha}, \theta_p) = \frac{1}{N} \sum_{i=1}^{I} \sum_{t=1}^{T} \big(m_{it}(\hat{\alpha}, \theta_p) - \bar{m}(\hat{\alpha}, \theta_p))(m_{it}(\hat{\alpha}, \theta_p) - \bar{m}(\hat{\alpha}, \theta_p))', \tag{C.17}$$

where both $m_{it}(\hat{\alpha}, \theta_p)$ and $\bar{m}(\hat{\alpha}, \theta_p)$ are $K \times 1$ vectors

$$\begin{aligned} m_{it}(\hat{\alpha}, \theta_p) &= (m_{1,it}(\hat{\alpha}, \theta_p), \ldots, m_{K,it}(\hat{\alpha}, \theta_p)), \\ \bar{m}(\hat{\alpha}, \theta_p) &= (\bar{m}_1(\hat{\alpha}, \theta_p), \ldots, \bar{m}_K(\hat{\alpha}, \theta_p)). \end{aligned}$$

Let's denote the element in row k and column k' of matrix $\hat{\Sigma}_2(\hat{\alpha}, \theta_p)$ as $\hat{\Sigma}_{2,kk'}(\hat{\alpha}, \theta_p)$. Then,

$$\hat{\Sigma}_{2,kk'}(\hat{\alpha}, \theta_p) = \frac{1}{N} \sum_{i=1}^{I} \sum_{t=1}^{T} \Big\{ \delta_{k,it}(\hat{\alpha}, \theta_p)' \Big(\frac{1}{N_{obs}} \sum_{i'=1}^{I} \sum_{t'=1}^{T} \sum_{j'=1}^{J} d_{i'j't'} t_{i'j't'}(\hat{\alpha}) t_{i'j't'}(\hat{\alpha})' \Big) \delta_{k',it}(\hat{\alpha}, \theta_p) \Big\}, \tag{C.18}$$

where $\delta_{k,it}$ denotes the Jacobian of the function $m_{k,it}(\alpha, \theta_p)$ with respect to α:

$$\begin{aligned} \delta_{k,it} &= \frac{\partial m_{k,it}(\alpha, \theta_p)}{\partial \alpha} \bigg|_{\alpha = \hat{\alpha}} = \\ &\sum_{j=1}^{J} \sum_{j' \in \mathcal{A}_{ijt}} \Psi_k(Z_{ijt}, Z_{ij't}) d_{ijt}(1 - d_{ij't}) \tilde{\eta}^{-1} \big(X^r_{ijt} \exp(X^r_{ijt}\hat{\alpha}) - X^r_{ij't} \exp(X^r_{ij't}\hat{\alpha}) \big). \end{aligned} \tag{C.19}$$

This expression takes into account that the function $(1/\gamma^S_{all})\pi_{ijj't} + \delta\pi_{ijj't+1}$ depends on the parameter vector α only through the term $\tilde{\eta}^{-1}(\exp(X^r_{ijt}\hat{\alpha}) - \exp(X^r_{ij't}\hat{\alpha}))$; see equation (C.6) in Appendix C.3 for more details on the function $(1/\gamma^S_{all})\pi_{ijj't} + \delta\pi_{ijj't+1}$.

The function $t_{ijt}(\hat{\alpha})$ in equation (C.18) is such that we can write the following asymptotic expansion for our estimator of α:

$$(\hat{\alpha} - \alpha) = \frac{1}{N_{obs}} \sum_{i=1}^{I} \sum_{t=1}^{T} \sum_{j=1}^{J} t_{ijt}(\hat{\alpha}) + o_p(N_{obs}^{-\frac{1}{2}}),$$

where N_{obs} denotes the total number of firm-country-year triples for which we observe positive exports; i.e. $N_{obs} \equiv \sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{J} d_{ijt}$, which corresponds to the total number of observations used to compute the estimate $\hat{\alpha}$. Given that $\hat{\alpha}$ is a nonlinear least squares estimate of α (see Appendix B.2 for more details), the term $t_{ijt}(\hat{\alpha})$ is equal to

$$t_{ijt}(\hat{\alpha}) = \Big[\frac{1}{N_{obs}}\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{J}\big(X_{ijt}^{R}\exp(X_{ijt}^{R}\hat{\alpha})\big)\big(X_{ijt}^{R}\exp(X_{ijt}^{R}\hat{\alpha})\big)'\Big]^{-1}\big(X_{ijt}^{R}\exp(X_{ijt}^{R}\hat{\alpha})\big)\hat{\varepsilon}_{ijt}^{R}, \tag{C.20}$$

where $\hat{\varepsilon}_{ijt}^{R} \equiv r_{ijt}^{obs} - \exp(X_{ijt}^{R}\hat{\alpha})$. See page 153 in Cameron and Trivedi (2005) for additional details on the shape of the function $t_{ijt}(\hat{\alpha})$ for a nonlinear least squares estimator of α.

Putting together the expressions in equations (C.16) to (C.20), we construct $\hat{\Sigma}(\hat{\alpha}, \theta_p)$. Using equation (C.15), we construct the correlation matrix of moments evaluated at θ_p, $\hat{\Omega}(\hat{\alpha}, \theta_p)$.

Step 5: simulate the asymptotic distribution of $Q(\theta_p)$. Take $r = 1, \ldots, R$ draws from the multivariate normal distribution $\mathbb{N}(0_K, I_K)$ where 0_K is a vector of 0s of dimension K and I_K is the identity matrix of dimension K. Denote each of these draws as ζ_r. Define the simulated criterion function $Q_r(\theta_p)$ as

$$Q_r(\hat{\alpha}, \theta_p) = \sum_{k=1}^{K}\Big\{(\min\{[\hat{\Omega}^{\frac{1}{2}}(\hat{\alpha}, \theta_p)\zeta_r]_k, 0\})^2 \times \mathbb{1}\{\sqrt{N}\frac{\bar{m}_k(\hat{\alpha}, \theta_p)}{\hat{\sigma}_k(\hat{\alpha}, \theta_p)} \leq \sqrt{\ln N}\}\Big\}$$

where $[\hat{\Omega}^{\frac{1}{2}}(\hat{\alpha}, \theta_p)\zeta_r]_k$ is the kth element of the vector $\hat{\Omega}^{\frac{1}{2}}(\hat{\alpha}, \theta_p)\zeta_r$.

Step 6: compute the critical value. The critical value $\hat{c}(\hat{\alpha}, \theta_p, 1-\beta)$ is the $(1-\beta)$-quantile of the distribution of $Q_r(\hat{\alpha}, \theta_p)$ across the R draws taken in the previous step.

Step 7: accept/reject θ_p. Include θ_p in the estimated $(1-\beta)\%$ confidence set, $\hat{\Theta}^{1-\beta}$, if

$$Q(\hat{\alpha}, \theta_p) \leq \hat{c}(\hat{\alpha}, \theta_p, 1-\beta).$$

Step 8: repeat steps 2 to 7 for every θ_p in the grid Θ_g.

Step 9: compare the points included in the set $\hat{\Theta}^{1-\beta}$ to those in the set Θ_g. If (a) some of the points included in the set $\hat{\Theta}^{1-\beta}$ are at the boundary of the set Θ_g, expand the limits of Θ_g and repeat steps 2 to 9. If (b) the set of points included in $\hat{\Theta}^{1-\beta}$ is only a small fraction of those included in Θ_g, redefine a set Θ_g that is again a 5-dimensional orthotope whose limits are the result of adding a small number to the corresponding limits of the set $\hat{\Theta}^{1-\beta}$ and repeat steps 2 to 9. If neither (a) nor (b) applies, define $\hat{\Theta}^{1-\beta}$ as the 95% confidence set for $(\kappa, \tilde{\eta})$.

C.6.2 Specification Tests

We describe here the procedure we follow to test the validity of the model defined by our moment inequalities. This procedure implements the "test BP" and the "test RS", as described in Bugni et al. (2015). We use the same notation as in Appendix C.6.1 and thus describe the set of moment inequalities we use for identification as

$$\mathbb{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) \geq 0, \qquad k = 1, \ldots, K,$$

where θ_κ denotes the unknown parameter vector whose true value is κ and $\theta_{\tilde{\eta}}$ denotes the unknown scalar parameter whose true value is $\tilde{\eta}$. For each $k = 1, \ldots, K$, the moment $\mathbb{m}_k(\cdot)$ is defined in equation (28) as:

$$\mathbb{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}}) = \mathbb{E}\Big[\sum_{j=1}^{J}\sum_{j' \in \mathcal{A}_{ijt}}\Psi_k(Z_{ijt}, Z_{ij't})d_{ijt}(1-d_{ij't})(1/\gamma_{all}^{S})(\pi_{ijj't} + \delta\pi_{ijj't+1})\Big],$$

where the difference in profits $\pi_{ijj't} + \delta\pi_{ijj't+1}$ depends on the estimated vector $\hat{\alpha}$ and the parameter vector $(\theta_\kappa, \theta_{\tilde{\eta}})$. We denote the identified set as Θ_I, which, by definition, includes all values of the parameter vector $(\theta_\kappa, \theta_{\tilde{\eta}})$ consistent with these K moment inequalities. The model defined by these inequalities is said to be correctly specified (or statistically adequate) when Θ_I is non-empty. Thus, formally, we want to test

$$H_0 : \Theta_I \neq \emptyset \qquad \text{vs.} \qquad H_1 : \Theta_I = \emptyset.$$

Therefore, the null hypothesis in our specification test is that the model is correct; i.e. Θ_I is non-empty. The traditional approach to perform this test checks whether a confidence set for θ is empty: Bugni et al. (2015) denote this test as the "test BP". Bugni et al. (2015) suggest two additional specification tests that dominate the "test BP" in terms of power. They denote them "test RS" and "test RC". Among these two, we compute only the "test RS", as it has better power properties than the "test RC".

Test BP This test has been proposed by Romano and Shaikh (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010). This test arises as a by-product of the confidence sets described in Appendix C.6.1. Specifically, we reject the model in a test with size β if the $(1-\beta)\%$ confidence set for the true value of the parameter vector is empty (see Appendix C.6.1 for a description of how we compute such confidence set). As pointed out by Andrews and Guggenberger (2009), Andrews and Soares (2010), and Bugni et al. (2015), this test is conservative, i.e., if the model defined by moment inequalities being tested is correctly specified, asymptotically, the test with size β will actually reject the null hypothesis less than $100\beta\%$ of times. Additionally, Bugni et al. (2015) show that this test can have strictly less power than the "test RS".

Test RS This test has been proposed by Bugni et al. (2015). The steps we follow to compute this test are the following:

Step 1: define a grid Θ_g on the parameter space. This step is identical to that in Appendix C.6.1.

Step 2: evaluate the MMM test statistic defined in equation (C.11) **at every $\theta_p \in \Theta_g$:**

$$\{Q(\hat{\alpha}, \theta_p); \forall \theta_p \in \Theta_g\}$$

where

$$Q(\hat{\alpha}, \theta_p) = \sum_{k=1}^{K} \Big(\min \Big\{ \frac{\bar{m}_k(\hat{\alpha}, \theta_p)}{\hat{\sigma}_k(\hat{\alpha}, \theta_p)}, 0 \Big\} \Big)^2, \tag{C.21}$$

$\bar{m}_k(\hat{\alpha}, \theta_p)$ is equal to the moment $\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}})$ evaluated at $(\theta_\kappa, \theta_{\tilde{\eta}}) = \theta_p$, and analogously for $\hat{\sigma}_k(\hat{\alpha}, \theta_p)$. The functions $\bar{m}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}})$ and $\hat{\sigma}_k(\hat{\alpha}, \theta_\kappa, \theta_{\tilde{\eta}})$ are defined in equations (C.10) and (C.12).

Step 3: compute the minimum of the set of MMM test statistics computed in step 2:

$$T(\hat{\alpha}) = \inf_{\theta_p \in \Theta_g} Q(\hat{\alpha}, \theta_p). \tag{C.22}$$

Step 4: for every θ_p in Θ_g, compute the correlation matrix of the K moments, $\hat{\Omega}(\hat{\alpha}, \theta_p)$. For each θ_p, this step is identical to that in Appendix C.6.1.

Step 5: simulate the asymptotic distribution of $T(\hat{\alpha})$. Take $r = 1, \ldots, R$ draws from the multivariate normal distribution $\mathbb{N}(0_K, I_K)$ where 0_K is a vector of 0s of dimension K and I_K is the identity matrix of dimension K. Denote each of these draws as ζ_r. Conditional on a value θ_p in Θ_g, define the simulated criterion function corresponding to the draw ζ_r as

$$Q_r(\hat{\alpha}, \theta_p) = \sum_{k=1}^{K} \Big\{ (\min\{[\hat{\Omega}^{\frac{1}{2}}(\hat{\alpha}, \theta_p)\zeta_r]_k, 0\})^2 \times \mathbb{1}\{\sqrt{N} \frac{\bar{m}_k(\hat{\alpha}, \theta_p)}{\hat{\sigma}_k(\hat{\alpha}, \theta_p)} \leq \sqrt{\ln N}\} \Big\}$$

where $[\hat{\Omega}^{\frac{1}{2}}(\hat{\alpha}, \theta_p)\zeta_r]_k$ is the kth element of the vector $\hat{\Omega}^{\frac{1}{2}}(\hat{\alpha}, \theta_p)\zeta_r$. For each r, compute $Q_r(\hat{\alpha}, \theta_p)$ for every $\theta_p \in \Theta_g$:

$$\{Q_r(\hat{\alpha}, \theta_p); \forall \theta_p \in \Theta_g\},$$

and compute the minimum of all these values as

$$T_r(\hat{\alpha}) = \inf_{\theta_p \in \Theta_g} Q_r(\hat{\alpha}, \theta_p). \tag{C.23}$$

Step 6: compute critical value. The critical value $\hat{c}^{RS}(\hat{\alpha}, 1-\beta)$ is the $(1-\beta)$-quantile of the distribution of $T_r(\hat{\alpha})$ across the R draws taken in the previous step.

Step 7: accept/reject H_0. Reject H_0 if

$$T(\hat{\alpha}) > \hat{c}^{RS}(\hat{\alpha}, 1-\beta).$$

C.7 Additional Results

We report here results that rely on the "short" revenue regression described in column VI of Table B.1. We do so with the purpose of comparing the moment inequality estimates in tables C.2 and C.3 to those computed using the "long" revenue regression described in column I of Table B.1, reported in tables 4 and 5.

The "long" revenue regression includes: firm and year fixed effects; the firm's value added per worker, share and average wages of skilled and unskilled workers; a large set of distance measures between foreign and home countries; this same set interacted with a dummy for first year of exports to a country; extended gravity covariates; and a measure of the foreign market's size. The "short" revenue regression includes only: firm and year fixed effects; the firm's domestic sales; the physical distance between foreign and home countries; and the aggregate imports in the foreign market. Besides fixed effects, the "long" regression includes 28 regressors, while the "short" one only 3.

A comparison of tables C.2 and C.3 to tables 4 and 5 shows that, no matter which specification we use to construct $\hat{r}_{ijt}$, the moment inequality estimates are very similar. This is consistent with Table B.2 in Appendix B.2: the pairwise correlation coefficients among the $\hat{r}_{ijt}$ computed using different revenue regression specifications are always very close to one.

Table C.2: Bounds on Individual Extended Gravity Parameters

Border	Continent	Language	GDPpc
[5.71%, 13.33%]	[17.14%, 28.57%]	[30.48%, 34.29%]	[0%, 30.48%]

Notes: This table report bounds on the vector κ defined in equation (21). It reports results for the revenue regression described in column VI of Table B.1. The confidence intervals are projections of a confidence set for $(\kappa, \tilde{\eta})$ computed following the procedure in Section 10.2 of Andrews and Soares (2010).

Table C.3: Bounds on Combinations of Extended Gravity Parameters

Border + Continent	Language + GDPpc	Continent + GDPpc	Continent + GDPpc + Border	Continent + Language	Continent + Language + Border	Continent + Language + GDPpc	All
[24.76%, 38.10%]	[32.38%, 62.86%]	[22.86%, 49.52%]	[34.29%, 59.05%]	[47.62%, 62.86%]	[55.24%, 72.38%]	[57.14%, 81.90%]	[68.57%] 91.43%]

Notes: This table reports bounds on sums of elements of the vector κ defined in equation (21). It reports results conditional on the revenue regression described in column VI in Table B.1. The confidence intervals reported in this table are projections of a 5-dimensional confidence set for $(\kappa, \tilde{\eta})$ computed following the procedure in Section 10.2 of Andrews and Soares (2010).

D Simulation

The model described in Section 3 imposes only necessary conditions on the planning horizon, L_{it}, information set, $\mathcal{J}_{it}$, and consideration set, $\mathcal{B}_{it}$, of every firm and year in the sample. It also imposes only weak assumptions on the distribution of the unobserved determinants of export revenues ε^R_{ijt}, fixed costs ε^F_{ijt}, and sunk costs ε^S_{ijt}: they are mean independent of any variable affecting the export participation decision of the firm. Similarly, no restriction is imposed on the distribution of the unobserved determinant of export fixed costs u_{ic_jt}.

In this section, we simulate data from several models that are consistent with the assumptions on planning horizons, information and consideration sets, and distributions of unobserved terms introduced in Section 3. The different models we simulate differ thus on the exact assumptions we impose on the firm's planning horizon, L_{it} the firm's information set, $\mathcal{J}_{it}$, and the distribution of the unobserved determinants of export fixed costs, u_{ic_jt}. We use these different models to explore in Appendix D.2 how sensitive maximum likelihood estimates of fixed and sunk export costs are to misspecifications of the firm's planning horizon, information and consideration set, and the functional form of the distribution of u_{ic_jt}.

Additionally, to explore the sensitivity of our moment inequality estimator to the presence of firm-, year-, and country-specific unobserved heterogeneity in export profits that is known to the firm when determining its set of export destinations, we extend the fixed costs specification in equation (6) to account for such unobserved heterogeneity through a term ν_{ijt}. We show in Appendix D.3 how sensitive our moment inequality estimates are to the variance of such firm-, year-, and country-specific unobserved determinant of export participation.

D.1 Simulated Models

Basic characteristics. In every model we simulate, we generate data for $N = 500$ firms, $T = 20$ years, and $J = 5$ foreign markets. The total number of observations in our simulated datasets is thus 50,000. Neither of the five simulated countries shares neither continent, nor language, nor similar income per capita with the firm's home market; i.e.

$$f^o_j = \gamma^F_{all} = \gamma^F_0 + \gamma^F_c + \gamma^F_l + \gamma^F_g, \tag{D.1}$$

$$s^o_j = \gamma^S_{all} = \gamma^S_0 + \gamma^S_c + \gamma^S_l + \gamma^S_g. \tag{D.2}$$

In terms of extended gravity relationships in border, continent, and language, we assume that the indicator functions border(j, j'), continent(j, j'), and language(j, j'), are equal to

$$\text{border}(j, j') = \begin{cases} 1 & \text{if } (j, j') \in \{(2, 3), (3, 2)\}, \\ 0 & \text{otherwise}; \end{cases} \tag{D.3}$$

$$\text{continent}(j, j') = \begin{cases} 1 & \text{if } (j, j') \in \{(2, 3), (3, 2), (2, 4), (4, 2), (3, 4), (4, 3)\}, \\ 0 & \text{otherwise}; \end{cases} \tag{D.4}$$

$$\text{language}(j, j') = \begin{cases} 1 & \text{if } (j, j') \in \{(3, 4), (4, 3), (3, 5), (5, 3), (4, 5), (5, 4)\}, \\ 0 & \text{otherwise}, \end{cases} \tag{D.5}$$

where, as a reminder, any of these functions equals 1 for a pair of countries (j, j') if and only if these two countries share the corresponding characteristics. For simplicity, we ignore any extended gravity relationship due to continent.

Fixed and sunk export costs. Fixed and sunk export costs in each of the five simulated countries are

$$f_{ijt} = \gamma^F_{all} + u_{ic_jt} + \sigma_\nu \nu_{ijt}, \tag{D.6}$$

$$s_{ijt} = \gamma^S_{all} + e^o_{ijt}, \tag{D.7}$$

with ν_{ijt} distributed logistically with location parameter equal to zero and scale parameter equal to one, and e^o_{ijt} depends on equations (D.3), (D.4), and (D.5). We define u_{ic_jt} as a variable that is common to all countries located in the same continent ($u_{ic_jt} = u_{ic_{j'}t}$ if continent(j, j')=1) and, to explore how sensitive the maximum likelihood estimates are to the assumed distribution of u_{ic_jt}, we generate different datasets in which we explore different distributions for it. The key fixed and sunk costs parameters in our simulated models are thus

$$(\gamma^F_{all}, \gamma^S_{all}, \gamma^E_b, \gamma^E_c, \gamma^E_l, \sigma_\nu).$$

In all simulated datasets, we fix

$$\gamma^F_{all} = 9; \qquad \gamma^S_{all} = 6.1; \qquad \gamma^E_b = 0.93; \qquad \gamma^E_c = 1.86; \qquad \gamma^E_l = 1.2, \tag{D.8}$$

and, thus, the relative reduction in sunk export costs due to each of the extended gravity effects equals

$$\kappa_b = \gamma^E_b/\gamma^S_{all} = 0.15; \qquad \kappa_c = \gamma^E_c/\gamma^S_{all} = 0.30; \qquad \kappa_l = \gamma^E_l/\gamma^S_{all} = 0.20, \tag{D.9}$$

and that due to their combined effect is

$$\kappa_b + \kappa_c + \kappa_l = 0.65. \tag{D.10}$$

Potential export revenues. Across all simulated models, we assume that

$$r_{ijt} = \exp(\alpha_0 + \alpha_1 \tilde{r}_{ijt}) \quad \text{with} \quad \tilde{r}_{ijt} = 0.5\tilde{r}_{ijt-1} + \omega_{ijt}, \tag{D.11}$$

and ω_{ijt} is assumed to be independent over time and distributed normally with mean zero and variance σ^2_ω. Except otherwise indicated, we build simulated series for $\tilde{r}_{ijt}$ and r_{ijt} under the assumption that $\alpha_0 = 14$, $\alpha_1 = 0.55$ and $\sigma_\omega = 4$.

Elasticity of demand. We assume that $\eta^{-1} = 0.2$.

Consideration set. All $2^J = 32$ possible bundles of countries belong to the consideration set $\mathcal{B}_{it}$ of all firms in every time period.

Planning horizon. We impose in all simulated datasets that all firms have the same planning horizon, $L_{it} = L$. We vary the value of L across the different datasets we generate.

Information set. We impose in all simulated datasets that firms know f_{ijt} and s_{ijt}. While in some datasets we also assume that firms know r_{ijt}, in others we assume that they only know r_{ijt-1} when deciding on their optimal set of export destinations.

State vector. Given that generating some of the simulated datasets we create requires solving a dynamic discrete choice problem, we discretize r_{ijt}. Specifically, we assume that r_{ijt} can take six values. Consequently, the exogenous state space of a firm has $6^5 = 7,776$ elements and the transition matrix of this exogenous state vector thus has $(7,776^2 - 7,776)/2 = 30,229,200$ distinct elements. The endogenous state space determines all possible lagged export status of the firm in the five destination markets and, thus, includes $2^5 = 32$ elements. As a result, the complete state space of the firm incorporates $7,776 \times 32 = 248,832$ elements. Extending either the set of countries J beyond 5, or the number of distinct values that r_{ijt} can take beyond 6, increases significantly the size of the state space, causing significant computational complications.

Data available to the researcher. For each firm, country and year in the simulated sample, the researcher only observes: (a) a dummy determining the export status, d_{ijt}; (b) the exogenous determinant of export revenues, $\tilde{r}_{ijt}$; and (c) realized export revenues, $r_{ijt}d_{ijt}$.

D.2 Impact of Model Misspecification on Maximum Likelihood Estimates

In this section, we explore the sensitivity of the maximum likelihood estimator of $(\eta, \gamma^F_{all}, \gamma^S_{all}, \gamma^E_b, \gamma^E_c, \gamma^E_l)$ to the assumptions imposed on: (a) the planning horizon, L_{it}; (b) the information set, $\mathcal{J}_{it}$; (c) the consideration set, $\mathcal{B}_{it}$; and (d) the parametric distribution of u_{ic_jt}. We consider the case of a researcher that correctly assumes that ν_{ijt} is distributed logistically with location parameter equal to zero and scale parameter equal to one.

Misspecification of planning horizon, L_{it}. To explore how sensitive maximum likelihood estimates are to the researcher's assumptions on the firm's planning horizon, we assume that the researcher correctly specifies the firm's information set, $\mathcal{J}_{it}$, the firm's consideration set, $\mathcal{B}_{it}$, and the distribution of u_{ic_jt}. Specifically, the researcher assumes an information set $\mathcal{J}_{it}$ such that,

$$\mathbb{E}[r_{ijt}|\mathcal{J}_{it}] = r_{ijt}, \tag{D.12}$$

Table D.1: Robustness to Different Planning Horizons

	Correct		Wrong	
	$L_{it}=0$	$L_{it}=1$	$L_{it}=2$	$L_{it}=3$
η^{-1}	0.201	0.209	0.206	0.209
	(0.001)	(0.001)	(0.001)	(0.001)
γ^F_{all}	9.032	8.063	7.133	6.989
	(0.157)	(0.249)	(0.335)	(0.394)
γ^S_{all}	6.121	5.016	5.077	5.390
	(0.064)	(0.096)	(0.135)	(0.189)
γ^E_b	0.989	0.418	0.344	0.458
	(0.081)	(0.142)	(0.194)	(0.224)
γ^E_c	1.818	1.063	0.813	0.846
	(0.072)	(0.132)	(0.191)	(0.242)
γ^E_l	1.226	0.535	0.309	0.248
	(0.057)	(0.099)	(0.144)	(0.191)
κ_b	0.162	0.083	0.067	0.085
κ_c	0.297	0.212	0.160	0.157
κ_l	0.200	0.107	0.061	0.046

Notes: standard errors appear in parenthesis. The underlying data is generated under the assumption that $L_{it} = 0$. The model reported in each column is estimated under the assumption on the value of L_{it} indicated in the corresponding column.

a consideration set $\mathcal{B}_{it}$ such that,

$$\mathcal{B}_{it} = \{\text{all possible bundles of the } J = 5 \text{ countries}\}, \tag{D.13}$$

a planning horizon L_{it} such that

$$L_{it} = 0, \qquad \text{for all firms and time periods}, \tag{D.14}$$

and sets u_{ic_jt}

$$u_{ic_jt} = 0, \tag{D.15}$$

for every firm, country and year. In words, assuming equation D.12 is equivalent to assuming that firms have perfect foresight; assuming equation (D.13) is equivalent to assuming that firms consider exporting to all countries (i.e. the consideration set and the potential choice set of the firm are identical); and assuming equation (D.14) is equivalent to assuming that firms are not forward looking. We simulate four different datasets that are consistent with the assumptions in equations (D.12), (D.13), and (D.15), but in which L_{it} takes values zero, one, two, and three, respectively, for every firm and time period. We present the results in Table D.1.

As the results in Table D.1 show, the maximum likelihood estimates of a model that corresponds exactly to the data generating process except for the assumption imposed on the firm's planning horizon may still be quite different from the true parameter values. As the difference between the true planning horizon ($L_{it} = 0$) and the assumed planning horizon (indicated at the top of each column in Table D.1) increases, all estimates other than η^{-1} become smaller in absolute value. The intuitive explanation for why this happens is that the difference between the true value function of the firm and that assumed by the researcher increases with the difference between the true and the assumed planning horizon. Misspecifying the planning horizon not only causes a bias in the scale of all parameter estimates, but it also causes a bias in the relative magnitude of these estimates. This is illustrated by the relative extended gravity parameters, κ_b, κ_c, and κ_l, which also become smaller as the difference between the the true and the assumed planning horizon increases. For example, while

Table D.2: Robustness to Different Information Sets

	Correct	Wrong With		
		$\sigma_\omega = 2$	$\sigma_\omega = 4$	$\sigma_\omega = 6$
η^{-1}	0.201	0.038	0.033	0.028
	(0.001)	(0.001)	(0.001)	(0.001)
γ^F_{all}	9.032	-1.933	-1.045	-0.269
	(0.157)	(0.326)	(0.019)	(0.006)
γ^S_{all}	6.121	5.919	5.879	5.663
	(0.064)	(0.014)	(0.004)	(0.002)
γ^E_b	0.989	0.563	0.846	0.831
	(0.081)	(0.083)	(0.009)	(0.004)
γ^E_c	1.818	1.777	1.721	1.722
	(0.072)	(0.065)	(0.007)	(0.003)
γ^E_l	1.226	1.089	1.143	1.079
	(0.057)	(0.023)	(0.005)	(0.003)
κ_b	0.162	0.095	0.144	0.147
κ_c	0.297	0.300	0.293	0.304
κ_l	0.200	0.184	0.194	0.190

Notes: standard errors appear in parenthesis. The data is generated under the assumption that $\mathbb{E}[r_{ijt}|\mathcal{J}_{it}] = r_{ijt}$, which is also the assumed imposed by the researcher compute the estimates in column 1. The estimates in columns 2 to 4 are computed under the wrong assumption that $\mathbb{E}[r_{ijt}|\mathcal{J}_{it}] = \mathbb{E}[r_{ijt}|r_{ijt-1}]$. From equation (D.11), as we increase the value of σ_ω we also increase the difference between r_{ijt} and $\mathbb{E}[r_{ijt}|r_{ijt-1}]$.

the total reduction in sunk costs due to all extended gravity effects in the simulated model is 65% according to equation (D.10), the estimated value of this parameter is 40% when the planning horizon is underestimated by a year, and approximately 29% when is underestimated by either two or three years.

Misspecification of information set, $\mathcal{J}_{it}$. To explore how sensitive maximum likelihood estimates are to the researcher's assumptions on the firm's information set, we assume that the researcher correctly specifies the firm's planning horizon, L_{it}, the firm's consideration set, $\mathcal{B}_{it}$, and the distribution of the unobserved term u_{ic_jt}. Specifically, the researcher maximizes a likelihood function that relies on the assumptions on information set, consideration set, planning horizon and distribution of u_{ic_jt} indicated in equations (D.12) to (D.15).

We simulate four different datasets that respect the assumptions in equations (D.13), (D.14) and (D.15), but that differ in the assumed content of the firm's information set. While the first dataset is consistent with equation (D.12), $\mathbb{E}[r_{ijt}|\mathcal{J}_{it}] = r_{ijt}$, datasets two to four are simulated assuming instead that $\mathbb{E}[r_{ijt}|\mathcal{J}_{it}] = \mathbb{E}[r_{ijt}|r_{ijt-1}]$. The difference between r_{ijt} and $\mathbb{E}[r_{ijt}|r_{ijt-1}]$ increases in the variance of the export revenue shock ω_{ijt} defined in equation (D.11). To show how sensitive the maximum likelihood estimates are to differences between the true firms' expectations, $\mathbb{E}[r_{ijt}|r_{ijt-1}]$, and the researcher's assumed ones, r_{ijt}, datasets two to four are different only in the value of the variance of the revenue shock ω_{ijt}, σ^2_ω. We present the results in Table D.2.

As the results in Table D.2 show, the maximum likelihood estimates of the coefficient on the expected potential export revenues, η^{-1}, are very sensitive to the researcher's assumption on the information set used by the firm to construct such expectations. If the firm has imperfect information about the potential export revenue that it would obtain in a country j and period t if it were to export to it, r_{ijt}, but the researcher assumes that its information is perfect, then the maximum likelihood estimate of the coefficient on r_{ijt} will be downward biased. Comparing the results across columns two to four in Table D.2, we can see that wrongly

Table D.3: Robustness to Different Consideration Sets

	Correct	Wrongly Excludes Country:				
		1	2	3	4	5
η^{-1}	0.201	0.188	0.178	0.182	0.179	0.176
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
γ^F_{all}	9.032	9.296	8.504	8.470	8.442	8.573
	(0.157)	(0.033)	(0.029)	(0.029)	(0.029)	(0.029)
γ^S_{all}	6.121	3.598	4.252	4.435	4.293	3.867
	(0.064)	(0.001)	(0.002)	(0.002)	(0.018)	(0.001)
γ^E_b	0.989	0.315			0.430	0.279
	(0.081)	(0.003)			(0.001)	(0.003)
γ^E_c	1.818	0.736	0.715	0.832	0.594	0.872
	(0.072)	(0.002)	(0.002)	(0.001)	(0.001)	(0.004)
γ^E_l	1.226	0.396	0.547	0.321	0.418	0.335
	(0.057)	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)
κ_b	0.162	0.088			0.100	0.072
κ_c	0.297	0.205	0.168	0.187	0.138	0.225
κ_l	0.200	0.110	0.128	0.072	0.097	0.086

Notes: standard errors appear in parenthesis. The underlying data is generated under the assumption that all possible bundles of the five countries described in equations (D.3) to (D.5) are included in the firm's consideration set $\mathcal{B}_{it}$. The model reported in the first column assumes the correct consideration set. The models reported in columns 2 to 7 assume a consideration set $\mathcal{B}_{it}$ formed by all possible bundles of a subset of four countries extracted from the actual set of five countries. The country dropped in each case is indicated in the corresponding column.

assuming that firms have perfect foresight generates estimates of η^{-1} that are significantly below the true value of this parameter, 0.2, and the downward bias increases as we increase the variance of the revenue shock, ω_{ijt}. Consistent with the results in Dickstein and Morales (2016), the misspecification of the firm's information set also biases significantly the maximum likelihood estimate of the export fixed costs, γ^F_{all}.

Surprisingly, while the estimates of the inverse elasticity of demand, η^{-1}, and fixed export costs, γ^F_{all}, suffer from sever bias when the researcher misspecifies the firm's information set, the maximum likelihood estimates of the sunk costs parameters γ^S_{all}, γ^E_b, γ^E_c, γ^E_l are not affected by this misspecification. As a consequence, the estimates of the relative extended gravity parameters κ_b, κ_c, and κ_l are invariant to the assumptions imposed on the information set used by firms to predict their potential export revenue upon entry.

Misspecification of consideration set, $\mathcal{B}_{it}$. To explore how sensitive maximum likelihood estimates are to the researcher's assumptions on the firm's consideration set, we assume that the researcher correctly specifies both the firm's planning horizon, L_{it}, the firm's information set, $\mathcal{J}_{it}$, and the distribution of the unobserved term u_{ic_jt}. Specifically, the researcher imposes on the firm's information set, the firm's planning horizon, and the distribution of u_{ic_jt} the same assumptions as in equations (D.12), (D.14), and (D.15), respectively. Concerning the specification of the firm's consideration set, we simulate a dataset that imposes the consideration set described in equation (D.13), but estimate models in which the consideration set assumed by the researcher is smaller than the actual one. Specifically, we estimate five misspecified models, each of them characterized by a consideration set that ignores one of the five countries included the actual choice set of the firm. We present the results in Table D.3.

As the results in Table D.3 show, misspecifications of the firm's consideration set do not have a large impact neither on the coefficient on potential export revenues, η^{-1}, nor on the fixed export costs parameter, γ^F_{all}. Conversely, they have a large impact on the estimates of the sunk costs parameters. It is interesting to remark that this bias in the estimated sunk costs parameters caused by misspecification of the firm's consideration set happens even when the excluded country does not share any extended gravity characteristic with any other destinations; as shown in equations (D.3) to (D.5), the country $j = 1$ does not share neither border, not continent, nor language with any of the other four destinations in the simulated data.

When we exclude either country $j = 2$ or country $j = 3$, the coefficient on extended gravity due to border is not identified, as countries two and three are the only ones to share border; i.e. if one of them is excluded, no firm can enter a country having previously exported to a bordering market. Looking at the estimates of κ_c and κ_l, one can observe that the downward bias in these estimates is larger when the excluded country shares the corresponding extended gravity covariate with at least one of the other destinations. Specifically, as equation (D.4) indicates, countries 2, 3 and 4 share continent and, as Table D.3 shows, the estimates of κ_c are lowest when one of these three countries is excluded from the consideration set. Similarly, as equation (D.5) indicates, countries 3, 4, and 5 share language and, as Table D.3 shows, the estimates of κ_l are lowest when one of these three countries is excluded from the consideration set.

Misspecification of the distribution of the random effect, u_{ic_jt}. To explore how sensitive maximum likelihood estimates are to the researcher's assumptions on the distribution of the firm-continent random effects u_{ic_jt}, we assume that the researcher correctly specifies the firm's planning horizon, L_{it}, the firm's information set, $\mathcal{J}_{it}$, and the consideration set, $\mathcal{B}_{it}$. Specifically, the researcher imposes the assumptions in equations (D.12), (D.13), and (D.14), respectively.

Concerning the researcher's assumptions on the distribution of the firm-continent random effects u_{ic_jt}, we consider two cases. First, a case in which the researcher assumes that $u_{ic_jt} = 0$ for all firms, countries and years (consistently with equation (D.15)) and, thus, estimates a binary logit model; the results corresponding to this case are in Table D.4. Second, a case in which the researcher assumes that $u_{ic_jt} = u_{ic_j}$ and that u_{ic_j} follows a normal distribution independent across firms and continents; the results corresponding to this case are in Table D.5. Each of the different columns in tables D.4 and D.5 show the estimates of the corresponding statistical model for a different dataset, and each of these datasets differ in the actual distribution of u_{ic_j}, which may be the same or different from that assumed by the researcher. In all of these datasets, we generate the random variable u_{ic_j} as a recentered at zero transformation $g(\cdot)$ of an independently and normally distributed random variable ξ_{ic_j} that has mean zero and a standard deviation equal to σ_ξ; i.e. $u_{ic_j} = g(\xi_{ic_j}) - \mathbb{E}[g(\xi_{ic_j})]$ with $\xi_{ic_j} \sim \mathbb{N}(0, \sigma_\xi^2)$. The different datasets we generate to compute the estimates reported in tables D.4 and D.5 differ in the transformation function $g(\cdot)$ and the standard deviation parameter σ_ξ. Specifically, we generate two vectors $\{\xi_{ic_j}, \forall\ i \text{ and } c_j\}$ that correspond to values of σ_ξ in the set $\{1, 3\}$ and, for each of them, we generate two different vectors $\{u_{ic_j}, \forall\ i \text{ and } c_j\}$ that correspond to the following two functions $g(\cdot)$: the normal case

$$g(\xi_{ic_j}) = \xi_{ic_j}; \tag{D.16}$$

and the Bernoulli case,

$$g(\xi_{ic_j}) = \mathbb{1}\{\xi_{ic_j} \geq 0\} \times percentile_{75}(\xi_{ic_j}) + \mathbb{1}\{\xi_{ic_j} < 0\} \times percentile_{25}(\xi_{ic_j}), \tag{D.17}$$

where $percentile_q(\xi_{ic_j})$ denotes the percentile q of ξ_{ic_j}. Additionally, for the specific case in which $\sigma_\xi = 1$, we also generate two different vectors $\{u_{ic_j}, \forall\ i \text{ and } c_j\}$ that correspond the log-normal case,

$$g(\xi_{ic_j}) = \exp(\xi_{ic_j}); \tag{D.18}$$

the "chi-squared" case,

$$g(\xi_{ic_j}) = (\xi_{ic_j})^2. \tag{D.19}$$

As Table D.4 shows, ignoring the presence of firm-continent random effects by estimating binary logit models results in an upward bias in the estimate of the reduction in the export entry costs due to extended gravity effects in continent, γ_c^E. Simultaneously, the baseline value of these export entry costs, γ_{all}^S is also upward biased. The consequence is that the *relative* reduction in export sunk costs due to extended gravity continent, κ_c, is only slightly biased upwards, but the relative reductions due to extended gravity in border, κ_b, and language, κ_l, are severely biased downwards. The size of these biases increase as the variance of the firm-continent random effect increases, as illustrated by: (a) the comparison of the normal and Bernoulli cases with $\sigma_\xi = 3$ to those with $\sigma_\xi = 1$; (b) the comparison of the log-normal and Chi-squared cases to the normal case.

Table D.4: Robustness to Firm-Continent Random Effects: Binary Logit

Distribution	Normal		Bernoulli		Ln-Nrm	Chi-Sq.
σ_ξ	1	3	1	3	1	1
η^{-1}	0.225	0.454	0.209	0.289	0.334	0.259
	(0.003)	(0.010)	(0.003)	(0.005)	(0.006)	(0.004)
γ^F_{all}	9.816	15.747	9.231	11.190	13.323	10.680
	(0.215)	(0.616)	(0.193)	(0.345)	(0.357)	(0.265)
γ^S_{all}	7.734	23.339	6.829	13.398	12.879	9.627
	(0.071)	(0.201)	(0.064)	(0.115)	(0.118)	(0.089)
γ^E_b	0.704	1.278	0.732	0.745	0.557	0.708
	(0.117)	(0.345)	(0.110)	(0.174)	(0.201)	(0.148)
γ^E_c	2.559	8.245	2.175	3.322	4.983	3.696
	(0.098)	(0.268)	(0.089)	(0.136)	(0.163)	(0.119)
γ^E_l	1.032	0.511	1.181	1.366	0.567	0.530
	(0.070)	(0.186)	(0.066)	(0.103)	(0.110)	(0.082)
κ_b	0.091	0.054	0.107	0.055	0.043	0.073
κ_c	0.331	0.353	0.318	0.247	0.386	0.384
κ_l	0.133	0.022	0.173	0.102	0.044	0.055

Notes: standard errors appear in parenthesis. All columns estimate a binary logit model and, to facilitate the comparability across models, all estimates are scaled by the true standard deviation of $u_{ic_j} + \sigma_\nu \nu_{ijt}$.

Table D.5: Robustness to Firm-Continent Random Effects: Mixed Logit

Distribution	Normal		Bernoulli		Ln-Nrm	Chi-Sq.
σ_ξ	1	3	1	3	1	1
η^{-1}	0.213	0.195	0.217	0.213	0.205	0.211
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
γ^F_{all}	11.404	12.245	11.327	13.820	10.986	11.205
	(0.222)	(0.263)	(0.237)	(0.259)	(0.234)	(0.233)
γ^S_{all}	6.241	6.272	6.273	6.325	6.225	6.338
	(0.065)	(0.076)	(0.066)	(0.073)	(0.070)	(0.068)
γ^E_b	0.856	0.831	0.871	0.924	0.767	0.867
	(0.104)	(0.119)	(0.108)	(0.096)	(0.114)	(0.116)
γ^E_c	1.933	2.242	1.867	1.859	2.051	1.985
	(0.090)	(0.101)	(0.095)	(0.081)	(0.101)	(0.100)
γ^E_l	1.357	0.963	1.395	1.377	1.158	1.256
	(0.065)	(0.071)	(0.068)	(0.060)	(0.072)	(0.072)
κ_b	0.137	0.132	0.138	0.146	0.123	0.136
κ_c	0.309	0.357	0.297	0.294	0.329	0.313
κ_l	0.217	0.226	0.222	0.217	0.186	0.198
Num. Obs. per (i, c_j)	32	32	32	32	32	32

Notes: standard errors appear in parenthesis. All columns estimate a mixed logit model with normally distributed firm-continent random effects.

As Table D.5 shows, estimating mixed logit models allows to deal with firm-continent random effects. Specifically, while the fixed cost component γ^F_{all} is severely upward biased, the estimates of both the baseline sunk costs γ^S_{all} and all the extended gravity covariates, γ^E_b, γ^E_c and γ^E_l, are large unbiased. This results in the estimates of the relative extended gravity parameters, κ_b, κ_c, and κ_l being also very close to their true parameter values. Importantly, this seems to be generally true even in those cases in which the true distribution of the firm-continent random effect does not follow the distribution imposed when estimating the mixed logit model: even in the cases in which the firm-continent unobservable is simulated from a log-normal distribution or from a Chi-squared distribution (both recentered to have mean zero), the mixed logit estimates of κ_b, κ_c, and κ_l are very close to their true values.

D.3 Impact of Model Misspecification on Moment Inequality Estimates

In this section, we explore how the confidence set for the parameter vector $(\kappa_b, \kappa_c, \kappa_l)$ is affected when there are unobserved (to the researcher) components that determine the optimal set of export destinations of a firm. We consider two cases. First, we study the effect of ignoring the presence of a country-, firm-, year-specific unobserved component in our inequalities; we denote this case below as *country-specific unobserved heterogeneity* and model it through the term ν_{ijt} in equation (D.6). Second, we study the effect of ignoring the presence of a continent-, firm-, year-specific unobserved component; we refer to this case below as *continent-specific unobserved heterogeneity* and model it through the term u_{ic_jt} in equation (D.6), with the index c_j defined as the continent to which country j belongs.

In order to focus on the sensitivity of the moment inequality estimator to the assumptions imposed on the distribution of the unobserved terms ν_{ijt} and u_{ic_jt}, we assume that the researcher's assumptions on the firm's planning horizon, L_{it}, information set, $\mathcal{J}_{it}$, and consideration set, $\mathcal{B}_{it}$ are correct. Specifically, we simulate data from the simplest possible model that is compatible with the restrictions on the planning horizon and information set imposed in assumptions 2 and 3 in Section 3.4: firms make decisions at any period t taking into account the impact of these decisions only on periods t and $t+1$ (i.e. $L_{it} = 1$) and have perfect foresight about the potential revenues of exporting to any country in the choice set in both periods t and $t+1$ (i.e. $(r_{ijt}, r_{ijt+1}) \in \mathcal{J}_{it}$). Concerning the consideration set, we assume that the set $\mathcal{A}_{it}$ imposed by the researcher when deriving her moment inequalities coincides with the true consideration set, $\mathcal{B}_{it}$, which includes all combinations of the five countries described in Appendix D.

Country-specific unobserved heterogeneity. To explore how sensitive moment inequality estimates are to the assumption, imposed in the model described in Section 3, that there are no country-, firm-, year-specific unobserved determinants of export profits, we apply the moment inequality estimator described in Section 4 (which is derived under the assumption that $\sigma_\nu = 0$) to datasets simulated under different values of σ_ν. The general pattern we observe is that, as the value of σ_ν increases, the estimated confidence set for the parameter vector $(\kappa_b, \kappa_c, \kappa_l)$ becomes smaller and, eventually, it becomes empty. However, interestingly, whenever the value of σ_ν is such that the moment inequality confidence set is nonempty, this one does contain the true value of the parameter vector. The values of σ_ν that we use to generate simulated data are such $\sigma^r_\nu = \sigma_\nu/\sqrt{var(\eta^{-1} r_{ijt})}$ is equal to either 0 (the case corresponding to the model described in Section 3), 0.4, 0.6, 1, or 1.4. The resulting confidence sets are represented in figures D.1 to D.5.

Continent-specific unobserved heterogeneity. To explore the sensitivity of our moment inequality estimates to wrongly ignoring the presence of firm-, year-, continent-specific unobserved determinants of export profits, we apply a moment inequality estimator that assumes that the standard deviation of u_{ic_jt} equals zero, $\sigma_u = 0$, to datasets simulated under different values of σ_u. The general pattern we observe is that, as the true value of σ_u increases, the estimated confidence set for the parameter vector $(\kappa_b, \kappa_c, \kappa_l)$ changes its shape in a way such that the estimated bounds for κ_c become tighter and the estimated bounds for κ_b and κ_l become wider. However, interestingly, for a very large range of values of σ_u, the confidence set for $(\kappa_b, \kappa_c, \kappa_l)$ still contains the true value of the parameter vector. The values of σ_u that we use to generate simulated data are such that the ratio between σ_u and the standard deviation of the remaining part of export profits $\eta^{-1} r_{ijt}$, σ^r_u, is equal to either 0 (the case corresponding to the model described in Section 3), 0.4, 0.6, 1, 1.4, or 2. The resulting confidence sets are represented in figures D.6 to D.10.

Number of countries and identification power. The confidence sets in Figure D.1 result from applying our moment inequality estimator to a dataset that is consistent with the assumptions under which such estimator is valid; i.e. $(\sigma_\nu^r, \sigma_u^r) = (0, 0)$. However, the resulting confidence sets are much wider than those reported in Section 6. The reason why this happens is that, in our simulated sample, the set of observed and alternative export bundles we can choose from to form our moments is much smaller than in the actual data. This limitation is due to the small number of export destinations ($J = 5$) in our simulated data. The reason for limiting ourselves to such a small number of destinations is purely due to computational costs. Simulating a model consistent with our moment inequality framework requires computing the firm's value function for each of the elements in its potential choice set and for each possible value of the state vector. As indicated above, the state space of our simulated firms already includes 248,832 points. If we had allowed for one more country in the choice set (i.e. $J = 6$), then the state space would have included $6^6 \times 2^6$=2,985,984 elements. In fact, these computational limitations are the main reason why we opted for using a moment inequality estimator on the actual data in the first place.

What Figure D.1 illustrates is that the identification power of our moment inequality approach increases significantly with the size of the firm's choice set, as this facilitates finding alternative export destinations that, when compared to the firm's observed destination, help identify the model parameters. To illustrate this point through an example, notice that, as indicated in Appendix C.4, one of the instrument functions that we use to define our moment inequalities is

$$\Psi_1(Z_{ijt}, Z_{ij't}) = \mathbb{1}\{s_j^o = s_{j'}^o = \gamma_{all}^S, d_{ijt-1} = d_{ij't-1} = 0, (\text{Ext. Grav. Border})_j = 1, (\text{Ext. Grav. Border})_{j'} = 0\}.$$

This function will take value one only when we observe a firm entering a new destination j in year t that shares border with a previous destination of the firm and, at the same time, not entering an alternative destination j' that does not share any border with any prior export destination of this firm. In our simulated data, only countries 2 and 3 share border and, thus, the function $\Psi_1(Z_{ijt}, Z_{ij't})$ will equal one only for those firms that happen to enter country 2 when they were previously exporting to country 3 (or vice versa). Only a very small number of firms in our simulated data will have this particular export pattern, meaning that the moment inequality defined by the instrument function $\Psi_1(Z_{ijt}, Z_{ij't})$ will average only over very few observations. It is this small number of observations per moment inequality in our simulation that explains the large size of the confidence sets in Figure D.1.

Figure D.1: Projected Confidence Set: $(\sigma_\nu^r, \sigma_u^r) = (0, 0)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

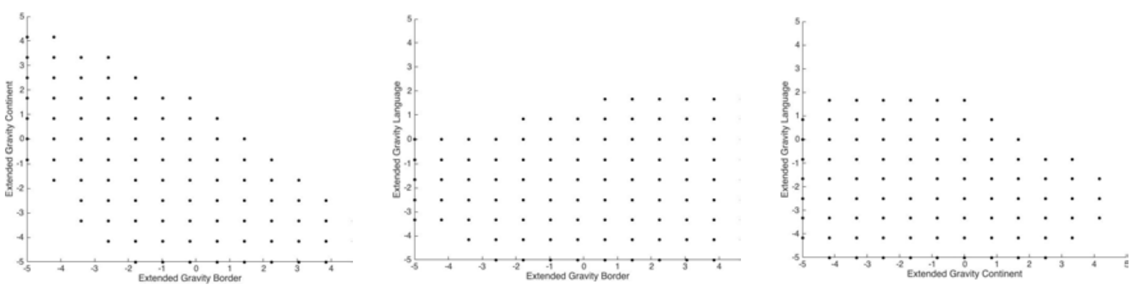


Figure D.2: Projected Confidence Set: $(\sigma_\nu^r, \sigma_u^r) = (0.4, 0)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

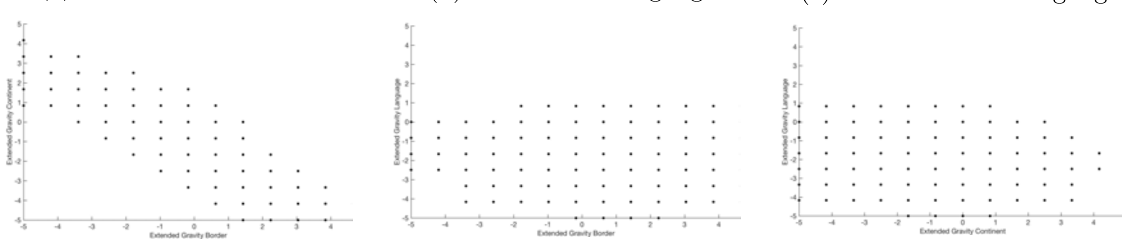


Figure D.3: Projected Confidence Set: $(\sigma_{\nu}^{r}, \sigma_{u}^{r}) = (0.6, 0)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

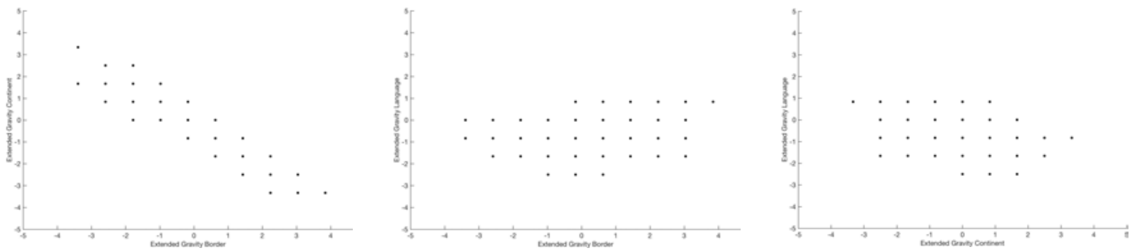


Figure D.4: Projected Confidence Set: $(\sigma_{\nu}^{r}, \sigma_{u}^{r}) = (1, 0)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

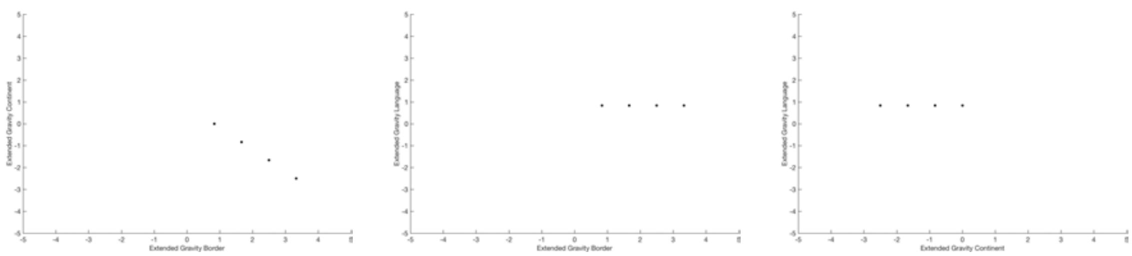


Figure D.5: Projected Confidence Set: $(\sigma_{\nu}^{r}, \sigma_{u}^{r}) = (1.4, 0)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

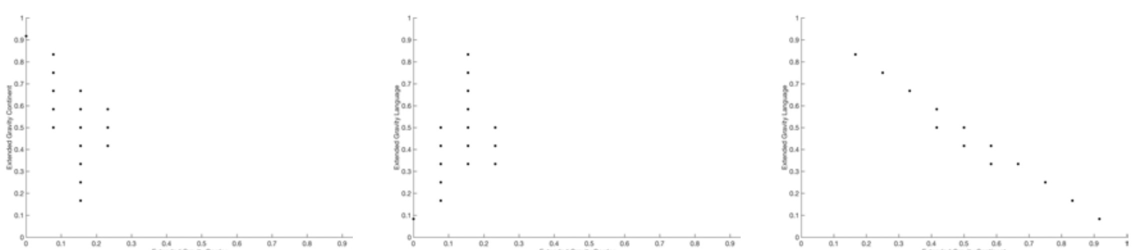


Figure D.6: Projected Confidence Set: $(\sigma_{\nu}^{r}, \sigma_{u}^{r}) = (0, 0.4)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

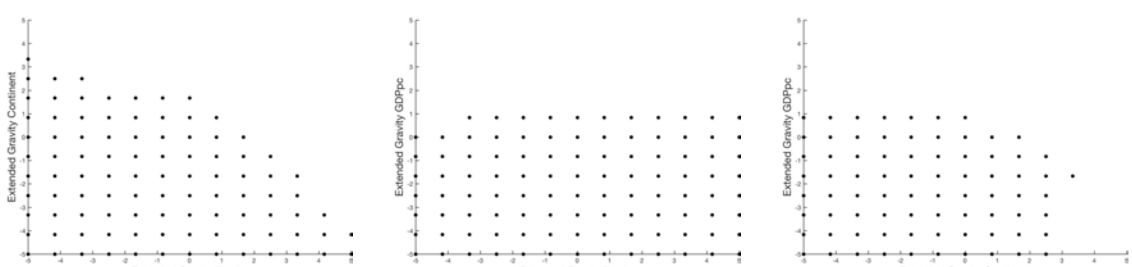


Figure D.7: Projected Confidence Set: $(\sigma_\nu^r, \sigma_u^r) = (0, 0.6)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

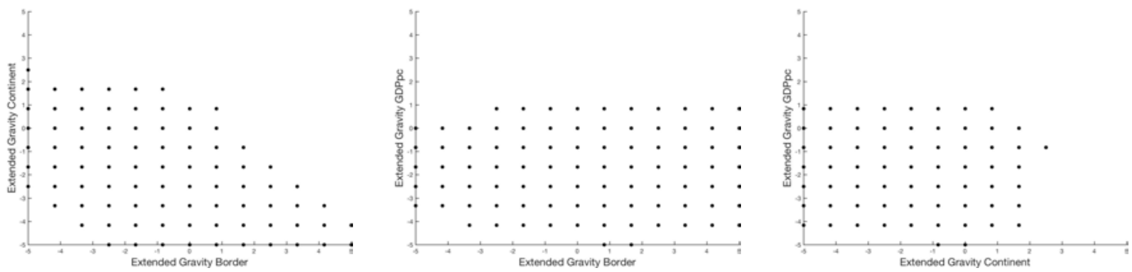


Figure D.8: Projected Confidence Set: $(\sigma_\nu^r, \sigma_u^r) = (0, 1)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

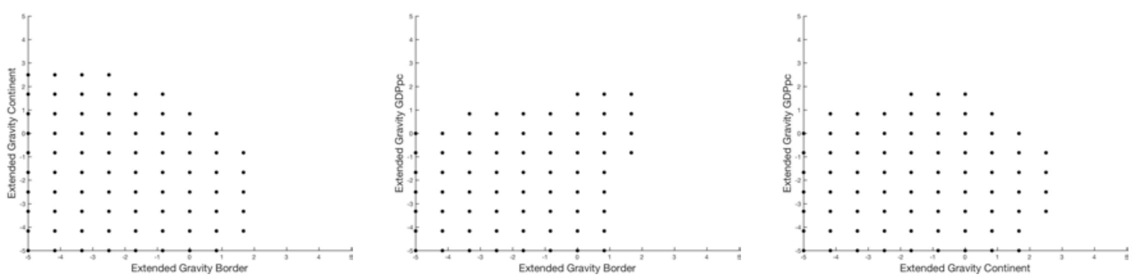


Figure D.9: Projected Confidence Set: $(\sigma_\nu^r, \sigma_u^r) = (0, 1.4)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language

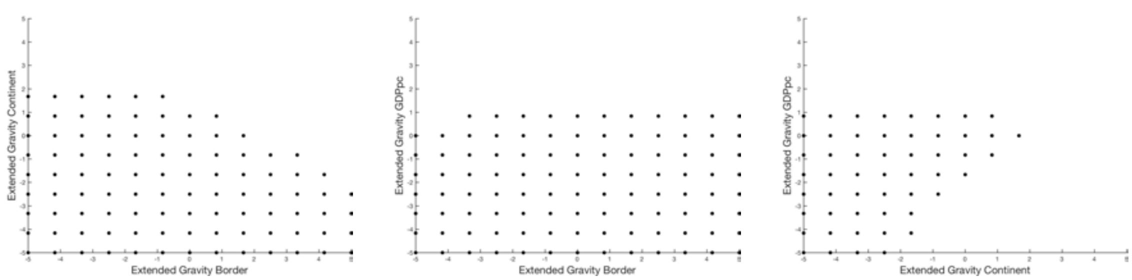


Figure D.10: Projected Confidence Set: $(\sigma_\nu^r, \sigma_u^r) = (0, 2)$

(a) Border vs. Continent (b) Border vs. Language (c) Continent vs. Language