What do Exporters Know?*

Michael J. Dickstein

New York University and NBER

Eduardo Morales

Princeton University and NBER

May 5, 2018

Abstract

Much of the variation in international trade volume is driven by firms' extensive margin decisions of whether to participate in export markets. We evaluate how the information potential exporters possess influences their decisions. To do so, we estimate a model of export participation in which firms weigh the fixed costs of exporting against the forecasted profits from serving a foreign market. We adopt a moment inequality approach, placing weak assumptions on firms' expectations. The framework allows us to test whether firms differ in the information they have about foreign markets. We find that larger firms possess better knowledge of market conditions in foreign countries, even when those firms have not exported in the past. Quantifying the value of information, we show that, in a typical destination, total exports rise while the number of exporters falls when firms have access to better information to forecast export revenues.

Keywords: export participation, demand under uncertainty, discrete choice methods, moment inequalities

---

*We are grateful to the editor and five anonymous referees for helpful comments and suggestions. We also thank Tim Bresnahan, Lorenzo Caliendo, Juan Manuel Castro, Jan De Loecker, Dave Donaldson, Jonathan Eaton, Liran Einav, Alon Eizenberg, Guido Imbens, Michal Kolesar, Ariel Pakes, Esteban Rossi-Hansberg, James Tybout and seminar participants at the CEPR-JIE conference on Applied Industrial Organization, Columbia University, Dallas Fed, Dartmouth College, LMU, MIT, McGill University, New York University, the NBER ITI meeting, University of Notre Dame, Northwestern University, Pennsylvania State University, Princeton University, the Stanford/Berkeley IO Fest, Stanford University, UCLA, University of Maryland, University of Minnesota, University of Pennsylvania, University of Virginia, and Yale University for helpful suggestions. The Appendix referenced in the text is available as an Online Appendix. Email: michael.dickstein@nyu.edu, ecmorale@princeton.edu.

1 Introduction

In 2014, approximately 300,000 US firms chose to export to foreign markets (Department of Commerce, 2016). The decision of these firms to sell abroad drives much of the variation in trade volume from the US (Bernard et al., 2010). Thus, to predict how exports may change with lower trade costs, exchange rate movements, or other policy or market fluctuations, researchers need to understand firms' decisions of whether to participate in export markets.

A large literature in international trade focuses on modeling firms' export decisions.[1] Empirical analyses of these decisions, however, face a data obstacle: the decision to export depends on a firm's expectation of the profits it will earn when serving a foreign market, which the researcher rarely observes. Absent direct data on firms' expectations, researchers must impose assumptions on how firms form these expectations. For example, researchers commonly assume firms' expectations are rational and depend on a set of variables observed in the data. The precise specification of agents' information, however, can influence the overall measurement, as Manski (1993, 2004), and Cunha and Heckman (2007) show in the context of evaluating the returns to schooling. In an export setting, the assumptions on expectations may affect both the estimates of the costs firms incur when exporting and predictions of how firms will respond to counterfactual changes in these trade costs.

In this paper, we first document that estimates of the parameters underlying firms' export decisions depend heavily on how researchers specify the firm's expectations. We compare the predictions of a standard model in the international trade literature (Melitz, 2003) under two specifications: the "perfect foresight" case, under which we assume firms perfectly predict observed profits from exporting, and a minimal information case, under which we assume firms use a specific set of observed variables to predict their export profits. For each case, we recover the fixed costs of exporting and the mean profits of firms predicted to export. Finding important differences in the predictions from the two models, we then estimate an empirical model of export participation that places fewer restrictions on firms' expectations.

Under our alternative approach, we do not require the researcher to have full knowledge of an exporter's information set. Instead, the researcher need only specify a subset of the variables that agents use to form their expectations. The researcher must observe this subset, but need not observe any remaining variables that affect the firm's expectations. The set of unobserved variables may vary flexibly across firms, markets, and years.

The trade-off from specifying only a subset of the firm's information is that we can only partially identify the parameters of interest. To do so, we develop a new type of moment inequality, the odds-based inequality, and combine it with inequalities based on revealed preference. Using these inequalities, we show that placing fewer assumptions on expectations

[1] See Melitz (2003), Das et al. (2007), Eaton et al. (2008), Arkolakis (2010), Moxnes (2010), Eaton et al. (2011), Eaton et al. (2014), Arkolakis et al. (2015), Cherkashin et al. (2015) and Ruhl and Willis (2017). The literature has also recently focused on the decisions of importers (Blaum et al., 2017; Antràs et al., 2017).

affects the measurement of the parameters of the exporter's problem. Further, we show our approach generates bounds on these parameters that are tight enough to be informative.

This paper makes four main contributions. First, we demonstrate the sensitivity of the estimated export fixed costs to assumptions the researcher imposes on firms' export profit forecasts. Second, we employ moment inequalities to partially identify the exporter's fixed costs under weak assumptions, applying insights from Pakes (2010) and Pakes et al. (2015). Third, we address the question of "what do exporters know?". We show that, under rational expectations, our moment inequality framework allows us to test whether potential exporters know and use specific variables to predict their export profits. Finally, fourth, we use our model's estimates to quantify the value of information.

To illustrate the sensitivity of export fixed estimates to the researcher's assumptions on exporters' information, we start by estimating a perfect foresight model under which firms perfectly predict the profits they will earn upon entry. Using maximum likelihood, we find fixed costs in the chemicals sector from Chile to Argentina, Japan, and the United States to equal \$868,000, \$2.6 million, and \$1.6 million, respectively. We compare these estimates to an alternative approach, suggested in Manski (1991) and Ahn and Manski (1993), in which we assume that firms' expectations are rational and we specify the variables firms use to form their expectations. Specifically, we assume that firms know only three variables: distance to the export market, aggregate exports from Chile to that market in the prior year, and their own domestic sales in the prior year. We estimate fixed costs of exporting under this approach that are 40-60% smaller than those found under the perfect foresight assumption.

That the fixed cost estimates differ under the two approaches reflects a bias in the estimation. Both require the researcher to specify the content of the agent's information set. If firms actually employ a different set of variables–containing more information or less– to predict their potential export profits, the estimates of the model parameters will generally be biased. Specifically, if the researcher wrongly assumes that firms have perfect foresight, the bias arises for a similar reason to the bias affecting Ordinary Least Squares estimates in linear models when a covariate contains classical measurement error; we show that, in our setting, this bias leads the researcher to overestimate fixed export costs. Thus, we move to employ moment inequalities to partially identify the exporter's fixed costs under weaker assumptions.

Here, we again assume that firms know the distance to the export market, the aggregate exports to that market in the prior year, and their own domestic sales from the prior year. However, unlike the minimal information approach described earlier, the inequalities we define do not restrict firms to use only these three variables when forecasting their potential export profits. We require only that firms know at least these variables. Using our inequalities, we find much lower fixed costs, representing only 10-15% of the perfect foresight values.

Comparing these findings to those in the existing literature is not simple. Our baseline model abstracts from other possible sources of bias, including marketing costs (Arkolakis, 2010), adjustment costs (Ruhl and Willis, 2017), persistent unobserved heterogeneity (Roberts and Tybout, 1997; Das et al., 2007), and buyer-seller relationships (Eaton et al., 2016; Bernard et al., 2017). In extensions to our baseline model, we account for path dependence in export status, as in Das et al. (2007), and allow firms to decide which markets to enter in reaction to unobserved (to the researcher) firm-country specific revenue shocks, as in Eaton et al. (2011).

We next employ our framework to investigate the set of information potential exporters use to forecast export revenues. We run alternative versions of our moment inequality model, holding fixed the model and data but varying the firm's presumed information set. Using the specification tests described in Bugni et al. (2015), we look for evidence against the null that potential exporters use particular variables in their forecasts.

We begin by testing our baseline assumption that exporters know *at least* distance, their own lagged domestic sales, and lagged aggregate exports when making their export decisions. Using data from both the chemicals and food sectors, we cannot reject this null hypothesis. We then test (a) whether firms have perfect foresight about their potential export profits in every country, and (b) whether firms have information on last period's realizations of a destination-time period specific shifter of firms' export revenues that, according to our model, is a sufficient statistic for the effect of all foreign market characteristics (i.e. market size, price index, trade costs and demand shifters) on these revenues. In both sectors, we reject the null that firms have perfect foresight. For the market-specific revenue shifters, we find interesting heterogeneity: we fail to reject that large firms know these shocks, but reject that small firms do. This distinction is not driven by prior export experience. That is, even when we focus only on large firms that did not to export in the previous year, we nonetheless fail to reject the null that these firms use knowledge of past revenue shifters when forecasting their potential export revenue. Large firms therefore have not only a productivity advantage over small firms, but also an informational advantage in foreign markets.

Finally, we use our model's estimates to quantify the value of information. Using our estimated bounds on fixed export costs, we compute counterfactual entry decisions, firm-level profits, and aggregate exports to each destination and in each year under different firm information sets. We find that, as we provide information to potential exporters, these firms choose to export to fewer markets: in the chemicals sector, the expected number of firm-destination pairs with positive export flows *decreases* between 3.5% and 5.7%. Interestingly, although the total number of firm-destination pairs decreases, the overall (aggregated across firms and destinations) export revenue in the sector *increases* between 6.4% and 9.5%. Were all firms able to access information on past export revenue shifters, fewer firms would make mistakes when choosing export markets and, consequently, the average firm's ex post profits in a typical market would increase between 17.5% and 20.6%. In comparison, with information to predict export revenues perfectly, the average firm's ex post profits in a typical market would increase between 46% and 52.9%.

We demonstrate our contributions using the exporter's problem. However, our estimation approach can apply more broadly to discrete choice decisions that depend on agents' forecasts of key payoff-relevant variables. For example, to determine whether to invest in R&D projects, firms must form expectations about the success of the research activity (Aw et al., 2011; Doraszleski and Jaumandreu, 2013; Bilir and Morales, 2016). When a firm develops a new product, it must form expectations of its future demand (Bernard et al., 2010; Bilbiie et al., 2012; Arkolakis et al., 2015). Firms deciding whether to enter health insurance markets must form expectations about the type of health risks that will enroll in their plans (Dickstein et al., 2015). Firms paying fixed costs to import from foreign markets must form expectations about the sourcing potential of these markets (Blaum et al., 2017; Antràs et al., 2017). Finally, in education, the decision to attend college crucially depends on potential students' expectations about earnings with and without a college education (Freeman, 1971; Willis and Rosen, 1979; Manski and Wise, 1983). In these settings, even without direct elicitation of agents' preferences (Manski, 2004), our approach allows the researcher both to test whether certain covariates belong to the agent's information set and to recover bounds on the economic primitives of the agent's problem without imposing strong assumptions on her expectations.

Our estimation approach contributes to a growing empirical literature that employs moment inequalities derived from revealed preference arguments, including Ho (2009), Holmes (2011), Crawford and Yurukoglu (2012), Ho and Pakes (2014), Eizenberg (2014), Morales et al. (2017), Wollman (2017), and Maini and Pammolli (2017). We follow a methodology closest to Morales et al. (2017), but add two features. First, we introduce inequalities in a setting with structural errors that are specific to each observation. The cost of allowing this flexibility is that we must assume a distribution for these structural errors, up to a scale parameter. We also cannot handle large choice sets, such as those considered in Morales et al. (2017). Second, we combine the revealed preference inequalities employed in the prior literature with our new odds-based inequalities, to gain identification power.

We proceed in this paper by first describing our model of firm exports in Section 2. We describe our data in Section 3. In sections 4 and 5, we discuss three alternative estimation approaches and compare the resulting parameter estimates. In sections 6 and 7, we use our moment inequalities both to test alternative information sets and to conduct counterfactuals on the value of information. In Section 8, we discuss extensions of our baseline model. Section 9 concludes. All appendix sections referenced below appear in an Online Appendix.

2 Empirical Model

We model firms' export decisions. All firms located in a country h choose whether to sell in each export market j. We index the firms located in h and active at period t by $i = 1, \dots, N_t$.[2]

[2] We eliminate the subindex for the country of origin h when possible to simplify notation.

We index the potential destination countries by $j = 1, \ldots, J$.

We model firms' export decisions using a two-period model. In the first period, firms choose the set of countries to which they wish to export. To participate in a market, firms must pay a fixed export cost.[3] When choosing among export destinations, firms may differ in their degree of uncertainty about the profits they will obtain upon exporting. In the second period, conditional on entering a foreign market, all firms acquire the information needed to set their prices optimally and obtain the corresponding export profits.

2.1 Demand, Supply, Market Structure, and Information

Firms face isoelastic demand in every country: $x_{ijt} = \zeta_{ijt}^{1-\eta} p_{ijt}^{-\eta} P_{jt}^{\eta-1} Y_{jt}$. Here, the quantity demanded x_{ijt} depends on: p_{ijt}, the price firm i sets in destination j at t; Y_{jt}, the total expenditure in the sector in which i operates; P_{jt}, a price index that captures the competition firm i faces in market j from other firms selling in the market; and, ζ_{ijt}, a demand shifter.

Firm i produces one unit of output with a constant marginal cost c_{it}.[4] When firm i chooses to sell in a market j, it must pay two export costs: a variable cost, τ_{ijt}, and a fixed cost, f_{ijt}. We adopt the "iceberg" specification of variable export costs and assume that firm i must ship τ_{ijt} units of output to country j for one unit to arrive. The total marginal cost for firm i of selling one unit in country j at period t is thus $\tau_{ijt} c_{it}$. Fixed costs f_{ijt} are paid by firms selling a positive amount in market j at period t, independently of the actual quantity exported.

We denote the firm's potential sales revenue in market j and period t as $r_{ijt} \equiv x_{ijt} p_{ijt}$, and use $\mathcal{J}_{ijt}$ to denote the information firm i possesses about its potential revenue r_{ijt} when deciding whether to participate in market j at t. We assume firm i knows the determinants of fixed costs f_{ijt} for every country j when deciding whether to export. Therefore, if relevant to predict r_{ijt}, these determinants of fixed costs will also enter $\mathcal{J}_{ijt}$.

2.2 Export Revenue

Upon entering a market, a firm observes both η and its marginal cost of selling in this market, and sets its price optimally taking other sellers' prices as given: $p_{ijt} = (\eta/(\eta - 1)) \tau_{ijt} c_{it}$. Thus, the revenue firm i would obtain if it were to sell in market j at period t is:

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

We can write an analogous expression for the sales revenue in the domestic market h. As we

[3] In Section 8.1, we consider a fully dynamic export model in which forward-looking firms must also pay a sunk export entry cost as in Das et al. (2007).

[4] The assumption of constant marginal costs is necessary for the export decision to be independent across markets. See Vannoorenberghe (2012), Blum et al. (2013), and Almunia et al. (2018) for models of firms' export decisions with increasing marginal costs.

show in Appendix A.1, taking the ratio of export revenue to domestic revenue for each firm in year t, we can rewrite potential export revenues in a destination market j as

$$r_{ijt} = \alpha_{ijt} r_{iht}, \qquad \text{with} \qquad \alpha_{ijt} \equiv \left(\frac{\zeta_{iht}\tau_{ijt}}{\zeta_{ijt}\tau_{iht}} \frac{P_{ht}}{P_{jt}} \right)^{1-\eta} \frac{Y_{jt}}{Y_{ht}}.$$

Here, α_{ijt} is a firm-destination-year specific shifter of export revenues that accounts for the destination's market size, price index, and the effect of variable trade costs and demand shocks across firms. We can further split this shifter into a component common to firms in a given market and year, and a component that varies across firms:

$$r_{ijt} = \alpha_{jt} r_{iht} + e_{ijt}, \qquad \text{where} \qquad \alpha_{jt} \equiv \mathbb{E}_{jt} \left[\left(\frac{\zeta_{iht}\tau_{ijt}}{\zeta_{ijt}\tau_{iht}} \frac{P_{ht}}{P_{jt}} \right)^{1-\eta} \frac{Y_{jt}}{Y_{ht}} \right], \tag{2}$$

where $\mathbb{E}_{jt}[\cdot]$ denotes the mean across firms in a given country-year pair jt. The term e_{ijt} accounts for firm-market-year specific relative revenue shocks. We assume firms do not know these shocks when deciding whether to export to market j at period t:[5]

$$\mathbb{E}_{jt}[e_{ijt} | \mathcal{J}_{ijt}, r_{iht}, f_{ijt}] = 0. \tag{3}$$

Conversely, we do not restrict the relationship between the information set $\mathcal{J}_{ijt}$ and the component of revenue $\alpha_{jt} r_{iht}$. Thus, for example, more productive firms may be systematically better informed than less productive firms about variables affecting their future domestic sales, r_{iht}, or about the country-year export shifters accounted for by the term α_{jt}. Similarly, we allow firms to have more information about markets that are closer to the domestic market.

2.3 Export Profits

We model the export profits that i would obtain in j if it were to export at t as

$$\pi_{ijt} = \eta^{-1} r_{ijt} - f_{ijt}. \tag{4}$$

We model fixed export costs as

$$f_{ijt} = \beta_0 + \beta_1 dist_j + \nu_{ijt}, \tag{5}$$

where $dist_j$ denotes the distance from country h to country j, and the term ν_{ijt} represents determinants of f_{ijt} that the researcher does not observe. As discussed in Section 2.1, we

[5] Appendix A.2 describes a set of assumptions on the distribution of demand shifters ζ_{ijt} and variable costs τ_{ijt} under which the mean independence condition in equation (3) holds. Furthermore, we extend the model in Section 8.2 to allow for firm-country-year specific export revenue shocks that are known to the firm when it decides whether to export: $r_{ijt} = \alpha_{jt} r_{iht} + e_{ijt} + \omega_{ijt}$ with $\mathbb{E}_{jt}[\omega_{ijt} | \mathcal{J}_{ijt}] = \omega_{ijt}$.

assume that firms know f_{ijt} when deciding whether to export to j at t.[6]

The estimation procedure introduced in Section 4.2 requires ν_{ijt} to be distributed independently of $\mathcal{J}_{ijt}$, and its distribution to be known up to a scale parameter. To match one typical binary choice model, we assume ν_{ijt} follows a normal distribution and is independent of other export determinants:[7]

$$\nu_{ijt}|(\mathcal{J}_{ijt}, dist_j) \sim \mathbb{N}(0, \sigma^2). \tag{6}$$

The assumed independence between ν_{ijt} and $\mathcal{J}_{ijt}$ implies that knowledge of ν_{ijt} is irrelevant to compute the firm's expected export revenue. However, we impose no assumption on the relationship between $\mathcal{J}_{ijt}$ and the observed determinants of fixed export costs, here $dist_j$.

2.4 Decision to Export

A risk-neutral firm i will decide to export to j in year t if and only if $\mathbb{E}[\pi_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] \geq 0$, where the vector $(\mathcal{J}_{ijt}, dist_j, \nu_{ijt})$ includes any variables firm i uses to predict potential export profits in country j. Combining equations (4) and (5), we can write

$$\mathbb{E}[\pi_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt}. \tag{7}$$

Here, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, following our definition of $\mathcal{J}_{ijt}$ as the set of variables firm i uses to predict r_{ijt}. Given the expression for r_{ijt} in equations (2) and (3), we write:

$$\mathbb{E}[\pi_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = \eta^{-1}\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt}. \tag{8}$$

Let $d_{ijt} = \mathbb{1}\{\mathbb{E}[\pi_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] \geq 0\}$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function. From equation (8), we can write:

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\}, \tag{9}$$

and, given equations (6) and (9), we can write the probability that i exports to j at t conditional on $\mathcal{J}_{ijt}$ and $dist_j$:

$$\mathcal{P}(d_{ijt} = 1|\mathcal{J}_{ijt}, dist_j) = \int_\nu \mathbb{1}\{\eta^{-1}\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu \geq 0\}(1/\sigma)\phi(\nu/\sigma)d\nu$$

[6] At a computational cost, we can allow fixed export costs to depend on additional variables the firm knows when deciding whether to export to a market, such as shared language (Morales et al., 2017) and the quality of institutions (Antràs et al., 2017). In Appendix B.3, we generalize the specification in equation (5) and present estimates for a model in which we assume $f_{ijt} = \beta_j + \nu_{ijt}$, where β_j varies freely across countries. In Appendix A.11, we discuss an extension in which firms face unexpected shocks to fixed costs.

[7] The assumption that ν_{ijt} is distributed normally is a sufficient but not a necessary condition to derive our moment inequalities. We provide the precise requirements for the distribution of ν_{ijt} when we derive the inequalities in Sections 4.2.1 and 4.2.2.

$$= \Phi\big(\sigma^{-1}\big(\eta^{-1}\mathbb{E}[\alpha_{jt} r_{iht} | \mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big), \tag{10}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the standard normal probability density function and cumulative distribution function.[8] Equation (10) indicates that, after integrating over the unobserved heterogeneity in fixed costs, ν_{ijt}, we can write the probability that firm i exports to country j at period t as a probit model whose index depends on firm i's expectation of the revenue it will earn in j at t upon entry. The key hurdle in estimation, which we discuss in Section 4, is that researchers rarely observe these expectations.

From equation (10), even if the researcher were to observe firms' actual expectations, data on export choices alone would not allow us to identify the scale of the remaining parameter vector $(\sigma, \eta, \beta_0, \beta_1)$. To normalize for scale in export models, researchers typically use additional data to estimate the demand elasticity η (Das et al., 2007). In our estimation, we set $\eta = 5$.[9] For simplicity of notation, we use $\theta \equiv (\theta_0, \theta_1, \theta_2)$ to denote the remaining parameter vector and $\theta^* \equiv (\beta_0, \beta_1, \sigma)$ to denote its true value, as determined by equation (10).

3 Data

Our data come from two separate sources. The first is an extract of the Chilean customs database, which covers the universe of exports of Chilean firms from 1995 to 2005. The second is the Chilean Annual Industrial Survey (*Encuesta Nacional Industrial Anual*, or ENIA), which surveys all manufacturing plants with at least 10 workers. We merge these two datasets using firm identifiers, allowing us to observe both the export and domestic activity of each firm.[10]

The firms in our dataset operate in one of two sectors: the manufacture of chemicals and the food products sector.[11] For each sector, we estimate our model restricting the set of countries to those served by at least five Chilean firms in all years of our data. This restriction leaves 22 countries in the chemicals sector and 34 countries in the food sector.

We observe 266 unique firms across all years in the chemicals sector; on average, 38% of these firms participate in at least one export market in a given year. In Table 1, we report

[8] If knowledge of $dist_j$ helps predict r_{ijt}, then $dist_j \in \mathcal{J}_{ijt}$ and $\mathcal{P}(d_{ijt} = 1 | \mathcal{J}_{ijt}, dist_j) = \mathcal{P}(d_{ijt} = 1 | \mathcal{J}_{ijt})$.

[9] This value is within the range of values in the literature (Simonovska and Waugh, 2014; Head and Mayer, 2014). Given our model, one can estimate η using data on firms' total sales and variable costs (Das et al., 2007; Antràs et al., 2017). We do not implement this estimation approach given limitations in our measure of variable costs. When presenting our estimates, we indicate which conclusions are sensitive to η.

[10] We aggregate the information from ENIA across plants to obtain firm-level information to match to the customs data. ENIA sometimes identifies firms as exporters when we do not observe exports in the customs data; in these cases, we follow the customs database and treat these firms as non-exporters. We lose a number of small firms in the merging process because, as indicated in the main text, ENIA only covers plants with more than 10 workers. The remaining firms account for roughly 80% of total export flows.

[11] The chemicals sector (sector 24 of the ISIC rev. 3.1) includes firms involved in the manufacture of chemicals and chemical products, including basic chemicals, fertilizers and nitrogen compounds, plastics, synthetic rubber, pesticides, paints, soap and detergents, and manmade fibers. The food sector (sector 151 of the ISIC rev. 3.1) includes the production, processing, and preservation of meat, fish, fruit, vegetables, oils, and fats.

the mean firm-level exports in this sector, which are \$2.18 million in 1996 and grow to \$3.58 million in 2005, with a dip in 2001 and 2002.[12] The median level of exports is much lower, at around \$150,000. In the food sector, we observe 372 unique firms, 30% of which export in a typical year. The mean exporter in this sector sells \$7.7 million, while the median exporter sells approximately \$2.24 million abroad. In the chemicals sector, the average exporter serves 4-5 countries. Firms in the food sector typically export to 6-7 markets on average.

Table 1: Summary Statistics

Year	Share of exporters	Exports per exporter (mean)	Exports per exporter (med)	Domestic sales per firm (mean)	Domestic sales per exporter (mean)	Destinations per exporter (mean)
			Chemical Products			
1996	35.7%	2.18	0.15	13.23	23.10	4.24
1997	36.1%	2.40	0.19	13.29	22.99	4.54
1998	42.5%	2.41	0.17	14.31	22.25	4.35
1999	38.7%	2.60	0.19	14.43	23.95	4.53
2000	37.6%	2.55	0.21	14.41	25.93	4.94
2001	39.8%	2.35	0.12	12.89	21.92	4.68
2002	38.7%	2.37	0.15	13.25	23.73	4.95
2003	38.0%	3.08	0.17	10.41	19.54	5.11
2004	37.6%	3.27	0.15	10.05	18.70	5.17
2005	38.0%	3.58	0.11	12.50	21.65	5.19
			Food			
1996	30.1%	7.47	2.59	9.86	13.68	5.93
1997	33.1%	6.97	2.82	10.56	15.32	6.23
1998	33.3%	7.49	2.86	10.05	14.80	6.34
1999	32.3%	6.71	2.37	9.67	14.88	6.74
2000	30.6%	6.49	2.21	8.44	13.33	5.93
2001	28.0%	6.48	1.74	8.70	14.08	6.09
2002	27.2%	7.82	2.01	7.83	13.59	6.86
2003	29.8%	7.60	1.68	7.15	12.79	6.15
2004	28.5%	9.25	1.68	8.05	13.85	6.69
2005	25.8%	10.72	2.43	9.88	16.27	7.05

Notes: All variables (except “share of exporters”) are reported in millions of year 2000 US dollars.

Our data set includes both exporters and non-exporters. Furthermore, we use an unbalanced panel that includes not only those firms that appear in ENIA in every year between 1995 and 2005 but also those that were created or disappeared during this period. Finally, we obtain information on the distance from Chile to each destination market from CEPII.[13]

4 Empirical Approach

In the model we describe in Section 2, r_{ijt}, firm i's potential export revenue to market j at t, is a function of its own marginal costs and demand shifter, and of country j's market size

[12]The revenue values we report are in year 2000 US dollars.
[13]Mayer and Zignago (2011) provide a detailed explanation of the content of this database.

and price index. In Section 2.2, we split these determinants into two terms: $\alpha_{jt} r_{iht}$ and e_{ijt}, where the latter reflects idiosyncratic shifters of firm i's demand and variable trade costs in j. Crucially, while the data described in Section 3 allow us to compute a consistent estimate of $\alpha_{jt} r_{iht}$ for every firm, market and year (see Appendix A.3), e_{ijt} is not observed for all firm-market-year triplets. We thus henceforth refer to the first term as the *observable* determinant of export revenue and label it $r^o_{ijt} = \alpha_{jt} r_{iht}$. We label e_{ijt} as the *unobservable* determinant.[14]

Our model implies that $\mathbb{E}[e_{ijt}|\mathcal{J}_{ijt}] = 0$ and, thus, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$. The model does not restrict the relationship between $\mathcal{J}_{ijt}$ and r^o_{ijt}. However, identifying the parameter vector θ underlying the fixed export costs f_{ijt} requires additional assumptions (Manski, 1993).

First, we consider a perfect foresight model. With this model, researchers assume an information set $\mathcal{J}^a_{ijt}$ for potential exporters such that $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] = r^o_{ijt}$. That is, firms are assumed to have ex ante (before deciding whether to enter a foreign market) the same information that the researcher has ex post (when data becomes available). Thus, firms predict r^o_{ijt} perfectly.[15]

Second, we consider a model in which we allow firms to face uncertainty when predicting r^o_{ijt}—for example, they may lack perfect knowledge of the size of the market or the degree of competition they will face. In this model, potential exporters forecast their export revenues in every foreign market using information on three variables: (1) their own domestic sales in the previous year, r_{iht-1}; (2) sectoral aggregate exports to destination j in the previous year, R_{jt-1}; and (3) distance from the home country to j, $dist_j$. That is, we assume that the actual information set $\mathcal{J}_{ijt}$ is identical to a vector of covariates $\mathcal{J}^a_{ijt}$ observed in our data: $\mathcal{J}^a_{ijt} = (r_{iht-1}, R_{jt-1}, dist_j)$. In practice, firms can easily access these three variables in any year. However, this information set is likely to be strictly smaller than the actual information set firms possess when deciding whether to export.[16] Furthermore, specifying $\mathcal{J}_{ijt}$ as in this second model implies that all firms base their entry decision on the same set of covariates. It does not permit firms to differ in the information they use.

Third, we discuss how to identify the model parameters imposing weaker assumptions on the information firms use to predict r^o_{ijt}. We propose a moment inequality estimator that can handle settings in which the econometrician observes only a *subset* of the elements contained in firms' true information sets. That is, we assume that the researcher observes a vector Z_{ijt} such that $Z_{ijt} \subseteq \mathcal{J}_{ijt}$. The researcher need not observe the remaining elements in $\mathcal{J}_{ijt}$. Those unobserved elements of firms' information sets can vary flexibly by firm and by export market.

[14] As an alternative micro-foundation for this structure, one can rule out firm-specific shifters of demand and variable trade costs, and instead assume e_{ijt} reflects error in the researcher's observation of r^o_{ijt}.

[15] Although we denote this case as "perfect foresight", perfectly predicting export revenues only refers to the observable component, r^o_{ijt}. Firms' information sets are still orthogonal to the unobserved component e_{ijt}.

[16] When we indicate that information set $\mathcal{J}^a_{ijt}$ is smaller than information set $\mathcal{J}_{ijt}$, we formally mean that the distribution of $\mathcal{J}^a_{ijt}$ conditional on $\mathcal{J}_{ijt}$ is degenerate.

4.1 Perfect Knowledge of Exporters' Information Sets

Under the assumption that the econometrician's specified information set, $\mathcal{J}^a_{ijt}$, equals the firm's true information set, $\mathcal{J}_{ijt}$, $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] = \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ and one can estimate θ^* as the value of the unknown parameter θ that maximizes the log-likelihood function

$$\mathcal{L}(\theta|d, \mathcal{J}^a, dist) =$$
$$\sum_{i,j,t} d_{ijt} \ln(\mathcal{P}(d_{jt} = 1|\mathcal{J}^a_{ijt}, dist_j; \theta)) + (1 - d_{ijt}) \ln(\mathcal{P}(d_{jt} = 0|\mathcal{J}^a_{ijt}, dist_j; \theta)), \tag{11}$$

where the vector $(d, \mathcal{J}^a, dist)$ includes all values of the corresponding covariates for every firm, country and year in the sample, and, according to equation (10) and the definition of r^o_{ijt},

$$\mathcal{P}(d_{jt} = 1|\mathcal{J}^a_{ijt}, dist_j; \theta) = \Phi\big(\theta_2^{-1}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \theta_0 - \theta_1 dist_j\big)\big). \tag{12}$$

To use equations (11) and (12) to estimate θ^*, one first needs to compute $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}]$. When the researcher assumes perfect foresight, $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] = r^o_{ijt}$. When the researcher assumes $\mathcal{J}^a_{ijt}$ is equal to a set of observed covariates, one can consistently estimate $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}]$ as the non-parametric projection of r^o_{ijt} on $\mathcal{J}^a_{ijt}$.[17] The key assumption underlying these two procedures is that the researcher correctly specifies the agent's information set.

Bias in estimation will generally arise when the agent's true information set, $\mathcal{J}_{ijt}$, differs from the researcher's specification, $\mathcal{J}^a_{ijt}$, for some firms, countries or years in the sample. To characterize this bias, we begin by defining two types of errors: the agent's expectational error and the researcher's specification error. For the agent, we define $\varepsilon_{ijt} \equiv r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ as the true expectational error that firm i makes when predicting the observed component of its export revenue. This error reflects the firm's uncertainty about r^o_{ijt}.[18] In contrast, we denote the difference between firms' true expectations and the researcher's proxy as ξ_{ijt}:

$$\xi_{ijt} \equiv \mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]. \tag{13}$$

Whenever this error term differs from zero, estimates based on equations (11) and (12) will be biased. In Appendix D, we present simulation results that illustrate the direction and magnitude of the bias that arise in three cases: when the researcher assumes perfect foresight, when the researcher's information set is larger than the firm's information set, and when the researcher's information set is smaller than the firm's information set.

To provide intuition on the direction of the bias, we focus here on the perfect foresight case. In this case, we find an upward bias in the estimates of the fixed costs parameters β_0, β_1 and σ. The upward bias arises for a similar reason to the attenuation bias that

[17] See Manski (1991) and Ahn and Manski (1993) for additional details on this two-step estimation approach.
[18] The total expectational error that the firm makes when forecasting export revenue r_{ijt} is $\varepsilon_{ijt} + e_{ijt}$.

affects Ordinary Least Squares estimates in linear models when a covariate contains classical measurement error (see Wooldridge, 2002). Under perfect foresight, the researcher assumes the firm perfectly predicts the observable part of its export revenue, such that $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] = r^o_{ijt}$. Thus, the measurement error affecting the researcher's specification, $\xi_{ijt} \equiv r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, is the same as the firm's true expectational error, ε_{ijt}. Rational expectations implies that firms' expectational errors are mean independent of their true expectation and, therefore, correlated with the ex-post realization of the variable being predicted; i.e. rational expectations implies that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}] = 0$ and $cov(\varepsilon_{ijt}, r^o_{ijt}) \neq 0$. Thus, if we were in a linear regression setting, wrongly assuming perfect foresight and using r^o_{ijt} as a regressor instead of the unobserved expectation, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, would generate a downward bias on the coefficient on r^o_{ijt}.

The probit model in equation (12) differs from this linear setting in two dimensions. First, our normalization by scale $\eta = 5$ sets the coefficient on the covariate measured with error, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, to a given value. Thus, the bias generated by the correlation between the expectational error, ε_{ijt}, and the realized export revenue, r^o_{ijt}, will be reflected in an upward bias in the estimates of the remaining parameters β_0, β_1 and σ. Second, the direction of the bias depends not only on the correlation between ε_{ijt} and r^o_{ijt} but also on the functional form of the distribution of unobserved expectations and the expectational error.[19]

4.2 Partial Knowledge of Exporters' Information Sets

In most empirical settings, researchers rarely observe the exact covariates that form the firm's information set. However, they can typically find a vector of covariates in their data that represents a subset of the firm's information set. For example, in each year, exporters will likely know past values of both their domestic sales, r_{iht-1}, and the aggregate exports from their home country to each destination market, R_{jt-1}; the former appears in firms' accounting statements, while the latter appears in publicly available trade data. Similarly, firms can easily obtain information on the distance to each destination, $dist_j$. Thus, while $(r_{iht-1}, R_{it-1}, dist_j)$ might not reflect firms' complete information, they likely know at least this vector.

In this section, we show how to proceed in estimation using a vector of observed covariates Z_{ijt} that is a subset of the information firms use to forecast export revenues, i.e. $Z_{ijt} \subseteq \mathcal{J}_{ijt}$. We show how to test formally whether firms possess this information in Section 6. We form two types of moment inequalities to partially identify θ^*.[20]

[19]If both firms' true expectations and expectational errors are normally distributed, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] \sim \mathbb{N}(0, \sigma^2_e)$ and $\varepsilon_{ijt}|(\mathcal{J}_{ijt}, \nu_{ijt}) \sim \mathbb{N}(0, \sigma^2_\varepsilon)$, one can apply the results in Yatchew and Griliches (1985) and conclude that there is an upward bias in the estimates of β_0, β_1 and σ. This bias increases in the variance of the expectational error, σ^2_ε, relative to the variance of the true expectations, σ^2_e. When either firms' true expectations, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, or the expectational error, ε_{ijt}, are not normally distributed, there is no analytic expression for the bias. However, our simulations in Appendix D illustrate that the upward bias in the estimates of all elements of θ^* generally persists under different distributions of $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ and ε_{ijt}.

[20]As shown in Appendix A.4, given the model described in Section 2, the assumption that the researcher observes a subset of a firm's true information set is not strong enough to point-identify θ^*. Whether the bounds

4.2.1 Odds-based Moment Inequalities

For any $Z_{ijt} \subseteq (\mathcal{J}_{ijt}, dist_j)$, we define the conditional odds-based moment inequalities as

$$\mathcal{M}^{ob}(Z_{ijt};\theta) = \mathbb{E}\left[\begin{array}{c} m_l^{ob}(d_{ijt}, r_{ijt}^o, dist_j;\theta) \\ m_u^{ob}(d_{ijt}, r_{ijt}^o, dist_j;\theta) \end{array} \middle| Z_{ijt}\right] \geq 0, \tag{14a}$$

where the two moment functions are defined as

$$m_l^{ob}(\cdot) = d_{ijt}\frac{1-\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}^o - \theta_0 - \theta_1 dist_j\big)\big)}{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}^o - \theta_0 - \theta_1 dist_j\big)\big)} - (1-d_{ijt}), \tag{14b}$$

$$m_u^{ob}(\cdot) = (1-d_{ijt})\frac{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}^o - \theta_0 - \theta_1 dist_j\big)\big)}{1-\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}^o - \theta_0 - \theta_1 dist_j\big)\big)} - d_{ijt}. \tag{14c}$$

We denote the set of all possible values of the parameter vector θ as Θ and the subset of those values consistent with the conditional moment inequalities described in equation (14) as Θ_0^{ob}.

Theorem 1 *Let $\theta^* = \big(\beta_0, \beta_1, \sigma\big)$ be the parameter defined by equation* (10). *Then $\theta^* \in \Theta_0^{ob}$.*

Theorem 1 indicates that the odds-based inequalities are consistent with the true value of the parameter vector, θ^*. We provide here an intuitive explanation of Theorem 1. The formal proof appears in Appendix C.1.

We focus on the intuition behind the moment function in equation (14c); the intuition for equation (14b) is analogous. From the definition of d_{ijt} in equation (9) and the definition of r_{ijt}^o, we can write

$$\mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}^o|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\} - d_{ijt} = 0. \tag{15}$$

This equation, using revealed preference, implies the condition that expected export profits are positive, $\eta^{-1}\mathbb{E}[r_{ijt}^o|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0$, is both necessary and sufficient for observing firm i exporting to country j in year t, $d_{ijt} = 1$. Equation (15) cannot be used directly for identification, as it depends on the unobserved terms ν_{ijt} and $\mathcal{J}_{ijt}$. To account for the term ν_{ijt}, we take the expectation of equation (15) conditional on $(\mathcal{J}_{ijt}, dist_j)$. Given the distributional assumption in equation (6), we use simple algebraic transformations to rewrite the resulting equality as

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}^o|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}^o|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))} - d_{ijt}\middle|\mathcal{J}_{ijt}, dist_j\right] = 0. \tag{16}$$

If we write this equality as a function of the unknown parameter θ, it would only hold at

defined by our inequalities are sharp is left for future research. However, as the results in Section 5 show, in our empirical application, they generate bounds that are tight enough to be informative.

its true value θ^*. This equality, however, still depends on the unknown true information set, $\mathcal{J}_{ijt}$, through the unobserved expectation, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$. We exploit the property that the moment function in equation (14c) is convex in the unobserved expectation $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$; i.e. $\Phi(\cdot)/(1-\Phi(\cdot))$ is convex. Thus, applying Jensen's inequality, equation (16) becomes an inequality if we introduce the observed proxy, r^o_{ijt}, in place of the unobserved expectation $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ and take the expectation of the resulting expression conditional on an observed vector $Z_{ijt} \subseteq \mathcal{J}_{ijt}$. Consequently, if the equality in equation (16) holds at the true value of the parameter vector, the inequality defined in equations (14) and (14c) will also hold at $\theta = \theta^*$.[21]

The moment functions in equations (14b) and (14c) are not redundant. For example, consider the identification of the parameter θ_0. Given observed values of d_{ijt}, r^o_{ijt}, and $dist_j$, and given any arbitrary value of the parameters θ_1 and θ_2, the moment function $m^{ob}_l(\cdot)$ in equation (14b) is increasing in θ_0 and, therefore, will identify a lower bound on θ_0. With the same observed values, $m^{ob}_u(\cdot)$ in equation (14c) is decreasing in θ_0 and will thus identify an upper bound on θ_0. The same intuition applies for identifying bounds for θ_1 and θ_2.

4.2.2 Revealed Preference Moment Inequalities

For any $Z_{ijt} \subseteq (\mathcal{J}_{ijt}, dist_j)$, we define a conditional revealed preference moment inequality as

$$\mathcal{M}^r(Z_{ijt};\theta) = \mathbb{E}\left[\begin{array}{c} m^r_l(d_{ijt}, r^o_{ijt}, dist_j;\theta) \\ m^r_u(d_{ijt}, r^o_{ijt}, dist_j;\theta) \end{array} \middle| Z_{ijt}\right] \geq 0, \tag{17a}$$

where the two moment functions are defined as

$$m^r_l(\cdot) = -(1-d_{ijt})\big(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j\big) + d_{ijt}\theta_2 \frac{\phi\big(\theta_2^{-1}(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j)\big)}{\Phi\big(\theta_2^{-1}(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j)\big)}, \tag{17b}$$

$$m^r_u(\cdot) = d_{ijt}\big(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j\big) + (1-d_{ijt})\theta_2 \frac{\phi\big(\theta_2^{-1}(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j)\big)}{1-\Phi\big(\theta_2^{-1}(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j)\big)}. \tag{17c}$$

We denote the values of θ consistent with the moment inequalities in equation (17) as Θ^r_0.

Theorem 2 *Let $\theta^* = \big(\beta_0, \beta_1, \sigma\big)$ be the parameter defined by equation* (10)*. Then $\theta^* \in \Theta^r_0$.*

We provide a formal proof of Theorem 2 in Appendix C.2. Theorem 2 indicates that the revealed preference inequalities are consistent with the true value of the parameter vector, θ^*.

Heuristically, the two moment functions in equations (17b) and (17c) are derived using standard revealed preference arguments. We focus our discussion on the moment function

[21] The assumption that ν_{ijt} follows a normal distribution is sufficient but not necessary to derive the odds-based inequalities. For any distribution of ν_{ijt} with cumulative distribution function $F_\nu(\cdot)$, we need simply that $F_\nu(\cdot)/(1-F_\nu(\cdot))$ and $(1-F_\nu(\cdot))/F_\nu(\cdot)$ are globally convex. This condition will be satisfied if the distribution of ν is log-concave. Both the normal and the logistic distributions are log-concave, as are the uniform, exponential, type I extreme value, and Laplace distributions. Heckman and Honoré (1990), and Bagnoli and Bergstrom (2005) provide more information on the properties of log-concave distributions.

in equation (17c); the intuition behind the derivation of the moment in equation (17b) is analogous. If firm i decides to export to country j in period t, so that $d_{ijt} = 1$, then by revealed preference, it must expect to earn positive returns; i.e. $d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt}\big) \geq 0$. Taking the expectation of this inequality conditional on $(d_{ijt}, \mathcal{J}_{ijt}, dist_j)$ and taking into account that $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, we obtain

$$d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + S_{ijt} \geq 0, \tag{18}$$

where $S_{ijt} = \mathbb{E}[-d_{ijt}\nu_{ijt}|d_{ijt}, \mathcal{J}_{ijt}, dist_j]$. The term S_{ijt} is a selection correction that accounts for how ν_{ijt} affects the firm's decision to export, where again ν_{ijt} captures determinants of profits that the researcher does not observe.[22] We cannot directly use the inequality in equation (18) because it depends on the unobserved agents' expectations, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, both directly and through the term S_{ijt}. However, the inequality in equation (18) becomes weaker if we introduce the observed covariate, r^o_{ijt}, in the place of the unobserved expectations, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, and take the expectation of the resulting expression conditional on Z_{ijt}. As in the case of the odds-based inequalities, we need the moment function in equation (17c) to be globally convex in the unobserved expectation $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$; i.e. $\phi(\cdot)/(1-\Phi(\cdot))$ is convex. Consequently, if the inequality in equation (18) holds at the true value of the parameter vector, the inequality in equations (17) and (17c) will also hold at $\theta = \theta^*$.[23]

The inequalities in equation (17) follow the revealed preference inequalities introduced in Pakes (2010) and Pakes et al. (2015). In our setting, our inequalities feature structural errors ν_{ijt} that may vary across (i, j, t) and that have unbounded support. The cost of allowing this flexibility is that we must assume a distribution for ν_{ijt}, up to a scale parameter.[24,25]

4.2.3 Combining Inequalities for Estimation

We combine the odds-based and revealed preference moment inequalities described in equations (14) and (17) for estimation. The set defined by the odds-based inequalities is a singleton

[22] Appendix C.2 shows that, under the assumptions in Section 2,

$$S_{ijt} = (1 - d_{ijt})\sigma \frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1 - \Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}.$$

[23] As in footnote 21, the assumption of normality of ν_{ijt} is sufficient but not necessary. For the inequality equations (17) and (17c) to hold, we need a distribution for ν_{ijt} such that $\mathbb{E}[\nu_{ijt}|\nu_{ijt} < \kappa]$ is globally convex in the constant κ. An analogous condition is needed to derive equation (17b). In addition to the normal distribution, the logistic distribution also satisfies this condition.

[24] In our empirical application, we find σ, the standard deviation of ν_{ijt}, to be greater than zero. Therefore, including the selection correction term S_{ijt} in our inequalities is important: given that $S_{ijt} \geq 0$ whenever $\sigma > 0$, if we had generated revealed preference inequalities without S_{ijt}, we would have obtained weakly smaller identified sets than those found using the inequalities in equation (17).

[25] Pakes and Porter (2015) and Shi et al. (2017) show how to estimate discrete choice models in panel data settings without imposing distributional assumptions on ν_{ijt}. Both models, however, impose a restriction that agents make no errors in their expectations.

only when firms make no expectational errors and the vector of instruments Z_{ijt} is identical to the set of variables firms use to form their expectations. In this very specific case, the revealed preference inequalities do not have any additional identification power beyond that of the odds-based inequalities. However, in all other settings, the revealed preference moments can provide additional identifying power.

The set of inequalities we define in equations (14) and (17) condition on particular values of the instrument vector, Z_{ijt}. Exploiting all the information contained in these conditional moment inequalities can be computationally challenging.[26] In this paper, we base our inference on a fixed number of unconditional moment inequalities implied by the conditional moment inequalities in equations (14) and (17). We describe in Appendix A.5 the unconditional moments we use to compute the estimates discussed in Section 5. We denote the set of values of θ consistent with our unconditional odds-based and revealed-preference inequalities as Θ_0.

Conditioning on a fixed set of moments, while convenient, entails a loss of information. Thus, the identified set defined by our unconditional moment inequalities may be larger than that implied by their conditional counterparts. However, as the empirical results in sections 5, 6 and 7 show, the moment inequalities we employ nonetheless generate economically meaningful bounds on our parameters and on counterfactual choice probabilities, and also allow us to explore hypotheses about the information firms use to forecast export revenue.

4.2.4 Characterizing the Identified Set

Theorems 1 and 2 imply that θ^* will be contained in the set Θ_0 defined by our odds-based and revealed-preference moment inequalities when the instrument vector Z_{ijt} used to define these inequalities satisfies $Z_{ijt} \subseteq \mathcal{J}_{ijt}$ for all i, j, and t. However, these theorems do not fully characterize the set Θ_0. That is, they do not indicate the values of θ other than θ^* that are also included in this set. A full characterization is beyond the scope of this paper, but we conduct a simulation, with full results reported in Appendix E, to explore the content of Θ_0.

In particular, we design a simulation in which the researcher has access to three possible information sets: (1) a small information set, $\mathcal{J}^s_{ijt}$, that contains too few variables relative to the true information set; (2) a medium-sized set, $\mathcal{J}^m_{ijt}$, that coincides with the true information set; and (3) a large information set, $\mathcal{J}^l_{ijt}$, that contains more information than the firm actually possesses. Here, under $\mathcal{J}^l_{ijt}$, every firm can predict perfectly the observable component of its potential export revenues; i.e. $\mathbb{E}[r^o_{ijt}|\mathcal{J}^l_{ijt}] = r^o_{ijt}$. We denote the probability limits of the corresponding maximum likelihood estimators as θ_s, θ_m, and θ_l. For example, the maximum likelihood estimator with probability limit θ_s is computed under the incorrect assumption that the true information set equals $\mathcal{J}^s_{ijt}$. We similarly denote the corresponding identified

[26] Recent theoretical work, including Andrews and Shi (2013), Chernozhukov et al. (2013), Chetverikov (2013), Armstrong (2014), Armstrong (2015), and Armstrong and Chan (2016), provide estimation procedures that exploit all information contained in conditional moment inequality models.

sets defined by our moment inequalities as Θ_s^0, Θ_m^0, and Θ_l^0. For example, the identified set Θ_s^0 is computed under the correct assumption that the true information set includes $\mathcal{J}_{ijt}^s$.

Using our moment inequalities, both the assumptions that exporters know at least the variables in $\mathcal{J}_i^s$ and $\mathcal{J}_i^m$ are compatible with the data generating process. Thus, as discussed in Section 4.2, Θ_s^0 and Θ_m^0 will both contain θ^*. Of the maximum likelihood estimators, only $\mathcal{J}_i^m$ is compatible with the data generating process and, consequently, as discussed in Section 4.1, only θ_m coincides with the true parameter θ^*.

The informational assumptions imposed to compute θ_s and θ_l are compatible with the weaker information assumption imposed to compute Θ_s^0. However, as we show in Appendix E, it need not be the case that Θ_s^0 contains θ_s and θ_l. Their inclusion depends on (a) how different θ_s and θ_l are from the true value θ^* and (b) the span of points in the identified set Θ_s^0 around θ^*.

The distance between θ_s and θ^* depends on the importance of those predictors of export revenues contained in the true information set, $\mathcal{J}_{ijt}^m$, and excluded from the assumed one, $\mathcal{J}_{ijt}^s$; i.e. θ_s and θ^* move further apart as the variance of $\xi_{ijt}^s \equiv \mathbb{E}[r_{ijt}^o|\mathcal{J}_{ijt}^s] - \mathbb{E}[r_{ijt}^o|\mathcal{J}_{ijt}^m]$ increases. The distance between θ_l and θ^* increases in the importance of the variables included in the assumed information set, $\mathcal{J}_{ijt}^l$, and excluded from the true one, $\mathcal{J}_{ijt}^m$. Here, the distance increases in the variance of the firm's true expectational error, $\varepsilon_{ijt} \equiv r_{ijt}^0 - \mathbb{E}[r_{ijt}^o|\mathcal{J}_{ijt}^m]$.

The identified set Θ_s^0 will be larger when $\mathcal{J}_i^s$ excludes important predictors of potential export profits, r_{ijt}^o. Specifically, as the variance of $r_{ijt}^o - \mathbb{E}[r_{ijt}^o|\mathcal{J}_{ijt}^s] = \varepsilon_{ijt} - \xi_{ijt}^l$ increases, the identified set grows larger. Therefore, the same factors that increase the difference between both θ_s and θ_l and the true parameter vector θ^* will also make the identified set Θ_s larger. However, as we show in Appendix E, these factors have a larger effect on the bias of the misspecified maximum likelihood estimators than on the size of the identified set. Consequently, θ_s and θ_l will tend to belong to Θ_s^0 when the two chosen information sets, respectively, are close to the true information set.

5 Results

We estimate the parameters of exporters' participation decisions using the three different empirical approaches discussed in sections 4.1 and 4.2. First, we use maximum likelihood when we assume perfect foresight. Second, we again use maximum likelihood methods, but under the two-step procedure in which we project realized revenues on a set of observable covariates that we assume form a firm's information set. Third, we carry out our moment inequality approach under the assumption that the firm knows the same observed variables as in the two-step approach, but may also use additional variables to forecast revenues.

Before implementing these three procedures, we first need to compute our proxy for the observable component of export revenue, r_{ijt}^o. We describe in Appendix A.3 how to obtain

Table 2: Parameter estimates

Estimator	Chemicals σ	Chemicals β_0	Chemicals β_1	Food σ	Food β_0	Food β_1
Perfect Foresight (MLE)	1,038.6 (393)	745.2 (280)	1,087.8 (394)	1,578.1 (225)	2,025.1 (292)	214.5 (26)
Minimal Information (MLE)	395.5 (83)	298.3 (56)	447.1 (102)	959.9 (146)	1,259.3 (188)	129.4 (16)
Moment Inequality (OB and RP)	[85.1, 115.9]	[62.8, 81.1]	[142.5, 194.2]	[114.9, 160.0]	[167.1, 264.0]	[36.4, 80.2]
Moment Inequality (OB only)	[85.1, 133.3]	[34.8, 129.3]	[101.3, 293.4]	[114.9, 1,000]	[115.1, 1,310]	[6.8, 485]
Moment Inequality (RP only)	[80.0, 133.3]	[44.0, 133.3]	[110.0, 333.3]	[66.7, 487.8]	[64.7, 659.7]	[3.3, 731.4]

Notes: All estimates are reported in thousands of year 2000 USD and their values scale proportionally with η, which is set equal to 5 (see Section 2.4). For the two ML estimators, bootstrap standard errors are computed according to the procedure described in Appendix A.6 and reported in parentheses. For the three moment inequality estimates, extreme points of the 95% confidence set are reported in square brackets. These confidence sets are projections of a confidence set for $(\beta_0, \beta_1, \sigma)$ computed according to the procedure described in Appendix A.7. The "OB and RP" confidence sets exploit both odds-based and revealed-preference moment inequalities. The "OB only" use only odds-based moment inequalities. The "RP only" use only revealed-preference moment inequalities.

this proxy, which requires estimating revenue shifters, α_{jt}, for each market j and period t. We report the estimates $\{\hat{\alpha}_{jt}; \forall j, t\}$ for both the chemicals and food sectors in Appendix B.1.[27]

5.1 Average Fixed Export Costs

In Table 2, we report the estimates and confidence regions for our model parameters. The first coefficient, σ, is the standard deviation of the structural error ν_{ijt}. It controls the heterogeneity across firms and time periods in the fixed costs of exporting to a particular destination j. The remaining coefficients, β_0 and β_1, represent a constant component and the contribution of distance to the level of the fixed costs. We normalize the demand elasticity, η, to equal five.

From Table 2, we see the models that assume researchers have full knowledge of the exporter's information set produce much larger average fixed export costs than does our moment inequality approach. For example, consider the coefficient on the distance variable in models estimated using data from the chemicals sector. Under the moment inequality approach, we find an added cost of $142,500 to $194,200 when the export destination is 10,000 kilometers farther in distance. Under the two maximum likelihood procedures, estimates of the added cost equal $1,087,800 and $447,100 for the same added distance.

The moment inequality bounds on each of the elements of the parameter vector θ reported in Table 2 arise from projecting a three-dimensional 95% confidence set for the vector

[27] When computing standard errors for the maximum likelihood estimates of θ and computing moment inequality confidence sets for this parameter, we take into account the sampling error affecting our estimates of α_{jt}. See appendices A.6 and A.7 for details.

$(\beta_0, \beta_1, \sigma)$, computed following the procedure in Appendix A.7.[28] The results in Table 2 illustrate the value of using the revealed-preference and odds-based inequalities jointly. Re-running our estimation using each set of inequalities separately, we obtain much larger bounds on the fixed export costs than when we combine both types of inequalities.

We translate the coefficients reported in Table 2 into estimates of the average fixed costs of exporting by country. We report the results in Table 3 for three countries: Argentina, Japan, and the United States. Total exports to these countries account for 29% of total exports of the Chilean chemicals sector and 56% of the food sector in the sample period. In addition, these three countries span a wide range of possible distances to Chile and thus provide an illustration of the impact of distance on average fixed export costs.[29]

Table 3: Average fixed export costs

		Chemicals			Food	
Estimator	Argentina	Japan	United States	Argentina	Japan	United States
Perfect Foresight (MLE)	868.0 (150.1)	2,621.4 (468.2)	1,645.0 (290.7)	2,049.3 (213.8)	2,395.1 (259.4)	2,202.5 (233.8)
Minimal Information (MLE)	348.7 (49.9)	1,069.4 (142.2)	668.1 (90)	1,273.9 (221.8)	1,482.4 (272.9)	1,366.3 (244.2)
Moment Inequality	[79.5, 102.6]	[309.2, 414.3]	[181.3, 240.1]	[175.6, 270.1]	[269.1, 361.0]	[227.3, 308.9]

Notes: All estimates are reported in thousands of year 2000 USD and their values scale proportionally with η, which is set equal to 5 (see Section 2.4). For the two ML estimators, bootstrap standard errors are computed according to the procedure described in Appendix A.6 and reported in parentheses. For the three moment inequality estimates, extreme points of the 95% confidence set are reported in square brackets. These confidence sets are projections of a confidence set for $(\beta_0, \beta_1, \sigma)$ computed according to the procedure described in Appendix A.7.

Table 4: Average fixed export costs relative to perfect foresight estimates

		Chemicals			Food	
Estimator	Argentina	Japan	United States	Argentina	Japan	United States
Minimal Info.	40.2%	40.8%	40.6%	62.2%	61.9%	62.0%
Moment Ineq.	[9.1%, 11.9%]	[11.8%, 15.8%]	[11.1%, 14.6%]	[8.6%, 13.2%]	[11.3%, 15.1%]	[10.3%, 14.1%]

Notes: This table reports the ratio of (a) the minimal information ML point estimates and (b) extremes of the moment inequality confidence set, compared to the perfect foresight ML point estimate. All numbers reported in this table are independent of the value of η chosen as normalizing constant.

[28]Formally, denoting $\hat{\Theta}^{95\%}$ as the 95% confidence set for the vector $(\beta_0, \beta_1, \sigma)$, the confidence set for β_0, for example, contains all values of the unknown parameter θ_0 such that there exists values of θ_1 and θ_2 for which the triplet $(\theta_0, \theta_1, \theta_2)$ is included in $\hat{\Theta}^{95\%}$. Bugni et al. (2016) introduce a new inference procedure that dominates this projection-based inference in terms of power. We report here confidence sets based on the projection of $\hat{\Theta}^{95\%}$ because (a) these one-dimensional confidence sets are nonetheless small enough to illustrate the difference between the maximum likelihood and the moment inequality estimates and (b) they do not require additional computation once we have computed $\hat{\Theta}^{95\%}$. We use $\hat{\Theta}^{95\%}$ directly to compute the results in sections 6 and 7.

[29]In Appendix B.2, we also report quantiles of the distribution of fixed export costs across firms. In Appendix B.3, we relax the assumptions in equation (5) and instead estimate average fixed costs for each country j as a country fixed effect. Moment inequality confidence sets and maximum likelihood confidence intervals are wider in this case, reflecting the larger number of parameters to estimate. The qualitative results are similar.

Under perfect foresight, we estimate the average fixed costs in Argentina, Japan, and the United States in the chemicals sector to equal $868,000, $2.62 million, and $1.64 million, respectively. In the food sector, the average fixed cost estimates in these three countries equal $2.05 million, $2.40 million, and $2.20 million, respectively. As we show in Table 4, when comparing the estimates under perfect foresight to the estimates that assume a minimal information set with only three variables, the latter produces estimates that are about 60% smaller in the chemicals sector and 38% smaller in the food sector. Under our moment inequality estimator, we find 95% confidence sets for the fixed costs of exporting in the chemicals sector between $79,500 and $102,600 for Argentina, $309,200 and $414,300 for Japan, and $181,300 and $240,100 for the United States.[30] In all cases, the estimated bounds we find from the inequalities represent only a fraction of the perfect foresight estimates and the estimates from the two-step approach. Taken together, these results reflect the discussion in Section 4.1 and in Appendix D of the bias that arises if the researcher incorrectly specifies the exporter's information set.[31]

It may seem counterintuitive that the maximum likelihood estimates obtained under the assumption $\mathcal{J}^a_{ijt} = (r_{iht-1}, R_{it-1}, dist_j)$ are not contained in the confidence set computed under the assumption that $(r_{iht-1}, R_{it-1}, dist_j) \subseteq \mathcal{J}_{ijt}$. However, as we discuss in Section 4.2.4 and illustrate in Appendix E in detail, not every information set $\mathcal{J}^a_{ijt}$ consistent with our assumption that $(r_{iht-1}, R_{jt-1}, dist_j) \subseteq \mathcal{J}_{ijt}$ generates a likelihood function whose maximand is contained in the identified set defined by our moment inequalities.

6 Testing Content of Exporters' Information Sets

What do exporters know? We use the moment inequalities introduced in Section 4.2.3 to provide some answers. To do so, we exploit an implication of our empirical model: under rational expectations, any variable in the information set the firm uses to predict export revenues serves as an instrument in our moment inequalities. Thus, we can test whether a set of observed variables, Z_{ijt}, belongs to the firm's information set using the model specification test in Bugni et al. (2015) to test the null hypothesis that there exists a value of the parameter vector that rationalizes the resulting set of moment inequalities.

If we reject that there is a value of the parameter vector at which all our moment inequalities hold, we can conclude either that (a) one of the assumptions embedded in the export model described in Section 2 does not hold in the data or that (b) the set of observed variables Z_{ijt} we specify are not contained in the firm's information set, $\mathcal{J}_{ijt}$. To distinguish between

[30] To find confidence sets for the average fixed costs in j, $\bar{f}_j = \beta_0 + \beta_1 dist_j$, we compute the lower bound on $\bar{f}_j$ as $\min_{\theta \in \hat{\Theta}^{95\%}} \theta_0 + \theta_1 dist_j$ and the upper bound as $\max_{\theta \in \hat{\Theta}^{95\%}} \theta_0 + \theta_1 dist_j$.

[31] Specifically, the upward bias in the minimal information set case is consistent with a simulation in which the distribution of the difference between the true expectation and the one implied by the minimal information set, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$, is not symmetric. See Table D.3 for details.

these two conclusions, we repeat our test with the same underlying model but different Z_{ijt}.[32]

The p-values for the tests we perform appear in Table 5. In panel A, we test our main specification in which Z_{ijt} contains three covariates: the aggregate exports from Chile to each destination market in the previous year, R_{jt-1}; the distance to each market, $dist_j$; and the firm's own domestic sales in the previous year, r_{iht-1}. We fail to reject, at conventional significance levels, the null hypothesis that potential exporters know at least these three covariates when predicting export revenue.[33] In panel B, we run our moment inequality procedure under the assumption of perfect foresight. Here, we presume the firm knows r^o_{ijt} when it chooses whether to export. We can reject, at conventional significance levels, that firms know this future revenue when deciding whether to export.

In the remaining panels of Table 5, we re-run the same test as in Panel A adding an additional variable to the vector of instruments. In panel C, we add the lagged value of the country-year revenue shifter, α_{jt-1}. From the model in Section 2, this shifter is a sufficient statistic for how destination-specific supply and demand factors affect export revenues. The results in panel C support two broad conclusions: (a) large firms have more information about α_{jt} than small firms; and (b) the information that a firm has about this shifter appears independent of prior export experience to a market and the popularity of the market. Specifically, at the 5% significance level, we cannot reject that α_{jt-1} is in the information set of large firms (defined as firms with above median domestic sales in the previous year) when exporting to either popular or unpopular markets (defined as markets with above or below the median number of Chilean exporters in the previous year). We further rule out that this finding on large firms' information is a result of past export experience: we cannot reject that large firms know α_{jt-1} even if they did not export to j at $t-1$. We perform the same tests for small firms. For most tests, we can reject that small firms have information on α_{jt-1}.

In panel D, we explore whether firms differ in the information they have about a quantity that may be simpler to observe—the number of exporters to a destination in the previous year, N_{jt-1}. Although acquiring information about N_{jt-1} is likely easier than acquiring information about α_{jt-1}, it is not trivial.[34] According to our tests, we cannot reject that large firms, including those without prior export experience, are likely to have information on N_{jt-1} for any destination j. The results are similar for both popular and unpopular destinations. Given

[32] We simultaneously test multiple hypotheses. We describe in Appendix A.8 how we compute individual p-values for each test and how we use the procedure in Holm (1979) to compute family-adjusted p-values.

[33] We will not reject our null hypothesis as long as the expectational error in the firm's revenue forecast, $\varepsilon_{ijt} \equiv r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, satisfies the condition $\mathbb{E}[\varepsilon_{ijt}|Z_{ijt}] = 0$. This condition will hold when Z_{ijt} is (a) irrelevant to predict r^o_{ijt} or, (b) if relevant, when Z_{ijt} is in the information set $\mathcal{J}_{ijt}$. To make the conclusion from our test clearer, we rule out the "irrelevant" explanation to our findings by running a pre-test on every variable included in any vector Z_{ijt} whose validity as instruments we test. In this pre-test, we check that these variables have predictive power for r^o_{ijt}. The results from this pre-test are included in Appendix B.4.

[34] The annual and monthly reports published by the Chilean Customs Agency (`http://www.aduana.cl/anuarios-compendios-y-reportes-estadisticos/aduana/2016-09-20/165452.html`) include information on total volume of exports by destination country and product, but not on the number of exporters.

Table 5: Testing Content of Information Sets

Set of Firms	Set of Export Destinations	Chemicals Individual p-value	Chemicals Adjusted p-value	Chemicals Reject at 5%	Food Individual p-value	Food Adjusted p-value	Food Reject at 5%
		Panel A: Minimal Information					
All	All	0.111	0.111	No	0.980	0.980	No
		Panel B: Perfect Foresight					
All	All	0.023	0.023	Yes	< 0.001	< 0.001	Yes
		Panel C: Minimal Information & Country Shifter					
Large	Popular	0.144	0.418	No	0.974	1	No
Large	Unpopular	0.114	0.418	No	0.981	1	No
Small	Popular	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
Small	Unpopular	0.024	0.118	No	0.004	0.021	Yes
Large Exporter	All	0.104	0.418	No	0.990	1	No
Large Non-exporter	All	0.140	0.418	No	0.048	0.190	No
Small Exporter	All	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
Small Non-exporter	All	< 0.001	< 0.001	Yes	0.015	0.075	No
		Panel D: Minimal Information & Number of Exporters					
Large	Popular	0.104	0.311	No	0.978	1	No
Large	Unpopular	0.114	0.311	No	0.981	1	No
Small	Popular	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
Small	Unpopular	0.116	0.311	No	0.003	0.015	Yes
Large Exporter	All	0.018	0.080	No	0.988	1	No
Large Non-exporter	All	0.016	0.080	No	0.717	1	No
Small Exporter	All	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
Small Non-exporter	All	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
		Panel E: Minimal Information & Country Group Avg. Shifter					
		(a) Continent Avg. Shifter					
Large	All	0.109	0.828	No	0.986	1	No
Small	All	0.112	0.828	No	0.470	1	No
		(b) Language Group Avg. Shifter					
Large	All	0.116	0.828	No	0.980	1	No
Small	All	0.115	0.828	No	0.003	0.0180	Yes
		(c) Income p.c. Group Avg. Shifter					
Large	All	0.104	0.828	No	0.980	1	No
Small	All	0.152	0.828	No	0.991	1	No
		(d) Border Group Avg. Shifter					
Large	All	0.119	0.828	No	0.981	1	No
Small	All	0.114	0.828	No	0.001	0.008	Yes

Notes: Each panel differs in the content of the information set being tested and is a separate family for the purpose of adjusting p-values. Panel A tests that $(dist_j, r_{iht-1}, R_{jt-1}) \subseteq \mathcal{J}_{ijt}$; panel B tests $\alpha_{jt} r_{iht} \subseteq \mathcal{J}_{ijt}$; panel C tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1}) \subseteq \mathcal{J}_{ijt}$; panel D tests $(dist_j, r_{iht-1}, R_{jt-1}, N_{jt-1}) \subseteq \mathcal{J}_{ijt}$; panel E(a) tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{c(j)t-1}) \subseteq \mathcal{J}_{ijt}$; panel E(b) tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{l(j)t-1}) \subseteq \mathcal{J}_{ijt}$; panel E(c) tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{i(j)t-1}) \subseteq \mathcal{J}_{ijt}$; panel E(d) tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{b(j)t-1}) \subseteq \mathcal{J}_{ijt}$. The variable $\alpha_{c(j)t-1}$ is the average value of α_{jt-1} across all countries that share continent with country j. The variables $\alpha_{l(j)t-1}$, $\alpha_{i(j)t-1}$, and $\alpha_{b(j)t-1}$ are analogous averages across countries that share language, similar income per capita and border, respectively, with j. *Large* firms are those with above median domestic sales in the previous year. Conversely, firm i is *Small* if its domestic sales fall below the median. *Popular* export destinations are those with above median number of exporters in the previous year. Firm i at period t as an *Exporter* to country j if $d_{ijt-1} = 1$ and as a *Non-exporter* if $d_{ijt-1} = 0$. All reported p-values correspond to the test RC; for details on how to compute these p-values, see Appendix A.8. All numbers reported in this table are independent of the value of η chosen as the normalizing constant.

that large firms were informed about α_{jt-1}, it is not surprising that they also have access to information on N_{jt-1}. For small firms, we find evidence to reject the null hypothesis that these firms have access to information on the past number of exporters by destination.

Finally, in panel E, we relax further the informational requirements. Rather than testing whether firms know country-specific information, we test whether they know variables that help predict exports for large groups of countries. Specifically, we test whether firms know average values of α_{jt-1} across groups of countries that share a continent, language, similar income per capita, or a border. Given the results in panels C and D, we focus on testing heterogeneity between large and small firms. For the chemicals sector, we cannot reject the null hypothesis that all firms, large and small firms alike, know these aggregate shifters. For the food sector, we can still reject that small firms have access to average revenue shifters when these averages are computed across countries that share income per capita or a border.

Overall, we find no evidence that firms learn from other exporters or from their prior export experience. However, firms that are either more productive or sell higher quality products tend to have an informational advantage when forecasting market conditions in foreign countries.[35] While we do not investigate why large firms have better information to predict their revenue in foreign markets, the prior literature offers some insights. Weiss (2008) documents that large firms are more likely to participate in international trade fairs and Álvarez and Crespi (2000) note that firms with larger domestic sales are more likely to gather information on foreign markets via programs sponsored by the Chilean Agency for Export Promotion.[36]

7 Counterfactuals

Finally, we use our model and the estimates in Section 5 to explore how changes in firms' information sets and changes in the fixed costs of exporting affect export decisions.

7.1 Changes in Information Sets

We first compute how export profits for the average firm and aggregate exports vary in a counterfactual in which we endow firms with more information about the determinants of export revenues. Specifically, we consider firms whose initial information set includes only three variables: the firm's own lagged domestic sales, r_{iht-1}; the distance to a destination market, $dist_j$; and Chile's lagged aggregate exports to this market, R_{jt-1}. This is the minimal

[35]Our approach allows us to test *passive learning* about a destination-year aggregate shifter of export revenues; we do not test whether firms learn about a firm-specific demand shock (as in Albornoz et al., 2012) or about the demand shifter in a particular buyer-seller relationship (as in Eaton et al., 2014).

[36]The Chilean National Agency for Export Promotion "manages a system that provides information to firms. It is used by companies interested in obtaining information about international markets, for example: external prices, transport costs, entrance regulations and trade barriers." (Álvarez and Crespi, 2000)

Table 6: Effect of Improving Information

Sector:		Chemicals			Food		
		Percentage Change in:			Percentage Change in:		
Firms	Markets	Number of Exporters	Mean Export Profits	Aggregate Exports	Number of Exporters	Mean Export Profits	Aggregate Exports
Panel A: Impact of Adding Information on Aggregate Revenue Shocks to Minimal Information							
All	All	[-5.7, -3.5]	[17.5, 20.6]	[6.4, 9.5]	[-23.2, -17.2]	[57.4, 71.6]	[-2.1, 4.3]
Large	All	[-9.0, -7.1]	[22.0, 24.9]	[0.6, 2.5]	[-20.1, -17.2]	[42.5, 56.8]	[-6.1, -5.3]
Small	All	[0.0, 0.0]	[0.0, 0.1]	[2.0, 2.7]	[0.2, 0.4]	[0.0, 0.1]	[7.4, 12.8]
All	Large	[-2.4, -1.4]	[5.8, 7.0]	[1.1, 2.0]	[-22.2, -17.0]	[5.8, 7.0]	[-3.9, -0.7]
All	Small	[0.3, 0.3]	[6.3, 9.0]	[21.6, 27.3]	[0.8, 1.8]	[16.0, 24.1]	[23.9, 33.4]
Panel B: Impact of Switching from Minimal Information to Perfect Foresight							
All	All	[-10.2, -6.1]	[46.0, 52.9]	[25.1, 33.5]	[-28.0, -20.0]	[90.6, 111.6]	[6.7, 16.8]
Large	All	[-17.3, -12.7]	[59.2, 67.4]	[13.5, 20.3]	[-26.1, -23.9]	[80.8, 97.2]	[1.5, 3.2]
Small	All	[0.3, 0.5]	[0.1, 0.4]	[21.1, 30.8]	[1.3, 2.8]	[2.7, 6.7]	[112.1, 175.1]
All	Large	[-7.5, -4.5]	[29.6, 32.8]	[14.4, 19.0]	[-29.7, -22.8]	[29.6, 32.8]	[1.0, 6.7]
All	Small	[1.7, 2.3]	[16.2, 23.6]	[68.7, 81.6]	[3.9, 6.2]	[53.2, 76.4]	[77.6, 106.8]

Notes: *Large* firms are those with above median domestic sales in the previous year. Conversely, firm i at period t is defined as *Small* if its domestic sales fall below the median. *Large* export destinations are those with above median aggregate exports from Chile in the previous year. Conversely, *Small* export destinations are those with below median aggregate exports from Chile in the previous year. Panel A studies the impact of switching the information set from $(r_{iht-1}, R_{jt-1}, dist_j)$ to $(r_{iht-1}, R_{jt-1}, dist_j, \alpha_{jt-1})$. Panel B studies the impact of switching the information set from $(r_{iht-1}, R_{jt-1}, dist_j)$ to $\alpha_{jt} r_{iht}$. Extreme points of 95% confidence sets computed according to the procedure described in Appendix A.7 are reported in square brackets.

information set we use to compute the moment inequality bounds in Section 5. As indicated in panel A of Table 5, we cannot rule out that every firm knows these three variables.

For this pool of minimally informed firms, we consider extending their information sets in two ways. First, we provide them with additional information on every destination's lagged aggregate revenue shock, α_{jt-1}. According to Table 5, only large firms have access to this information. This counterfactual thus represents an outcome in which all firms gain the same information known by the most informed firms in the economy. Second, we endow firms with all the information necessary to predict perfectly the observable component of potential export revenues.[37] This perfect foresight case allows us to evaluate the overall importance of the informational frictions firms face when predicting their potential export profits.

As we discuss in Section 2, firms in our model obtain all relevant information once they enter a market and, thus, set their prices optimally upon entry. Consequently, the counterfactual change in information sets that we consider here can only affect aggregate export revenues and average export profits by changing the set of firms that self-select into each destination market. We describe in Appendix A.9 the procedure we follow to compute these counterfactual changes in export participation, and report the results in Tables 6 and 7.

The first row in panel A of Table 6 illustrates that, as we provide information on α_{jt-1} for all markets to all firms, the total number of firm-destination pairs with positive exports

[37]Firms remain uncertain about the unobservable component of revenues e_{ijt} (see equations (2) and (3)).

decreases between 3.5% and 5.7% in the chemicals sector. Interestingly, although the total number of firm-destinations decreases, the overall (aggregated across firms and destinations) export revenue in the sector *increases* between 6.4% and 9.5%. Therefore, across all firms, destinations and years, information frictions operate as barriers to trade.

The value to the average firm of acquiring information on α_{jt-1} is quantitatively important. With this information, firms improve their export revenue forecasts and, thus, fewer firms make mistakes in their entry decisions. As a consequence, the realized ex post profits of the average firm in the average market to which it exports increases between 17.5% and 20.6% in the chemical sector.

Panel A in Table 7 provides a detailed accounting of the basis for these counterfactual changes in exports. In the second row in Table 7, we document that export flows for between 708 and 914 firm-destination pairs present in the baseline case would cease if firms acquired information on α_{jt-1}. This would increase average export profits, as mean ex post export profits in these specific firm-destination pairs are negative; the mean export losses among these observations is between 47,000 and 52,000 USD. In short, as information on destination markets improves for firms in the chemical sector, these firms realize that their expectations of export revenues were too optimistic. Accounting only for reductions due to overly optimistic forecasts, overall exports would fall 209 to 219 million USD. However, as information on α_{jt-1} becomes available to all firms, there are also between 512 and 572 new firm-destination pairs with positive export flows. Average export profits and total export revenues in these new destinations are, respectively, between 62,000 and 77,000 USD and between 456 and 512 million USD. Consequently, although firm-destination pairs are lost on net as information increases, export revenue aggregated over all firms and destinations increases, as does average export profits per firm and market.

The results in tables 6 and 7 also allow us to compare the importance of information to firms of different sizes, and to compare against the perfect foresight benchmark. First, increasing the information firms can access always increases realized average ex post profits, particularly for large firms. This is an implication of better informed firms being less likely to make mistakes. Small firms generally benefit less from improving their information because, for most of these firms, their optimal decision is not to export, both before and after acquiring the extra information. Second, increasing access to information for a subset of firms has ambiguous effects on the total number of firm-destination pairs with positive export flows and on the aggregate exports of these firms. For example, in the food sector, informing large firms of the value of lagged aggregate export shocks α_{jt-1} for all destinations leads to a drop in the number of export markets they enter. Aggregate exports of these large exporters end up falling. Third, the information firms may acquire from lagged variables is fairly limited relative to the perfect foresight benchmark. Specifically, comparing the results in panels A and B in tables 6 and 7, we observe that the predicted effect of acquiring information about

Table 7: Decomposing the Effect of Improving Information for All Firms and Destinations

	Chemicals			Food		
Firm Groups by Export Status	Number of Firm-Dest.-Year	Mean Export Profits	Aggregate Exports	Number of Firm-Dest.-Year	Mean Export Profits	Aggregate Exports
Panel A: Impact of Adding Information on Aggregate Revenue Shocks to Minimal Information						
Always export	[4,830, 5,146]	[121, 126]	[3,021, 3,506]	[5,147, 6,606]	[245, 270]	[9,319, 11,407]
Switch out	[708, 914]	[-52, -47]	[209, 219]	[2,199, 3,114]	[-119, -108]	[1,230, 1,411]
Switch in	[512, 572]	[62, 77]	[456, 512]	[919, 1,034]	[61, 114]	[1,136, 1,679]
Never export	[37,403, 37,968]	[-205, -157]	[3,148, 3,586]	[74,311, 76,708]	[-312, -229]	[8,598, 10,324]
Panel B: Impact of Switching from Minimal Information to Perfect Foresight						
Always export	[4,096, 4,267]	[158, 162]	[2,978, 3,442]	[4,481, 5,611]	[307, 344]	[9,322, 11,341]
Switch out	[1,387, 1,814]	[-63, -55]	[250, 282]	[2,865, 4,127]	[-129, -113]	[1,227, 1,477]
Switch in	[1,046, 1,196]	[95, 119]	[1,215, 1,331]	[1,375, 1,589]	[121, 183]	[2,338, 2,995]
Never export	[36,781, 37,438]	[-210, -162]	[2,389, 2,769]	[73,754, 76,252]	[-316, -233]	[7,396, 9,007]

Notes: *Mean export profits* and *Total export revenues* indicate the model-predicted potential mean export profits and aggregate export revenues, respectively, for all firms, countries and years belonging to the group indicated in the corresponding row. *Mean export profits* are reported in thousands of year 2000 USD and are conditional on the assumption that $\eta = 5$. *Total export revenues* are reported in millions of year 2000 USD and do not depend on the value of η. Panel A studies the impact of switching the information set of every firm i in every country j and year t from $(r_{iht-1}, R_{jt-1}, dist_j)$ to $(r_{iht-1}, R_{jt-1}, dist_j, \alpha_{jt-1})$. Panel B studies the impact of switching the information set of every firm i in every country j and year t from $(r_{iht-1}, R_{jt-1}, dist_j)$ to $\alpha_{jt} r_{iht}$. Extreme points of 95% confidence sets computed according to the procedure described in Appendix A.7 are reported in square brackets.

$r^o_{ijt} \equiv \alpha_{jt} r_{iht}$ is generally larger than the effect of acquiring information about α_{jt-1} only.

7.2 Changes in Fixed Export Costs

Here, we conduct a counterfactual exercise in which we simulate a reduction in exporters' fixed costs of 40%. With this counterfactual, we aim to capture in a stylized way the effect of export promotion programs, such as those instituted in Canada (Van Biesebroeck et al., 2015), Peru (Volpe Martincus and Carballo, 2008), or Uruguay (Volpe Martincus et al., 2010). It is difficult to quantify the precise savings in fixed costs that these measures imply; our choice of a 40% reduction illustrates one possible level.

Counterfactual export probabilities in our setting are partially identified for two reasons: (a) our parameter of interest θ is partially identified; and, (b) we do not want to impose assumptions on the content of the firm's information set $\mathcal{J}_{ijt}$ beyond those we imposed for estimation. Thus, even given a value of θ, export probabilities are not point identified because we only observe a subset Z_{ijt} of the variables firms use to predict export revenues. Thus, we cannot compute firms' expectations, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, exactly and therefore cannot compute the export probabilities in equation (10) directly. We provide details of our algorithm in Appendix A.10. Here, we show the theorem that allows us to bound the probability of export participation given a value of θ and a set of variables $Z_{ijt} \subseteq \mathcal{J}_{ijt}$.

Theorem 3 *Suppose $Z_{ijt} \subseteq \mathcal{J}_{ijt}$ and, for any $\theta \in \Theta$, define $\mathcal{P}(Z_{ijt}; \theta) \equiv \mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]$, with*

$\mathcal{P}_{ijt}(\theta)$ defined as

$$\mathcal{P}_{ijt}(\theta) \equiv \Phi\big(\theta_2^{-1}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j\big)\big). \tag{19}$$

Then,

$$\mathcal{P}^l(Z_{ijt};\theta) \leq \mathcal{P}(Z_{ijt};\theta) \leq \mathcal{P}^u(Z_{ijt};\theta), \tag{20}$$

where

$$\mathcal{P}^l(Z_{ijt};\theta) = \frac{1}{1 + B^l(Z_{ijt};\theta)}, \tag{21a}$$

$$\mathcal{P}^u(Z_{ijt};\theta) = \frac{B^u(Z_{ijt};\theta)}{1 + B^u(Z_{ijt};\theta)}, \tag{21b}$$

and

$$B^l(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1 - \Phi\big(\theta_2^{-1}\big(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)}{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)}\middle| Z_{ijt}\right], \tag{22a}$$

$$B^u(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)}{1 - \Phi\big(\theta_2^{-1}\big(\eta^{-1}r^o_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)}\middle| Z_{ijt}\right]. \tag{22b}$$

The proof of Theorem 3 appears in Appendix A.10.1. If $\theta = \theta^*$, $\mathcal{P}_{ijt}(\theta)$ in equation (19) equals the true export probability in equation (10) and, thus, $\mathcal{P}^l(Z_{ijt};\theta^*)$ and $\mathcal{P}^u(Z_{ijt};\theta^*)$ bound the conditional probability that firm i exports to j at time t.

We use data from the chemicals sector and compare the counterfactual predictions from our moment inequality approach and from the models that require the researcher to specify the covariates included in firms' information sets.[38] Three elements of our model dictate how a change in fixed export costs translates into a change in the number of firms participating in export markets: (a) the initial level of average fixed costs, (b) the heterogeneity across firms in fixed export costs, and (c) firms' expectations of potential export revenues.

First, from equation (10), the level of fixed export costs, $\beta_0 + \beta_1 dist_j$, affects the number of firms that export. Since we reduce fixed costs by a fixed percentage in the counterfactual, the larger the initial estimate of average fixed costs, the larger the reduction in the *level* of fixed export costs. In our setting, the average fixed export costs we recover are largest under the perfect foresight assumption, and thus our counterfactual change in the number of exporters would be largest under that assumption, holding all else equal.

[38]When computing the effect of the reduction in fixed export costs, we assume the parameters $\{\alpha_{jt}; \forall j \text{ and } t\}$ remain invariant. In Appendix B.5, we provide support for this partial equilibrium assumption in our setting.

Table 8: Impact of 40% Reduction in Fixed Costs in Chemicals

Estimator	1996 Argentina	1996 Japan	1996 United States	2005 Argentina	2005 Japan	2005 United States
	Counterfactual Number of Exporters					
Perfect Foresight	67	38	51	70	37	72
Minimal Info.	68	29	38	71	43	56
Moment Inequality	[68, 72]	[12, 95]	[91, 106]	[68, 72]	[5, 89]	[131, 152]

Notes: For the moment inequality estimates, the minimum and maximum predicted values obtained by projecting the 95% confidence set for θ are reported in squared brackets. We compute counterfactual exporter counts by multiplying the observed number of exporters by the counterfactual changes predicted by each of the three models, rounding to the nearest exporter. For the chemicals sector in 2005, we observe 46, 5 and 24 exporters to Argentina, Japan and United States, respectively. For 1996, we observe 44, 5, and 17. All numbers reported in this table are independent of the value of η chosen as normalizing constant.

Second, the joint distribution of firms' heterogeneity in fixed export costs and expectations, $\{(\nu_{ijt}, \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]); \forall i\}$ will also affect the participation decision of firms in reaction to a decrease in fixed export costs. Specifically, for a given $100(1-\lambda)\%$ reduction in average fixed export costs, the firms that will start exporting will be those for which $\lambda(\beta_0 + \beta_1 dist_j) < \eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \nu_{ijt} < \beta_0 + \beta_1 dist_j$. Different features of the distributions of $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ and ν_{ijt} will thus impact the mass of switchers. As an example, consider the case in which there is no heterogeneity across firms in predicted export revenues—i.e. $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] = r^*_{jt}$ for all i—and ν_{ijt} is equal to 0. In this case, the response depends on the level of r^*_{jt}. If r^*_{jt} is less than the baseline fixed costs but greater than the counterfactual ones, all firms will stay out in the baseline and all firms will export in the counterfactual. If $r^*_{jt} < \eta\lambda(\beta_0 + \beta_1 dist_j)$, no firm will export in the baseline or counterfactual.

We report our counterfactual estimates in Table 8 for 1996 and 2005. The counterfactual results differ importantly across our three example markets. For Argentina, the three estimation procedures yield very similar answers. Two features of the market explain this similarity. First, given that Argentina is very close to Chile, changes in the distance coefficient β_1 have little impact on entry into Argentina. Therefore, differences across models in the estimate of β_1 will not translate into large differences in predicted export participation. Second, revenues predicted using the minimal information set approach do not differ much from the predicted revenue under perfect foresight. Thus, with similar predicted revenues entering the export participation decision in equation (9), both the perfect foresight and the minimal information models should generate similar counterfactual predictions. For Japan, the two maximum likelihood estimators yield different predictions, and our moment inequality estimator yields predictions that are wide and thus not very informative. The lack of precision in our predictions in this market relates to the relatively few firms we observe exporting to Japan in the data. Finally, for the United States, the moment inequality approach produces predictions that are informative and larger than both maximum likelihood approaches.

8 Extensions

In this section, we extend the model presented in Section 2 in two directions. First, we relax the assumption that a firm's export decision is static and independent of past export participation. To do so, in Section 8.1 we build on Das et al. (2007) and Morales et al. (2017) to allow for sunk export entry costs and forward-looking exporters. Second, in Section 8.2, we relax the assumption, captured in equations (2) and (3), that all firm-country-year specific export revenue shocks are mean independent of exporters' information sets, $\mathcal{J}_{ijt}$. The extensions we discuss here involve larger dimensional parameter vectors than our benchmark specification and, thus, require more computing time to estimate than our benchmark model (see Ho and Rosen, 2017); we thus restrict our estimation below to the chemicals sector.

8.1 Dynamics

The model introduced in Section 2 is static: the export profits of firm i in country j at period t are independent of the previous export path of i in j. Here we extend this model to allow for dynamics. In this extension, exporting firms must still pay fixed costs f_{ijt} in every period in which they choose to export, but they must also pay sunk costs s_{ijt} if they export to j at t and did not export to j at period $t-1$. Therefore, the potential export profits are

$$\pi_{ijt} = \eta_j^{-1} r_{ijt} - f_{ijt} - (1 - d_{ijt-1}) s_{ijt}. \tag{23}$$

We model sunk export costs as:

$$s_{ijt} = \gamma_0 + \gamma_1 dist_j, \tag{24}$$

and assume that firms know these costs when deciding whether to export to a destination.[39] We further assume that information sets evolve independently of past export decisions:

$$(\mathcal{J}_{ijt+1}, f_{ijt+1}, s_{ijt+1}) | (\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt}) \sim (\mathcal{J}_{ijt+1}, f_{ijt+1}, s_{ijt+1}) | (\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}). \tag{25}$$

If firms are forward-looking, the export dummy d_{ijt} becomes:

$$\begin{aligned} d_{ijt} = \mathbb{1}\{ & \eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - (1 - d_{ijt-1})(\gamma_0 + \gamma_1 dist_j) - \nu_{ijt} \\ & + \rho \mathbb{E}[V(\mathcal{J}_{ijt+1}, f_{ijt+1}, s_{ijt+1}, d_{ijt}) | \mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt} = 1] \\ & - \rho \mathbb{E}[V(\mathcal{J}_{ijt+1}, f_{ijt+1}, s_{ijt+1}, d_{ijt}) | \mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt} = 0] \geq 0\}, \end{aligned} \tag{26}$$

[39] Contrary to Das et al. (2007), we do not allow for unobserved heterogeneity in sunk costs. Our moment inequality approach may be generalized, at a loss of identification power, to allow for sunk costs $s_{ijt} = \gamma_0 + \gamma_1 dist_j + \nu^s_{ijt}$, with ν^s_{ijt} independent over time and normally distributed with mean zero and constant variance.

Table 9: Export fixed and sunk costs: firm average

Estimator	Cost	Argentina	Chemicals Japan	United States
Benchmark	Fixed	[79.5, 102.6]	[309.2, 414.3]	[181.3, 240.1]
Dynamics	Fixed	[55.8, 109.3]	[853.3, 1,670.0]	[409.2, 800.8]
	Sunk	[384,2, 734,3]	[5,874.4, 11,224.5]	[2,816.6, 5,382.7]

Notes: All estimates are reported in thousands of year 2000 USD and their values scale proportionally with η, which is set equal to 5 (see Section 2.4). Extreme points of 95% confidence sets computed according to the procedure described in Appendix A.7 are reported in square brackets.

where $V(\cdot)$ denotes the value function, ρ is the discount factor and $(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1})$ is the state vector on which firm i conditions its entry decision in country j at period t. The parameter to estimate is $\theta_D^* \equiv (\beta_0, \beta_1, \sigma, \gamma_0, \gamma_1)$, and we normalize $\eta = 5$ as in the static case.

The firm's export decision now depends on the firm's expectations of both the observable component of static revenues, r_{ijt}^o, and the difference in the value function depending on whether firm i exports to j in t. We follow the approach from the static case to find a measure of r_{ijt}^o, but finding a measure of the difference in value functions is impossible: $V(\cdot)$ at $t+1$ depends on the observed choice at $t+1$, d_{ijt+1}, which is a function of the observed choice at t, d_{ijt}. Therefore, even if firms were only to account for profits at periods t and $t+1$ when making a decision at t, we can only find a measure of either $\mathbb{E}[V(\cdot)|\cdot, d_{ijt} = 1]$ or $\mathbb{E}[V(\cdot)|\cdot, d_{ijt} = 0]$. To solve this lack of measurement, we adjust the Euler approach in Morales et al. (2017). This approach follows the methodology developed in Hansen and Singleton (1982) and Luttmer (1999) for continuous controls, but adapted for our model with discrete controls.[40]

The moment inequalities we employ to compute a confidence set on θ_D^* are the equivalent of the odds-based and revealed-preference inequalities introduced in Section 4.2, adjusted to account for the forward-looking behavior of firms. In Table 9 we report confidence sets for the fixed and sunk costs of exporting. Sunk entry costs are significantly larger than fixed export costs, consistent with Das et al. (2007). Fixed and sunk export costs are also increasing in distance. Furthermore, the sensitivity of these costs to distance is very similar for both types: relative to the bounds for Argentina, the bounds on fixed and sunk costs for the United States and Japan are approximately eight and fifteen times larger.

Comparing the estimated fixed costs in the static model to those from the dynamic model, we find two key differences. First, the bounds are wider; when we estimate fixed and sunk costs simultaneously, we face difficulties in separately identifying both types of costs. Second,

[40] Appendix F shows how to adapt the Euler approach in Morales et al. (2017) to the model described in Section 2 and in equations (23) to (26). Morales et al. (2017) consider models in which the unobserved component ν_{ijt} is constant across groups of countries for each firm-year specific pair. The Euler approach in Morales et al. (2017) has the advantage that it allows us to partially identify the parameter vector of interest without taking a stand on the information set of each exporter, as in Pakes et al. (2015). We also need not specify the number of periods ahead that each firm takes into account when deciding whether to export.

fixed costs for the United States and Japan are larger in the dynamic model. This difference is due to the parameter β_1, the effect of distance on fixed export costs, which we estimate to be larger when accounting for dynamics.[41,42]

8.2 Expected Firm-Country Export Revenue Shocks

In this section, we generalize the model described in Section 2 and allow the firm to observe determinants of export revenue that the researcher does not observe. Specifically, we assume:

$$r_{ijt} = \alpha_{jt} r_{iht} + \omega_{ijt} + e_{ijt}, \qquad \mathbb{E}[e_{ijt} | \mathcal{J}_{ijt}, r_{iht}, f_{ijt}] = 0, \qquad \mathbb{E}[\omega_{ijt} | \mathcal{J}_{ijt}] = \omega_{ijt}; \tag{27}$$

and define a subset $\mathcal{W}_{ijt}$ of the information set of the firm $\mathcal{J}_{ijt}$ such that the firm's expectations of the observable component of export revenues satisfy $\mathbb{E}[\alpha_{jt} r_{iht} | \mathcal{J}_{ijt}] = \mathbb{E}[\alpha_{jt} r_{iht} | \mathcal{W}_{ijt}]$. The export dummy d_{ijt} therefore becomes

$$d_{ijt} = \mathbb{1}\{\eta^{-1} \mathbb{E}[\alpha_{jt} r_{iht} | \mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j - (\nu_{ijt} - \eta^{-1} \omega_{ijt}) \geq 0\}, \tag{28}$$

and the components of revenue known to the firm but not the researcher follow the distribution:

$$\left(\begin{array}{c} \omega_{ijt} \\ \nu_{ijt} \end{array} \right) \Bigg| (\mathcal{W}_{ijt}, dist_j) \sim \mathbb{N} \left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma_\omega^2 & \sigma_{\omega\nu} \\ \sigma_{\omega\nu} & \sigma^2 \end{array} \right) \right). \tag{29}$$

If we assume both that potential exporters have perfect foresight over the component $\alpha_{jt} r_{iht}$ of their export revenues, $\mathbb{E}[\alpha_{jt} r_{iht} | \mathcal{W}_{ijt}] = \alpha_{jt} r_{iht}$, and that the firm observes all components of export revenue, i.e. $e_{ijt} = 0$ for all i, j, and t, then we can estimate the parameter vector $(\{\alpha_{jt}\}_{j,t}, \beta_0, \beta_1, \sigma_\omega, \sigma_{\omega\nu}, \sigma_\nu)$ using the procedure introduced in Heckman (1979).[43] Appendix G.1 shows how to use moment inequalities to estimate a confidence set for this parameter vector when we both allow e_{ijt} to differ from zero and impose weaker assumptions on the firm's information. We require only that the researcher observes a vector $Z_{ijt} \subseteq \mathcal{W}_{ijt}$.

[41] It may seem counterintuitive that accounting for sunk export costs increases the estimates of fixed export costs. This pattern would not arise if exporters were to decide whether to export at period t by comparing the static profits at t with the sum of fixed and sunk export costs. However, the presence of the value function in equation (26) makes the pattern we observe more likely, as firms in the dynamic model decide whether to export at any given period t taking into account the effect their decision has on subsequent periods' potential export profits. Specifically, when exiting an export destination, exporters take into account that they would have to repay the sunk costs if they were to re-enter in subsequent periods. This implies that, if fixed costs in the dynamic model were to remain at the values estimated in the static model, firms would be less likely to exit than in the static model. Therefore, rationalizing the observed exit behavior in the data requires larger fixed export costs in the dynamic model with forward-looking firms than in the static case.

[42] While our estimates of export sunk costs for Argentina are similar to those in Das et al. (2007), those for the U.S. and Japan are significantly larger. The differences in these estimates could be due to the differences in the datasets or in the specification of sunk costs and information sets.

[43] When estimating the model introduced in Heckman (1979), it is typical to fix one of the components of the variance matrix in equation (29) as a normalization. In our case, we opt to maintain the normalization $\eta = 5$.

Table 10: Export fixed costs: specification with expected unobserved revenue shocks

		Chemicals		
Specification	Estimator	Argentina	Japan	United States
Baseline: $\omega_{ijt}=0$	Perfect Foresight (MLE)	868.0 (106.5)	2,621.4 (315.9)	1,645.0 (199.3)
	Moment Inequality	[79.5, 102.6]	[309.2, 414.3]	[181.3, 240.1]
Extension: $\omega_{ijt}\neq 0$	Perfect Foresight (MLE)	323.4 (34.8)	983.6 (107.5)	615.9 (66.5)
	Moment Inequality	[75.2, 767.7]	[435.9, 3,449.4]	[284.7, 1,654.2]

Notes: All estimates are reported in thousands of year 2000 USD and their values scale proportionally with η, which is set equal to 5. Estimates reported for the extension assume $\alpha_{jt}=\alpha\ \forall\ j$ and t . Extreme points of 95% confidence sets computed according to the procedure described in Appendix G.1.3 appear in square brackets.

When we allow firms to account for ω_{ijt} in their export entry decision while imposing only that we observe a vector $Z_{ijt}\subseteq\mathcal{W}_{ijt}$, the export revenue coefficients $\{\alpha_{jt}\}_{j,t}$ are only partially identified and must be estimated jointly with the remaining parameters $(\beta_0,\beta_1,\sigma_\omega,\sigma_{\omega\nu},\sigma_\nu)$. Given that our sample period covers 10 years and 22 countries, we would have to estimate jointly a confidence set for over 200 parameters. While this is theoretically possible, as far as we know, it is infeasible to compute using our chosen inference approach. Therefore, we simplify the problem by assuming $\alpha_{jt}=\alpha$ for every j and t and, thus, estimate the parameter vector $\theta_S\equiv(\alpha,\beta_0,\beta_1,\sigma_\omega,\sigma_{\omega\nu},\sigma_\nu)$.

We report the results in Table 10. To facilitate comparison, the top portion of this table repeats the results for the baseline case. Allowing the export revenue shock ω_{ijt} to differ from zero yields perfect foresight estimates of the export fixed costs that are smaller than the parallel estimates computed in the baseline case. They are, however, still larger than the baseline moment inequality estimates. For example, the perfect foresight estimates computed in the extended model with $\omega_{ijt}\neq 0$ yields estimates of fixed export costs in Argentina equal to \$323,400, significantly smaller than the baseline perfect foresight estimate of \$868,000. However, they remain larger than the upper bound of the moment inequality baseline estimate, which equals \$102,600. Unfortunately, our moment inequality bounds for the case in which we allow ω_{ijt} to differ from zero are too wide to be informative. A consequence of our bounds being uninformative is that they always contain the corresponding maximum likelihood estimates.[44]

9 Conclusion

We study the extensive margin decision of firms to enter foreign export markets. This participation decision drives much of the variation in trade volume. Thus, to predict how trade

[44]Equations (27) to (29) may be compatible with restrictions on θ_S not accounted for by the inequalities we describe in Appendix G. These additional restrictions may further reduce the size of the confidence set for θ_S. We leave this exploration for future work.

flows will adjust to changes in the economic environment, policymakers first need a measure of the determinants of firms' decisions to engage in exporting. In this paper, we measure these determinants using a moment inequality approach that exploits relatively weak assumptions on the content of exporters' information sets. We show how to use our moment inequalities to recover the fixed costs of exporting, to quantify how firms will react to counterfactual changes in the information they access and in export trade costs, and also to test whether firms use certain key variables to forecast their potential export revenues.

The estimated fixed costs from our inequality model are between ten and thirty percent of the size of the costs found using approaches that require the researcher to specify fully the content of exporters' information sets. When evaluating the effect of endowing potential exporters with better information, we find average export profits of large firms increase significantly while those of small firms remain largely invariant. The overall share of firm-destination pairs with observed positive exports decreases, and this reduction is concentrated among large firms and in large markets. The total volume of exports may increase or decrease.

Finally, we test alternative assumptions on the content of the information sets firms use in their export decision—that is, we test what exporters know. We find important heterogeneity in information sets by firm size: large firms have better information on foreign markets than small firms. While large firms have access to country-by-country information on market-specific demand and trade costs shifters, small firms do not.

References

Ahn, Hyungtaik and Charles F. Manski, "Distribution Theory for the Analysis of Binary Choice under Uncertainty with Nonparametric Estimation of Expectations," *Journal of Econometrics*, 1993, *56*, 291–321. [2, 11]

Albornoz, Facundo, Héctor F. Calvo Pardo, Gregory Corcos, and Emanuel Ornelas, "Sequential Exporting," *Journal of International Economics*, 2012, *88* (1), 1–24. [23]

Almunia, Miguel, Pol Antràs, David Lopez-Rodriguez, and Eduardo Morales, "Venting Out: Exports during a Domestic Slump," *mimeo*, 2018. [5]

Álvarez, Roberto and Gustavo Crespi, "Exporter Performance and Promotion Instruments: Chilean Empirical Evidence," *Estudios de Economía*, 2000, *27* (2), 225–241. [23]

Andrews, Donald W.K. and Gustavo Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78*, 119–157. [8, 10, 24, 25, 28, 68]

— **and Patrik Guggenberger**, "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 2009, *25*, 669–709. [10, 28]

— **and Xiaoxia Shi**, "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, *81* (2), 609–666. [16]

Antràs, Pol, Teresa Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S. Firms," *American Economic Review*, 2017, *107* (9), 2514–2564. [1, 4, 7, 8]

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151–1199. [1, 2]

— **, Sharat Ganapati, and Marc-Andreas Muendler**, "The Extensive Margin of Exporting Goods: A Firm-level Analysis," *mimeo*, November 2015. [1, 4]

Armstrong, Tim B., "Weighted KS Statistic for Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 2014, *182* (2), 92–116. [16]

— , "Asymptotically Exact Inference in Conditional Moment Inequality Models," *Journal of Econometrics*, 2015, *186* (1), 51–65. [16]

— **and Hock Peng Chan**, "Multiscale Adaptive Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 2016, *194* (1), 24–43. [16]

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. [4]

Bagnoli, Mark and Theodore C. Bergstrom, "Log-Concave Probability and Its Applications," *Economic Theory*, 2005, *26* (2), 445–469. [14]

Bernard, Andrew B., Andreas Moxnes, and Karen Helen Ulltveit-Moe, "Two-Sided Heterogeneity and Trade," *Review of Economics and Statistics*, 2017, *forthcoming.* [3]

_ , J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott, "The Margins of US Trade," *American Economic Review*, 2010, *99*, 487–493. [1, 4]

Biesebroeck, Johannes Van, Emily Yu, and Shenjie Chen, "The Impact of Trade Promotion Services on Canadian Exporter Performance," *Canadian Journal of Economics*, 2015, *48* (4), 1481–1512. [26]

Bilbiie, Florian, Fabio Ghironi, and Marc Melitz, "Endogenous Entry, Product Variety and Business Cycles," *Journal of Political Economy*, 2012, *120* (2), 304–345. [4]

Bilir, Kamran L. and Eduardo Morales, "Innovation in the Global Firm," *mimeo*, November 2016. [4]

Blaum, Joaquín, Claire Lelarge, and Michael Peters, "The Gains from Input Trade with Heterogeneous Importers," *American Economic Journal: Macroeconomics*, 2017, *forthcoming.* [1, 4]

Blum, Bernard S., Sebastiàn Claro, and Ignatius J. Horstmann, "Occasional and Perennial Exporters," *Journal of International Economics*, 2013, *90* (1), 65–74. [5]

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. [3, 20, 10, 24, 25, 28, 29]

_ , _ , and _ , "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 2016, *8* (1), 1–38. [19]

Cherkashin, Ivan, Svetlana Demidova, Hiau Looi Kee, and Kala Krishna, "Firm Heterogeneity and Costly Trade: A New Estimation Strategy and Policy Experiments," *Journal of International Economics*, 2015, *96* (1), 18–36. [1]

Chernozhukov, Victor, Sokbae Lee, and Adam M. Rosen, "Intersection Bounds: Estimation and Inference," *Econometrica*, 2013, *81* (2), 667–737. [16]

Chetverikov, Denis, "Adaptive Test of Conditional Moment Inequalities," *Econometric Theory*, 2013, *forthcoming.* [16]

Coşar, Kerem A., Paul L. E. Grieco, Shengyu Li, and Felix Tintelnot, "What Drives Home Market Advantage?," *Journal of International Economics*, 2016, *110*, 135–150. [18]

Crawford, Gregory S. and Ali Yurukoglu, "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 2012, *102* (2), 643–685. [4]

Cunha, Flavio and James J Heckman, "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 2007, *14*, 870–893. [1]

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout, "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 2007, *75* (3), 837–873. [1, 3, 5, 8, 29, 30, 31, 65, 66]

Department of Commerce, "A Profile of U.S. Importing and Exporting Companies, 2013-2014," *Department of Commerce*, 2016. [1]

Dickstein, Michael J., Mark Duggan, Joe Orsini, and Pietro Tebaldi, "The Impact of Market Size and Composition on Health Insurance Premiums: Evidence from the First Year of the ACA," *American Economic Review Papers and Proceedings*, 2015, *105* (5), 120–125. [4]

Doraszleski, Ulrich and Jordi Jaumandreu, "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies*, 2013, *80*, 1338–1383. [4]

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

_ , Marcela Eslava, C.J. Krizan, Maurice Kugler, and James R. Tybout, "A Search and Learning Model of Export Dynamics," *mimeo*, February 2014. [1, 23]

_ , _ , 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. [1]

_ , Samuel Kortum, and Francis Kramraz, "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 2011, *79* (5), 1453–1498. [1, 3]

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

Freeman, Richard, *The Market for College-Trained Manpower*, Cambridge MA: Harvard University Press, 1971. [4]

Hansen, Lars Peter and Kenneth J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 1982, *50* (5), 1269–1286. [30]

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. [8]

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161. [31]

_ and Bo E. Honoré, "The Empirical Content of the Roy Model," *Econometrica*, 1990, *58* (5), 1121–1149. [14]

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," in Bo Honoré, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, eds., *Advances in Economics and Econometrics: Eleventh World Congress (Econometric Society Monographs) (Volume II)*, Cambridge University Press, 2017, pp. 307–359. [29]

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

Holm, Sture, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 1979, *6* (2), 65–70. [21, 10, 11, 28]

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

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. [30]

Maini, Luca and Fabio Pammolli, "Reference Pricing as a Deterrent to Entry: Evidence from the European Pharmaceutical Market," *mimeo*, 2017. [4]

Manski, Charles F., "Nonparametric Estimation of Expectations in the Analysis of Discrete Choice Under Uncertainty," in William Barnett, James Powell, and George Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge: Cambridge University Press, 1991. [2, 11]

— , "Adolescent econometricians: How do youth infer the returns to schooling?," in Charles Clotfelter and Michael Rothschild, eds., *Studies of Supply and Demand in Higher Education*, Chicago: University of Chicago Press, 1993, pp. 43–60. [1, 10]

— , "Measuring Expectations," *Econometrica*, 2004, *5* (9), 1329–1376. [1, 4]

— **and David A. Wise**, *College Choice in America*, Cambridge MA: Harvard University Press, 1983. [4]

Martincus, Christian Volpe and Jerónimo Carballo, "Is Export Promotion Effective in Developing Countries? Firm-Level Evidence on the Intensive and the Extensive Margins of Exports," *Journal of International Economics*, 2008, *76* (1), 89–106. [26]

— **,** — **, and Pablo Garcá**, "Entering New Country and Product Markets: Does Export Promotion Help?," *Review of World Economics*, 2010, *146* (3), 437–467. [26]

Mayer, Thierry and Soledad Zignago, "Notes on CEPII's Distances Measures," *mimeo*, December 2011. [9]

Melitz, Marc, "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71*, 1695–1725. [1]

Morales, Eduardo, Gloria Sheu, and Andrés Zahler, "Extended Gravity," *mimeo*, June 2017. [4, 7, 29, 30]

Moxnes, Andreas, "Are Sunk Costs in Exporting Country-Specific?," *Canadian Journal of Economics*, 2010, *43* (2), 467–493. [1]

Pakes, Ariel, "Alternative Models for Moment Inequalities," *Econometrica*, 2010, *78* (6), 1783–1822. [2, 15]

— and Jack R. Porter, "Moment Inequalities for Multinomial Choice with Fixed Effects," *mimeo*, December 2015. [15]

—, —, Katherine Ho, and Joy Ishii, "Moment Inequalities and their Application," *Econometrica*, 2015, *83* (1), 315–334. [2, 15, 30]

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]

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. [10, 28]

Ruhl, Kim J. and Jonathan L. Willis, "New Exporter Dynamics," *International Economic Review*, 2017, *58* (3), 703–725. [1, 3]

Shi, Xiaoxia, Matthew Shum, and Wei Song, "Estimating Semi-parametric Panel Multinomial Choice Models using Cyclic Monotonicity," *Econometrica*, 2017, *forthcoming.* [15]

Simonovska, Ina and Michael Waugh, "The Elasticity of Trade: Estimates and Evidence," *Journal of International Economics*, 2014, *92* (1), 34–50. [8]

Vannoorenberghe, Gonzague, "Firm-level Volatility and Exports," *Journal of International Economics*, 2012, *86* (1), 57–67. [5]

Weiss, Kenneth D., *Building an Import-Export Business*, Hoboken NJ: John Wiley & Sons, Inc., 2008. [23]

Willis, Robert J. and Sherwin Rosen, "Education and Self-Selection," *Journal of Political Economy*, 1979, *87* (5), 7–36. [4]

Wollman, Thomas G., "Trucks Without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles," *American Economic Review*, 2017, *forthcoming.* [4]

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA: MIT Press, 2002. [12]

Yatchew, Adonis and Zvi Griliches, "Specification Error in Probit Models," *The Review of Economics and Statistics*, 1985, *67* (1), 134–139. [12, 47]

Online Appendix for "What do Exporters Know?"

Michael J. Dickstein
New York University

Eduardo Morales
Princeton University

April 4, 2018

Contents

A Model and Estimation Strategy: Details

A.1 Export Revenue: Details

We exploit the structure introduced in Section 2.1 to derive the expression for the potential export revenue in equation (2). In sections 2.1 and 2.2, we assume potential exporters face: (a) a constant elasticity of substitution demand function; (b) a constant marginal cost; and (c) a monopolistically competitive market in every destination. These three assumptions imply that we can write the potential revenue that firm i would obtain in market j at period t as in equation (1); i.e.

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

Assuming that the same three assumptions hold in the domestic market, we can similarly write the potential revenue that i will obtain in the home market at period t as

$$r_{iht} = \left[\frac{\eta}{\eta - 1} \frac{\tau_{iht} c_{it}}{\zeta_{iht} P_{ht}} \right]^{1-\eta} Y_{ht}. \tag{A.1}$$

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

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

Multiplying by r_{iht} on both sides of the equality and defining

$$\alpha_{ijt} \equiv \left(\frac{\zeta_{iht} \tau_{ijt}}{\zeta_{ijt} \tau_{iht}} \frac{P_{ht}}{P_{jt}} \right)^{1-\eta} \frac{Y_{jt}}{Y_{ht}} \tag{A.3}$$

we obtain:

$$r_{ijt} = \alpha_{ijt} r_{iht}. \tag{A.4}$$

Splitting α_{ijt} into a component that is market-year specific and a component that varies across firms within each market-year pair,

$$\alpha_{jt} \equiv \mathbb{E}_{jt} \left[\left(\frac{\zeta_{iht} \tau_{ijt}}{\zeta_{ijt} \tau_{iht}} \frac{P_{ht}}{P_{jt}} \right)^{1-\eta} \frac{Y_{jt}}{Y_{ht}} \right], \tag{A.5}$$

$$\tilde{\alpha}_{ijt} \equiv \alpha_{ijt} - \alpha_{jt}, \tag{A.6}$$

we can rewrite equation (A.4) as

$$r_{ijt} = \alpha_{jt} r_{iht} + \tilde{\alpha}_{ijt} r_{iht}. \tag{A.7}$$

Finally, defining $e_{ijt} \equiv \tilde{\alpha}_{ijt} r_{iht}$, we obtain equation (2).

A.2 Mean Independence of Firm-Specific Export Revenue Shocks: Details

Assume that, for every firm i, foreign market j, and period t, it holds that

$$\zeta_{ijt} = \zeta_{it} \zeta_{jt} \tilde{\zeta}_{ijt}, \tag{A.8}$$

$$\tau_{ijt} = \tau_{it} \tau_{jt} \tilde{\tau}_{ijt}, \tag{A.9}$$

with

$$\mathbb{E}_{jt} \left[\left(\frac{\tilde{\tau}_{ijt}}{\tilde{\zeta}_{ijt}} \right)^{1-\eta} \middle| \mathcal{J}_{ijt}, \tau_{it}, \zeta_{it}, c_{it}, f_{ijt} \right] = 1, \tag{A.10}$$

and that, for every firm i and period t, the analogous demand and variable costs in the home market h are

$$\zeta_{iht} \equiv \zeta_{it}\zeta_{ht} \tag{A.11}$$
$$\tau_{iht} \equiv \tau_{it}\tau_{ht}. \tag{A.12}$$

Under these assumptions, α_{jt} and e_{ijt} in equation (2) may be rewritten as

$$\alpha_{jt} = \left(\frac{\zeta_{ht}\tau_{jt}}{\zeta_{jt}\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}, \tag{A.13}$$
$$e_{ijt} = \left(\frac{\zeta_{ht}\tau_{jt}}{\zeta_{jt}\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}\left(\left(\frac{\tilde{\tau}_{ijt}}{\tilde{\zeta}_{ijt}}\right)^{1-\eta} - 1\right) r_{iht}, \tag{A.14}$$

and r_{iht} in equation (A.1) may be rewritten as

$$r_{iht} = \left[\frac{\eta}{\eta - 1}\frac{\tau_{it}\tau_{ht}c_{it}}{\zeta_{it}\zeta_{ht}P_{ht}}\right]^{1-\eta} Y_{ht}. \tag{A.15}$$

Except for the composite term $(\tau_{it}c_{it}/\zeta_{it})$, r_{iht} depends exclusively on variables that are constant across firms within each period t. Therefore, using the Law of Iterating Expectations, the mean independence condition in equation (A.10) implies that

$$\mathbb{E}_{jt}\left[\left(\frac{\tilde{\tau}_{ijt}}{\tilde{\zeta}_{ijt}}\right)^{1-\eta}\middle|\mathcal{J}_{ijt}, r_{iht}, f_{ijt}\right] = 1. \tag{A.16}$$

Combining this mean independence condition with the definition of e_{ijt} in equation (A.14), it holds that

$$\begin{aligned}
\mathbb{E}_{jt}[e_{ijt}|\mathcal{J}_{ijt}, r_{iht}, f_{ijt}] &= \mathbb{E}_{jt}\left[\left(\frac{\zeta_{ht}\tau_{jt}}{\zeta_{jt}\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}\left(\left(\frac{\tilde{\tau}_{ijt}}{\tilde{\zeta}_{ijt}}\right)^{1-\eta} - 1\right) r_{iht}\middle|\mathcal{J}_{ijt}, r_{iht}, f_{ijt}\right] \\
&= \left(\frac{\zeta_{ht}\tau_{jt}}{\zeta_{jt}\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}} r_{iht}\mathbb{E}_{jt}\left[\left(\frac{\tilde{\tau}_{ijt}}{\tilde{\zeta}_{ijt}}\right)^{1-\eta} - 1\middle|\mathcal{J}_{ijt}, r_{iht}, f_{ijt}\right] \\
&= \left(\frac{\zeta_{ht}\tau_{jt}}{\zeta_{jt}\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}} r_{iht}\left(\mathbb{E}_{jt}\left[\left(\frac{\tilde{\tau}_{ijt}}{\tilde{\zeta}_{ijt}}\right)^{1-\eta}\middle|\mathcal{J}_{ijt}, r_{iht}, f_{ijt}\right] - 1\right) \\
&= 0,
\end{aligned}$$

where the last equality is implied by equation (A.16). Thus, equations (A.8) to (A.12) impose a set of assumptions on the distribution of the firm-market-year specific demand ζ_{ijt} and variable cost τ_{ijt} under which the mean independence condition in equation (3) holds.

Intuition. Equations (A.8) and (A.11) indicate that the demand shock in every foreign market j is the product of a firm-year specific component ζ_{it}, a market-year specific component ζ_{jt}, and an idiosyncratic term $\tilde{\zeta}_{ijt}$ that captures the firm-specific component of the demand in the foreign market j relative to the home market h. Equations (A.9) and (A.12) indicate that the market-specific variable cost τ_{ijt} has an analogous structure. Given equations (A.8), (A.9), (A.11), and (A.12), we can rewrite the ratio of sales in equation (A.2) as

$$\frac{r_{ijt}}{r_{iht}} = \left[\frac{\tilde{\tau}_{ijt}}{\tilde{\zeta}_{ijt}}\right]^{1-\eta}\left[\frac{\zeta_{ht}\tau_{jt}}{\zeta_{jt}\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right]^{1-\eta}\frac{Y_{jt}}{Y_{ht}}. \tag{A.17}$$

Only the first term in the expression on the right hand side is firm specific and, thus, equation (A.10) establishes a mean independence condition on the composite of the firm-specific components affecting the ratio of export sales in market j to domestic sales; i.e. $(\tilde{\tau}_{ijt}/\tilde{\zeta}_{ijt})^{1-\eta}$. Specifically, equation (A.10) imposes that this ratio is mean independent of: (a) the information the firm has when deciding whether to enter market j; (b) the firm's marginal production cost, c_{it}; (c) firm-specific shifters of variable trade costs and demand that impact equally every market, τ_{it} and ζ_{it}; and, (d) the firm's fixed cost in market j, f_{ijt}. Condition (a) can be relaxed as shown in Section 8.2. Conditions (b), (c) and (d) are necessary to obtain a consistent estimate of α_{jt} by regressing potential export revenue r_{ijt} on domestic sales r_{iht} for the selected subset of firms-destination-year triplets for which we observe positive exports in the data (see Section 4 and Appendix A.3 for additional details on this point).

A.3 Estimation of Export Revenue Shifters

We describe a procedure to consistently estimate the parameter vector $\{\alpha_{jt}; \forall j \text{ and } t\}$. For each destination country j and year t, we use information on the covariates (r_{ijt}, r_{iht}) for every exporting firm–i.e. where $d_{ijt} = 1$. With this data, for each country-year pair jt, we compute the OLS estimate of the coefficient on domestic sales in a regression of observed export revenues on this covariate; i.e.

$$\hat{\alpha}_{jt} \equiv \frac{c\hat{o}v_{jt}(r_{ijt}, r_{iht}|d_{ijt} = 1)}{v\hat{a}r_{jt}(r_{iht}|d_{ijt} = 1)}, \tag{A.18}$$

where $c\hat{o}v_{jt}(r_{ijt}, r_{iht}|d_{ijt} = 1)$ denotes the sample covariance between r_{ijt} and r_{iht} across observations with observed positive exports in the corresponding country-year pair jt; and $v\hat{a}r_{jt}(r_{ijt}, r_{iht}|d_{ijt} = 1)$ denotes the sample variance of r_{iht} across these same firms with observed positive exports.

Given: (a) the definition of α_{jt} and e_{ijt} in equation (2); (b) the mean independence condition in equation (3); (c) the definition of the fixed export costs f_{ijt} in equation (5); and (d) the definition of the export participation dummy d_{ijt} in equation (9):

$$cov(e_{ijt}, r_{iht}|d_{ijt} = 1) = 0$$

and, therefore,

$$plim(\hat{\alpha}_{jt}) = \alpha_{jt}.$$

A.4 Partial Identification: Example

We prove that the model described in Section 2, combined with the assumption that the econometrician only observes a vector $(d_{ijt}, Z_{ijt}, r^o_{ijt})$ such that $Z_{ijt} \subseteq \mathcal{J}_{ijt}$, is not enough to point identify the parameter vector of interest θ^*. To simplify notation, we assume in this section that $\theta_1 = 0$ and that Z_{ijt} is a scalar. None of the conclusions in this section depend on these assumptions.

The data are informative about the joint distribution of $(d_{ijt}, Z_{ijt}, r^o_{ijt})$ across i, j, and t. We denote the joint distribution of the vector $(d_{ijt}, Z_{ijt}, r^o_{ijt})$ as $\mathbb{P}(d_{ijt}, Z_{ijt}, r^o_{ijt})$. In this section, we use $\mathbb{P}(\cdot)$ to denote distributions that may be directly estimated given the available data on $(d_{ijt}, Z_{ijt}, r^o_{ijt})$. To simplify notation, we use r^e_{ijt} to denote $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ in this section. Without loss of generality, we can write

$$\mathbb{P}(d_{ijt}, Z_{ijt}, r^o_{ijt}) = \int f(d_{ijt}, Z_{ijt}, r^o_{ijt}, r^e_{ijt}) dr^e_{ijt},$$

where, for any vector $(x_1, \dots, x_K)$, we use $f(x_1, \dots, x_K)$ to denote the joint distribution of $(x_1, \dots, x_K)$. Here, we use $f(\cdot)$ to denote distributions that involve a variable that is not directly observed in the data, such as r^e_{ijt}. Using rules of conditional distributions, we can write:

$$\mathbb{P}(d_{ijt}, Z_{ijt}, r^o_{ijt}) = \int f^y(d_{ijt}|r^e_{ijt}, r^o_{ijt}, Z_{ijt}) f^y(r^o_{ijt}|r^e_{ijt}, Z_{ijt}) f^y(r^e_{ijt}|Z_{ijt}) \mathbb{P}(Z_{ijt}) dr^e_{ijt}, \tag{A.19}$$

where we use $\mathbb{P}(Z_{ijt})$ to denote the marginal distribution of Z_{ijt} that is directly observable in the data. Any structure $S^y \equiv \{f^y(d_{ijt}|r^e_{ijt}, r^o_{ijt}, Z_{ijt}),\ f^y(r^o_{ijt}|r^e_{ijt}, Z_{ijt}),\ f^y(r^e_{ijt}|Z_{ijt})\}$ is admissible as long as it verifies the restrictions imposed in Section 2 and equation (A.19). Given the assumption that $\theta_1 = 0$, the model in Section 2 imposes the following restriction on the elements of equation (A.19):

$$\begin{gathered} f^y(d_{ijt}|r^e_{ijt}, r^o_{ijt}, Z_{ijt}) = f(d_{ijt}|r^e_{ijt}, Z_{ijt}; \theta) = \\ \Big(\Phi\big((\theta_2)^{-1}\big(\eta^{-1} r^e_{ijt} - \theta_0\big)\big)\Big)^{d_{ijt}} \Big(1 - \Phi\big((\theta_2)^{-1}\big(\eta^{-1} r^e_{ijt} - \theta_0\big)\big)\Big)^{1-d_{ijt}}. \end{gathered} \tag{A.20}$$

The only parameters that are left to identify are (θ_0, θ_2). Here, we show that (θ_0, θ_2) is partially identified in a model that imposes restrictions that are stronger than those in Section 2. Specifically, we impose the following additional restrictions on the elements of equation (A.19)

$$Z_{ijt} = r^e_{ijt} + \xi_{ijt}, \qquad \xi_{ijt}|r^e_{ijt} \sim \mathbb{N}((\sigma_\xi/\sigma_{r^e})\rho_{\xi r^e}(r^e_{ijt} - \mu_{r^e}), (1 - \rho^2_{\xi r^e})\sigma^2_\xi), \tag{A.21a}$$

$$r^o_{ijt} = r^e_{ijt} + \varepsilon_{ijt}, \qquad \varepsilon_{ijt}|(r^e_{ijt}, \xi_{ijt}) \sim \mathbb{N}(0, \sigma^2_\varepsilon), \tag{A.21b}$$

$$r^e_{ijt} \sim \mathbb{N}(\mu_{r^e}, \sigma^2_{r^e}). \tag{A.21c}$$

Equation (A.21a) imposes a particular assumption on the joint distribution of firms' unobserved true expectations r^e_{ijt} and the subset of the variables used by firms to form those expectations that are observed to the researcher, Z_{ijt}. The model in Section 2 does not impose any assumption on this relationship. Equation (A.21b) assumes that firms' expectational error is normally distributed and independent of both firms' unobserved expectations and the difference between the instrument Z_{ijt} and the unobserved expectations, ξ_{ijt}. By contrast, the model in Section 2 only imposes mean independence between ε_{ijt} and r^e_{ijt}. Finally, equation (A.21c) imposes that firms' unobserved expectations are normally distributed; in the model in the main text, we do not impose a distributional assumption. Therefore, it is clear that equation (A.21) defines a model that is more restrictive than that defined in Section 2. However, as we show below, even after imposing the assumptions in equation (A.21), we can still find at least two structures

$$S^{y_1} \equiv \{(\theta_0^{y_1}, \theta_2^{y_1}), f^{y_1}(r^o_{ijt}|r^e_{ijt}, Z_{ijt}), f^{y_1}(r^e_{ijt}|Z_{ijt})\},$$
$$S^{y_2} \equiv \{(\theta_0^{y_2}, \theta_2^{y_2}), f^{y_2}(r^o_{ijt}|r^e_{ijt}, Z_{ijt}), f^{y_2}(r^e_{ijt}|Z_{ijt})\},$$

that verify: (1) equations (A.19), (A.20) and (A.21); and (2) $\theta^{y_1} \neq \theta^{y_2}$. If θ is partially identified in this stricter model, it will also be partially identified in the more general model described in Section 2.

Equation (A.21b) assumes that the expectational error not only has mean zero and finite variance but is also normally distributed. This implies that the conditional density $f(r^o_{ijt}|r^e_{ijt}, Z_{ijt})$ is normal:

$$f(r^o_{ijt}|r^e_{ijt}, Z_{ijt}) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\Big[-\frac{1}{2}\Big(\frac{r_{ijt} - r^e_{ijt}}{\sigma_\varepsilon}\Big)^2\Big].$$

By applying Bayes' rule, both equations (A.21a) and (A.21c) jointly determine the conditional density $f(r^e_{ijt}|Z_{ijt})$ entering equation (A.19).

Result A.4.1 *There exists empirical distributions of the vector of observable variables* (d, Z, X), $\mathbb{P}(d, Z, X)$, *such that there are at least two structures* S^{y_1} *and* S^{y_2} *for which*

1. *both* S^{y_1} *and* S^{y_2} *verify equations* (A.19)*,* (A.20)*, and* (A.21)*;*
2. $\theta^{y_1} \neq \theta^{y_2}$.

This result can be proved by combining the following two lemmas.

Lemma A.4.1 *The parameter vector* (θ_0, θ_2) *is point-identified only if the parameter* $\sigma_{r^e} = var(r^e_{ijt})$ *is point-identified.*

Proof: Define $r^e_{ijt} = \sigma_{r^e}\tilde{r}^e_{ijt}$, such that $var(\tilde{r}^e_{ijt}) = 1$. We can then rewrite equation (A.20) as

$$\Big(\Phi\big(\eta^{-1}\frac{\sigma_{r^e}}{\theta_2}\tilde{r}^e_{ijt} - \frac{\theta_0}{\theta_2}\big)\Big)^{d_{ijt}}\Big(1 - \Phi\big(\eta^{-1}\frac{\sigma_{r^e}}{\theta_2}\tilde{r}^e_{ijt} - \frac{\theta_0}{\theta_2}\big)\Big)^{1-d_{ijt}}.$$

The parameter θ_2 only enters likelihood function in equation (A.19) either dividing σ_{r^e} or dividing θ_0. Therefore, we can only separately identify θ_0 and θ_2 if we know σ_{r^e}. ■

Lemma A.4.2 *The parameter vector* σ_{r^e} *is point-identified if and only if the parameter* $\rho_{\xi r^e}$ *is assumed to be equal to zero.*

Proof: From equations (A.21a), (A.21b), and (A.21c), we can conclude that r_{ijt} and Z_{ijt} are jointly normal. Therefore, the information arising from observing their joint distribution is summarized in three moments:

$$\begin{aligned}\sigma_r^2 &= \sigma_{r^e}^2 + \sigma_\varepsilon^2,\\ \sigma_z^2 &= \sigma_{r^e}^2 + \sigma_\xi^2 + 2\rho_{\xi r^e}\sigma_{r^e}\sigma_\xi,\\ \sigma_{rz} &= \sigma_{r^e}^2 + \rho_{\xi r^e}\sigma_{r^e}\sigma_\xi \end{aligned} \tag{A.22}$$

The left hand side of these three equations is directly observed in the data. If we impose the assumption that $\rho_{\xi r^e} = 0$, then $\sigma_{rz} = \sigma^2_{r^e}$ and, therefore, from Lemma A.4.1, the vector θ is point identified. If we allow $\rho_{\xi r^e}$

to be different from zero, the system of equations in equation (A.22) only allows us to define bounds on $\sigma^2_{r^e}$. We can rewrite the system of equations in equation (A.22) as

$$\begin{aligned}\sigma^2_r &= \sigma^2_{r^e} + \sigma^2_{\varepsilon},\\ \sigma^2_z &= \sigma^2_{r^e} + \sigma^2_{\xi} + 2\sigma_{\xi r^e}\\ \sigma_{rz} &= \sigma^2_{r^e} + \sigma_{\xi r^e}.\end{aligned} \tag{A.23}$$

This is a linear system with 3 equations and 4 unknowns, $(\sigma^2_{r^e}, \sigma^2_{\varepsilon}, \sigma^2_{\xi}, \sigma_{\xi r^e})$. Therefore, the system is underidentified and does not have a unique solution for $\sigma^2_{r^e}$.

A.5 Deriving Unconditional Moments

The moment inequalities described in equations (14) and (17) condition on particular values of the instrument vector, Z. From these conditional moments, we can derive unconditional moment inequalities. Each of these unconditional moments is defined by an *instrument function*. Specifically, given a positive-valued instrument function $g(\cdot)$, we derive unconditional moments that are consistent with our conditional moments:

$$\mathbb{E}\left[\left\{\begin{array}{c} m^{ob}_l(d_{ijt}, r_{ijt}, dist_j; \gamma)\\ m^{ob}_u(d_{ijt}, r_{ijt}, dist_j; \gamma)\\ m^{r}_l(d_{ijt}, r_{ijt}, dist_j; \gamma)\\ m^{r}_u(d_{ijt}, r_{ijt}, dist_j; \gamma)\end{array}\right\} \times g(Z_{ijt})\right] \geq 0,$$

where $m^{ob}_l(\cdot)$, $m^{ob}_u(\cdot)$, $m^{r}_l(\cdot)$, $m^{r}_u(\cdot)$, and Z_{ijt} are defined in equations (14) and (17).

In Section 5, we present results based on a set of instrument functions $g_a(\cdot)$ such that, for each scalar random variable Z_{kijt} included in the instrument vector Z_{ijt}

$$g_a(Z_{kijt}) = \left\{\begin{array}{c} \mathbb{1}\{Z_{kijt} > med(Z_{kijt})\}\\ \mathbb{1}\{Z_{kijt} \leq med(Z_{kijt})\}\end{array}\right\} \times (|Z_{kijt} - med(Z_{kijt})|)^a.$$

In words, for each of scalar random variable Z_{kijt} included in the instrument vector $Z_{ijt} = (Z_{1ijt}, \ldots, Z_{kijt}, \ldots, Z_{Kijt})$, the function $g_a(\cdot)$ builds two moments by splitting the observations into two groups depending on whether the value of the instrument variable for that observation is above or below its median. Within each moment, each observation is weighted differently depending on the value of a and on the absolute value of the distance between the value of Z_{kijt} and its median value in the sample. Specifically, in Section 5, we assume that $Z_{ijt} = (r_{iht-1}, R_{jt-1}, dist_j)$ and, for a given value of a, we construct the following instruments

$$g_a(Z_{ijt}) = \left\{\begin{array}{l} \mathbb{1}\{r_{iht-1} > med(r_{iht-1})\} \times (|r_{iht-1} - med(r_{iht-1})|)^a,\\ \mathbb{1}\{r_{iht-1} \leq med(r_{iht-1})\} \times (|r_{iht-1} - med(r_{iht-1})|)^a,\\ \mathbb{1}\{R_{jt-1} > med(R_{jt-1})\} \times (|R_{jt-1} - med(R_{jt-1})|)^a,\\ \mathbb{1}\{R_{jt-1} \leq med(R_{jt-1})\} \times (|R_{jt-1} - med(R_{jt-1})|)^a,\\ \mathbb{1}\{dist_j > med(dist_j)\} \times (|dist_j - med(dist_j)|)^a,\\ \mathbb{1}\{dist_j \leq med(dist_j)\} \times (|dist_j - med(dist_j)|)^a.\end{array}\right.$$

Given that each particular instrument function $g_a(Z_{ijt})$ contains six instruments and there are four basic odds-based and revealed preference inequalities (in equations (14) and (17)), the total number of moments used in the estimation is equal to twenty-four for a given value of a. In the benchmark case we simultaneously use two different instrument functions, $g_a(Z_{ijt})$, for $a = \{0, 1\}$, to compute the 95% confidence set $\hat{\Theta}^{95\%}$.[45]

A.6 Bootstrap Standard Errors of Maximum Likelihood Estimates: Details

We describe here the procedure we follow to compute the standard errors for the maximum likelihood estimate of θ. These estimates are a function of the proxy for the expected observed component of the potential revenues from exporting, $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}]$, which are themselves a function of the estimates of the export revenue parameters $\{\alpha_{jt}; \forall j, t\}$. To account for the sampling noise in our measure of the expected potential revenues when we

[45] We have recomputed the tables presented in Section 5 using alternative definitions of the instrument function $g_a(Z_{ijt})$. Even though the boundaries of the confidence sets depend on the instrument functions, the main conclusions are robust.

compute our maximum likelihood estimates of θ, we implement the following bootstrap procedure:

Step 1: for each destination-year pair jt, draw a bootstrap sample of firms of size N_{jt}. We construct a bootstrap sample for a destination-year jt by drawing N_t firms with replacement from the original set of N_t firms active at period t.[46] For each firm i that we include in our bootstrap sample for the destination-year pair jt, we collect information on its domestic sales, r_{iht}, and on its exports (which may equal 0) to country j, r_{ijt}.

Step 2: combine all jt-specific bootstrap samples drawn in step 1 into a single bootstrap sample. The size of the resulting sample will thus be $\sum_j \sum_t N_{jt}$, identical to the size of the original sample.

Step 3: obtain bootstrap estimates of the export revenue shifters. Using the bootstrap sample in step 2, we apply the estimation procedure described in Appendix A.3 to compute the estimated export revenue shifters $\{\hat{\alpha}_{jt,b}; \forall j, t\}$, where the additional subscript b indexes the bootstrapped sample used to compute these estimates.

Step 4: given an assumed information set for every observation, we obtain bootstrap estimates of the fixed export costs parameters. Given an assumed information set $\mathcal{J}^a_{ijt}$ for every firm, country and year in the bootstrap sample defined in step 2, and given the estimates $\{\hat{\alpha}_{jt,b}; \forall j, t\}$ defined in step 3, we use all observations in the bootstrap sample defined in step 2 to construct a measure of predicted export revenues for every firm, country and year in it as a non-parametric projection of $\hat{\alpha}_{jt,b} r_{iht}$ on $\mathcal{J}^a_{ijt}$. Given the outcome of this non-parametric projection, we obtain an estimate of the fixed export cost parameter θ by maximizing the log-likelihood function defined in equations (11) and (12). We denote this estimate as $\hat{\theta}_b \equiv (\hat{\theta}_{0,b}, \hat{\theta}_{1,b}, \hat{\theta}_{2,b})$.

Step 5: repeat steps 1 to 4 1000 times. For each $k = 0, 1, 2$, we obtain thus a set of 1000 bootstrap estimates of θ_k, $\{\hat{\theta}_{k,b}; b = 1, \ldots, 1000\}$.

Step 6: for each $k = 0, 1, 2$, compute the bootstrap standard error of the maximum likelihood estimate of θ_k as the standard deviation of $\{\hat{\theta}_{k,b}; b = 1, \ldots, 1000\}$. Specifically, for each $k = 0, 1, 2$, we compute $s.e.(\hat{\theta}_k)$ as

$$s.e.(\hat{\theta}_k) = \sqrt{\frac{1}{1000} \sum_{b=1}^{1000} (\hat{\theta}_{k,b} - \bar{\hat{\theta}}_k)^2}, \qquad \text{with} \qquad \bar{\hat{\theta}}_k = \frac{1}{1000} \sum_{b=1}^{1000} \hat{\theta}_{k,b}. \tag{A.24}$$

In order to compute the standard errors reported in Table 2, we repeat steps 1 to 6 above for the information sets $\{\mathcal{J}^a_{ijt}; \forall i, j, t\}$ corresponding to the perfect foresight case as well as for those corresponding to the minimal information case. In both cases, we use the same 1000 bootstrap samples. A problem with the bootstrap procedure described above is that some of the bootstrap samples do not contain any firm exporting to a market j in a period t and, consequently, we cannot compute the estimate $\hat{\alpha}_{jt}$ for these bootstrap samples. Whenever this happens, we discard the observations corresponding to this destination-year pair jt from the combined bootstrap sample we construct in step 2 above. In large samples, there should be no bootstrap draws that miss all actual exporting firms for a single destination-year pair; therefore, our discarding procedure should not affect the asymptotic validity of our bootstrap procedure.

In order to compute the standard errors reported in Table 3 for a given assumption on the information sets $\{\mathcal{J}^a_{ijt}; \forall i, j, t\}$, we additionally need to compute an estimate of the covariance between the maximum likelihood estimators of θ_0 and θ_1 computed under such informational assumption. Using the corresponding bootstrap estimates $\{(\hat{\theta}_{0,b}, \hat{\theta}_{1,b}); b = 1, \ldots, 1000\}$, we compute such covariance as

$$cov(\hat{\theta}_0, \hat{\theta}_1) = \frac{1}{1000} \sum_{b=1}^{1000} (\hat{\theta}_{0,b} - \bar{\hat{\theta}}_0)(\hat{\theta}_{1,b} - \bar{\hat{\theta}}_1), \qquad \text{with} \qquad \bar{\hat{\theta}}_k = \frac{1}{1000} \sum_{b=1}^{1000} \hat{\theta}_{k,b}, \text{ for } k = 0, 1. \tag{A.25}$$

Using this covariance, we compute the bootstrap standard error of the maximum likelihood estimate of the

[46] We define a firm as active in year t if we observe that it has positive domestic sales in the home market, Chile. A firm need not sell in any foreign market to be categorized as active in year t.

average fixed cost of exporting to a particular destination j as

$$\begin{aligned} s.e.(\hat{\theta}_0 + \hat{\theta}_1 \times dist_j) &= \sqrt{var(\hat{\theta}_0 + \hat{\theta}_1 \times dist_j)} \\ &= \sqrt{(s.e.(\hat{\theta}_0))^2 + (dist_j)^2 \times (s.e.(\hat{\theta}_1))^2 + 2 \times dist_j \times cov(\hat{\theta}_0, \hat{\theta}_1)}, \end{aligned} \tag{A.26}$$

where $s.e.(\hat{\theta}_0)$, $s.e.(\hat{\theta}_1)$, and $cov(\hat{\theta}_0, \hat{\theta}_1)$ are computed as indicated in equations (A.24) and (A.25) using the bootstrap estimates $\{(\hat{\theta}_{0,b}, \hat{\theta}_{1,b}); b = 1, \ldots, 1000\}$ computed under the corresponding informational assumption. We compute the standard error in equation (A.26) for every country j both under the assumption of perfect foresight and under the minimal information set assumption.

A.7 Confidence Sets for True Parameter: Details

A.7.1 Computation

We implement the asymptotic version of the Generalized Moment Selection (GMS) test described on page 135 of Andrews and Soares (2010) to compute the confidence set for the true parameter vector θ^*. Our confidence set for θ^*, however, is a function of our measure of the potential export profits, $\hat{r}_{ijt} \equiv \hat{\alpha}_{jt} r_{iht}$, and, consequently, is also a function of the preliminary estimates of the export revenue parameters $\{\alpha_{jt}; \forall j, t\}$. We adapt the procedure in Section 10.2 of Andrews and Soares (2010) to account for the fact that our moment inequalities depend on (consistent and asymptotically normal) estimators $\{\hat{\alpha}_{jt}; \forall j, t\}$.

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

$$\bar{m}_k(\theta; \hat{\alpha}) \geq 0, \qquad k = 1, \ldots, K,$$

where $\hat{\alpha} \equiv \{\hat{\alpha}_{jt}; \forall j, t\}$ and, for every $k = 1, \ldots, K$,

$$\bar{m}_k(\theta; \hat{\alpha}) \equiv \frac{1}{n} \sum_i \sum_j \sum_t m_k(X_{ijt}, Z_{ijt}, \theta; \hat{\alpha}),$$

with n denoting the sample size (i.e. sum of distinct ijt triplets included in our sample). The MMM statistic is therefore defined as

$$Q(\theta; \hat{\alpha}) = \sum_{k=1}^{K} (\min\{\frac{\bar{m}_k(\theta; \hat{\alpha})}{\hat{\sigma}_k(\theta; \hat{\alpha})}, 0\})^2, \tag{A.27}$$

where $\hat{\sigma}_k(\theta; \hat{\alpha}) = \sqrt{\hat{\sigma}_k^2(\theta; \hat{\alpha})}$ and

$$\hat{\sigma}_k^2(\theta; \hat{\alpha}) = \frac{1}{n} \sum_i \sum_j \sum_t (m_k(X_{ijt}, Z_{ijt}, \theta; \hat{\alpha}) - \bar{m}_k(X_{ijt}, Z_{ijt}; \theta; \hat{\alpha}))^2.$$

In the notation introduced in sections 4.2.1 and 4.2.2, $X_{ijt} \equiv (d_{ijt}, r^o_{ijt}, dist_j)$ and $m_k(\cdot)$ may be either an odds-based or a revealed-preference moment function. 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 odds-based and revealed-preference moment inequalities described in sections 4.2.1 and 4.2.2; Appendix A.5 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 (A.27), we compute confidence sets for the true parameter value θ^* 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 θ. In the case of the confidence set for the parameter vector $\theta^* \equiv (\beta_0, \beta_1, \sigma)$, Θ_g is a 3-dimensional orthotope. To define the limits of Θ_g, we solve the following nonlinear optimization

$$\begin{aligned} &min_\theta \ \mathbb{d} \cdot \theta \\ &\text{subject to} \end{aligned}$$

$$\frac{1}{n}\sum_i\sum_j\sum_t m(X_{ijt}, Z_{ijt}, \theta; \hat{\alpha}) + \ln n \geq 0, \tag{A.28}$$

where n denotes the sample size (i.e. sum of distinct ijt triplets included in our sample), $m(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}) \equiv (m_1(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}), m_2(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}), \ldots, m_K(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}))$, and $\mathbb{d}$ is one of the elements of the matrix

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

Given the six elements of $\mathbb{D}$, we solve six nonlinear optimizations as in equation (A.28). Denoting the six 3-dimensional vectors θ that solve each of these optimizations as $(\theta_{1+}, \theta_{1-}, \theta_{2+}, \theta_{2-}, \theta_{3+}, \theta_{3-})'$, we compute the six boundaries of the 3-dimensional orthotope Θ_g as

$$\begin{pmatrix} \mathbb{d}_{1+} \cdot \theta_{1+} & \mathbb{d}_{1-} \cdot \theta_{1-} \\ \mathbb{d}_{2+} \cdot \theta_{2+} & \mathbb{d}_{2-} \cdot \theta_{2-} \\ \mathbb{d}_{3+} \cdot \theta_{3+} & \mathbb{d}_{3-} \cdot \theta_{3-} \end{pmatrix}$$

where the first column contains the minimum value of the element of θ indicated by the corresponding row and the second column contains the corresponding maximum. Once we have these six limits of Θ_g, we fill it with 64,000 equidistant points.

Step 2: choose a point $\theta_p \in \Theta_g$. With θ_p, we test the null hypothesis that the vector θ_p equals the true value of θ:

$$H_0 : \theta^* = \theta_p \qquad \text{vs.} \qquad H_0 : \theta^* \neq \theta_p.$$

Step 3: evaluate the MMM test statistic at θ_p:

$$Q(\theta_p; \hat{\alpha}) = \sum_{k=1}^{K}(\min\{\frac{\bar{m}_k(\theta_p; \hat{\alpha})}{\hat{\sigma}_k(\theta_p; \hat{\alpha})}, 0\})^2, \tag{A.29}$$

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

$$\hat{\Omega}(\theta_p) = Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))\hat{\Sigma}(\theta_p)Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p)),$$

where $Diag(\hat{\Sigma}(\theta_p))$ is the $K \times K$ diagonal matrix whose diagonal elements are equal to those of $\hat{\Sigma}(\theta_p)$, $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))$ is a matrix such that $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p)) = Diag^{-1}(\hat{\Sigma}(\theta_p))$ and

$$\hat{\Sigma}(\theta_p) = \frac{1}{n}\frac{1}{B}\sum_{b=1}^{B}\sum_i\sum_j\sum_t(m(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}_b) - \bar{m}(\theta_p))(m(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}_b) - \bar{m}(\theta_p))',$$

where $\hat{\alpha}_b$ is a draw from the asymptotic distribution of $\hat{\alpha}$, B is the total number of draws we use to simulate the asymptotic distribution of $\hat{\alpha}$ and $\bar{m}(\theta_p) = (\bar{m}_1(\theta_p), \ldots, \bar{m}_K(\theta_p))$, with

$$\bar{m}_k(\theta_p) \equiv \frac{1}{n}\frac{1}{B}\sum_{b=1}^{B}\sum_i\sum_j\sum_t m_k(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}_b), \qquad \text{for all } k = 1, \ldots, K.$$

In practice, for computational reasons, we set $B = 1000$. The correlation matrix $\hat{\Omega}(\theta_p)$ accounts for the effect of sampling variation in $\hat{\alpha}$ on the correlation of the moment inequalities we use for estimation.

Step 5: simulate the asymptotic distribution of $Q(\theta_p; \hat{\alpha})$. Take 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 criterion function $Q^{AA}_{n,r}(\theta_p;\hat{\alpha})$ as

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

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

Step 6: compute critical value. The critical $\hat{c}^{AA}_n(\theta_p, 1-\delta;\hat{\alpha})$ is the $(1-\delta)$-quantile of the distribution of $Q^{AA}_{n,r}(\theta_p;\hat{\alpha})$ across the R draws taken in the previous step.

Step 7: accept/reject θ_p. Include θ_p in the estimated $(1-\delta)\%$ confidence set, $\hat{\Theta}^{1-\delta}$, if $Q(\theta_p;\hat{\alpha}) \leq \hat{c}^{AA}_n(\theta_p, 1-\delta;\hat{\alpha})$.

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 $\hat{\Theta}^{1-\alpha}$ to those in the set Θ_g. If (a) some of the points included in the set $\hat{\Theta}^{1-\alpha}$ 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 $\hat{\Theta}^{1-\alpha}$ is only a small fraction of those included in Θ_g, redefine a set Θ_g that is again a 3-dimensional orthotope whose limits are the result of adding a small number to the corresponding limits of the set $\hat{\Theta}^{1-\alpha}$ and repeat steps 2 to 9. If neither (a) nor (b) applies, define $\hat{\Theta}^{1-\alpha}$ as the 95% confidence set for θ^*.

A.8 Specification Tests: Details

We follow a two-part procedure to test the validity of the model defined by our moment inequalities when we assume a particular vector of observed covariates Z_{ijt} belongs to the firm's information set. First, we compute appropriate p-values for each of the tests we perform. Second, given these p-values, we define families of tests and compute family-adjusted p-values using the procedure in Holm (1979).

A.8.1 Computing P-values for Each Test

We use here the same notation as in Appendix A.7.1 and, thus, write the set of moment inequalities we use for identification as

$$\bar{m}_k(\theta;\hat{\alpha}) \geq 0, \qquad k = 1, \ldots, K,$$

where θ denotes the unknown parameter vector whose true value is θ^*, $\hat{\alpha} \equiv \{\hat{\alpha}_{jt};\forall j,t\}$ and, for every $k = 1,\ldots,K$,

$$\bar{m}_k(\theta;\hat{\alpha}) \equiv \frac{1}{n}\sum_i\sum_j\sum_t m_k(X_{ijt}, Z_{ijt}, \theta;\hat{\alpha}),$$

with n and X_{ijt} defined as in Appendix A.7.1. The identified set Θ_0 includes all values of the parameter vector θ consistent with these K moment inequalities. The model defined by these inequalities is said to be correctly specified (or statistically adequate) when Θ_0 is non-empty. Thus, formally, we want to test

$$H_0 : \Theta_0 \neq \emptyset \qquad \text{vs.} \qquad H_1 : \Theta_0 = \emptyset.$$

Therefore, the null hypothesis in our specification test is that the model is correct; i.e. Θ_0 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 RC", as it is computationally simpler than the "test RS".

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 A.7.1. Specifically, we reject the model in a test with size δ if the $(1-\delta)\%$ confidence set for the true value of the parameter vector is empty. As Andrews and Guggenberger (2009), Andrews and Soares (2010), and Bugni et al. (2015) point out, this test is conservative—i.e., if the model defined by moment inequalities being tested is correctly specified, asymptotically, the test with size δ will actually reject the null hypothesis less than $100\delta\%$ of times.

Test RC We compute this test using the following steps:

Steps 1 to 3. Identical to steps 1 to 3 in Appendix A.7.1.

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

$$\hat{\Omega}(\theta_p) = Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))\hat{\Sigma}(\theta_p)Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p)),$$

where $Diag(\hat{\Sigma}(\theta_p))$ is the $K \times K$ diagonal matrix whose diagonal elements are equal to those of $\hat{\Sigma}(\theta_p)$, $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))$ is a matrix such that $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p))Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_p)) = Diag^{-1}(\hat{\Sigma}(\theta_p))$ and

$$\hat{\Sigma}(\theta_p) = \frac{1}{n}\sum_i\sum_j\sum_t(m(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}) - \bar{m}(\theta_p; \hat{\alpha}))(m(X_{ijt}, Z_{ijt}, \theta_p; \hat{\alpha}) - \bar{m}(\theta_p; \hat{\alpha}))',$$

where $\bar{m}(\theta_p; \hat{\alpha}) = (\bar{m}_1(\theta_p; \hat{\alpha}), \ldots, \bar{m}_K(\theta_p; \hat{\alpha}))$. This correlation matrix $\hat{\Omega}(\theta_p)$ conditions on the first-stage estimate $\hat{\alpha}$.

Step 5: simulate the asymptotic distribution of $Q(\theta_p; \hat{\alpha})$. Take 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 criterion function $Q^{AA}_{n,r}(\theta_p; \hat{\alpha})$ as

$$Q^{AA}_{n,r}(\theta_p; \hat{\alpha}) = \sum_{k=1}^{K}\left\{(\min\{[\hat{\Omega}^{\frac{1}{2}}(\theta_p)\zeta_r]_k, 0\})^2 \times \mathbb{1}\{\sqrt{n}\frac{\bar{m}_k(\theta_p; \hat{\alpha})}{\hat{\sigma}_k(\theta_p; \hat{\alpha})} \leq \sqrt{\ln n}\}\right\}$$

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

Step 6: compute critical value. The critical $\hat{c}^{AA}_n(\theta_p, 1-\delta; \hat{\alpha})$ is the $(1-\delta)$-quantile of the distribution of $Q^{AA}_{n,r}(\theta_p; \hat{\alpha})$ across the R draws taken in the previous step.

Step 7: compute the minimum of the MMM test statistic across all $\theta_p \in \Theta_g$:

$$T(\hat{\alpha}) = \inf_{\theta_p \in \Theta_g} Q(\theta_p; \hat{\alpha}), \tag{A.30}$$

where $Q(\theta_p; \hat{\alpha})$ is defined in equation (G.16).

Step 8: compute the minimum of the critical values $\hat{c}^{AA}_n(\theta_p, 1-\delta; \hat{\alpha})$ across all $\theta_p \in \Theta_g$:

$$\hat{c}^{RC}_n(1-\delta; \hat{\alpha}) = \inf_{\theta_p \in \Theta_g} \hat{c}^{AA}_n(\theta_p, 1-\delta; \hat{\alpha}), \tag{A.31}$$

where $\hat{c}^{AA}_n(\theta_p, 1-\delta; \hat{\alpha})$ is defined in step 6.

Step 9: accept/reject H_0. Reject H_0 if

$$T(\hat{\alpha}) > \hat{c}^{RC}_n(1-\delta; \hat{\alpha}).$$

A.8.2 Adjusting P-values Following Holm (1979)

Given the p-values for the individual tests, computed as indicated in Appendix A.8.1, we apply the procedure in Holm (1979) to adjust these p-values for the fact that we simultaneously test multiple hypotheses. With a family of tests $H_1, H_2, \ldots, H_S$ with individual p-values $p_1, p_2, \ldots, p_S$, we proceed as follows:

Step 1: accept/reject H_0. Rank the S hypothesis in increasing order of their individual p-values. Denote this index as (i).

Step 2: adjust individual p-values. Denoting as $\tilde{p}_{(i)}$ the adjusted p-value for the (i)-th smallest individual p-value, we compute:

$$\tilde{p}_{(i)} = \max_{j \leq i} \big\{ \min\{(S-j+1)p_{(j)}, 1\} \big\}.$$

For the test with the smallest p-value (i.e. $p_{(1)}$), the adjusted p-value is simply the corresponding unadjusted p-value times the number of tests being performed (i.e. $\tilde{p}_{(1)} = S \times p_{(1)}$). For the test with the second smallest p-value (i.e. $p_{(2)}$), the adjusted p-value is the maximum of $\tilde{p}_{(1)}$ and the product of the corresponding unadjusted p-value times the number of tests being performed minus one (i.e. $\tilde{p}_{(1)} = \max\{\tilde{p}_{(1)}, (S-1) \times p_{(2)}\}$). Generally, for the i-th smallest p-value (i.e. $p_{(i)}$), the adjusted p-value is the maximum of $\{\tilde{p}_{(1)}, \ldots, \tilde{p}_{(i-1)}\}$ and the product of the corresponding unadjusted p-value times the number of tests being performed minus $i-1$ (i.e. $\tilde{p}_{(i)} = \max\{\tilde{p}_{(1)}, \ldots, p_{(i-1)}, (S-i+1) \times p_{(i)}\}$).

A.9 Counterfactual Changes in Information Sets: Details

In Section 7.1, we compute the value of acquiring extra information for a firm with an initial information set that includes three variables: (1) the distance to each destination market, $dist_j$, (2) last year's aggregate exports from the home country to the destination market, R_{jt-1}, and (3) the firm's own domestic sales, r_{iht-1}, from last year. We denote this initial information set as

$$\mathcal{J}^s_{ijt} \equiv (dist_j, R_{jt-1}, r_{iht-1}), \tag{A.32}$$

where the superscript s stands for "small" information set. To a firm with this specific information set, we explore the value of acquiring information on each destination's demand shock, α_{jt-1}:

$$\mathcal{J}^m_{ijt} \equiv (dist_j, R_{jt-1}, r_{iht-1}, \alpha_{jt-1}), \tag{A.33}$$

We also consider the value of acquiring all additional information necessary to predict the observable determinants of potential export revenues perfectly:

$$\mathcal{J}^l_{ijt} \equiv r^o_{jt}, \tag{A.34}$$

where the superscripts m and l respectively stand for "medium" and "large" information sets. For each of these three information sets, we can define the associated predicted export revenues as

$$\hat{r}^s_{ijt} \equiv \mathbb{E}[r^o_{ijt}|\mathcal{J}^s_{ijt}], \qquad \hat{r}^m_{ijt} \equiv \mathbb{E}[r^o_{ijt}|\mathcal{J}^m_{ijt}], \qquad \hat{r}^l_{ijt} \equiv \mathbb{E}[r^o_{ijt}|\mathcal{J}^l_{ijt}] = r_{ijt},$$

The associated expectational errors that firms make when trying to predict r^o_{ijt} are:

$$\hat{\varepsilon}^s_{ijt} \equiv r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}^s_{ijt}], \qquad \hat{\varepsilon}^m_{ijt} \equiv r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}^m_{ijt}], \qquad \hat{\varepsilon}^l_{ijt} \equiv r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}^l_{ijt}] = 0.$$

To compute the vector of predicted revenues $\{\hat{r}^s_{ijt}; \forall(i,j,t)\}$ for a particular subset of firms, countries and years $\mathcal{S}$ (e.g. firms with above median domestic sales, all destinations and all years), we run a linear regression of r^o_{ijt} on $(dist_j, R_{jt-1}, r_{iht-1})$ using the observations belonging to $\mathcal{S}$ and use the OLS estimated coefficients to form the predicted values. Similarly, to compute the vector of predicted revenues $\{\hat{r}^m_{ijt}; \forall(i,j,t)\}$, we run a linear regression of r^o_{ijt} on $(dist_j, R_{jt-1}, r_{iht-1}, \alpha_{jt-1})$ and use the OLS estimated coefficients to form the corresponding predicted values.

Conditional on a particular value of the parameter vector θ, an information set $\mathcal{J}^x_{ijt}$ -for x equal to s, m or l-, potential export revenues r_{ijt}, and a fixed export cost shock ν_{ijt}, the potential ex post profits that a firm i may make in period t in a destination j are

$$\pi_{ijt}(\theta, \nu_{ijt}) = \eta^{-1} r_{ijt} - \beta_0 - \beta_1 dist_j - \nu_{ijt},$$

the actual export participation decision is

$$d_{ijt}(\theta, \mathcal{J}^x_{ijt}, \nu_{ijt}) = \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^x_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\},$$

the realized export sales are

$$d_{ijt}(\theta, \mathcal{J}_{ijt}^x, \nu_{ijt}) \times r_{ijt} =$$
$$\mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}^o | \mathcal{J}_{ijt}^x] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\} \times r_{ijt},$$

and the realized profits are

$$d_{ijt}(\theta, \mathcal{J}_{ijt}^x, \nu_{ijt}) \times \pi_{ijt}(\theta, \nu_{ijt}) =$$
$$\mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}^o | \mathcal{J}_{ijt}^x] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\} \times (\eta^{-1} r_{ijt} - \beta_0 - \beta_1 dist_j - \nu_{ijt}).$$

As we do not observe the actual value of the fixed export cost shock ν_{ijt} nor the actual value of the unobservable component of export revenues e_{ijt}, we compute instead the mean potential ex post profits of a firm, taking the mean over the distribution of ν_{ijt} and e_{ijt}. Specifically, the probability that a firm i decides to export to j at t, integrating over the distribution of ν_{ijt}, is:

$$\int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}_{ijt}^x, \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt} =$$
$$\int_{\nu_{ijt}} \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}^o | \mathcal{J}_{ijt}^x] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\}(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}.$$

The mean realized ex post export sales, averaging over the distribution of ν_{ijt} and e_{ijt}, is:

$$r_{ijt}^o \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}_{ijt}^x, \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt},$$

The mean realized ex post profits, averaging over the distribution of ν_{ijt} and e_{ijt}, is:

$$\int_{\nu_{ijt}} \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}^o | \mathcal{J}_{ijt}^x] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\} \times (\eta^{-1} r_{ijt}^o - \beta_0 - \beta_1 dist_j - \nu_{ijt}))(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}.$$

Once we have computed these expressions for every firm i, period t, and destination market j included in the set of observations of interest $\mathcal{S}$, we can aggregate these expressions to obtain the predicted total number of firm-destination-year triplets with positive exports,

$$N^{ex}(\theta, \mathcal{J}_{ijt}^x, \mathcal{S}) \equiv \sum_{(i,j,t)\in\mathcal{S}} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}_{ijt}^x, \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt},$$

the predicted average export profits,

$$\bar{\pi}(\theta, \mathcal{J}_{ijt}^x, \mathcal{S}) \equiv \frac{\sum_{(i,j,t)\in\mathcal{S}} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}_{ijt}^x, \nu_{ijt})(\eta^{-1} r_{ijt}^o - \beta_0 - \beta_1 dist_j - \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}}{\sum_{(i,j,t)\in\mathcal{S}} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}_{ijt}^x, \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}},$$

and the predicted aggregate exports,

$$R(\theta, \mathcal{J}_{ijt}^x, \mathcal{S}) \equiv \sum_{(i,j,t)\in\mathcal{S}} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}_{ijt}^x, \nu_{ijt}) r_{ijt}^o (1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}.$$

The column under the heading "Number of Exporters" in panel A of Table 6 reports bounds on the relative change in $N^{ex}(\theta, \mathcal{J}_{ijt}^x, \mathcal{S})$ as we change information sets from $\mathcal{J}_{ijt}^s$ to $\mathcal{J}_{ijt}^m$:

$$\left[\min_{\theta\in\hat{\Theta}^{95\%}} \left\{\frac{N^{ex}(\theta, \mathcal{J}_{ijt}^m, \mathcal{S})}{N^{ex}(\theta, \mathcal{J}_{ijt}^s, \mathcal{S})}\right\}, \max_{\theta\in\hat{\Theta}^{95\%}} \left\{\frac{N^{ex}(\theta, \mathcal{J}_{ijt}^m, \mathcal{S})}{N^{ex}(\theta, \mathcal{J}_{ijt}^s, \mathcal{S})}\right\}\right]. \tag{A.35}$$

The column under the heading "Mean Export Profits" in panel A of Table 6 reports bounds on the relative change in $\bar{\pi}(\theta, \mathcal{J}_{ijt}^x, \mathcal{S})$ as we change information sets from $\mathcal{J}_{ijt}^s$ to $\mathcal{J}_{ijt}^m$:

$$\left[\min_{\theta\in\hat{\Theta}^{95\%}} \left\{\frac{\bar{\pi}(\theta, \mathcal{J}_{ijt}^m, \mathcal{S})}{\bar{\pi}(\theta, \mathcal{J}_{ijt}^s, \mathcal{S})}\right\}, \max_{\theta\in\hat{\Theta}^{95\%}} \left\{\frac{\bar{\pi}(\theta, \mathcal{J}_{ijt}^m, \mathcal{S})}{\bar{\pi}(\theta, \mathcal{J}_{ijt}^s, \mathcal{S})}\right\}\right]. \tag{A.36}$$

The column under the heading "Aggregate Exports" in panel A of Table 6 reports bounds on the relative change in $R(\theta, \mathcal{J}^x_{ijt}, \mathcal{S})$ as we change information sets from $\mathcal{J}^s_{ijt}$ to $\mathcal{J}^m_{ijt}$:

$$\left[\min_{\theta \in \hat{\Theta}^{95\%}} \left\{ \frac{R(\theta, \mathcal{J}^m_{ijt}, \mathcal{S})}{R(\theta, \mathcal{J}^s_{ijt}, \mathcal{S})} \right\}, \max_{\theta \in \hat{\Theta}^{95\%}} \left\{ \frac{R(\theta, \mathcal{J}^m_{ijt}, \mathcal{S})}{R(\theta, \mathcal{J}^s_{ijt}, \mathcal{S})} \right\} \right]. \tag{A.37}$$

The five different rows in panel A of Table 6 differ by the definition of the set $\mathcal{S}$. The first row includes all firms, destinations and years in the sample. Rows two and three split all the firm-year pairs into two groups, including all destination markets in each of them. Rows four and five split all destination markets into two groups, including all firms and years in each of them. Panel B in Table 6 is analogous to Panel A, but uses $\mathcal{J}^l_{ijt}$ instead of $\mathcal{J}^m_{ijt}$ in equations (A.35) to (A.37).

The bounds reported in Table 7 are computed following a procedure analogous to that described above. As an example, we describe how to compute the "number of exporters", "mean export profits" and "aggregate exports" for those firms that decide to export under both information sets considered in panel A ($\mathcal{J}^s_{ijt}$ and $\mathcal{J}^m_{ijt}$). This subgroup of firms is defined in Table 7 as the firms that "always export". Given a value of the parameter vector θ, the number of "always exporters" is defined as

$$N_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt}) \equiv \sum_{(i,j,t)} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}^s_{ijt}, \nu_{ijt}) d_{ijt}(\theta, \mathcal{J}^m_{ijt}, \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}.$$

Their average export profits are

$$\bar{\pi}_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt}) \equiv \frac{\sum_{(i,j,t)} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}^s_{ijt}, \nu_{ijt}) d_{ijt}(\theta, \mathcal{J}^m_{ijt}, \nu_{ijt})(\eta^{-1} r^o_{ijt} - \beta_0 - \beta_1 dist_j - \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}}{\sum_{(i,j,t)} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}^s_{ijt}, \nu_{ijt}) d_{ijt}(\theta, \mathcal{J}^m_{ijt}, \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}}.$$

Their total aggregate exports are

$$R_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt}, \mathcal{S}) \equiv \sum_{(i,j,t) \in \mathcal{S}} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}^s_{ijt}, \nu_{ijt}) d_{ijt}(\theta, \mathcal{J}^m_{ijt}, \nu_{ijt}) r^o_{ijt}(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}.$$

Thus, for the "always exporters", the bounds on the "Number of Exporters" reported in panel A of Table 7 are computed as

$$\left[\min_{\theta \in \hat{\Theta}^{95\%}} \{N_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}, \max_{\theta \in \hat{\Theta}^{95\%}} \{N_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\} \right];$$

the bounds on the "Mean Export Profits" are computed as

$$\left[\min_{\theta \in \hat{\Theta}^{95\%}} \{\bar{\pi}_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}, \max_{\theta \in \hat{\Theta}^{95\%}} \{\bar{\pi}_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\} \right];$$

and the bounds on the "Aggregate Exports" are computed as

$$\left[\min_{\theta \in \hat{\Theta}^{95\%}} \{R_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}, \max_{\theta \in \hat{\Theta}^{95\%}} \{R_{ax}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\} \right].$$

Similarly, given a value of the parameter vector θ, we define the number of firms that "switch out" as

$$N_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt}) \equiv \sum_{(i,j,t)} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}^s_{ijt}, \nu_{ijt})(1 - d_{ijt}(\theta, \mathcal{J}^m_{ijt}, \nu_{ijt}))(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt},$$

their average export profits are

$$\bar{\pi}_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt}) \equiv \frac{\sum_{(i,j,t)} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}^s_{ijt}, \nu_{ijt})(1 - d_{ijt}(\theta, \mathcal{J}^m_{ijt}, \nu_{ijt}))(\eta^{-1} r^o_{ijt} - \beta_0 - \beta_1 dist_j - \nu_{ijt})(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}}{\sum_{(i,j,t)} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}^s_{ijt}, \nu_{ijt})(1 - d_{ijt}(\theta, \mathcal{J}^m_{ijt}, \nu_{ijt}))(1/\sigma)\phi(\nu_{ijt}/\sigma) d\nu_{ijt}},$$

and their total aggregate exports are

$$R_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt}, \mathcal{S}) \equiv \sum_{(i,j,t)\in\mathcal{S}} \int_{\nu_{ijt}} d_{ijt}(\theta, \mathcal{J}^s_{ijt}, \nu_{ijt})(1 - d_{ijt}(\theta, \mathcal{J}^m_{ijt}, \nu_{ijt}))r^o_{ijt}(1/\sigma)\phi(\nu_{ijt}/\sigma)d\nu_{ijt}.$$

Thus, for the firms that "switch out", the bounds on the "Number of Exporters" reported in Panel A of Table 7 are computed as

$$\Big[\min_{\theta\in\hat{\Theta}^{95\%}}\{N_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}, \max_{\theta\in\hat{\Theta}^{95\%}}\{N_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}\Big];$$

the bounds on the "Mean Export Profits" are computed as

$$\Big[\min_{\theta\in\hat{\Theta}^{95\%}}\{\bar{\pi}_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}, \max_{\theta\in\hat{\Theta}^{95\%}}\{\bar{\pi}_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}\Big];$$

and the bounds on the "Aggregate Exports" are computed as

$$\Big[\min_{\theta\in\hat{\Theta}^{95\%}}\{R_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}, \max_{\theta\in\hat{\Theta}^{95\%}}\{R_{so}(\theta, \mathcal{J}^s_{ijt}, \mathcal{J}^m_{ijt})\}\Big].$$

In Table 7, we compute the bounds for the firms that "switch in" and for the firms that "never export" following an analogous procedure. Similarly, computing the bounds reported in panel B of Table 7 requires following a procedure analogous to that followed to compute panel A, except we use $\mathcal{J}^l_{ijt}$ instead of $\mathcal{J}^m_{ijt}$.

A.10 Counterfactual Changes in Export Costs: Details

A.10.1 Proof of Theorem 3

Lemma 1 *Given the definition of $\mathcal{P}_{ijt}(\theta)$ in equation (19), for any $\theta \in \Theta$,*

$$\mathbb{E}\left[\frac{1-\Phi(\theta_2^{-1}(\eta^{-1}r^o_{ijt}-\theta_0-\theta_1 dist_j))}{\Phi(\theta_2^{-1}(\eta^{-1}r^o_{ijt}-\theta_0-\theta_1 dist_j))}\middle|\mathcal{J}_{ijt}, dist_j\right] \geq \mathbb{E}\left[\frac{1-\mathcal{P}_{ijt}(\theta)}{\mathcal{P}_{ijt}(\theta)}\middle|\mathcal{J}_{ijt}, dist_j\right]. \tag{A.38}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}] = 0$. The mean independence between ε_{ijt} and $dist_j$ is implied by the model described in Section 2. By definition, the set $\mathcal{J}_{ijt}$ includes all variables the firm uses to forecast its revenue when it decides whether to export to a particular destination. Our model additionally imposes that the firm knows $dist_j$ when determining its export decision. Thus, $dist_j$ is either independent of r^o_{ijt} and, consequently, of ε_{ijt}, or belongs to $\mathcal{J}_{ijt}$. In either scenario, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j] = 0$. Since

$$\frac{1-\Phi(y)}{\Phi(y)}$$

is convex for any value of y and $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j] = 0$, by Jensen's Inequality

$$\mathbb{E}\left[\frac{1-\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\theta_0-\theta_1 dist_j+\eta^{-1}\varepsilon_{ijt}))}{\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\theta_0-\theta_1 dist_j+\eta^{-1}\varepsilon_{ijt}))}\middle|\mathcal{J}_{ijt}, dist_j\right] \geq$$
$$\mathbb{E}\left[\frac{1-\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\theta_0-\theta_1 dist_j))}{\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\theta_0-\theta_1 dist_j))}\middle|\mathcal{J}_{ijt}, dist_j\right].$$

Equation (A.38) follows from the equality $\eta^{-1}r^o_{ijt} = \eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$ and the definition of $\mathcal{P}_{ijt}(\theta)$ in equation (19). ■

Lemma 2 *Given the definition of $\mathcal{P}_{ijt}(\theta)$ in equation (19), for any $\theta \in \Theta$,*

$$\mathbb{E}\left[\frac{\Phi(\theta_2^{-1}(\eta^{-1}r^o_{ijt}-\theta_0-\theta_1 dist_j))}{1-\Phi(\theta_2^{-1}(\eta^{-1}r^o_{ijt}-\theta_0-\theta_1 dist_j))}\middle|\mathcal{J}_{ijt}, dist_j\right] \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}(\theta)}{1-\mathcal{P}_{ijt}(\theta)}\middle|\mathcal{J}_{ijt}, dist_j\right]. \tag{A.39}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}] = 0$. The mean independence between ε_{ijt} and $dist_j$ is implied by the model described in Section 2. By definition, the set $\mathcal{J}_{ijt}$ includes all variables the firm uses to forecast its revenue when deciding whether to export to a particular destination. Our model additionally imposes that the firm knows $dist_j$ when determining its export decision. Thus, $dist_j$ is either independent of r^o_{ijt} and, consequently, of ε_{ijt}, or belongs to $\mathcal{J}_{ijt}$. In either scenario, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j] = 0$. Since

$$\frac{\Phi(y)}{1-\Phi(y)}$$

is convex for any value of y and $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j] = 0$, by Jensen's Inequality

$$\mathbb{E}\Big[\frac{\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}{1-\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}\Big|\mathcal{J}_{ijt}, dist_j\Big] \geq$$
$$\mathbb{E}\Big[\frac{\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j))}{1-\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j))}\Big|\mathcal{J}_{ijt}, dist_j\Big].$$

Equation (A.38) follows from the equality $\eta^{-1}r^o_{ijt} = \eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$ and the definition of $\mathcal{P}_{ijt}(\theta)$ in equation (19). ■

Lemma 3 *If the distribution of Z_{ijt} conditional on $(\mathcal{J}_{ijt}, dist_j)$ is degenerate, then, for any $\theta \in \Theta$,*

$$\mathbb{E}\Big[\frac{1-\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}{\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}\Big|Z_{ijt}\Big] \geq \mathbb{E}\Big[\frac{1-\mathcal{P}_{ijt}(\theta)}{\mathcal{P}_{ijt}(\theta)}\Big|Z_{ijt}\Big], \tag{A.40}$$

and

$$\mathbb{E}\Big[\frac{\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}{1-\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}\Big|Z_{ijt}\Big] \geq \mathbb{E}\Big[\frac{\mathcal{P}_{ijt}(\theta)}{1-\mathcal{P}_{ijt}(\theta)}\Big|Z_{ijt}\Big]. \tag{A.41}$$

Proof: It follows from lemmas 1 and 2 and the Law of Iterated Expectations. ■

Lemma 4 *Given the definition of $\mathcal{P}_{ijt}(\theta)$ in equation (19), for any $\theta \in \Theta$,*

$$\mathbb{E}\Big[\frac{1-\mathcal{P}_{ijt}(\theta)}{\mathcal{P}_{ijt}(\theta)}\Big|Z_{ijt}\Big] \geq \frac{1-\mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]}{\mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]}, \tag{A.42}$$

and

$$\mathbb{E}\Big[\frac{\mathcal{P}_{ijt}(\theta)}{1-\mathcal{P}_{ijt}(\theta)}\Big|Z_{ijt}\Big] \geq \frac{\mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]}{1-\mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]}. \tag{A.43}$$

Proof: Equation (19) implies that the support of $\mathcal{P}_{ijt}(\theta)$ is the interval $(0,1)$. Therefore, equations (A.42) and (A.43) are implied by Jensen's inequality. ■

Lemma 5 *If the distribution of Z_{ijt} conditional on $(\mathcal{J}_{ijt}, dist_j)$ is degenerate, then, for any $\theta \in \Theta$,*

$$\frac{1}{1+B^l(Z_{ijt};\theta)} \leq \mathcal{P}(Z_{ijt};\theta) \leq \frac{B^u(Z_{ijt};\theta)}{1+B^u(Z_{ijt};\theta)}, \tag{A.44}$$

where

$$B^l(Z_{ijt};\theta) = \mathbb{E}\Big[\frac{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}{\Phi\big(\sigma^{-1}\big(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}\Big|Z_{ijt}\Big]. \tag{A.45}$$

$$B^u(Z_{ijt};\theta) = \mathbb{E}\Big[\frac{\Phi\big(\sigma^{-1}\big(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}\Big|Z_{ijt}\Big], \tag{A.46}$$

Proof: Combining equations (A.40) and (A.42),

$$B^l(Z_{ijt};\theta) \geq \mathbb{E}\left[\frac{1-\mathcal{P}_{ijt}(\theta)}{\mathcal{P}_{ijt}(\theta)}\middle|Z_{ijt}\right] \geq \frac{1-\mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]}{\mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]},$$

and, reordering terms, we obtain the inequality

$$\frac{1}{1+B^l(Z_{ijt};\theta)} \leq \mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]. \tag{A.47}$$

Combining equations (A.41) and (A.43),

$$B^u(Z_{ijt};\theta) \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}(\theta)}{1-\mathcal{P}_{ijt}(\theta)}\middle|Z_{ijt}\right] \geq \frac{\mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]}{1-\mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]}$$

and, reordering terms, we obtain the inequality

$$\frac{B^u(Z_{ijt};\theta)}{1+B^u(Z_{ijt};\theta)} \geq \mathbb{E}[\mathcal{P}_{ijt}(\theta)|Z_{ijt}]. \tag{A.48}$$

Combining the inequalities in equations (A.47) and (A.48) we obtain equation (A.44). ■

A.11 Unexpected Shocks to Fixed Costs

There are two ways to model firms that are uncertain about the value of the fixed costs f_{ijt} when they decide whether to enter market j at period t. First, we could extend the specification of the fixed export cost f_{ijt} in equation (5) to incorporate a firm-country-year specific shock u_{ijt} that is unknown to firm i when it decides whether to export to destination j in year t. In this case,

$$f_{ijt} = \beta_0 + \beta_1 dist_j + \nu_{ijt} + u_{ijt},$$

where the unobserved (to the researcher) shocks ν_{ijt} and u_{ijt} differ in that firm i knows the former but not the latter when deciding whether to export to country j at t. Including the unexpected component u_{ijt} does not affect the expected export profits in equation (7) and, consequently, does not change the decision problem that we describe in equation (9). Therefore, the only effect of including u_{ijt} in the definition of the fixed export costs is the interpretation of the expression $\beta_0 + \beta_1 dist_j + \nu_{ijt}$. Rather than reflecting the total fixed cost that firm i has to pay in country j if it were to export to it at period t, that expression would reflect only the component of this fixed export cost that firm i expects to pay when deciding whether to export.

Second, we also could have allowed for uncertainty in the fixed export cost f_{ijt} by assuming that firms do not know the exact value of $dist_j$ when deciding whether to export to destination j. In this alternative, firms would make their export participation decision based on an unobserved (to the researcher) expectation of $dist_j$. However, firms have easy access to data on distances between countries. Thus, we assume in the model in Section 2 that $dist_j$ belongs to the information set of every potential exporter. If our specification of fixed costs contained observed (to the researcher) covariates that we did not believe firms knew when deciding which foreign markets to enter, we could account in our model for firms' unobserved expectations of these covariates. In this case, for each of the observed covariates entering the expression for fixed export costs and over which firms need to form an expectation, our estimation approach would require that we observe an instrument that belongs to firms' information sets.

B Additional Results

B.1 Estimates of Export Revenue Shifters

Figures B.1 and B.2 summarize the estimates of α_{jt}, for every country j and year t in our sample. We describe the estimation procedure in Appendix A.3. As described in Appendix A.1, according to the model introduced in Section 2, the parameter α_{jt} is a function of variable trade costs, a price index, and the market size in destination market j and year t relative to the same variables in the home market in the same period:

$$\alpha_{jt} \equiv \mathbb{E}_{jt}\left[\left(\frac{\zeta_{iht}\tau_{ijt}}{\zeta_{ijt}\tau_{iht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}\right].$$

Therefore, *ceteris paribus*, our model predicts α_{jt} to be larger in larger countries (as they are more likely to have a large value of Y_{jt}), in countries geographically close to Chile, and in Spanish-speaking countries (as they are more likely to have small values of τ_{jt}). In figures B.1 and B.2, we order countries from left to right according to distance to Chile and, for each country, we plot the distribution of α_{jt} across the 10 years of our sample. A few features of the distribution of the estimates $\{\hat{\alpha}_{jt}, \forall j, t\}$ stand out.

First, the estimates $\hat{\alpha}_{jt}$ are less than one in every country j and year t. This is consistent with τ_{jt} being significantly larger than τ_{ht}. Furthermore, given that τ may capture both supply-side and demand-side factors—i.e. both variable trade costs as well as demand shifters affecting all firms located in Chile—the estimates of $\hat{\alpha}_{jt}$ are consistent with consumers showing home bias in preferences.[47]

Second, the estimates $\hat{\alpha}_{jt}$ do not vary much with distance. This is true for both figures B.1 and B.2, where the distributions of $\hat{\alpha}_{jt}$ do not seem to vary systematically as we move along the horizontal axis from closer countries (on the left) to far away countries (on the right). Conversely, in both figures, the estimates for Spain (ESP) are larger than those of other European countries larger in size than Spain (e.g. Great Britain (GBR), France (FRA), and Italy (ITA)), suggesting that linguistic differences between Chile and destination markets are a significant determinant of the variable trade costs τ_{jt} and, consequently, of the parameters α_{jt}.

Third, the estimates $\hat{\alpha}_{jt}$ are significantly larger for those countries that are larger in size. Specifically, countries with larger GDP have larger values of $\hat{\alpha}_{jt}$. For example, Brazil (BRA), the United States (USA) and Japan (JAP) have estimates of α_{jt} that are, on average, significantly larger than those of their smaller neighboring countries.

In addition to the information contained in figures B.1 and B.2, Table B.1 contains moments of the distribution of $\hat{\alpha}_{jt}$ for Argentina, Japan and the United States, the three countries we use in the main text to illustrate our results. The larger size of the whiskers for the case of the United States in figures B.1 and B.2 is reflected in Table B.1 in a relatively large standard deviation of $\hat{\alpha}_{jt}$ for this country. Similarly, the large mean for both the United States and Japan relative to that of Argentina is consistent with the box plots for the former two countries appearing higher up in figures B.1 and B.2.

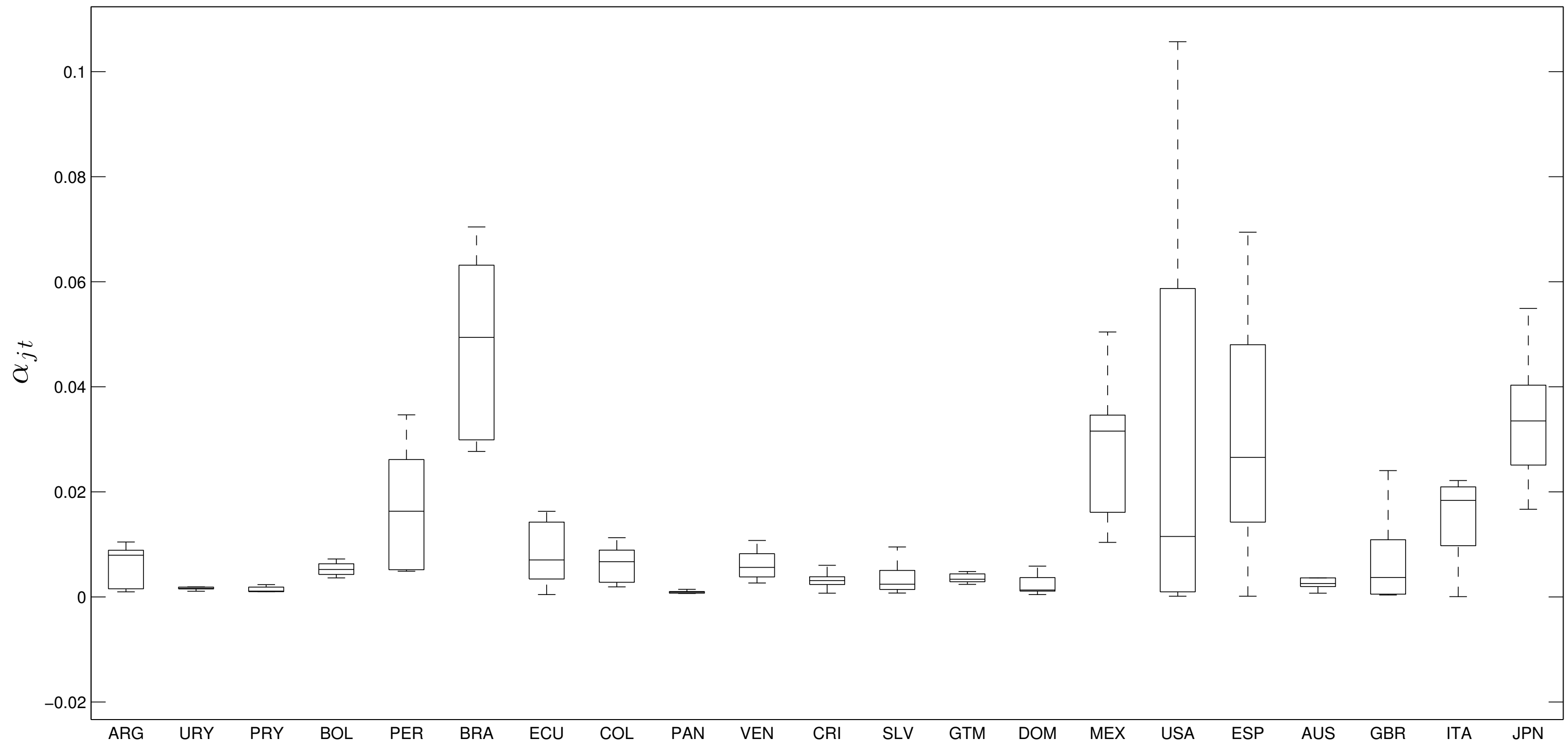
Table B.1: Moments of the distribution of α_{jt}

	Chemicals			Food		
	Argentina	Japan	United States	Argentina	Japan	United States
Mean	0.59%	3.27%	3.37%	1.22%	14.39%	19.45%
Standard Deviation	0.38%	1.16%	4.28%	0.84%	4.18%	14.35%

Notes: For the country-sector combination indicated by the first two rows, this table reports the mean and the standard deviation of the estimates of $\{\hat{\alpha}_{jt}\}_{t=1995}^{t=2005}$.

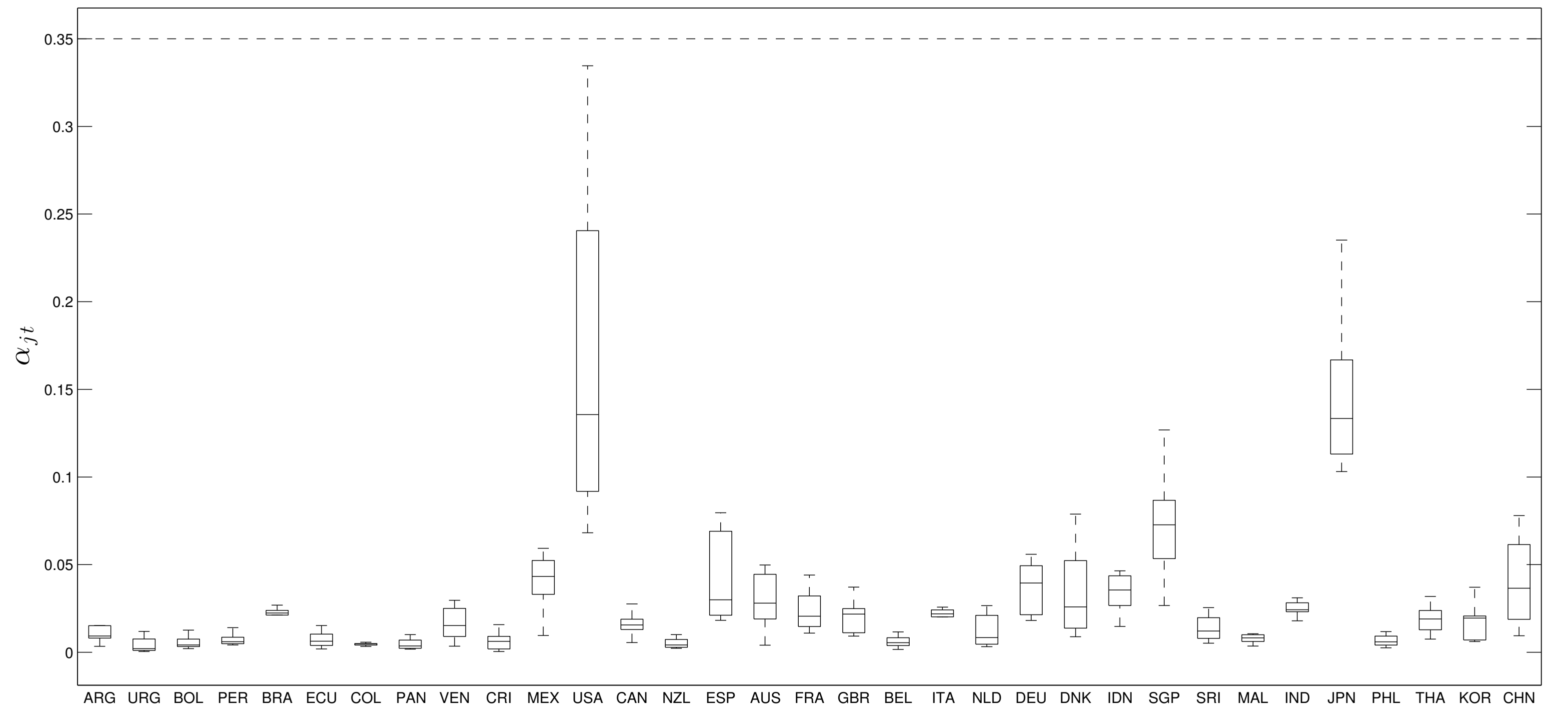
[47] Coşar et al. (2016) provide empirical evidence on the importance of home bias in consumption as a determinant of a firm's home market advantage.

Figure B.1: Distribution of α_{jt}: Chemicals



For each of the countries indicated in the horizontal axis, this figure represents the box plot of the corresponding vector of estimates of the parameter α_{jt}. For each country j and year t, the parameter α_{jt} captures the expected potential export revenue a firm might obtain if it exports to j in t relative to the potential revenue that same firm may obtain in the home market. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points the algorithm has not excluded as outliers. Points are considered outliers if they are larger than Q3+1.5×(Q3-Q1) or smaller than Q1−1.5×(Q3-Q1), where Q1 and Q3 are the 25th and 75th percentiles, respectively. If the data were normally distributed, the limits of the whiskers would contain 99.3% of the observations.

Figure B.2: Distribution of α_{jt}: Food



For each of the countries indicated in the horizontal axis, this figure represents the box plot of the corresponding vector of estimates of the parameter α_{jt}. For each country j and year t, the parameter α_{jt} captures the expected potential export revenue a firm might obtain if it exports to j in t relative to the potential revenue that same firm may obtain in the home market. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points the algorithm has not excluded as outliers. Points are considered outliers if they are larger than Q3+1.5×(Q3-Q1) or smaller than Q1−1.5×(Q3-Q1), where Q1 and Q3 are the 25th and 75th percentiles, respectively. If the data were normally distributed, the limits of the whiskers would contain 99.3% of the observations.

B.2 Quantiles of Distribution of Fixed Export Costs across Firms

We report quantiles of the distribution of fixed export costs. Given equations (5) and (6),

$$D_q(f_{ijt};(\beta_0,\beta_1,\sigma)) \equiv \beta_0 + \beta_1 dist_j + D_q(\nu_{ijt};\sigma),$$

where $D_q(\cdot)$ denotes the decile q function of the corresponding distribution for a given country j; e.g. $D_1(f_{ijt};\cdot)$ denotes the first decile of the distribution of f_{ijt} across firms and time periods for a given country j. Given equation (6) and a value of σ, we simulate $D_q(\nu_{ijt};\sigma)$ for every q using 10,000 draws from a normal distribution with mean zero and standard deviation σ. Specifically, given maximum likelihood estimates $(\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma})$ of (β_0,β_1,σ), we compute the maximum likelihood estimates of each decile as $D_q(f_{ijt};(\hat{\beta}_0,\hat{\beta}_1,\hat{\sigma})) \equiv \hat{\beta}_0+\hat{\beta}_1 dist_j + D_q(\nu_{ijt};\hat{\sigma})$. Given our moment inequality confidence set $\hat{\Theta}^{95\%}$ for θ, we compute the confidence interval for each decile as

$$[\min_{\theta\in\hat{\Theta}^{95\%}} \theta_0+\theta_1 dist_j + D_q(\nu_{ijt};\theta_2), \max_{\theta\in\hat{\Theta}^{95\%}} \theta_0+\theta_1 dist_j + D_q(\nu_{ijt};\theta_2)].$$

Table B.2 reports these deciles of the distribution of fixed costs in Argentina, Japan and the United States for both the chemical and food sectors.

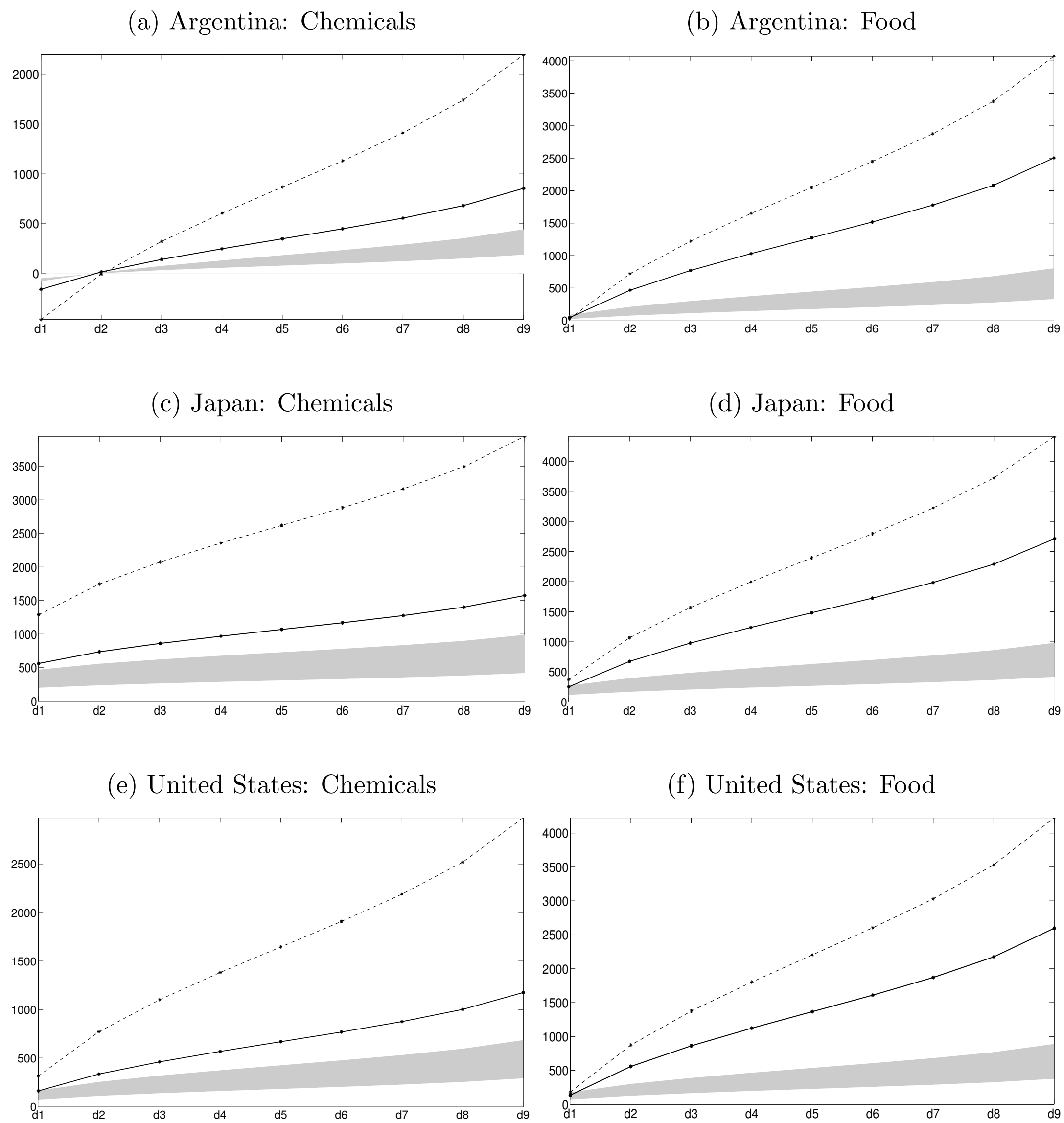
Table B.2: Fixed Export Costs: Deciles

			Chemicals			Food	
Decile	Estimator	Argentina	Japan	USA	Argentina	Japan	USA
	Perfect Fore.	-463.1	1,290.3	313.9	26.9	372.7	180.1
1	Minimal Info.	-158.2	562.4	161.1	43.6	252.2	136.1
	Mom. Ineq.	[-47.9, -29.6]	[200.2, 269.7]	[71.7, 92.8]	[22.1, 65.1]	[117.0, 156.0]	[73.9, 103.9]
	Perfect Fore.	-6.2	1,747.3	770.8	721.1	1,066.9	874.4
2	Minimal Info.	15.8	736.5	335.2	466.0	674.5	558.4
	Mom. Ineq.	[2.4, 9.2]	[237.6, 321.4]	[109.6, 144.6]	[74.8, 135.5]	[169.7, 226.4]	[126.6, 174.3]
	Perfect Fore.	323.3	2,076.7	1,100.3	1,221.7	1,567.5	1,375.0
3	Minimal Info.	141.3	862.0	460.6	770.5	979.0	862.9
	Mom. Ineq.	[33.7, 42.4]	[264.6, 358.8]	[136.7, 181.9]	[112.8, 186.2]	[164.6, 277.1]	[164.6, 225.1]
	Perfect Fore.	604.8	2,358.3	1,381.8	1,649.5	1,995.3	1,802.7
4	Minimal Info.	248.5	969.2	567.9	1,030.7	1,239.2	1,123.1
	Mom. Ineq.	[57.3, 74.3]	[287.7, 390.7]	[159.7, 213.8]	[145.2, 229.6]	[197.0, 320.5]	[197.0, 268.5]
	Perfect Fore.	868.0	2621.4	1645.0	2,049.3	2,395.1	2,202.5
5	Minimal Info.	348.7	1069.4	668.1	1,273.9	1,482.4	1,366.3
	Mom. Ineq.	[79.1, 104.1]	[309.2, 420.5]	[181.3, 243.6]	[175.6, 270.1]	[269.1, 361.0]	[227.3, 308.9]
	Perfect Fore.	1,131.1	2,884.5	1,908.1	2,449.1	2,794.8	2,602.3
6	Minimal Info.	449.0	1,169.6	768.3	1,517.1	1,725.6	1,609.5
	Mom. Ineq.	[100.8, 133.9]	[330.8, 450.3]	[202.9, 273.4]	[205.9, 310.7]	[298.3, 401.6]	[257.7, 349.5]
	Perfect Fore.	1,412.6	3,166.1	2,189.6	2,876.8	3,222.6	3,030.0
7	Minimal Info.	556.2	1,276.8	875.5	1,777.3	1,985.8	1,869.7
	Mom. Ineq.	[124.1, 165.8]	[353.9, 482.2]	[225.9, 305.3]	[238.4, 354.0]	[329.4, 444.9]	[289.1, 392.9]
	Perfect Fore.	1,742.1	3,495.6	2,519.1	3,377.4	3,723.2	3,530.6
8	Minimal Info.	681.7	1,402.3	1,001.0	2,081.8	2,290.3	2,174.2
	Mom. Ineq.	[151.1, 203.1]	[380.9, 519.5]	[252.9, 342.6]	[276.4, 404.8]	[365.9, 495.7]	[325.6, 443.7]
	Perfect Fore.	2,199.1	3,952.5	2,976.0	4,071.6	4,417.4	4,224.8
9	Minimal Info.	855.7	1,576.4	1,175.0	2,504.1	2,712.7	2,596.5
	Mom. Ineq.	[188.6, 254.9]	[418.3, 571.2]	[290.4, 394.3]	[329.1, 475.2]	[416.4, 566.1]	[376.2, 514.0]

Notes: All variables are reported in thousands of year 2000 USD. For the two ML estimators, standard errors are reported in parentheses. For the moment inequality estimates, extreme points of the confidence set are reported in square brackets. All the information in this table is reflected in Figure B.3.

In Figure B.3, we illustrate the estimates reported in Table B.2. This figure illustrates that both the distance between the two maximum likelihood estimates and the distance between these two point estimates and the moment inequality confidence set monotonically increases as we move towards higher quantiles of the distribution of fixed costs. This monotonicity reflects the relative estimates of σ for the three approaches, reported in Table 2.

Figure B.3: Distribution of Fixed Export Costs



In all figures, the vertical axis indicates fixed export costs in thousands of year 2000 USD and the horizontal axis indicates the deciles of the distribution. The shaded area corresponds to the confidence interval predicted by our moment inequality estimator. The continuous black line corresponds to the minimal information ML point estimates. The dotted black line corresponds to the perfect foresight ML point estimates. The underlying estimates reflected in these plots appear in Table B.2.

B.3 Alternative Specification of Fixed Export Costs

In this section, we report both maximum likelihood and moment inequality estimates of fixed export costs that now vary freely across countries. Specifically, we generalize the model described in Section 2 in two dimensions: (a) we substitute the specification of fixed export costs in equation (5) by the alternative specification $f_{ijt} = \beta_j + \nu_{ijt}$; and (b) we allow the dispersion in fixed export costs to be country-specific; i.e. σ_j may be different from $\sigma_{j'}$ for distinct countries j and j'. To estimate the parameter vector (β_j, σ_j) for each country j, we divide the samples of firms in the chemical and in the food sector into country-specific subsamples, and use each of them separately to compute the maximum likelihood and moment inequality estimates of β_j and σ_j. For each of these subsamples, we compute confidence sets using odds-based and revealed-preference moment inequalities analogous to those described in Appendix A.5, but here we use instrument functions $g_a(Z_{ijt})$ that depend on firms' lagged domestic sales, r_{iht-1} and on lagged aggregate exports to each destination market j, R_{jt-1}.

Table B.3 and Figure B.4 contain the estimates for the chemicals sector. Of the 21 countries considered, 18 of them have a non-empty 95% confidence set and 19 of them have larger maximum likelihood estimates under perfect foresight than under the minimal information assumptions. Of the 18 countries with non-empty confidence sets, both the perfect foresight and the minimal information maximum likelihood estimates are larger than the upper bound of the moment inequality confidence set in 16 of them. Of the three countries with empty 95% confidence intervals, only two have p-values below 1% the null that the identified set is empty.

Table B.4 and Figure B.5 contain the estimates for the food sector. Of the 33 countries considered, 27 of them have a non-empty 95% confidence set and 32 of them have larger maximum likelihood estimates under perfect foresight than under the minimal information assumptions. Of the 27 countries with non-empty confidence sets, all of them have perfect foresight maximum likelihood estimates larger than the upper bound of the moment inequality confidence set, and the same is true for 24 of the minimal information maximum likelihood estimates. Of the six countries with empty 95% confidence intervals, only three have p-values below 1% the null that the identified set is empty.

Table B.3: Alternative Specification of Fixed Export Costs: Chemicals

	Num. Export Obs.	Perfect Foresight	Minimal Information	Moment Inequality
Argentina	511	79.4	53.5	[12.5, 46.4]
		(7.7)	(3.4)	(0.102)
Australia	86	400.9	207.7	[26.5, 153.2]
		(98.9)	(30.8)	(0.146)
Bolivia	617	46.3	38.8	[26.4, 47.6]
		(3.7)	(2.6)	(0.054)
Brazil	269	1,863.8	1,387.0	[314.5, 817.5]
		(267.3)	(160.8)	(0.081)
Colombia	274	88.9	83.5	[42.4, 49.2]
		(6.3)	(5.4)	(0.012)
Costa Rica	163	194.0	142.9	[34.5, 67.9]
		(35.6)	(19.8)	(0.056)
Dominican Republic	103	236.8	123.1	[20.2, 49.7]
		(50.1)	(16.5)	(0.042)
Ecuador	393	186.1	121.9	[-, -]
		(22.3)	(10.3)	(0.003)
El Salvador	83	252.9	174.9	[31.0, 132.7]
		(42.9)	(22.2)	(0.184)
Great Britain	66	3,245.5	1,045.4	[-, -]
		(2,915.7)	(330.3)	(0.001)
Guatemala	126	160.0	135.1	[26.5, 86.5]
		(21.3)	(15.9)	(0.104)
Italy	58	1,866.0	1,779.7	[227.4, 918.4]
		(475.9)	(440.1)	(0.124)

Table B.3: Alternative Specification of Fixed Export Costs: Chemicals (cont.)

	Num. Export Obs.	Perfect Foresight	Minimal Information	Moment Inequality
Japan	59	2,861.6	2,534.6	[850.0, 4,002.8]
		(629.4)	(529.5)	(0.199)
Mexico	173	1,260.7	802.8	[200.5, 690.3]
		(170.1)	(81.3)	(0.067)
Panama	153	37.9	35.1	[13.2, 19.1]
		(4.3)	(3.8)	(0.030)
Peru	651	85.0	88.3	[45.8, 61.6]
		(5.2)	(5.0)	(0.023)
Paraguay	324	29.4	23.7	[10.1, 20.5]
		(2.9)	(2.0)	(0.092)
Spain	88	3,631.2	1,687.9	[215.6, 1,674.4]
		(730.3)	(203.3)	(0.120)
Uruguay	314	315.1	271.7	[106.4, 200.1]
		(29.9)	(22.6)	(0.043)
United States	201	6,123.9	4,562.4	[665.8, 1,154.3]
		(2,184.7)	(1,617.1)	(0.057)
Venezuela	215	149.2	178.0	[-, -]
		(19.8)	(19.5)	(0.001)

Notes: All variables are reported in thousands of year 2000 USD. For the two ML estimators, standard errors are reported in parentheses. For the moment inequality estimates, extreme points of the 95% confidence set are reported in square brackets and p-values are reported in parenthesis. The MI confidence sets are computed as in Andrews and Soares (2010), and the p-values are computed as in Bugni et al. (2015).

Table B.4: Alternative Specification of Fixed Export Costs: Food

	Num. Export Obs.	Perfect Foresight	Minimal Information	Moment Inequality
Argentina	363	322.0	211.6	[-, -]
		(42.4)	(19.5)	(0.001)
Australia	149	1,930.4	1,268.9	[220.0, 328.1]
		(52.4)	(30.8)	(0.020)
Belgium	123	480.6	227.7	[49.8, 228.9]
		(129.4)	(34.7)	(0.171)
Bolivia	149	322.3	187.9	[42.5, 105.3]
		(62.9)	(24.6)	(0.104)
Brazil	368	864.3	742.1	[243.0, 243.0]
		(194.2)	(147.7)	(0.010)
Canada	263	857.9	566.8	[-, -]
		(180.1)	(87.5)	(0.007)
China	265	991.3	614.6	[274.4, 532.6]
		(103.6)	(44.9)	(0.053)
Colombia	301	385.8	135.6	[32.7, 54.0]
		(82.1)	(12.6)	(0.066)
Costa Rica	105	555.9	254.4	[59.5, 156.3]
		(195.8)	(46.8)	(0.046)
Denmark	111	2,627.2	1,607.5	[276.6, 1,616.4]
		(102.8)	(367.5)	(0.098)

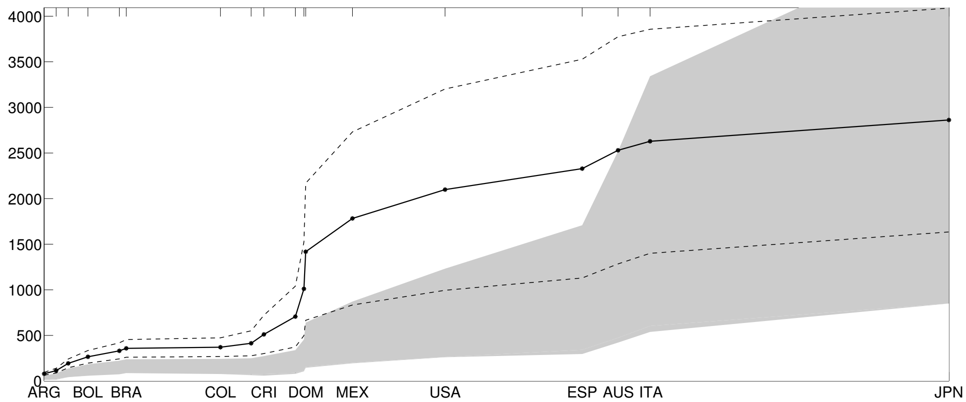
Table B.4: Alternative Specification of Fixed Export Costs: Food (cont.)

	Num. Export Obs.	Perfect Foresight	Minimal Information	Moment Inequality
Ecuador	185	210.3 (24.9)	130.8 (10.6)	[41.0, 83.8] (0.066)
France	257	1,153.7 (222.6)	838.7 (129.9)	[230.2, 569.4] (0.009)
Germany	319	886.6 (99.9)	721.6 (71.1)	[282.1, 620.4] (0.188)
Great Britain	214	598.0 (68.7)	497.9 (50.1)	[198.3, 222.3] (0.001)
Indonesia	122	1,289.2 (175.1)	1,106.7 (139.5)	[301.7, 476.7] (0.036)
India	72	1,355.6 (241.3)	838.3 (99.7)	[216.7, 494.1] (0.103)
Italy	167	1,195.6 (211.5)	684.0 (77.3)	[166.5, 420.2] (0.109)
Japan	636	2,205.1 (267.5)	1,885.6 (194.9)	[993.5, 1,339.0] (0.016)
Mexico	321	809.1 (80.7)	632.9 (52.9)	[2,39.1, 4,44.0] (0.072)
Malaysia	98	327.7 (49.6)	288.5 (41.0)	[83.9, 111.2] (0.017)
Netherlands	185	840.7 (180.2)	427.5 (59.5)	[77.7, 324.3] (0.174)
New Zealand	102	300.3 (58.3)	236.8 (39.6)	[61.2, 71.2] (0.011)
Panama	109	340.2 (89.6)	182.1 (30.5)	[35.8, 1,82.1] (0.138)
Peru	282	166.4 (17.3)	122.2 (99)	[38.4, 121.2] (0.191)
Philippines	116	355.4 (55.7)	225.1 (25.0)	[60.0, 144.2] (0.098)
Singapore	117	704.8 (130.9)	586.8 (100.3)	[121.8, 247.1] (0.016)
South Korea	207	648.6 (92.1)	521.2 (66.2)	[143.0, 353.8] (0.161)
Spain	258	3,645.6 (1,695.8)	2,256.1 (722.7)	[477.2, 677.0] (0.022)
Thailand	155	1,744.8 (358.7)	1,746.6 (390.9)	[286.6, 292.4] (0.013)
Uruguay	184	585.9 (282.6)	173.1 (29.3)	[-, -] (0.006)
United States	595	7,654.8 (2,143.6)	5,218.3 (1,121.1)	[-, -] (0.001)
Venezuela	231	849.5 (171.9)	682.1 (122.0)	[-, -] (0.007)

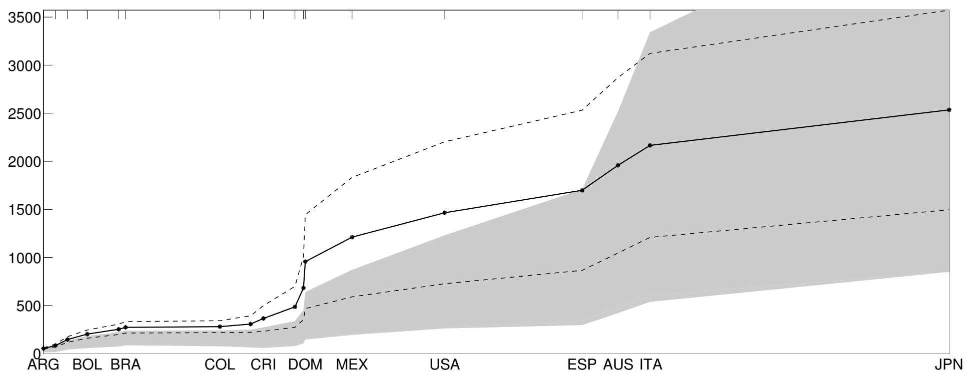
Notes: All variables are reported in thousands of year 2000 USD. For the two ML estimators, standard errors are reported in parentheses. For the moment inequality estimates, extreme points of the 95% confidence set are reported in square brackets and p-values are reported in parenthesis. The MI confidence sets are computed as in Andrews and Soares (2010), and the p-values are computed as in Bugni et al. (2015).

Figure B.4: Alternative Specification of Fixed Export Costs: Chemicals

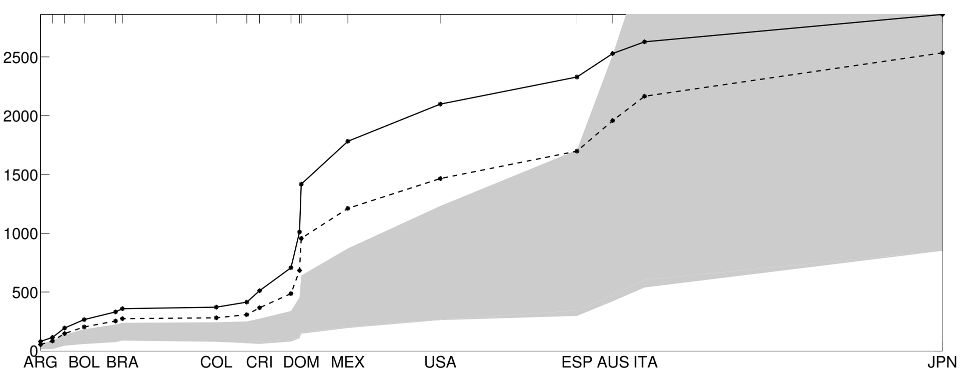
(a) MI & Perfect Foresight ML Confidence Sets



(b) MI & Minimal Information ML Confidence Sets



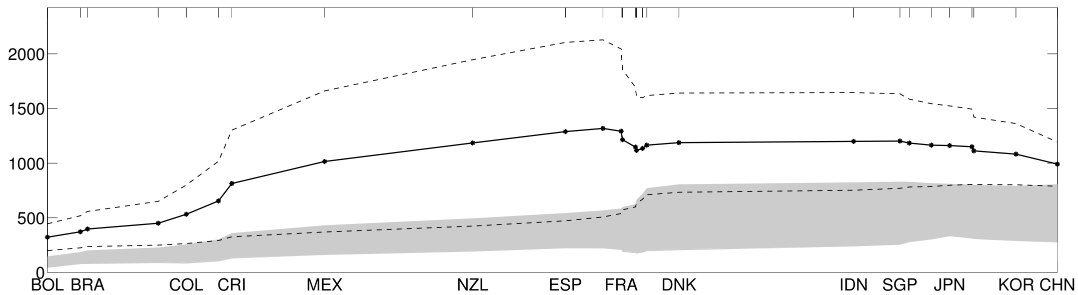
(c) MI Confidence Set & ML Point Estimates



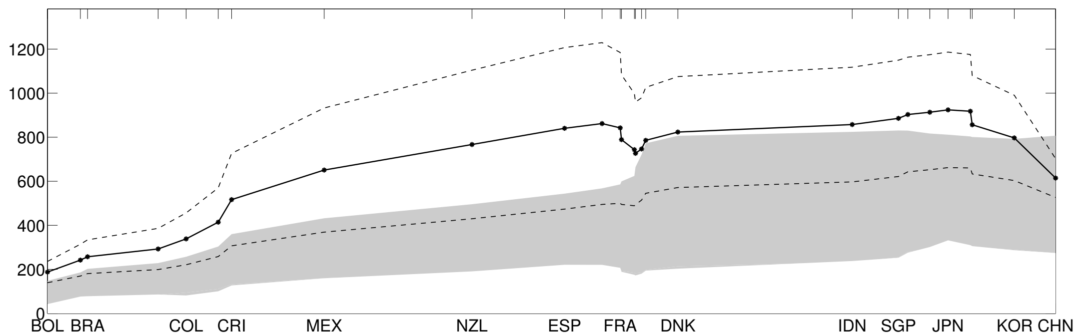
In all the three figures, the vertical axis indicates fixed export costs in thousands of year 2000 USD and the horizontal axis indicates different export destinations. The countries are placed along the horizontal axis according to their distance to Chile and we have limited the labeling to only a few countries for clarity. In all three figures, the light-grey shaded area denotes the 95% confidence interval generated by our moment inequalities. In panels (a) and (b), the continuous black line corresponds to the ML point estimates and the dotted black lines denotes the bounds of the corresponding 95% confidence interval. In panel (c), the continuous black line corresponds to the perfect foresight ML estimate and the dotted black line corresponds to the minimal information ML point estimate. We have smoothed for clarify the different lines reported in these figures.

Figure B.5: Alternative Specification of Fixed Export Costs: Food

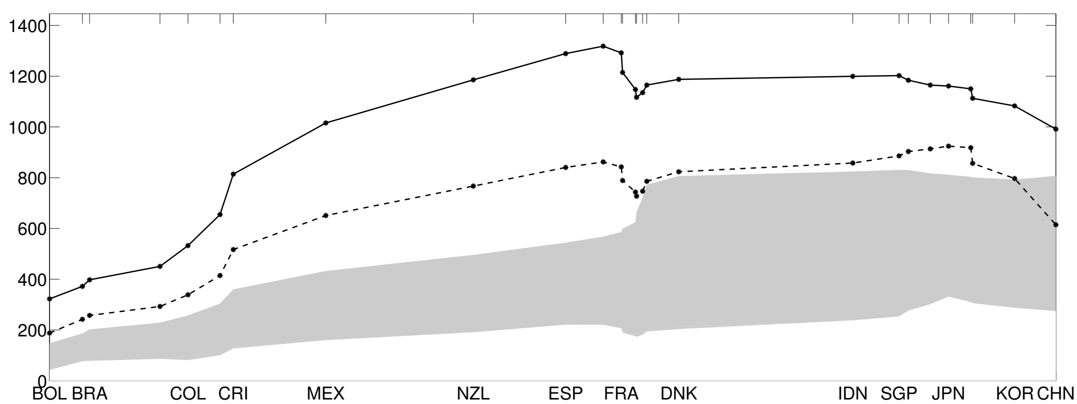
(a) MI & Perfect Foresight ML Confidence Sets



(b) MI & Minimal Information ML Confidence Sets



(c) MI Confidence Set & ML Point Estimates



In all the three figures, the vertical axis indicates fixed export costs in thousands of year 2000 USD and the horizontal axis indicates deciles different export destination countries. The countries are placed along the horizontal axis according to their distance to Chile and we have limited the labeling to only a few countries for clarity. In all three figures, the light-grey shaded area denotes the 95% confidence interval generated by our moment inequalities. In panels (a) and (b), the continuous black line corresponds to the ML point estimates and the dotted black lines denotes the bounds of the corresponding 95% confidence interval. In panel (c), the continuous black line corresponds to the perfect foresight ML estimate and the dotted black line corresponds to the minimal information ML point estimate. We have smoothed for clarify the different lines reported in these figures.

B.4 What Do Exporters Know? Additional Details

B.4.1 P-values for Test BP

Bugni et al. (2015) discuss alternative procedures to test the null hypothesis that the identified set defined by a finite set of moment inequalities is non-empty. Specifically, Bugni et al. (2015) introduce two novel specification tests, which they label test RS or re-sampling and test RC or re-cycling. As these authors show, both of these tests have better power properties than the BP or by-product test, studied previously in Romano and Shaikh (2008), Andrews and Guggenberger (2009), and Andrews and Soares (2010). For the different null hypothesis tested here, we report p-values for the RC test in Table 5 in Section 6 in the main text. We report here in Table B.5 the p-values for the BP test. The BP p-values are either identical or slightly above the RC p-values, consistent with the theoretical properties of these two tests discussed in Bugni et al. (2015). However, in those cases in which there are differences between both p-values, they are never large enough to modify our conclusions qualitatively.

Table B.5 also reports the family-adjusted p-values for the BP test. Given the individual BP p-values, we compute the adjusted ones following the procedure in Holm (1979), as described in Appendix A.8.2.

B.4.2 Instrument Relevance

In Section 6, we test whether a set of variables Z_{ijt} is contained in the firm's actual information set $\mathcal{J}_{ijt}$. In practice, our test asks whether, given a finite number of unconditional moment inequalities constructed using observed instruments Z_{ijt}, the corresponding identified set is non-empty; i.e. there exists a value of the parameter vector θ consistent with the corresponding set of moment inequalities.

If the model introduced in Section 2 is correct, the proofs of our odds-based and revealed-preference inequalities in Appendix C show that such moment inequalities must hold at the true value of the parameter vector, θ^*, if the distribution of the observed covariates Z_{ijt} is such that the true expectational error in revenue, $\varepsilon_{ijt} \equiv r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, satisfies $\mathbb{E}[\varepsilon_{ijt}|Z_{ijt}] = 0$. Put differently, if $\mathbb{E}[\varepsilon_{ijt}|Z_{ijt}] = 0$, then the set of parameter values consistent with our moment inequalities, conditioning on the vector Z_{ijt}, is necessarily non-empty, as it will always contain the true parameter value θ^*.

There are two sufficient conditions under which the mean independence condition $\mathbb{E}[\varepsilon_{ijt}|Z_{ijt}] = 0$ will hold. First, it will hold if the set of covariates Z_{ijt} is irrelevant to predict r^o_{ijt}. Second, the mean independence condition will hold if the set of covariates Z_{ijt} is relevant to predict r^o_{ijt} and the distribution of Z_{ijt} conditional on the information set $\mathcal{J}_{ijt}$ is degenerate. To rule out the possibility that we fail to reject a null hypothesis simply because Z_{ijt} is not relevant, we perform a pre-test on every vector Z_{ijt}. We test and show that the variables we include in Z_{ijt} have predictive power for the potential export revenues r_{ijt}. The results from these pre-tests appear in Tables B.6 and B.7. With relevancy confirmed, our moment inequality test will have a clear interpretation: for a set of relevant variables, we learn whether there's statistical evidence to reject the hypothesis that these variables are in the agent's information set.

The results show that, for all subsets of firms and countries considered in tables 5 and B.5, the covariates included in the vector Z_{ijt} are generally relevant as a predictor of the potential export revenues $r^o_{ijt} \equiv \hat{\alpha}_{jt} r_{iht}$. Given our choice of covariates, this is to be expected. First, lagged domestic sales, r_{iht-1}, are a good predictor of current domestic sales, r_{iht}. Second, for every period t, aggregate exports $R_{jt} \equiv \alpha_{jt} \sum_{i=1}^{N_t} r_{iht}$ depend on the value of the aggregate shifter α_{jt} in the same time period and, therefore, are a good proxy for it; as long as α_{jt} is serially correlated, then lagged aggregate exports R_{jt-1} will be a good predictor of it as well. Third, if the term α_{jt} is serially correlated, then lagged values of it α_{jt-1} will also be a good predictor of future values of it. Finally, if the supply or demand shocks captured in the term τ_{jt} are correlated with distance to Chile, then $dist_j$ will help predict the variation in r^o_{ijt} across destinations.

Our model in Section 2 does not impose assumptions on the functional form of the relationship between the observable component of revenue r^o_{ijt} and the set of covariates being tested, Z_{ijt}. Thus, to establish the relevance of instrument vector Z_{ijt} as a predictor of r^o_{ijt}, the researcher need only find at least one functional form that relates Z_{ijt} to r^o_{ijt}. In tables B.6 and B.7, we assume a linear relationship between r^o_{ijt} and each of the elements included in the vector Z_{ijt}, and found significant coefficients in this linear projection. This is enough to establish the relevance of the instruments included in Z_{ijt}. It does not, however, rule out that one could demonstrate a larger predictive capacity of the vector Z_{ijt} using a more flexible functional form.

Table B.5: Testing Content of Information Sets (BP Test)

Set of Firms	Set of Export Destinations	Chemicals Individual p-value	Chemicals Adjusted p-value	Chemicals Reject at 5%	Food Individual p-value	Food Adjusted p-value	Food Reject at 5%
		Panel A: Minimal Information					
All	All	0.111	0.111	No	0.980	0.980	No
		Panel B: Perfect Foresight					
All	All	0.009	0.009	Yes	< 0.001	< 0.001	Yes
		Panel C: Minimal Information & Country Shifter					
Large	Popular	0.144	0.418	No	0.974	1	No
Large	Unpopular	0.114	0.418	No	0.981	1	No
Small	Popular	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
Small	Unpopular	0.028	0.140	No	0.004	0.021	Yes
Large Exporter	All	0.104	0.418	No	0.990	1	No
Large Non-exporter	All	0.140	0.418	No	0.048	0.190	No
Small Exporter	All	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
Small Non-exporter	All	< 0.001	< 0.001	Yes	0.015	0.075	No
		Panel D: Minimal Information & Number of Exporters					
Large	Popular	0.108	0.324	No	0.978	1	No
Large	Unpopular	0.116	0.324	No	0.981	1	No
Small	Popular	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
Small	Unpopular	0.116	0.324	No	0.003	0.015	Yes
Large Exporter	All	0.019	0.080	No	0.988	1	No
Large Non-exporter	All	0.016	0.080	No	0.717	1	No
Small Exporter	All	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
Small Non-exporter	All	< 0.001	< 0.001	Yes	< 0.001	< 0.001	Yes
		Panel E: Minimal Information & Country Group Avg. Shifter					
		(a) Continent Avg. Shifter					
Large	All	0.109	0.860	No	0.986	1	No
Small	All	0.112	0.860	No	0.470	1	No
		(b) Language Group Avg. Shifter					
Large	All	0.116	0.860	No	0.980	1	No
Small	All	0.115	0.860	No	0.003	0.0180	No
		(c) Income p.c. Group Avg. Shifter					
Large	All	0.108	0.860	No	0.980	1	No
Small	All	0.152	0.860	No	0.991	1	No
		(d) Border Group Avg. Shifter					
Large	All	0.119	0.860	No	0.981	1	No
Small	All	0.114	0.860	No	0.001	0.008	No

Notes: Each panel differs in the content of the information set being tested and is a separate family for the purpose of adjusting p-values. Panel A tests that $(dist_j, r_{iht-1}, R_{jt-1}) \subseteq \mathcal{J}_{ijt}$; panel B tests $\alpha_{jt} r_{iht} \subseteq \mathcal{J}_{ijt}$; panel C tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1}) \subseteq \mathcal{J}_{ijt}$; panel D tests $(dist_j, r_{iht-1}, R_{jt-1}, N_{jt-1}) \subseteq \mathcal{J}_{ijt}$; panel E(a) tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{c(j)t-1}) \subseteq \mathcal{J}_{ijt}$; panel E(b) tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{l(j)t-1}) \subseteq \mathcal{J}_{ijt}$; panel E(c) tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{i(j)t-1}) \subseteq \mathcal{J}_{ijt}$; panel E(d) tests $(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{b(j)t-1}) \subseteq \mathcal{J}_{ijt}$. *Large* firms are those with above median domestic sales in the previous year. Conversely, firm i at period t is defined as *Small* if its domestic sales fall below the median. *Popular* export destinations are those with above median number of exporters in the previous year. We define a firm i at period t as *Exporter* with respect to a country j if $d_{ijt-1} = 1$ and as a *Non-exporter* if $d_{ijt-1} = 0$. All reported p-values correspond to the test BP (Bugni et al., 2015); for details on how to compute these p-values, see Appendix A.8. All numbers reported in this table are independent of the value of η chosen as the normalizing constant.

Table B.6: Instrument Relevance (Chemicals)

	Panel C: Minimal Information & Country Shifter							
R_{jt-1}	0.011^a	0.006^b	0.001^a	< 0.001	0.017^a	0.013^a	0.001^a	0.001^a
	(6.40)	(2.05)	(11.6)	(0.71)	(5.69)	(8.59)	(4.09)	(19.3)
r_{iht-1}	0.014^a	0.011^a	0.016^a	0.013^a	0.009^a	0.014^a	0.014^a	0.015^a
	(9.82)	(11.7)	(32.6)	(29.4)	(13.1)	(10.7)	(6.32)	(41.8)
$dist_j$	0.332^a	0.443^a	0.016^a	0.023^a	0.579^a	0.253^a	0.050^a	0.014^a
	(4.55)	(22.1)	(7.53)	(26.4)	(5.12)	(11.3)	(2.85)	(21.2)
α_{jt-1}	12.94^a	2.611^b	0.584^a	0.146^a	13.18^a	4.140^a	0.221	0.238^a
	(6.84)	(2.15)	(11.9)	(3.82)	(3.82)	(3.78)	(0.68)	(7.63)
Firms	Large	Large	Small	Small	Large Exporter	Large Non-Exporter	Small Exporter	Small Non-Exporter
Countries	Popular	Unpopular	Popular	Unpopular	All	All	All	All
Num. Obs.	10,940	11,089	11,089	10,812	4,142	17,887	664	21,344
R^2_u/R^2_r	5.9%	0.3%	10.9%	1.0%	4.3%	0.9%	1.1%	2.8%

	Panel D: Minimal Information & Number of Exporters							
R_{jt-1}	0.023^a	0.010^a	0.001^a	$< 0.001^a$	0.030^a	0.018^a	0.001^a	0.001^a
	(19.3)	(8.08)	(34.7)	(9.22)	(12.3)	(14.7)	(6.04)	(34.7)
r_{iht-1}	0.023^a	0.011^a	0.016^a	0.012^a	0.009^a	0.014^a	0.014^a	0.014^a
	(19.3)	(11.7)	(31.9)	(29.6)	(12.9)	(10.8)	(6.28)	(42.2)
$dist_j$	-0.129	0.475^a	-0.005	0.024^a	0.421^b	0.236^a	0.039	0.012^a
	(-1.54)	(18.5)	(-1.88)	(21.6)	(3.46)	(11.1)	(1.65)	(17.4)
N_{jt-1}	-10.80^a	2.633^a	-0.534^a	0.120	-8.388^a	-3.245^a	-0.372	-0.235^a
	(-12.8)	(1.02)	(-17.8)	(1.12)	(-5.21)	(-5.60)	(-1.16)	(-13.1)
Firms	Large	Large	Small	Small	Large Exporter	Large Non-Exporter	Small Exporter	Small Non-Exporter
Countries	Popular	Unpopular	Popular	Unpopular	All	All	All	All
Num. Obs.	10,940	11,089	11,089	10,812	4,142	17,887	664	21,344
R^2_u/R^2_r	3.6%	0.1%	8.2%	0.1%	2.3%	0.4%	2.9%	2.1%

	Panel E: Minimal Information & Country Group Avg. Shifter							
	(a) Continent		*(b) Language*		*(c) Income p.c.*		*(d) Border*	
R_{jt-1}	0.019^a	0.001^a	0.017^a	$< 0.001^a$	0.019^a	$< 0.001^a$	0.018^a	0.001^a
	(19.4)	(33.7)	(17.7)	(33.1)	(20.5)	(38.2)	(19.1)	(34.5)
r_{iht-1}	0.012^a	0.014^a	0.012^a	0.014^a	0.012^a	0.014^a	0.012^a	0.014^a
	(14.7)	(42.4)	(14.7)	(42.4)	(14.6)	(42.8)	(14.7)	(42.3)
$dist_j$	0.341^a	0.017^a	0.273^a	0.014^a	0.368^a	0.018^a	0.361^a	0.018^a
	(20.3)	(26.3)	(12.6)	(19.9)	(21.8)	(31.9)	(23.0)	(31.8)
$\alpha_{g(j)t-1}$	2.514^a	0.184^a	12.12^a	0.556^a	0.875^b	0.102^a	4.002^a	0.210^a
	(6.03)	(6.18)	(9.36)	(14.3)	(1.92)	(3.11)	(5.61)	(9.57)
Firms	Large	Small	Large	Small	Large	Small	Large	Small
Countries	All	All	All	All	All	All	All	All
Num. Obs.	22,029	22,008	22,029	22,008	22,029	22,008	22,029	22,008
R^2_u/R^2_r	0.4%	2.6%	1.7%	3.7%	0.1%	0.5%	0.5%	1.5%

Notes: t-stats appear in parenthesis. ^{a} denotes 1% significance; ^{b} denotes 5% significance. In all regressions, the dependent variable is $\alpha_{jt} r_{ijt}$. The rows *Firm* and *Countries* describe the subset of firms and countries used as observations in each regression. Specifically, *Large* firms are those with above median domestic sales in the previous year: $\mathbb{1}\{domsales_{jt-1} \geq median(domsales_{jt-1})\} = 1$. Conversely, *Small* denotes $\mathbb{1}\{domsales_{jt-1} < median(domsales_{jt-1})\} = 1$. *Popular* destinations are those with above median number of exporters: $\mathbb{1}\{N_{jt-1} \geq median(N_{jt-1})\} = 1$, where N_{jt-1} is the number of Chilean firms in the corresponding sector (chemicals or food) exporting to j at $t-1$. Conversely, *Unpopular* denotes $\mathbb{1}\{N_{jt-1} < median(N_{jt-1})\} = 1$. *Exporter* denotes $d_{ijt-1} = 1$ and *Non-exporter* is $d_{ijt-1} = 0$. R^2_u/R^2_r is the ratio between the R^2 of the corresponding regression and the R^2 of a *restricted* regression that includes only $(R_{jt-1}, r_{iht-1}, dist_j)$ as covariates.

Table B.7: Instrument Relevance (Food)

	Panel C: Minimal Information & Country Shifter							
R_{jt-1}	0.007^a	0.003	$< 0.001^a$	$< 0.001^a$	0.008^a	0.006^a	0.001^a	$< 0.001^a$
	(14.4)	(1.63)	(20.3)	(3.73)	(10.5)	(11.3)	(5.26)	(21.9)
r_{iht-1}	0.036^a	0.019^a	0.041^a	0.020^a	0.032^a	0.027^a	0.075^a	0.029^a
	(22.1)	(26.5)	(30.5)	(44.9)	(10.8)	(29.0)	(6.55)	(48.0)
$dist_j$	-0.059^a	0.099^a	-0.004^a	0.005^a	0.065^a	0.043^a	-0.044^a	0.004^a
	(-2.81)	(15.1)	(-4.05)	(16.2)	(1.66)	(3.04)	(-4.27)	(9.95)
α_{jt-1}	5.614^a	9.067^a	0.161^a	0.486^a	5.381^a	7.359^a	0.375^a	0.251^a
	(6.40)	(17.4)	(6.82)	(22.8)	(5.29)	(9.48)	(2.72)	(15.1)
Firms	Large	Large	Small	Small	Large Exporter	Large Non-Exporter	Small Exporter	Small Non-Exporter
Countries	Popular	Unpopular	Popular	Unpopular	All	All	All	All
Num. Obs.	21,587	20,917	21,579	20,892	5,768	36,736	1,250	41,221
R_u^2/R_r^2	2.6%	8.9%	1.6%	29.2%	3.1%	6.1%	3.4%	5.9%
	Panel D: Minimal Information & Number of Exporters							
R_{jt-1}	0.013^a	0.041^a	0.001^a	0.002^a	0.014^a	0.012^a	0.001^a	0.001^a
	(17.1)	(23.4)	(22.5)	(27.7)	(12.6)	(16.1)	(6.93)	(29.7)
r_{iht-1}	0.036^a	0.019^a	0.041^a	0.019^a	0.031^a	0.027^a	0.074^a	0.028^a
	(22.1)	(26.3)	(30.1)	(43.4)	(10.7)	(18.7)	(6.55)	(47.1)
$dist_j$	-0.141^a	0.072^a	-0.007^a	0.003^a	-0.045	0.039^b	-0.046^a	0.003^a
	(-4.03)	(9.29)	(-5.03)	(10.7)	(-0.77)	(1.98)	(-3.09)	(5.51)
N_{jt-1}	-12.31^a	-35.84^a	-0.416^a	-1.930^a	-12.88^a	-9.444^a	-2.092^a	-0.361^a
	(-4.69)	(-14.7)	(-4.41)	(-18.7)	(-4.20)	(-5.62)	(-3.33)	(-8.61)
Firms	Large	Large	Small	Small	Large Exporter	Large Non-Exporter	Small Exporter	Small Non-Exporter
Countries	Popular	Unpopular	Popular	Unpopular	All	All	All	All
Num. Obs.	21,587	20,917	21,579	20,892	5,768	36,736	1,250	41,221
R_u^2/R_r^2	0.6%	4.0%	0.4%	13.0%	0.8%	0.7%	3.8%	0.7%
	Panel E: Minimal Information & Country Group Avg. Shifter							
	(a) Continent		*(b) Language*		*(c) Income p.c.*		*(d) Border*	
R_{jt-1}	0.009^a	0.001^a	0.011^a	0.001^a	0.011^a	0.001^a	0.010^a	0.001^a
	(28.8)	(33.5)	(27.0)	(32.6)	(28.6)	(34.5)	(25.9)	(31.0)
r_{iht-1}	0.028^a	0.031^a	0.028^a	0.030^a	0.028^a	0.030^a	0.028^a	0.030^a
	(30.1)	(42.1)	(29.9)	(41.8)	(29.9)	(41.8)	(29.9)	(42.1)
$dist_j$	0.069^a	0.004^a	0.103^a	0.005^a	0.087^a	0.004^a	0.096^a	0.005^a
	(5.49)	(7.26)	(1.8)	(12.3)	(8.15)	(8.93)	(9.89)	(10.5)
$\alpha_{g(j)t-1}$	5.201^a	0.188^a	0.134	-0.004	1.403^a	0.054^a	1.118^a	0.036^a
	(9.83)	(12.6)	(0.98)	(-0.51)	(9.86)	(6.71)	(6.91)	(4.98)
Firms	Large	Small	Large	Small	Large	Small	Large	Small
Countries	All	All	All	All	All	All	All	All
Num. Obs.	42,504	42,471	42,504	42,471	42,504	42,471	42,504	42,471
R_u^2/R_r^2	1.7%	1.7%	0.0%	0.0%	0.3%	0.2%	0.2%	0.2%

Notes: t-stats appear in parenthesis. ^{a} denotes 1% significance; ^{b} denotes 5% significance. In all regressions, the dependent variable is $\alpha_{jt} r_{ijt}$. The rows *Firm* and *Countries* describe the subset of firms and countries used as observations in each regression. Specifically, *Large* firms are those with above median domestic sales in the previous year: $\mathbb{1}\{domsales_{jt-1} \geq median(domsales_{jt-1})\} = 1$. Conversely, *Small* denotes $\mathbb{1}\{domsales_{jt-1} < median(domsales_{jt-1})\} = 1$. *Popular* destinations are those with above median number of exporters: $\mathbb{1}\{N_{jt-1} \geq median(N_{jt-1})\} = 1$, where N_{jt-1} is the number of Chilean firms in the corresponding sector (chemicals or food) exporting to j at $t-1$. Conversely, *Unpopular* denotes $\mathbb{1}\{N_{jt-1} < median(N_{jt-1})\} = 1$. *Exporter* denotes $d_{ijt-1} = 1$ and *Non-exporter* is $d_{ijt-1} = 0$. R_u^2/R_r^2 is the ratio between the R^2 of the corresponding regression and the R^2 of a *restricted* regression that includes only $(R_{jt-1}, r_{iht-1}, dist_j)$ as covariates.

B.5 Partial Equilibrium Counterfactuals: Details

When computing the effect of the 40% reduction in fixed export costs, we assume the parameters $\{\alpha_{jt}; \forall j \text{ and } t\}$ remain invariant. In equation (2), α_{jt} is a function of variable trade costs, foreign demand shocks, price indices and aggregate market size in both country j and Chile. A sufficient condition for α_{jt} not to change in reaction to the change in the parameters β_0 and β_1 is that the components of α_{jt} themselves remains invariant.

The model described in Section 2 treats variable trade costs τ_{ijt} and demand shifters ζ_{ijt} as exogenous parameters and, therefore, within our theoretical framework, they are invariant to changes in fixed export costs. The assumed invariance of the price index P_{jt} and market size Y_{jt} in destination country j to changes in trade costs between Chile and j rules out general equilibrium effects linking the increase in the number of Chilean firms exporting to j to either average prices in j or total income in j. This assumption is likely to be a good approximation as long as the share of imports coming from Chile in destination market j is small. Table B.8 shows that this is the case for both sectors and all destination countries in our sample.

Table B.8: Share of Imports coming from Chile

	1999	2000	2001	2002	2003
			Chemicals		
Argentina	0.09%	1.23%	1.42%	1.09%	1.09%
Australia	0.11%	0.01%	0.01%	0.01%	0.12%
Bolivia	8.96%	9.70%	8.23%	8.23%	7.82%
Brazil	0.91%	1.03%	0.96%	1.13%	1.20%
Colombia	0.60%	0.87%	0.93%	1.09%	1.06%
Costa Rica	0.56%	0.72%	0.46%	0.66%	0.60%
Ecuador	2.40%	3.69%	3.96%	4.00%	3.47%
Spain	0.18%	0.26%	0.30%	0.25%	0.25%
Great Britain	0.01%	0.01%	0.01%	0.01%	0.01%
Guatemala	0.43%	0.41%	0.48%	0.21%	0.27%
Italy	0.03%	0.04%	0.04%	0.03%	0.04%
Japan	0.14%	0.18%	0.22%	0.13%	0.31%
Mexico	0.24%	0.32%	0.38%	0.41%	0.35%
Panama	0.26%	0.35%	0.56%	0.54%	0.69%
Peru	4.09%	5.20%	6.58%	6.14%	5.62%
Paraguay	0.74%	1.46%	1.51%	1.30%	1.18%
El Salvador	0.13%	0.29%	0.17%	0.19%	0.31%
Uruguay	0.98%	0.77%	0.84%	0.95%	0.76%
United States	0.30%	0.28%	0.35%	0.36%	0.25%
Venezuela	0.46%	0.61%	0.62%	0.67%	0.73%
			Food		
Argentina	12.29%	9.81%	9.33%	8.17%	5.15%
Australia	0.72%	0.77%	0.06%	0.68%	0.55%
Belgium	0.14%	0.17%	0.17%	0.19%	0.21%
Bolivia	26.92%	22.12%	21.18%	22.19%	20.97%
Brazil	2.49%	3.27%	3.14%	3.58%	3.70%
Canada	0.94%	0.91%	0.77%	0.92%	1.04%
China	0.77%	0.83%	1.37%	2.09%	1.67%
Colombia	6.75%	5.60%	5.65%	5.60%	6.23%
Costa Rica	3.86%	3.61%	3.50%	4.00%	5.85%
Ecuador	20.98%	22.54%	25.46%	20.13%	16.48%
Germany	0.35%	0.42%	0.44%	0.39%	0.46%

Table B.8: Share of Imports coming from Chile (cont.)

	1999	2000	2001	2002	2003
Denmark	0.91%	1.17%	1.55%	1.22%	1.32%
Spain	0.87%	0.87%	0.97%	0.92%	0.99%
France	0.43%	0.51%	0.50%	0.54%	0.48%
Great Britain	0.77%	1.23%	1.53%	1.26%	1.10%
Indonesia	0.26%	0.47%	0.15%	0.32%	0.33%
India	0.17%	0.04%	0.13%	0.27%	0.55%
Italy	0.13%	0.18%	0.30%	0.36%	0.33%
Japan	0.27%	0.27%	0.26%	0.26%	0.28%
South Korea	0.31%	0.46%	0.49%	0.49%	0.98%
Sri Lanka	3.07%	3.60%	3.40%	3.34%	3.03%
Mexico	1.91%	1.84%	2.15%	2.32%	2.63%
Malaysia	0.11%	0.18%	0.16%	0.15%	0.19%
Netherlands	0.37%	0.40%	0.39%	0.29%	0.27%
New Zeland	0.61%	0.76%	0.53%	0.38%	0.63%
Panama	2.23%	1.77%	1.90%	2.05%	2.54%
Peru	8.77%	9.92%	10.14%	11.78%	11.88%
Phillipines	0.10%	0.33%	0.20%	0.23%	0.35%
Singapore	0.63%	0.81%	0.72%	0.83%	0.86%
Thailand	0.92%	0.43%	0.75%	0.92%	1.11%
Uruguay	2.81%	5.42%	6.75%	5.88%	5.46%
United States	1.90%	2.17%	2.19%	2.18%	2.41%
Venezuela	4.03%	5.44%	6.20%	5.49%	4.20%

Notes: Data on trade flows from UN Comtrade.

C Odds-based and Revealed-Preference Inequalities: Proofs

C.1 Proof of Theorem 1

We present here two alternative proofs of Theorem 1. We present the first proof in Appendix C.1.1 and the second one in Appendix C.1.2. The first proof makes use of the score function corresponding to the model in Section 2. The derivation in Appendix C.1.2 makes use of the definition of the export dummy d_{ijt} in equation (9).

C.1.1 First Proof of Theorem 1

Lemma C.1 *Let $L(d_{ijt}|\mathcal{J}_{ijt}, dist_j; \theta)$ denote the log-likelihood conditional on $\mathcal{J}_{ijt}$ and $dist_j$. Then*

$$\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt};\theta)}{\partial\theta} = \mathbb{E}\Big[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))} - (1-d_{ijt})\Big|\mathcal{J}_{ijt}, dist_j\Big] = 0. \quad \text{(C.1)}$$

Proof: It follows from the model in Section 2 that the log-likelihood function conditional on $\mathcal{J}_{ijt}$ and $dist_j$ can be written as

$$\begin{aligned} L(d_{ijt}|\mathcal{J}_{ijt}, dist_j;\theta) = \mathbb{E}\Big[& d_{ijt}\ln(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))) \\ & + (1-d_{ijt})\ln(\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))\Big|\mathcal{J}_{ijt}, dist_j\Big]. \end{aligned}$$

The score function is given by

$$\begin{aligned} &\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt}, dist_j;\theta)}{\partial\theta} = \quad \text{(C.2)} \\ &\mathbb{E}\Big[d_{ijt}\frac{1}{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\frac{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))}{\partial\theta} \\ &+(1-d_{ijt})\frac{1}{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\frac{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\partial\theta}\Big|\mathcal{J}_{ijt}, dist_j\Big] = 0. \end{aligned}$$

Reordering terms

$$\begin{aligned} \frac{\partial L(d_{ijt}|\mathcal{J}_{ijt}, dist_j;\theta)}{\partial\theta} = \mathbb{E}\Big[&\frac{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta}{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\times \\ &\Big[d_{ijt}\frac{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\times \\ \times\frac{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta} &+ (1-d_{ijt})\Big]\Big|\mathcal{J}_{ijt}, dist_j\Big] = 0. \end{aligned} \quad \text{(C.3)}$$

Given that

$$\frac{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta}{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}$$

is a function of $(\mathcal{J}_{ijt}, dist_j)$ and different from 0 for any value of the index $\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)$, and

$$\frac{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta} = -1$$

we can simplify:

$$\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt}, dist_j;\theta)}{\partial\theta} = \mathbb{E}\Big[d_{ijt}\frac{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))} - (1-d_{ijt})\Big|\mathcal{J}_{ijt}, dist_j\Big] = 0.$$

Equation (C.1) follows by symmetry of the function $\Phi(\cdot)$. ■

Lemma C.2 *Given the definition of d_{ijt} in equation (9),*

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j))}\middle|\mathcal{J}_{ijt}, dist_j\right]$$
$$\geq$$
$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\middle|\mathcal{J}_{ijt}, dist_j\right]. \tag{C.4}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}] = 0$ where, as a reminder, the set $\mathcal{J}_{ijt}$ includes every covariate the firm uses to predict export revenue at the time it decides on export destinations. From equation (9), it follows that d_{ijt} may be written as a function of the vector $(\mathcal{J}_{ijt}, dist_j, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt}, dist_j, \nu_{ijt})$. In the model in Section 2, we assume firms know both $dist_j$ and ν_{ijt} when determining d_{ijt}. Therefore, either they are independent of r^o_{ijt} and, consequently, of ε_{ijt}, or they belong to $\mathcal{J}_{ijt}$. In any case, it will be true that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt}] = 0$. Since

$$\frac{1-\Phi(y)}{\Phi(y)}$$

is convex for any value of y and $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt}] = 0$, by Jensen's Inequality

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j+\eta^{-1}\varepsilon_{ijt}))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)+\eta^{-1}\varepsilon_{ijt})}\middle|\mathcal{J}_{ijt}, dist_j\right]$$
$$\geq$$
$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\middle|\mathcal{J}_{ijt}, dist_j\right].$$

Equation (C.4) follows from the equality $\eta^{-1}r^o_{ijt} = \eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$. ■

Corollary 1 *Suppose the distribution of Z_{ijt} conditional on $(\mathcal{J}_{ijt}, dist_j)$ is degenerate. Then*

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-(1-d_{ijt})\middle|Z_{ijt}\right] = 0. \tag{C.5}$$

and

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j))}\middle|Z_{ijt}\right]$$
$$\geq$$
$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\middle|Z_{ijt}\right]. \tag{C.6}$$

Proof: The result follows from Lemmas C.1 and C.2 and the application of the Law of Iterated Expectations. ■

Lemma C.3 *Let $L(d_{ijt}|\mathcal{J}_{ijt}, dist_j;\theta)$ denote the log-likelihood conditional on $(\mathcal{J}_{ijt}, dist_j)$. Then*

$$\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt}, dist_j;\theta)}{\partial\theta} = \mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-d_{ijt}\middle|\mathcal{J}_{ijt}, dist_j\right] = 0. \tag{C.7}$$

Proof: From equation (C.2), reordering terms

$$\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt}, dist_j;\theta)}{\partial\theta} = \mathbb{E}\Bigg[\frac{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\Bigg[d_{ijt}+(1-d_{ijt})\times$$

$$\frac{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\frac{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta}{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}\Bigg|\mathcal{J}_{ijt}, dist_j\Bigg]=0.$$

Given that

$$\frac{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}$$

is a function of $\mathcal{J}_{ijt}$ and $dist_j$, and different from 0 for any value of the index $\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)$, and

$$\frac{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta}{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}=-1$$

we can simplify:

$$\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt}, dist_j;\theta)}{\partial\theta}=\mathbb{E}\Bigg[(1-d_{ijt})\frac{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-d_{ijt}\Bigg|\mathcal{J}_{ijt}, dist_j\Bigg]=0.$$

Equation (C.7) follows by symmetry of the function $\Phi(\cdot)$. ■

Lemma C.4 *Given the definition of d_{ijt} in equation* (9)*,*

$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j))}\Bigg|\mathcal{J}_{ijt}, dist_j\Bigg]$$
$$\geq$$
$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\Bigg|\mathcal{J}_{ijt}, dist_j\Bigg]. \tag{C.8}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt}=r^o_{ijt}-\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}]=0$ where, as a reminder, the set $\mathcal{J}_{ijt}$ includes every covariate the firm uses to predict export revenue at the time it decides on export destinations. From equation (9), it follows that d_{ijt} may be written as a function of the vector $(\mathcal{J}_{ijt}, dist_j, \nu_{ijt})$; i.e. $d_{ijt}=d(\mathcal{J}_{ijt}, dist_j, \nu_{ijt})$. In the model in Section 2, we assume firms know both $dist_j$ and ν_{ijt} when determining d_{ijt}. Therefore, either they are independent of r^o_{ijt} and, consequently, of ε_{ijt}, or they belong to $\mathcal{J}_{ijt}$. In any case, it will be true that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt}]=0$. Since

$$\frac{\Phi(y)}{1-\Phi(y)}$$

is convex for any value of y and $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt}]=0$, by Jensen's Inequality

$$\mathbb{E}\Bigg[d_{ijt}\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j+\eta^{-1}\varepsilon_{ijt}))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)+\eta^{-1}\varepsilon_{ijt})}\Bigg|\mathcal{J}_{ijt}, dist_j\Bigg]$$
$$\geq$$
$$\mathbb{E}\Bigg[d_{ijt}\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\Bigg|\mathcal{J}_{ijt}, dist_j\Bigg].$$

Equation (C.8) follows from the equality $\eta^{-1}r^o_{ijt}=\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]+\eta^{-1}\varepsilon_{ijt}$. ■

Corollary 2 *Suppose the distribution of Z_{ijt} conditional on $(\mathcal{J}_{ijt}, dist_j)$ is degenerate. Then*

$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-d_{ijt}\Bigg|Z_{ijt}\Bigg]=0. \tag{C.9}$$

and

$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j))}\Bigg|Z_{ijt}\Bigg]$$

$$\geq$$

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\middle| Z_{ijt}\right]. \tag{C.10}$$

Proof: The results follow from Lemmas C.3 and C.4 and the application of the Law of Iterated Expectations. ■

First Proof of Theorem 1 Combining equations (C.5) and (C.6), we obtain the inequality defined by equations (14) and (14b). Combining equations (C.9) and (C.10), we obtain the inequality defined by equations (14) and (14c). ■

C.1.2 Second Proof of Theorem 1

Lemma C.5 *Equations* (6) *and* (9) *imply that*

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-(1-d_{ijt})\middle|\mathcal{J}_{ijt}, dist_j\right]\geq 0. \tag{C.11}$$

Proof: Equation (9) implies that

$$d_{ijt}-\mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt}\geq 0\}\geq 0,$$

or, equivalently,

$$1-\mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt}\geq 0\}-(1-d_{ijt})\geq 0,$$
$$\mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt}\leq 0\}-(1-d_{ijt})\geq 0,$$

for every i, j and t. Given that this inequality holds for every firm, country, and year, it will also hold on average (conditional on any set of variables) across firms, countries and years. We specifically condition on the set $(\mathcal{J}_{ijt}, dist_j)$:

$$\mathbb{E}[\mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt}\leq 0\}-(1-d_{ijt})|\mathcal{J}_{ijt}, dist_j]\geq 0.$$

Imposing the distributional assumption in equation (6),

$$\mathbb{E}[(1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))-(1-d_{ijt})|\mathcal{J}_{ijt}, dist_j]\geq 0.$$

Dividing by $\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))$, we obtain

$$\mathbb{E}\left[\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-\frac{1-d_{ijt}}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\middle|\mathcal{J}_{ijt}, dist_j\right]\geq 0.$$

Adding and subtracting $1-d_{ijt}$

$$\mathbb{E}\Bigg[\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}$$
$$-\left(1-1+\frac{1}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\right)(1-d_{ijt})\Bigg|\mathcal{J}_{ijt}, dist_j\Bigg]\geq 0,$$

and, doing some simple algebra, we obtain

$$\mathbb{E}\Bigg[\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}$$
$$-\left(1+\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\right)(1-d_{ijt})\Bigg|\mathcal{J}_{ijt}, dist_j\Bigg]\geq 0,$$

and, finally, regrouping terms,

$$\mathbb{E}\Bigg[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-(1-d_{ijt})\Bigg|\mathcal{J}_{ijt},dist_j\Bigg]\geq 0. \quad \blacksquare$$

Lemma C.6 *Equations* (6) *and* (9) *imply that*

$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-d_{ijt}\Bigg|\mathcal{J}_{ijt},dist_j\Bigg]\geq 0. \tag{C.12}$$

Proof: Equation (9) implies that

$$\mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt}\geq 0\}-d_{ijt}\geq 0,$$

for every i, j and t. Given that this inequality holds for every firm, country, and year, it will also hold on average (conditional on any set of variables) across firms, countries and years. We specifically condition on the set $(\mathcal{J}_{ijt}, dist_j)$:

$$\mathbb{E}[\mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt}\geq 0\}-d_{ijt}|\mathcal{J}_{ijt},dist_j]\geq 0.$$

or, equivalently,

$$\mathbb{E}[\mathbb{1}\{\nu_{ijt}\leq\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\}-d_{ijt}|\mathcal{J}_{ijt},dist_j]\geq 0.$$

Imposing the distributional assumption in equation (6),

$$\mathbb{E}[\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))-d_{ijt}|\mathcal{J}_{ijt},dist_j]\geq 0.$$

Dividing by $1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))$,

$$\mathbb{E}\Bigg[\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-\frac{d_{ijt}}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\Bigg|\mathcal{J}_{ijt},dist_j\Bigg]\geq 0.$$

Adding and subtracting d_{ijt}

$$\mathbb{E}\Bigg[\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))} \\ -\Big(1-1+\frac{1}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\Big)d_{ijt}\Bigg|\mathcal{J}_{ijt},dist_j\Bigg]\geq 0,$$

and, doing some simple algebra, we obtain

$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-d_{ijt}\Bigg|\mathcal{J}_{ijt},dist_j\Bigg]\geq 0. \quad \blacksquare$$

Second Proof of Theorem 1 Combining equations (C.4), (C.5), (C.6), and (C.11), we obtain the inequality defined by equations (14) and (14b). Combining equations (C.8), (C.9), (C.10), and (C.12), we obtain the inequality defined by equations (14) and (14c). ■

C.2 Proof of Theorem 2

Lemma C.7 *Equation* (9) *implies that*

$$\mathbb{E}[d_{ijt}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt})|\mathcal{J}_{ijt},dist_j]\geq 0. \tag{C.13}$$

Proof: Equation (9) implies that

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\},$$

and, therefore,

$$d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt}\big) \geq 0.$$

This inequality holds for every firm i, country j, and year t. Therefore, it will also hold in expectation conditional on $(\mathcal{J}_{ijt}, dist_j)$. ■

Lemma C.8 *Equations* (6) *and* (9) *imply that*

$$\begin{aligned}\mathbb{E}\Big[&d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) \\ &+ (1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}, dist_j\Big] \geq 0.\end{aligned} \tag{C.14}$$

Proof: Equation (C.13) implies

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}, dist_j\Big] - \mathbb{E}\big[d_{ijt}\nu_{ijt}|\mathcal{J}_{ijt}, dist_j\big] \geq 0. \tag{C.15}$$

Since the assumption in equation (6) implies that $\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j] = 0$, it follows that

$$\mathbb{E}[d_{ijt}\nu_{ijt} + (1-d_{ijt})\nu_{ijt}|\mathcal{J}_{ijt}, dist_j] = 0,$$

and we can rewrite equation (C.15) as

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)|\mathcal{J}_{ijt}, dist_j] + \mathbb{E}\big[(1-d_{ijt})\nu_{ijt}|\mathcal{J}_{ijt}, dist_j\big] \geq 0. \tag{C.16}$$

Applying the Law of Iterated Expectations, it follows that

$$\begin{aligned}\mathbb{E}\big[(1-d_{ijt})\nu_{ijt}|\mathcal{J}_{ijt}, dist_j\big] &= \mathbb{E}\big[\mathbb{E}[(1-d_{ijt})\nu_{ijt}|d_{ijt}, \mathcal{J}_{ijt}, dist_j]|\mathcal{J}_{ijt}, dist_j\big] \\ &= \mathbb{E}\big[(1-d_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt}, \mathcal{J}_{ijt}, dist_j]|\mathcal{J}_{ijt}, dist_j\big] \\ &= P(d_{ijt}=1|\mathcal{J}_{ijt}, dist_j) \times 0 \times \mathbb{E}[\nu_{ijt}|d_{ijt}=1, \mathcal{J}_{ijt}, dist_j] \\ &\qquad + P(d_{ijt}=0|\mathcal{J}_{ijt}, dist_j) \times 1 \times \mathbb{E}[\nu_{ijt}|d_{ijt}=0, \mathcal{J}_{ijt}, dist_j] \\ &= P(d_{ijt}=0|\mathcal{J}_{ijt}, dist_j)\mathbb{E}[\nu_{ijt}|d_{ijt}=0, \mathcal{J}_{ijt}, dist_j] \\ &= \mathbb{E}[(1-d_{ijt})|\mathcal{J}_{ijt}, dist_j]\mathbb{E}[\nu_{ijt}|d_{ijt}=0, \mathcal{J}_{ijt}, dist_j] \\ &= \mathbb{E}\big[(1-d_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt}=0, \mathcal{J}_{ijt}, dist_j]|\mathcal{J}_{ijt}, dist_j\big],\end{aligned}$$

and we can rewrite equation (C.16) as

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + (1-d_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt}=0, \mathcal{J}_{ijt}]|\mathcal{J}_{ijt}, dist_j\big] \geq 0. \tag{C.17}$$

Using the definition of d_{ijt} in equation (9), it follows

$$\mathbb{E}[\nu_{ijt}|d_{ijt}=0, \mathcal{J}_{ijt}, dist_j] = \mathbb{E}[\nu_{ijt}|(\nu_{ijt} \geq \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j), \mathcal{J}_{ijt}, dist_j]$$

and, following equation (6), we can rewrite

$$\mathbb{E}[\nu_{ijt}|d_{ijt}=0, \mathcal{J}_{ijt}, dist_j] = \sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}.$$

Equation (C.14) follows by applying this equality to equation (C.17). ■

Lemma C.9 *Equations* (6) *and* (9) *imply*

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}, dist_j\Big] = \mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}, dist_j\Big] \tag{C.18}$$

Proof: From: (a) the definition of ε_{ijt} as $\varepsilon_{ijt} = r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$; (b) the definition of $\mathcal{J}_{ijt}$ as any variable known to firm i at the time it decides whether to export to j at t that is relevant to predict r^o_{ijt}; (c) the assumption that all firms know $dist_j$ when deciding whether to export to j at t, we can conclude that

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}, dist_j\Big] =$$
$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}, dist_j\Big] + \mathbb{E}\Big[\eta^{-1}d_{ijt}\varepsilon_{ijt}\Big|\mathcal{J}_{ijt}, dist_j\Big]. \tag{C.19}$$

From equation (6) and the definition of $\mathcal{J}_{ijt}$ as encompassing any variable that is useful to predict r_{ijt}, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = 0$. Therefore, applying the Law of Iterated Expectations,

$$\mathbb{E}\big[\eta^{-1}d_{ijt}\varepsilon_{ijt}\big|\mathcal{J}_{ijt}, dist_j\big] = \mathbb{E}\big[\eta^{-1}d_{ijt}\mathbb{E}\big[\varepsilon_{ijt}\big|\mathcal{J}_{ijt}, dist_j, d_{ijt}\big]\big|\mathcal{J}_{ijt}, dist_j\big] = \mathbb{E}\big[\eta^{-1}d_{ijt}\times 0\big|\mathcal{J}_{ijt}, dist_j\big] = 0.$$

Applying this result to equation (C.19) yields equation (C.18).

Lemma C.10 *Equations* (6) *and* (9) *imply that*

$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}, dist_j\Big]$$
$$\geq$$
$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}, dist_j\Big] \tag{C.20}$$

Proof: From the definition of ε_{ijt} as $\varepsilon_{ijt} = r^o_{ijt} - \mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, the assumption in equation (6), and the definition of $\mathcal{J}_{ijt}$ as the information set firm i uses to predict revenue when it decides whether to export to j at t, it follows that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = 0$. From equation (9), it follows that d_{ijt} is a function of the vector $(\mathcal{J}_{ijt}, dist_j, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt}, dist_j, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt}] = 0$. Since

$$\frac{\phi(y)}{1-\Phi(y)}$$

is convex for any value of y and $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, d_{ijt}, dist_j] = 0$, by Jensen's Inequality

$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + \eta^{-1}\varepsilon_{ijt})\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + \eta^{-1}\varepsilon_{ijt})\big)}\Big|\mathcal{J}_{ijt}, dist_j\Big]$$
$$\geq$$
$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}, dist_j\Big]$$

Equation (C.20) follows from the equality $\eta^{-1}r^o_{ijt} = \eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$. ■

Corollary 3 *Suppose the distribution of* Z_{ijt} *conditional on* $(\mathcal{J}_{ijt}, dist_j)$ *is degenerate. Then*

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + (1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq 0, \tag{C.21}$$

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] = \mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|Z_{ijt}\Big], \tag{C.22}$$

and

$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}r^o_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big]$$

$$\geq$$

$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|Z_{ijt}\Big]. \tag{C.23}$$

Proof: The results follow from Lemmas C.8, C.9 and C.10 and the application of the Law of Iterated Expectations. ■

Lemma C.11 *Equation* (9) *implies*

$$\mathbb{E}[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt})|\mathcal{J}_{ijt},dist_j]\geq 0. \tag{C.24}$$

Proof: From equation (9),

$$d_{ijt}=\mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt}\geq 0\}.$$

This implies

$$-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt})\geq 0.$$

This inequality holds for every firm i, country j, and year t. Therefore, it will also hold in expectation conditional on $\mathcal{J}_{ijt}$ and $dist_j$. ■

Lemma C.12 *Equations* (6) *and* (9) *imply*

$$\begin{aligned}\mathbb{E}\Big[&-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\\&+d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt},dist_j\Big]\geq 0.\end{aligned} \tag{C.25}$$

Proof: From equation (C.24),

$$\mathbb{E}\big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt},dist_j\Big]+\mathbb{E}\big[(1-d_{ijt})\nu_{ijt}\big|\mathcal{J}_{ijt},dist_j\big]\geq 0. \tag{C.26}$$

Since the assumption in equation (6) implies that $\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt},dist_j]=0$, it follows that

$$\mathbb{E}[d_{ijt}\nu_{ijt}+(1-d_{ijt})\nu_{ijt}|\mathcal{J}_{ijt},dist_j]=0,$$

and we can rewrite equation (C.26) as

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\big|\mathcal{J}_{ijt}\big]-\mathbb{E}\big[d_{ijt}\nu_{ijt}\big|\mathcal{J}_{ijt},dist_j\big]\geq 0. \tag{C.27}$$

Applying the Law of Iterated Expectations, it follows that

$$\begin{aligned}\mathbb{E}\big[d_{ijt}\nu_{ijt}\big|\mathcal{J}_{ijt},dist_j\big]&=\mathbb{E}\big[\mathbb{E}[d_{ijt}\nu_{ijt}|d_{ijt},\mathcal{J}_{ijt},dist_j]\big|\mathcal{J}_{ijt},dist_j\big]\\&=\mathbb{E}\big[d_{ijt}\mathbb{E}[\nu_{ijt}|d_{ijt},\mathcal{J}_{ijt},dist_j]\big|\mathcal{J}_{ijt},dist_j\big]\\&=P(d_{ijt}=1|\mathcal{J}_{ijt},dist_j)\times 1\times\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{J}_{ijt},dist_j]\\&\qquad+P(d_{ijt}=0|\mathcal{J}_{ijt},dist_j)\times 0\times\mathbb{E}[\nu_{ijt}|d_{ijt}=0,\mathcal{J}_{ijt},dist_j]\\&=P(d_{ijt}=1|\mathcal{J}_{ijt},dist_j)\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{J}_{ijt},dist_j]\\&=\mathbb{E}[d_{ijt}|\mathcal{J}_{ijt},dist_j]\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{J}_{ijt},dist_j]\\&=\mathbb{E}\big[d_{ijt}\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{J}_{ijt},dist_j]\big|\mathcal{J}_{ijt},dist_j\big],\end{aligned}$$

and we can rewrite equation (C.27) as

$$\mathbb{E}\big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)-d_{ijt}\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{J}_{ijt}]\big|\mathcal{J}_{ijt},dist_j\big]\geq 0. \tag{C.28}$$

Using the definition of d_{ijt} in equation (9), it follows

$$\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{J}_{ijt},dist_j]=\mathbb{E}[\nu_{ijt}|\nu_{ijt}\leq\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j,\mathcal{J}_{ijt},dist_j]$$

and, following equation (6), we can rewrite

$$\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{J}_{ijt},dist_j]=-\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}.$$

Equation (C.25) follows by applying this equality to equation (C.28). ■

Lemma C.13 *Equations* (6) *and* (9) *imply*

$$\begin{aligned}&\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt},dist_j\Big]\\&\qquad=\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt},dist_j\Big]\end{aligned}\tag{C.29}$$

Proof: From the definition of ε_{ijt} as $\varepsilon_{ijt}=r^o_{ijt}-\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$,

$$\begin{aligned}&\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt},dist_j\Big]=\\&\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt},dist_j\Big]-\mathbb{E}\Big[\eta^{-1}(1-d_{ijt})\varepsilon_{ijt}\Big|\mathcal{J}_{ijt},dist_j\Big].\end{aligned}\tag{C.30}$$

From the definition of ε_{ijt} as $\varepsilon_{ijt}=r^o_{ijt}-\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, the assumption in equation (6), and the definition of $\mathcal{J}_{ijt}$ as the information set firm i uses to predict revenue when it decides whether to export to j at t, it follows that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt},dist_j,\nu_{ijt}]=0$. From equation (9), it follows that d_{ijt} is a function of the vector $(\mathcal{J}_{ijt},dist_j,\nu_{ijt})$; i.e. $d_{ijt}=d(\mathcal{J}_{ijt},dist_j,\nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt},dist_j,d_{ijt}]=0$ and, applying the Law of Iterated Expectations,

$$\begin{aligned}\mathbb{E}\big[\eta^{-1}(1-d_{ijt})\varepsilon_{ijt}\big|\mathcal{J}_{ijt},dist_j\big]&=\mathbb{E}\big[\eta^{-1}(1-d_{ijt})\mathbb{E}\big[\varepsilon_{ijt}\big|\mathcal{J}_{ijt},dist_j,d_{ijt}\big]\big|\mathcal{J}_{ijt},dist_j\big]\\&=\mathbb{E}\big[\eta^{-1}(1-d_{ijt})\times 0\big|\mathcal{J}_{ijt},dist_j\big]=0.\end{aligned}$$

Applying this result to equation (C.30) yields equation (C.29).

Lemma C.14 *Equations* (6) *and* (9) *imply*

$$\begin{aligned}&\mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt},dist_j\Big]\\&\qquad\geq\mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt},dist_j\Big]\end{aligned}\tag{C.31}$$

Proof: From the definition of ε_{ijt} as $\varepsilon_{ijt}=r^o_{ijt}-\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$, the assumption in equation (6), and the definition of $\mathcal{J}_{ijt}$ as the information set firm i uses to predict revenue when it decides whether to export to j at t, it follows that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt},dist_j,\nu_{ijt}]=0$. From equation (9), it follows that d_{ijt} is a function of the vector $(\mathcal{J}_{ijt},dist_j,\nu_{ijt})$; i.e. $d_{ijt}=d(\mathcal{J}_{ijt},dist_j,\nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt},dist_j,d_{ijt}]=0$. Since

$$\frac{\phi(y)}{\Phi(y)}$$

is convex for any value of y and $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt},dist_j,d_{ijt}]=0$, by Jensen's Inequality

$$\begin{aligned}&\mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j+\eta^{-1}\varepsilon_{ijt})\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j+\eta^{-1}\varepsilon_{ijt})\big)}\Big|\mathcal{J}_{ijt},dist_j\Big]\geq\\&\mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt},dist_j\Big]\end{aligned}$$

Equation (C.31) follows from the equality $\eta^{-1}r^o_{ijt}=\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]+\eta^{-1}\varepsilon_{ijt}$. ■

Corollary 4 *Suppose the distribution of Z_{ijt} conditional on $(\mathcal{J}_{ijt}, dist_j)$ is degenerate. Then*

$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big) + d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq 0. \tag{C.32}$$

$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j\big)\Big|Z_{ijt}\Big] = \mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|Z_{ijt}\Big] \tag{C.33}$$

and

$$\mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}r^o_{ijt}-\beta_0-\beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq \mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \tag{C.34}$$

Proof of Theorem 2 Combining equations (C.21), (C.22), and (C.23) we obtain the inequality defined by equations (17) and (17b). Combining equations (C.32), (C.33), and (C.34) we obtain the inequality defined by equations (17) and (17c). ■

D Bias in Maximum Likelihood Estimates

In Section D.1, we show theoretically the different sources of bias that might affect the maximum likelihood (henceforth, ML) estimator in those cases in which the researcher assumes an information set $\mathcal{J}^a_{ijt}$ that is different from the actual information set $\mathcal{J}_{ijt}$ firm i uses to predict its potential export revenue. In Section D.2, we report results from simulations that illustrate numerically the magnitude and sign of the bias in the ML estimator, depending on the relationship between the true information set, $\mathcal{J}_{ijt}$, and the assumed one, $\mathcal{J}^a_{ijt}$.

D.1 Theory

To estimate the parameter vector θ using maximum likelihood, the researcher must construct a proxy for the firm's expectations about the observable part of its potential export revenue, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$. If the researcher assumes perfect foresight, she sets the proxy equal to the observed export revenues, r^o_{ijt}; if the researcher opts for fully specifying the content of exporters' information sets, she projects observed export revenues on a vector of observed covariates $\mathcal{J}^a_{ijt}$, and uses the outcome of this projection as her proxy.

As in equation (13), we use ξ_{ijt} to denote the difference between the researcher's assumed proxy $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}]$ and the firm's true expectation $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$. Here, $\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]$ is the true unobserved covariate entering the firm's export decision (see equation (9)) and $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}]$ is the researcher's proxy for it. Therefore, ξ_{ijt} represents measurement error in the definition of the proxy.

Under the assumption that $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}]$ is a perfect proxy for firms' unobserved expectations, and using the model assumptions in equations (6) and (9), the researcher will conclude that:

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\}, \qquad \nu_{ijt}|(\mathcal{J}^a_{ijt}, dist_j) \sim \mathbb{N}(0, \sigma^2). \tag{D.1}$$

Therefore, for a given value for the normalizing constant η, the researcher constructs estimates for $(\beta_0, \beta_1, \sigma)$ using values of the unknown parameter vector $(\theta_0, \theta_1, \theta_2)$ that maximize the following log-likelihood function

$$\begin{aligned} &L_a(\theta|d, \mathcal{J}^a, dist) = \qquad\qquad \text{(D.2)}\\ &\sum_{i,j,t}\Big\{d_{ijt}\ln\Big(\int_\nu \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \theta_0 - \theta_1 dist_j - \nu \geq 0\} f_\nu(\nu|\mathcal{J}^a_{ijt}, dist_j;\theta_2)\Big) +\\ &(1-d_{ijt})\ln\Big(\int_\nu \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \theta_0 - \theta_1 dist_j - \nu \leq 0\} f_\nu(\nu|\mathcal{J}^a_{ijt}, dist_j;\theta_2)\Big)\Big\} =\\ &\sum_{i,j,t}\Big\{d_{ijt}\ln(\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \theta_0 - \theta_1 dist_j))) + (1-d_{ijt})\ln(1-\Phi(\theta_2^{-1}(\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \theta_0 - \theta_1 dist_j)))\Big\}, \end{aligned}$$

where $L_a(\cdot)$ stands for *assumed* log-likelihood function and $f_\nu(\nu|\mathcal{J}^a_{ijt}, dist_j;\theta_2)$ is the density function of ν_{ijt} conditional on the vector $(\mathcal{J}^a_{ijt}, dist_j)$. According to equation (D.1), $f_\nu(\nu|\mathcal{J}^a_{ijt}, dist_j;\theta_2)$ is simply the density of a normal random variable with mean zero and standard deviation θ_2.

However, if $\xi_{ijt} \neq 0$, then the actual decision rule that conditions on the observed proxy $\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}]$ is

$$\begin{aligned} d_{ijt} &= \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\},\\ &= \mathbb{1}\{\eta^{-1}(\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \xi_{ijt}) - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\},\\ &= \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \beta_0 - \beta_1 dist_j - (\nu_{ijt} + \eta^{-1}\xi_{ijt}) \geq 0\},\\ &= \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \beta_0 - \beta_1 dist_j - \chi_{ijt} \geq 0\}, \end{aligned} \tag{D.3}$$

where the last equality defines a random variable $\chi_{ijt} \equiv \nu_{ijt} + \eta^{-1}\xi_{ijt}$ that accounts for both the structural error ν_{ijt} and the measurement error ξ_{ijt}. Therefore, the correct log-likelihood function conditional on the assumed information set $\mathcal{J}^a_{ijt}$ for every firm, country and year is

$$\begin{aligned} &L(\theta|d, \mathcal{J}^a, dist) =\\ &\sum_{i,j,t}\Big\{d_{ijt}\ln\Big(\int_\chi \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \theta_0 - \theta_1 dist_j - \chi \geq 0\} f_\chi(\chi|\mathcal{J}^a_{ijt}, dist_j;\theta_2)\Big) +\\ &(1-d_{ijt})\ln\Big(\int_\chi \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}^a_{ijt}] - \theta_0 - \theta_1 dist_j - \chi \leq 0\} f_\chi(\chi|\mathcal{J}^a_{ijt}, dist_j;\theta_2)\Big)\Big\}, \end{aligned} \tag{D.4}$$

where $f_\chi(\chi|\mathcal{J}^a_{ijt}, dist_j; \theta_2)$ is the correct density function of χ_{ijt} conditional on the vector $(\mathcal{J}^a_{ijt}, dist_j)$.

The values of the parameter vector $(\theta_0, \theta_1, \theta_2)$ that maximize the correct log-likelihood function in equation (D.4) will generally be different from those maximizing the log-likelihood function assumed by the researcher and described in equation (D.2). As a comparison of equations (D.4) and (D.2) clearly reveals, this difference arises only because the conditional density function $f_\chi(\chi|\mathcal{J}^a_{ijt}, dist_j; \theta_2)$ entering the correct log-likelihood function in equation (D.4) is different from the conditional density function $f_\nu(\nu|\mathcal{J}^a_{ijt}, dist_j; \theta_2)$ entering the assumed log-likelihood function in equation (D.2). These two density functions may differ for three reasons.

First, statistical independence may fail. While equation (6) assumes that ν_{ijt} is independent of $(\mathcal{J}^a_{ijt}, dist_j)$, the distribution of χ_{ijt} may not be independent of $(\mathcal{J}^a_{ijt}, dist_j)$. In particular, we worry about statistical dependence between ξ_{ijt} and $\mathcal{J}^a_{ijt}$, which will arise when the assumed information set, $\mathcal{J}^a_{ijt}$, includes a covariate that is correlated with r^o_{ijt} and not measurable in the true information set, $\mathcal{J}_{ijt}$. In practice, this dependence arises when the researcher assumes that exporters know more than what they actually know – for example, when the researcher wrongly assumes that exporters have perfect foresight.

Second, functional forms may differ. Even in cases in which the measurement error ξ_{ijt} is statistically independent of $\mathcal{J}^a_{ijt}$, the distribution $f_\chi(\chi|\mathcal{J}^a_{ijt}, dist_j; \theta_2)$ will generally not have the same functional form as the distribution of $f_\nu(\nu|\mathcal{J}^a_{ijt}, dist_j; \theta_2)$. In our empirical application, equation (6) assumes that ν_{ijt} is normal. Therefore, the marginal density functions of χ_{ijt} and ν_{ijt} conditional on $(\mathcal{J}^a_{ijt}, dist_j)$ will have the same functional form if and only if ξ_{ijt} is also normally distributed.

Finally, third, the variance of the unobserved determinant of export participation may differ. Even in those cases in which χ_{ijt} is independent of $\mathcal{J}^a_{ijt}$ and distributed normally, the value of θ_2 that maximizes the correct log-likelihood function in equation (D.4) and the researcher's assumed log-likelihood function in equation (D.2) will be different when the variance of ν_{ijt} and that of χ_{ijt} are different. Specifically, one can rewrite the variance of χ_{ijt} as $var(\chi_{ijt}) = var(\nu_{ijt} + \eta^{-1}\xi_{ijt}) = var(\nu_{ijt}) + \eta^{-2}var(\xi_{ijt}) = \sigma^2 + \eta^{-2}var(\xi_{ijt})$. Therefore, if χ_{ijt} is independent of $\mathcal{J}^a_{ijt}$ and distributed normally, the parameter vector (θ_0, θ_1) that maximizes the log-likelihood function in equation (D.2) is a consistent estimator of the true parameter vector (β_0, β_1). In contrast, the parameter θ_2 that maximizes this same log-likelihood function will overestimate the variance of the structural error: it will converge to $\sigma^2 + \eta^{-2}var(\xi_{ijt})$ instead of converging to σ^2.

D.2 Simulated Model

In the following subsections, we simulate simplified versions of the model described in Section 2 and explore the bias of the ML estimator for three different kinds of misspecification of firms' information sets. First, we examine the case in which exporters are uncertain about the observable part of their export revenue but the researcher wrongly assumes they have perfect foresight. Second, we consider the case in which the researcher assumes an information set $\mathcal{J}^a_{ijt}$ such that the distribution of $\mathcal{J}_{ijt}$ conditional on $\mathcal{J}^a_{ijt}$ is degenerate; i.e. the researcher assumes that exporters know strictly more than what they actually know. Third, we study the case in which the researcher assumes that exporters know less than what they actually know; i.e. the distribution of the assumed information set $\mathcal{J}^a_{ijt}$ conditional on the true information set $\mathcal{J}_{ijt}$ is degenerate. Apart from the definition of $\mathcal{J}_{ijt}$ and $\mathcal{J}^a_{ijt}$, we keep all other attributes of these three simulated models the same.

In our simulations, we model the decision process of $N = 1,000,000$ potential exporters $i = 1, \ldots, N$ who decide at a single period t whether to export to a single market j. We therefore omit the subindices j and t here. Each firm i decides whether to export according to the decision rule

$$d_i = \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_i|\mathcal{J}_i] - \beta_0 - \nu_i\}, \tag{D.5}$$

where $\beta_0 + \nu_i$ denotes i's fixed costs of exporting. We fix $\eta^{-1} = \beta_0 = 0.5$ and simulate the vector $(\nu_1, \ldots, \nu_N)$ by taking independent random draws from a normal distribution with mean zero and variance $\sigma^2 = 1$:

$$\nu_i \sim \mathbb{N}(0, \sigma^2). \tag{D.6}$$

We simulate the observable component of the potential export revenue as:

$$r^o_i = x_{1i} + x_{2i} + x_{3i}, \tag{D.7}$$

where x_{1i}, x_{2i} and x_{3i} all independently distributed. We set $x_{3i} \sim \mathbb{N}(0, 0.5)$ for all models. As we define each of the three models, we will place different assumptions on which of x_{1i}, x_{2i} and x_{3i} are included in the true information set and in the researcher's proxy for the firm's information. We will also place different assumptions on the marginal distributions of x_{1i} and x_{2i}.

For each of our three models, we consider the inference problem of a researcher who observes $\{(d_i, x_{1i}, x_{2i}, x_{3i}); i = 1, \ldots, N\}$, fixes η^{-1} to its true value 0.5, and estimates the parameter vector (β_0, σ). To match the empirical setting discussed in the main draft, we assume the researcher does not observe the true information set of each firm i (the researcher does not observe $\{\mathcal{J}_i; i = 1, \ldots, N\}$) and must therefore assume an information set for each of these firms (the researcher assumes $\{\mathcal{J}_i^a; i = 1, \ldots, N\}$).[48]

D.3 Bias Under Perfect Foresight

We consider here the bias generated by wrongly assuming perfect foresight in cases in which exporters are uncertain about their export revenue upon entry. Specifically, we simulate a model in which, for every firm i, $\mathcal{J}_i = x_{1i}$ and $\mathcal{J}_i^a = (x_{1i}, x_{2i}, x_{3i})$. Therefore,

$$\mathbb{E}[r_i^o|\mathcal{J}_i] = x_{1i} \qquad \text{and} \qquad \mathbb{E}[r_i^o|\mathcal{J}_i^a] = r_i = x_{1i} + x_{2i} + x_{3i}, \tag{D.8}$$

and, from equation (13), the measurement error introduced by the misspecification of agents' expectations is thus

$$\xi_i = \mathbb{E}[r_i^o|\mathcal{J}_i^a] - \mathbb{E}[r_i^o|\mathcal{J}_i] = r_i - x_{1i} = x_{2i} + x_{3i}, \tag{D.9}$$

which is identical to the expectational error that firm i makes; i.e. $\xi_i = r_i^o - \mathbb{E}[r_i^o|\mathcal{J}_i]$. Given equations (D.5), (D.6), and (D.8), the researcher will estimate the parameter vector (β_0, σ) finding the values of (θ_0, θ_2) that maximize the following log-likelihood function

$$L_a(\theta|d, \mathcal{J}^a) = \sum_{i=1}^{N} \Big\{ d_i \ln(\Phi(\theta_2^{-1}(\eta^{-1} r_i^o - \theta_0))) + (1 - d_i) \ln(1 - \Phi(\theta_2^{-1}(\eta^{-1} r_i^o - \theta_0))) \Big\}. \tag{D.10}$$

However, the correct log-likelihood function is:

$$\begin{aligned} &L(\theta|d, \mathcal{J}) = \\ &\sum_{i=1}^{N} \Big\{ d_i \ln \Big(\int_{\nu+\eta^{-1}(x_{2i}+x_{3i})} \mathbb{1}\{\eta^{-1} r_i^o - \theta_0 - (\nu_i + \eta^{-1}(x_{2i} + x_{3i})) \geq 0\} f(\nu_i + \eta^{-1}(x_{2i} + x_{3i})|r_i^o; \theta_2) \Big) \\ &+ (1 - d_i) \ln \Big(\int_{\nu+\eta^{-1}(x_{2i}+x_{3i})} \mathbb{1}\{\eta^{-1} r_i^o - \theta_0 - (\nu_i + \eta^{-1}(x_{2i} + x_{3i})) < 0\} f(\nu_i + \eta^{-1}(x_{2i} + x_{3i})|r_i^o; \theta_2) \Big) \Big\}. \end{aligned} \tag{D.11}$$

The conditional densities $f(\nu_i|r_i^o; \theta_2)$ and $f(\nu_i + \eta^{-1}(x_{2i} + x_{3i})|r_i^o; \theta_2)$ differ in at least two dimensions. First, while ν_i is independent of r_i^o, $\nu_i + \eta^{-1}(x_{2i} + x_{3i})$ is not. From (D.8) and (D.9), the measurement error term ξ_i is positively correlated with the researcher's measure of the exporters' expectation, $\mathbb{E}[r_i^o|\mathcal{J}_i^a] = r_i^o$:

$$\begin{aligned} cov(\xi_i, \mathbb{E}[r_i^o|\mathcal{J}_i^a]) &= cov(\xi_i, r_i^o) \\ &= cov(x_{2i} + x_{3i}, x_{1i} + x_{2i} + x_{3i}) = var(x_{2i}) + var(x_{3i}). \end{aligned} \tag{D.12}$$

Therefore, the aggregate error term, $\chi_i = \nu_i + \eta^{-1}\xi_i$, is also correlated with $\mathbb{E}[r_i^o|\mathcal{J}_i^a]$. Second, the variance of χ_i will also be larger than the variance of ν_i:

$$var(\chi_i) = \sigma^2 + \eta^{-2}(var(x_{2i}) + var(x_{3i})). \tag{D.13}$$

Additionally, if either x_{2i} and x_{3i} are not normally distributed, then the shape of the density function $f(\nu_i|r_i^o; \theta_2)$ will also differ from that of $f(\nu_i + \eta^{-1}(x_{2i} + x_{3i})|r_i^o; \theta_2)$.

By definition, the value of the parameter (θ_0, θ_2) that maximizes the log-likelihood function in equation (D.11) is equal to the true parameter vector $(\beta_0, \sigma^2) = (0.5, 1)$. In Table D.1 below, for different distributions of x_{1i} and x_{2i}, we show the point estimates and standard errors for the parameter vector (θ_0, θ_2) that maximizes the likelihood function in equation (D.10).

From equation (D.8), we know the distribution of x_{1i} is identical to the distribution of the true unobserved

[48] Note that the parameter vector (η, β_0, σ) is identified only up to a scale parameter. Therefore, without loss of generality, a researcher would have to fix the value of one of these three scalars. We assume for simplicity that the researcher knows the true value of η; i.e. $\eta^{-1} = 0.5$. This simplifies the comparison of the researcher's estimates of (β_0, σ) and their true values $\beta_0 = 0.5$ and $\sigma = 1$.

expectations $\mathbb{E}[r_i^o|\mathcal{J}_i]$. As equation (D.9) shows, altering the distribution of x_{2i} is equivalent to altering the distribution of the measurement error in exporters' expectations, ξ_i. Specifically, from equations (D.9) and (D.12), as we increase the variance of x_{2i} we are increasing both the variance of the measurement error ξ_i and its covariance with the measured expected export revenue, r_i^o. Therefore, Table D.1 shows how the distribution of the expectational error and agents' true expectations affect the estimates obtained by a researcher when she assumes perfect foresight.

Table D.1: Estimates under Perfect Foresight

Model	Distribution of x_{1i}	Distribution of x_{2i}	θ_0	θ_2
1	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.25)$	0.6565 (0.0013)	1.3470 (0.0014)
2	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.5)$	0.7468 (0.0012)	1.5571 (0.0013)
3	$\mathbb{N}(0,1)$	$\mathbb{N}(0,1)$	1.1205 (0.0009)	2.3866 (0.0013)
4	t_2	t_2	1.7052 (0.0005)	4.0000 (0.0013)
5	t_5	t_5	1.1584 (0.0008)	2.5381 (0.0014)
6	t_{20}	t_{20}	1.1237 (0.0009)	2.4108 (0.0013)
7	t_{50}	t_{50}	1.1289 (0.0009)	2.4096 (0.0013)
8	$\textit{log-normal}(0,1)$	$\textit{log-normal}(0,1)$	1.8737 (0.0005)	3.4698 (0.0014)
9	$-\textit{log-normal}(0,1)$	$-\textit{log-normal}(0,1)$	1.4872 (0.0006)	4.4287 (0.0013)

Notes: All estimates in this table are normalized by scale by setting $\eta^{-1} = 0.5$. In order to estimate each of the models, we generate 1,000,000 observations from the distributions $\nu_i \sim \mathbb{N}(0,1)$, $x_{3i} \sim \mathbb{N}(0,0.5)$ and from the distributions of x_{1i}, and x_{2i} described in columns 2 and 3. Whenever draws are generated from the log-normal distribution, we re-center them at zero. For each of the nine cases considered, the difference between the values of the true parameter vector $(\beta_0,\sigma) = (0.5,1)$ and those reported in columns 4 and 5 show the asymptotic bias generated by the perfect foresight assumption.

The first three rows in Table D.1 are specific examples of the general model studied by Yatchew and Griliches (1985). They consider a statistical model that encompasses our model when both the true exporters' expectation, $\mathbb{E}[r_i^o|\mathcal{J}_i]$, and the exporters' expectational error, ξ_i, are normally distributed. Our simulation results show that the researcher's ML estimates of the unknown parameter vector, (θ_0,θ_2), converge to values that are larger than the true value of the parameter vector $(\beta_0,\sigma) = (0.5,1)$, and the bias is larger as we increase the variance of the exporters' expectational errors. The sign and magnitude of the bias we find is consistent with the analytical formula for the bias in Yatchew and Griliches (1985).

In rows 4 to 10, we explore settings that no longer correspond to the model studied in Yatchew and Griliches (1985). Specifically, we depart from the assumption that both the unobserved firm's expectation and the expectational error are normally distributed. In rows 4 to 7, we choose a distribution with fatter tails than the normal: the student t distribution. The upward bias in the estimates persists and is larger the higher the dispersion in the student t distribution. In rows 8 and 9, we choose distributions that are asymmetric: log-normal distributions re-centered at zero. Specifically, in row 8 we use distributions that are positively skewed, and in row 9 we use distributions that are negatively skewed. In all cases, θ_0 and θ_2 are larger than β_0 and σ, respectively.

D.4 Bias when the Researcher's Information Set is Too Large

We consider here the bias that affects the ML estimates in those cases in which a researcher does not assume perfect foresight but still assumes that exporters have an information set that is strictly larger than their true information set. Specifically, we simulate a model in which, for every firm i, $\mathcal{J}_i = x_{1i}$ and $\mathcal{J}_i^a = (x_{1i}, x_{2i})$. Therefore,

$$\mathbb{E}[r_i^o|\mathcal{J}_i] = x_{1i} \qquad \text{and} \qquad \mathbb{E}[r_i^o|\mathcal{J}_i^a] = x_{1i} + x_{2i}, \tag{D.14}$$

and, from equation (13), the measurement error introduced by the misspecification of agents' expectations is thus

$$\xi_i = \mathbb{E}[r_i^o|\mathcal{J}_i^a] - \mathbb{E}[r_i^o|\mathcal{J}_i] = (x_{1i} + x_{2i}) - x_{1i} = x_{2i}. \tag{D.15}$$

Given equations (D.5), (D.6), and (D.14), the researcher will estimate the parameter vector (β_0, σ) finding the values of (θ_0, θ_2) that maximize the following log-likelihood function

$$L_a(\theta|d, \mathcal{J}^a) =$$
$$\sum_{i=1}^{N} \Big\{ d_i \ln(\Phi(\theta_2^{-1}(\eta^{-1}(x_{1i} + x_{2i}) - \theta_0))) + (1 - d_i) \ln(1 - \Phi(\theta_2^{-1}(\eta^{-1}(x_{1i} + x_{2i}) - \theta_0))) \Big\}. \tag{D.16}$$

However, the correct log-likelihood function is:

$$L(\theta|d, \mathcal{J}) =$$
$$\sum_{i=1}^{N} \Big\{ d_i \ln\Big(\int_{\nu+\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}(x_{1i} + x_{2i}) - \theta_0 - (\nu_i + \eta^{-1}x_{2i}) \geq 0\} f(\nu_i + \eta^{-1}x_{2i}|x_{1i} + x_{2i}; \theta_2) \Big)$$
$$+ (1 - d_i) \ln\Big(\int_{\nu+\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}(x_{1i} + x_{2i}) - \theta_0 - (\nu_i + \eta^{-1}x_{2i}) < 0\} f(\nu_i + \eta^{-1}x_{2i}|x_{1i} + x_{2i}; \theta_2) \Big) \Big\}. \tag{D.17}$$

The conditional densities $f(\nu_i|x_{1i} + x_{2i}; \theta_2)$ and $f(\nu_i + \eta^{-1}x_{2i}|x_{1i} + x_{2i}; \theta_2)$ differ in the same dimensions in which they differ in the perfect foresight case. First, while ν_i is independent of $x_{1i} + x_{2i}$, $\nu_i + \eta^{-1}x_{2i}$ is not. Second, the variance of $\chi_i = \nu_i + \eta^{-1}x_{2i}$ will also be larger than the variance of ν_i:

$$var(\chi_i) = \sigma^2 + \eta^{-2} var(x_{2i}). \tag{D.18}$$

Additionally, if x_{2i} is not normally distributed, then the shape of the density function $f(\nu_i|x_{1i} + x_{2i}; \theta_2)$ will also differ from that of $f(\nu_i + \eta^{-1}x_{2i}|x_{1i} + x_{2i}; \theta_2)$. We should expect the bias of the ML estimator to be smaller here than in the perfect foresight case: both $cov(\chi_i, \mathbb{E}[r_i^o|\mathcal{J}_i^a])$ as well as $var(\chi_i)/\sigma^2$ are smaller.

By definition, the value of the parameter (θ_0, θ_2) that maximizes the log-likelihood function in equation (D.17) is equal to the true parameter vector $(\beta_0, \sigma^2) = (0.5, 1)$. In Table D.2, for different distributions of x_{1i} and x_{2i}, we show the point estimates and standard errors for the parameter vector (θ_0, θ_2) that maximizes the researcher's likelihood function in equation (D.16). As in Table D.1, the distribution of x_{1i} is identical to the distribution of the true unobserved expectations $\mathbb{E}[r_i^o|\mathcal{J}_i]$. From equation (D.15), the distribution of x_{2i} is now identical to the distribution of the measurement error in exporters' expectations, ξ_i. Therefore, as we increase the variance of x_{2i} we are increasing both the variance of the measurement error ξ_i and its covariance with the researcher's assumed proxy of firms' expectations, $\mathbb{E}[r_i^o|\mathcal{J}_i^a]$.

Comparing the results in Tables D.1 and D.2, the biases have the same sign but are smaller in absolute value in this case than in the perfect foresight case. While in the perfect foresight case the researcher wrongly assumes that both x_{2i} and x_{3i} are in the information set of the exporter, here we wrongly assume only that variable x_{2i} is in the information set.

D.5 Bias when the Researcher's Information Set is Too Small

Here we consider the case in which the researcher assumes an information set for exporters that is strictly smaller than the firm's true information set. Specifically, the researcher assumes that only x_{1i} is in the exporters' information set when the true information set includes both x_{1i} and x_{2i}; i.e. $\mathcal{J}_i^a = x_{1i}$ and $\mathcal{J}_i = (x_{1i}, x_{2i})$. This implies that

$$\mathbb{E}[r_i^o|\mathcal{J}_i] = x_{1i} + x_{2i} \qquad \text{and} \qquad \mathbb{E}[r_i^o|\mathcal{J}_i^a] = x_{1i}, \tag{D.19}$$

Table D.2: Estimates when Information Set is Too Large

Model	Distribution of x_{1i}	Distribution of x_{2i}	θ_0	θ_2
1	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.25)$	0.5308 (0.0014)	1.0668 (0.0014)
2	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.5)$	0.6243 (0.0013)	1.2817 (0.0014)
3	$\mathbb{N}(0,1)$	$\mathbb{N}(0,1)$	1.0068 (0.0009)	2.1395 (0.0013)
4	t_2	t_2	1.6167 (0.0005)	3.7651 (0.0013)
5	t_5	t_5	1.0736 (0.0008)	2.3364 (0.0013)
6	t_{20}	t_{20}	1.0105 (0.0009)	2.1487 (0.0013)
7	t_{50}	t_{50}	0.9935 (0.0009)	2.1061 (0.0013)
8	$log\text{-}normal(0,1)$	$log\text{-}normal(0,1)$	1.8189 (0.0005)	3.3602 (0.0014)
9	$-log\text{-}normal(0,1)$	$-log\text{-}normal(0,1)$	1.4204 (0.0006)	4.1876 (0.0013)

Notes: All estimates in this table are normalized by scale by setting $\eta^{-1} = 0.5$. In order to estimate each of the models, we generate 1,000,000 observations from the distributions $\nu_i \sim \mathbb{N}(0,1)$, $x_{3i} \sim \mathbb{N}(0,0.5)$ and from the distributions of x_{1i}, and x_{2i} described in columns 2 and 3. Whenever draws are generated from the log-normal distribution, we re-center them at zero. For each of the nine cases considered, the difference between the values of the true parameter vector $(\beta_0, \sigma) = (0.5, 1)$ and those reported in columns 4 and 5 show the asymptotic bias generated of the corresponding ML estimates.

and, therefore, the measurement error introduced by the misspecification of agents' expectations is

$$\xi_i = \mathbb{E}[r_i^o|\mathcal{J}_i^a] - \mathbb{E}[r_i^o|\mathcal{J}_i] = x_{1i} - (x_{1i} + x_{2i}) = -x_{2i}. \tag{D.20}$$

Given equations (D.5), (D.6), and (D.19), the researcher will estimate the parameter vector (β_0, σ) finding the values of (θ_0, θ_2) that maximize the following log-likelihood function

$$L_a(\theta|d, \mathcal{J}^a) = \sum_{i=1}^{N} \Big\{ d_i \ln(\Phi(\theta_2^{-1}(\eta^{-1}x_{1i} - \theta_0))) + (1 - d_i)\ln(1 - \Phi(\theta_2^{-1}(\eta^{-1}x_{1i} - \theta_0))) \Big\}. \tag{D.21}$$

However, the correct log-likelihood function is:

$$\begin{aligned} L(\theta|d, \mathcal{J}) = \\ \sum_{i=1}^{N} \Big\{ d_i \int_{\nu-\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}x_{1i} - \theta_0 - (\nu_i - \eta^{-1}x_{2i}) \geq 0\} f(\nu_i - \eta^{-1}x_{2i}|x_{1i}; \theta_2) \\ +(1 - d_i) \int_{\nu-\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}x_{1i} - \theta_0 - (\nu_i - \eta^{-1}x_{2i}) < 0\} f(\nu_i - \eta^{-1}x_{2i}|x_{1i}; \theta_2) \Big\} = \\ \sum_{i=1}^{N} \Big\{ d_i \int_{\nu-\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}x_{1i} - \theta_0 - (\nu_i - \eta^{-1}x_{2i}) \geq 0\} f(\nu_i - \eta^{-1}x_{2i}; \theta_2) \\ +(1 - d_i) \int_{\nu-\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}x_{1i} - \theta_0 - (\nu_i - \eta^{-1}x_{2i}) < 0\} f(\nu_i - \eta^{-1}x_{2i}; \theta_2) \Big\}, \end{aligned} \tag{D.22}$$

where the second equality applies the property that, in this case, the measurement error $\xi_i = -x_{2i}$ is independent of the information set assumed by the researcher $\mathcal{J}_i^a = x_{1i}$.

The biggest difference between this case and that considered in Sections D.3 and D.4 is that now the measurement error ξ_i is guaranteed to be mean independent of the researcher's measure of exporter i's expectation, $\mathbb{E}[r_i^o|\mathcal{J}_i^a]$. Specifically,

$$cov(\xi_i, \mathbb{E}[r_i^o|\mathcal{J}_i^a]) = cov(\xi_i, x_{1i}) = cov(-x_{2i}, x_{1i}) = 0. \tag{D.23}$$

Our simulation represents a very special case in which ξ_i is not only mean independent of $\mathbb{E}[r_i^o|\mathcal{J}_i^a]$ but also fully independent of $\mathbb{E}[r_i^o|\mathcal{J}_i^a]$ such that the shape of the distribution of ν_i is the same as the shape of the distribution of $\nu_i + \eta^{-1}\xi_i$ (i.e. both are normal). In this case, the functional form of the likelihood function in equation (D.21) is the same as that in (D.22). Therefore, the values of (θ_0, θ_2) that maximize the log-likelihood function specified by the researcher are:

$$(\theta_0, \theta_2) = (\beta_0, \sqrt{\sigma^2 + \eta^{-2} var(x_{2i})}).$$

In this special case, the ML estimate of θ_0 the researcher recovers is asymptotically unbiased for the parameter β_0; only the ML estimator of the variance of ν is biased upwards.

Outside this special case, if ξ_i is only mean independent of $\mathbb{E}[r_i^o|\mathcal{J}_i^a]$ (but not fully independent) or if the distribution of ξ_i is such that the distributions of the random variables ν_i and $\chi_i \equiv \nu_i - \eta^{-1}\xi_i$ do not belong to the same family, then the ML estimate of β_0 will also be biased. We illustrate these cases in Table D.3. The results in Table D.3 show that, if the distribution of x_{2i} is symmetric, then the ML estimates of β_0 are always approximately unbiased and those of σ are always upward biased. In those cases in which the distribution of ξ_i is not symmetric, the estimate of β_0 also becomes asymptotically biased.

Table D.3: Estimates when Information Set is Too Small

Model	Distribution of x_{1i}	Distribution of x_{2i}	θ_0	θ_2
1	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.25)$	0.5027 (0.0015)	1.0079 (0.0014)
2	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.5)$	0.5021 (0.0015)	1.0309 (0.0014)
3	$\mathbb{N}(0,1)$	$\mathbb{N}(0,1)$	0.5012 (0.0014)	1.1181 (0.0014)
4	t_2	t_2	0.5153 (0.0011)	1.3228 (0.0014)
5	t_5	t_5	0.5014 (0.0013)	1.1701 (0.0014)
6	t_{20}	t_{20}	0.5012 (0.0014)	1.1271 (0.0014)
7	t_{50}	t_{50}	0.4988 (0.0014)	1.1191 (0.0014)
8	$log\text{-}normal(0,1)$	$log\text{-}normal(0,1)$	0.6092 (0.0011)	1.2370 (0.0014)
9	$-log\text{-}normal(0,1)$	$-log\text{-}normal(0,1)$	0.3689 (0.0014)	1.1387 (0.0015)

Notes: All estimates in this table are normalized by scale by setting $\eta^{-1} = 0.5$. In order to estimate each of the models, we generate 1,000,000 observations from the distributions $\nu_i \sim \mathbb{N}(0,1)$, $x_{3i} \sim \mathbb{N}(0,0.5)$ and from the distributions of x_{1i}, and x_{2i} described in columns 2 and 3. Whenever draws are generated from the log-normal distribution, we re-center them at zero. For each of the nine cases considered, the difference between the values of the true parameter vector $(\beta_0, \sigma) = (0.5, 1)$ and those reported in columns 4 and 5 show the asymptotic bias of the corresponding ML estimates.

The results reported in rows 1 to 7 of Tables D.2 and D.3 are very different. In our simulation, the bias is much larger when the information set assumed by the researcher is too large than when it is too small. The key difference is that the measurement error in firms' expectations is more likely correlated with the assumed proxy for these expectations when the assumed information set is too large, making the estimates subject to classical measurement error. When the assumed information set is too small, the measurement error ξ_i is correlated with the true expectations but uncorrelated with the measured ones. The measure of the agents' true expectations is in this case exogenous and, thus, the bias in the parameter estimates becomes less severe.

E Relationship between ML Estimates and Confidence Sets

In Section 4.1 and Appendix D, we show that the probability limit of the maximum likelihood estimator of θ equals the true parameter vector, θ^*, if and only if for every agent, the expected export revenue implied by the researcher's information set equals the agent's true expectation; i.e. $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$ for every i, j, and t. In Section 4.2 and Appendix C, we show that the identified set defined by the odds-based and revealed-preference moment inequalities, Θ_0, includes the true value of the parameter vector, θ^*, if, for every agent, the set of variables Z_{ijt} included by the researcher in the conditioning set of the moment inequalities belong to the agent's true information set; i.e. $Z_{ijt} \subseteq \mathcal{J}_{ijt}$ for all i, j, and t.

However, if $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] \neq \mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$ for a nonzero measure of individuals in the population, then neither Section 4.1 nor Section 4.2 indicate whether the probability limit of the corresponding maximum likelihood estimator will be contained in the identified set defined by our moment inequalities that condition on Z_{ijt}. In this appendix section, we examine this relationship. In particular, we show that, for any given information sets $\{\mathcal{J}^a_{ijt}; \forall i, j, t\}$, the probability limit of the maximum likelihood estimator constructed under the assumption that $\mathcal{J}_{ijt} = \mathcal{J}^a_{ijt}$ for all i, j, and t, need not be included in the identified set defined by our moment inequalities constructed under the assumption that $\mathcal{J}^a_{ijt} \subseteq \mathcal{J}_{ijt}$ for all i, j, and t.

To illustrate the properties of the data generating process that determine whether the probability limit of the maximum likelihood estimator is included in the identified set defined by our moment inequalities, we compute these objects for several different data generating processes. We use a simplified version of the model described in Section 2, as used also in Appendix D.2, to illustrate the direction and magnitude of the bias of several maximum likelihood estimators.

E.1 Simulated Model and Estimators

We model the decision process of $N = 1,000,000$ potential exporters $i = 1, \ldots, N$ who decide at a single period t whether to export to a single market j. We therefore omit the subindices j and t here. Each firm i decides whether to export according to the decision rule

$$d_i = \mathbb{1}\{\eta^{-1}\mathbb{E}[r^o_i|\mathcal{J}_i] - \beta_0 - \nu_i\}, \tag{E.1}$$

where $\beta_0 + \nu_i$ denotes i's fixed costs of exporting. We fix $\eta^{-1} = \beta_0 = 0.5$ and simulate the vector $(\nu_1, \ldots, \nu_N)$ by taking independent random draws from a normal distribution with mean zero and variance $\sigma^2 = 1$: $\nu_i \sim \mathbb{N}(0, 1)$. We assume that $r^o_i = x_{1i} + x_{2i} + x_{3i}$ and we fix the true information set of every agent i to be $\mathcal{J}_i = \{x_{1i}, x_{2i}\}$.

The data generating processes that we use to simulate data differ only in the joint distribution of x_{1i}, x_{2i}, and x_{3i}. Specifically, in all of our simulations, we assume: (a) x_{1i}, x_{2i}, and x_{3i} are distributed independently across individuals i; (b) for each individual i, x_{1i}, x_{2i}, and x_{3i} are independent of each other; (c) the distributions of x_{1i}, x_{2i}, and x_{3i} belong to the same family (normal or log-normal); (d) $\mathbb{E}[x_{1i}] = \mathbb{E}[x_{2i}] = \mathbb{E}[x_{3i}] = 0$. The simulations differ in the particular shape of the distribution of x_{1i}, x_{2i}, and x_{3i} and the variance of each of these three random variables.

As in Section 2, we consider the estimation problem of a researcher that knows that: (a) the export participation decision is determined by equation (E.1); (b) $\eta^{-1} = 0.5$; (c) $(\nu_1, \ldots, \nu_N)$ are independently normally distributed with mean zero and common variance σ^2; and (d) the potential export revenues are determined by the equation $r^o_i = x_{1i} + x_{2i} + x_{3i}$. We also assume that the researcher observes the value of the vector $(d_i, x_{1i}, x_{2i}, x_{3i})$ for every potential exporter i. The researcher's goal is to estimate the parameter vector (β_0, σ) and must do so without knowing which of the three variables x_{1i}, x_{2i}, and x_{3i} are included in the information set of each potential exporter i.

We compute maximum likelihood estimators of (β_0, σ) under three different assumptions on the content of the firm's information set: (a) $\mathcal{J}_i = \mathcal{J}^s_i$ with $\mathcal{J}^s_i \equiv x_{1i}$ for every i; (b) $\mathcal{J}_i = \mathcal{J}^m_i$ with $\mathcal{J}^m_i \equiv \{x_{1i}, x_{2i}\}$ for every i; (c) $\mathcal{J}_i = \mathcal{J}^l_i$ with $\mathcal{J}^l_i \equiv \{x_{1i}, x_{2i}, x_{3i}\}$ for every i. We denote the corresponding maximum likelihood estimators as $\hat{\theta}_s$, $\hat{\theta}_m$, and $\hat{\theta}_l$. As indicated above, the information set $\mathcal{J}^m_i$ coincides with the true information set of every firm i, while $\mathcal{J}^s_i$ is too small and $\mathcal{J}^l_i$ is too large. Therefore, only assumption (b) is compatible with the data generating process and, consequently, only $\hat{\theta}_m$ is a consistent estimator of the true parameter θ^*.

We also compute moment inequality bounds on (β_0, σ) under three different assumptions on the content of the firm's information set: (a) $Z^s_i \subseteq \mathcal{J}_i$ with $Z^s_i \equiv x_{1i}$ for every i; (b) $Z^m_i \subseteq \mathcal{J}_i$ with $Z^m_i \equiv \{x_{1i}, x_{2i}\}$ for every i; (c) $Z^l_i \subseteq \mathcal{J}_i$ with $Z^l_i \equiv \{x_{1i}, x_{2i}, x_{3i}\}$ for every i. We denote the corresponding 95% confidence sets for the true parameter as $\hat{\Theta}^{95\%}_s$, $\hat{\Theta}^{95\%}_m$, and $\hat{\Theta}^{95\%}_l$. The sets of observed covariates in Z^s_i and Z^m_i are included in the true information set, while the set in Z^l_i is not. Therefore, assumptions (a) and (b) are compatible with the data generating process and, consequently, the confidence sets $\hat{\Theta}_s^{95\%}$ and $\hat{\Theta}_m^{95\%}$ will contain the true value of the parameter vector, θ^*, with at least 95% probability.

E.2 Results

In Table E.1, we report the value of the maximum likelihood estimates $\hat{\theta}_s$, $\hat{\theta}_m$, and $\hat{\theta}_l$ and the confidence sets $\hat{\Theta}_s^{95\%}$, $\hat{\Theta}_m^{95\%}$, and $\hat{\Theta}_l^{95\%}$ for different assumptions on the marginal distributions of x_{1i}, x_{2i}, and x_{3i}. We consider two sets of cases. First, we assume x_{1i}, x_{2i}, and x_{3i} are normally distributed (panel A in Table E.1). Second, we assume that these three variables are log-normally distributed re-centered at zero (panel B in Table E.1). For each of these two sets, we consider six different cases. In case 1, our benchmark case, the variances of x_{1i}, x_{2i}, and x_{3i} are set to one. In cases 2 and 3, we change the variance of x_{2i} with respect to the benchmark case. In cases 4 and 5, we change the variance of x_{3i}. In case 6, we simultaneously change the variance of x_{2i} and the variance of x_{3i}. The results in panels A and B are very similar and, thus, we discuss them jointly.[49]

Discussion of maximum likelihood estimates. As discussed in detail in Appendix D, the maximum likelihood estimator that imposes the correct assumption on the agent's information set, $\hat{\theta}_m$, is a consistent estimator of θ^*. The maximum likelihood estimator that assumes an information set that is too small, $\hat{\theta}_s$, is consistent as an estimator of β_0 but biased upwards as an estimator of σ; the upward bias increases in the variance of the variable that the researcher has mistakenly left out of the information set of the agent. We see this result when comparing cases 1, 2 and 3 in Table E.1. Finally, the maximum likelihood estimator that assumes an information set that is too large, $\hat{\theta}_l$, is biased upwards as an estimator of both β_0 and σ, and the upward bias increases in the variance of variable that the researcher has mistakenly included in the information set of the agent. We can see this result by comparing cases 1 to 5 in Table E.1.

Discussion of moment inequality confidence sets. As we discuss in Section 4.2, the confidence set that correctly assumes that a set of variables belongs to the agent's information set, $\hat{\Theta}_s^{95\%}$ and $\hat{\Theta}_m^{95\%}$, will contain θ^* with the correct probability.

Table E.1 also illustrates that the set of values of θ contained in the moment inequality confidence set gets smaller the smaller is the variance of the variables in the agent's true information set and not exploited in the moment inequalities; i.e. the smaller is the variance of the variables in $\mathcal{J}_i$ and not included in Z_i. This is why $\hat{\Theta}_m^{95\%}$ is strictly smaller than $\hat{\Theta}_s^{95\%}$ for all six cases considered in Table E.1, and why the set of points included in $\hat{\Theta}_s^{95\%}$ increases as the variance of x_{2i} increases (see cases 1 to 3 in Table E.1).

Table E.1 also shows that including variables in Z_i that do not belong to the actual information set of the agent, $\mathcal{J}_i$, may generate a confidence set that is empty; i.e. $\hat{\Theta}_l^{95\%}$ is empty for all cases considered in Table E.1. This is important, as it illustrates that the tests described in Section 6 may indeed be able to detect that certain covariates do not belong to the information set of the potential exporters.

Finally, Table E.1 illustrates that the size of the moment inequality confidence sets increases in the variance of the true expectational error of the agents. In the model described in Appendix E.1, the agent's expectational error corresponds to the variable x_{3i} and, by comparing cases 1, 4, 5, and 6 in Table E.1, we observe that the sets of points included in the confidence sets $\hat{\Theta}_s^{95\%}$ and $\hat{\Theta}_m^{95\%}$ are larger, the larger is the variance of x_{3i}.

Relationship between maximum likelihood estimates and moment inequality confidence sets. The last four columns in Table E.1 illustrate how the maximum likelihood estimates relate to the confidence sets.

The column with the heading "$\hat{\theta}_s \in \hat{\Theta}_s^{95\%}$" shows that, although both $\hat{\theta}_s$ and $\hat{\Theta}_s^{95\%}$ are computed only using information on x_{1i} as a determinant of the firm's expectations of r_i^o, there are some data generating processes for which the estimate $\hat{\theta}_s$ is not included in the set $\hat{\Theta}_s^{95\%}$. Specifically, cases 3, 4 and 6 in Table E.1 indicate that this is more likely to happen when the variance of the content of the firm's true information set not accounted for by the researcher, x_{2i}, is large (i.e. which both increases the bias in $\hat{\theta}_s$ and the size of $\hat{\Theta}_s^{95\%}$) and when the variance of the firm's true expectational error, x_{3i}, is small (i.e. which does not affect the bias in $\hat{\theta}_s$ but reduces the size of $\hat{\Theta}_s^{95\%}$). Consequently, among the cases considered in Table E.1, case 6 is the case in which it is more likely that $\hat{\theta}_s$ is not included in $\hat{\Theta}_s^{95\%}$.

A comparison of cases 2 and 3 is informative about the role played by the variance of x_{2i}. In case 2, the variance of the variable x_{2i} is too small and, thus, the misspecification of the firm's information set is not

[49] In a robustness analysis, we have confirmed that the same patterns illustrated in panels A and B in Table E.1 arise when we assume x_{1i}, x_{2i}, and x_{3i} are uniformly distributed.

Table E.1: Simulation Results

Case	σ_1^2	σ_2^2	σ_3^2	θ^*	$\hat{\theta}_s$	$\hat{\theta}_m$	$\hat{\theta}_l$	$\hat{\Theta}_s^{95\%}$	$\hat{\Theta}_m^{95\%}$	$\hat{\Theta}_l^{95\%}$	$\hat{\theta}_s \in \hat{\Theta}_s^{95\%}$	$\hat{\theta}_l \in \hat{\Theta}_s^{95\%}$	$\hat{\theta}_s \in \hat{\Theta}_m^{95\%}$	$\hat{\theta}_l \in \hat{\Theta}_m^{95\%}$
Panel A: Determinants of Export Revenue Are Distributed Normally														
1	1	1	1	0.5 1	0.50 (0.003) 1.11 (0.005)	0.50 (0.002) 0.99 (0.003)	0.75 (0.003) 1.61 (0.006)	[0.14, 0.60] [0.56, 1.19]	[0.19, 0.60] [0.57, 1.19]	$\emptyset$	Yes	No	Yes	No
2	1	0.25	1	0.5 1	0.50 (0.002) 1.02 (0.004)	0.50 (0.002) 1.00 (0.004)	0.89 (0.005) 1.91 (0.010)	[0.20, 0.60] [0.66, 1.19]	[0.20, 0.60] [0.66, 1.19]	$\emptyset$	Yes	No	Yes	No
3	1	4	1	0.5 1	0.50 (0.003) 1.40 (0.008)	0.50 (0.002) 1.00 (0.002)	0.60 (0.003) 1.32 (0.004)	[0.08, 0.60] [0.45, 1.19]	[0.28, 0.56] [0.81, 1.14]	$\emptyset$	No	No	No	No
4	1	1	0.25	0.5 1	0.50 (0.003) 1.11 (0.005)	0.50 (0.002) 0.99 (0.003)	0.56 (0.003) 1.15 (0.005)	[0.28, 0.52] [0.64, 1.04]	[0.43, 0.52] [0.89, 1.04]	$\emptyset$	No	No	No	No
5	1	1	4	0.5 1	0.50 (0.003) 1.11 (0.005)	0.50 (0.002) 0.99 (0.003)	1.49 (0.010) 3.44 (0.019)	[0.11, 0.81] [0.45, 1.66]	[0.12, 0.78] [0.57, 1.19]	$\emptyset$	Yes	No	Yes	No
6	1	4	0.25	0.5 1	0.50 (0.004) 1.40 (0.008)	0.50 (0.002) 1.00 (0.003)	0.53 (0.002) 1.08 (0.003)	[0.08, 0.52] [0.45, 1.04]	[0.45, 0.52] [0.96, 1.04]	$\emptyset$	No	No	No	No
Panel B: Determinants of Export Revenue Are Distributed Log-Normally (Re-centered at 0)														
1	1	1	1	0.5 1	0.53 (0.004) 1.07 (0.006)	0.50 (0.003) 1.00 (0.004)	0.95 (0.006) 1.85 (0.009)	[0.39, 0.63] [0.86, 1.25]	[0.40, 0.63] [0.86, 1.25]	$\emptyset$	Yes	No	Yes	No
2	1	0.25	1	0.5 1	0.51 (0.003) 1.02 (0.005)	0.50 (0.003) 1.00 (0.005)	1.13 (0.008) 2.24 (0.014)	[0.40, 0.63] [0.86, 1.25]	[0.39, 0.66] [0.86, 1.32]	$\emptyset$	Yes	No	Yes	No
3	1	4	1	0.5 1	0.60 (0.004) 1.19 (0.007)	0.50 (0.003) 1.00 (0.003)	0.79 (0.004) 1.50 (0.005)	[0.37, 0.59] [0.73, 1.19]	[0.39, 0.57] [0.93, 1.14]	$\emptyset$	No	No	No	No
4	1	1	0.25	0.5 1	0.53 (0.004) 1.07 (0.006)	0.50 (0.003) 1.00 (0.004)	0.63 (0.004) 1.23 (0.005)	[0.46, 0.54] [0.93, 1.09]	[0.46, 0.54] [0.96, 1.09]	$\emptyset$	Yes	No	Yes	No
5	1	1	4	0.5 1	0.53 (0.004) 1.07 (0.006)	0.50 (0.003) 1.00 (0.004)	2.07 (0.016) 4.10 (0.029)	[0.22, 0.78] [0.58, 1.57]	[0.22, 0.78] [0.60, 1.56]	$\emptyset$	Yes	No	Yes	No
6	1	4	0.25	0.5 1	0.60 (0.007) 1.19 (0.004)	0.50 (0.003) 1.00 (0.003)	0.58 (0.003) 1.14 (0.004)	[0.44, 0.52] [0.89, 1.04]	[0.47, 0.52] [0.96, 1.04]	$\emptyset$	No	No	No	No

Notes: The columns σ_1^2, σ_2^2 and σ_3^2 indicate the variance of x_{1i}, x_{2i} and x_{3i}, respectively. The column θ^* indicates the true value of the two parameters being estimated. See Appendix E.1 for a precise definition of the maximum likelihood estimators $\hat{\theta}_s$, $\hat{\theta}_m$ and $\hat{\theta}_l$, and of the confidence sets $\hat{\Theta}_s^{95\%}$, $\hat{\Theta}_m^{95\%}$ and $\hat{\Theta}_l^{95\%}$. The terms in parenthesis in the columns with headings $\hat{\theta}_s$, $\hat{\theta}_m$ and $\hat{\theta}_l$ are the standard errors of the corresponding maximum likelihood estimators. The last four columns indicate whether one of the three maximum likelihood estimates reported in columns 6 to 8 is included in one of the confidence sets reported in columns 9 to 11. For example, the column with heading $\hat{\theta}_s \in \hat{\Theta}_s^{95\%}$ indicates whether the maximum likelihood estimate $\hat{\theta}_s$ (reported in column 6) is included in the confidence set $\hat{\Theta}_s^{95\%}$ (reported in column 9) on all dimensions of θ.

very relevant. Consequently, the bias in the maximum likelihood estimator is not large enough to generate an estimate outside of the confidence set $\hat{\Theta}_s^{95\%}$.

A comparison of cases 4 and 5 is informative about the role played by the variance of x_{3i}. The maximum likelihood estimate $\hat{\theta}_s$ is identical in both cases; however, in case 5, the variance of the expectational error is too large, making the confidence set $\hat{\Theta}_s^{95\%}$ large enough to include $\hat{\theta}_s$.

As $\hat{\Theta}_m^{95\%} \subseteq \hat{\Theta}_s^{95\%}$ always, $\hat{\theta}_s \notin \hat{\Theta}_s^{95\%}$ implies $\hat{\theta}_s \notin \hat{\Theta}_m^{95\%}$. Thus, whenever a "No" appears in the column with the heading "$\hat{\theta}_s \in \hat{\Theta}_s^{95\%}$", a "No" also appears in the column with the heading "$\hat{\theta}_s \in \hat{\Theta}_m^{95\%}$".

The fact that $\hat{\theta}_s$ is not included in $\hat{\Theta}_s^{95\%}$ or $\hat{\Theta}_m^{95\%}$ for certain data generating processes has two important implications. First, a confidence set does not necessarily include any maximum likelihood estimate that one could generate by assuming as true an information set compatible with the assumption used to construct that confidence set. Restating this idea more formally, a confidence set constructed using moment inequalities that condition on a vector of covariates Z_i such that $Z_i \subseteq \mathcal{J}_i$ does not necessarily include any maximum likelihood estimate constructed using the assumption that a set $\mathcal{J}_i^a$ such that $Z_i \subseteq \mathcal{J}_i^a$ is the actual information set of the agent. Importantly, this is true when $\mathcal{J}_i^a \subseteq \mathcal{J}_i$ as well as when $\mathcal{J}_i \subseteq \mathcal{J}_i^a$. Second, by comparing the maximum likelihood estimate computed under the assumption that a set of observed covariates $\mathcal{J}_i^a$ equals the true information set of the agent to a moment inequality confidence set computed under the assumption that $Z_i \subseteq \mathcal{J}_i$, one may be able to design a test of validity of information sets different from that employed in Section 6. This potential alternative test would have the advantage that it could be used not only to test whether a certain vector of covariates is included in the firm's true information set, but also to test whether a certain vector of covariates coincides exactly with the firm's true information set. As far as we know, such a test has not been formalized in the literature yet. It is also beyond the scope of this paper.

The columns with the headings "$\hat{\theta}_l \in \hat{\Theta}_s^{95\%}$" and "$\hat{\theta}_l \in \hat{\Theta}_m^{95\%}$" show that the maximum likelihood estimate $\hat{\theta}_l$ is out of the confidence sets $\hat{\Theta}_s^{95\%}$ and $\hat{\Theta}_m^{95\%}$ in all data generating processes considered in Table E.1. A comparison of the content of these two columns to that under the headings "$\hat{\theta}_s \in \hat{\Theta}_s^{95\%}$" and "$\hat{\theta}_s \in \hat{\Theta}_m^{95\%}$" reflects the degree of bias of the maximum likelihood estimator. The bias appears larger when the assumed information is too large rather than too small. Appendix D provides additional details on this comparison.

F Sunk Costs of Exporting and Forward-Looking Firms

F.1 Introduction to Dynamic Model

We extend our benchmark model in Section 2 to allow firms to account for how the decision to export to j at t will affect the firm's potential profits from exporting to j in subsequent periods, $\{\pi_{ijt'}\}_{t+1}^{\infty}$. In this dynamic model, we will recover both the firm's fixed costs of exporting and sunk costs of exporting. We show how to compute both odds-based and revealed-preference moment inequalities that identify the parameter vector $(\beta_0, \beta_1, \gamma_0, \gamma_1, \sigma)$ under the assumptions of the dynamic model introduced in Section 8.1.

In this appendix, we will differentiate between the path of export participation choices that would be optimal in periods beyond t if firm i decides to export to country j in period t, $\{d(1_t)_{ijt'}\}_{t+1}^{\infty}$, and the path of export participation choices that would be optimal in periods beyond t if firm i decides not to export to country j in period t, $\{d(0_t)_{ijt'}\}_{t+1}^{\infty}$. We will also differentiate between the firm's optimal export participation decision at t, d_{ijt}, and the actual choice firm i makes in country j in year t, a_{ijt}.

To form the odds-based and revealed-preference moment inequalities for our dynamic model, we need to compute four objects. The first two objects are straightforward: we need the expected discounted sum of profits of firm i in market j when (a) the firm exported to j at t and then chose the optimal path from $t' > t$ given that the firm exported at t and (b) the firm did not export to j at t and then chose the optimal path from $t' > t$ given that the firm chose not to export at t. The second two objects represent 'counterfactual' objects. We need the expected discounted sum of profits of firm i in market j when (c) the firm exported to j at t and then chose the optimal path from $t' > t$ as if the firm chose not to export at t and (d) the firm did not export to j at t and then chose the optimal path from $t' > t$ as if the firm chose to export at t.

In notation, we compute the four objects as follows. First, the expected discounted sum of profits of firm i in market j conditional on choosing to export to j at t, $a_{ijt} = 1$, and choosing the optimal path in every period $t' > t$,

$$\begin{aligned} V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) &= \eta_j^{-1}\mathbb{E}\big[r^o_{ijt}\big|\mathcal{J}_{ijt}\big] - f_{ijt} - (1 - d_{ijt-1})s_{ijt} \\ &+ \rho\mathbb{E}\big[d(1_t)_{ijt+1}\big(\eta_j^{-1}r_{ijt+1} - f_{ijt+1}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1\big] \\ &+ \sum_{t'=t+2} \rho^{t'-t}\mathbb{E}\big[d(1_t)_{ijt'}\big(\eta_j^{-1}r_{ijt'} - f_{ijt'} - (1 - d(1_t)_{ijt'-1})s_{ijt'}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1\big]. \end{aligned} \tag{F.1}$$

Second, we compute the expected discounted sum of profits of firm i in market j conditional on choosing not to export to j at t, $a_{ijt} = 0$, and choosing the optimal path in every period $t' > t$ is

$$\begin{aligned} V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) &= \rho\mathbb{E}\big[d(0_t)_{ijt+1}\big(\eta_j^{-1}r_{ijt+1} - f_{ijt+1} - s_{ijt+1}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0\big] \\ &+ \sum_{t'=t+2} \rho^{t'-t}\mathbb{E}\big[d(0_t)_{ijt'}\big(\eta_j^{-1}r_{ijt'} - f_{ijt'} - (1 - d(0_t)_{ijt'-1})s_{ijt'}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0\big]. \end{aligned} \tag{F.2}$$

Third, we compute the expected discounted sum of profits of firm i in market j conditional on choosing to export to j at t, $a_{ijt} = 1$, and choosing in every period $t' > t$ the path that would have been optimal if firm i had not exported to j at period t

$$\begin{aligned} W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, a_{ijt} = 1) &= \eta_j^{-1}\mathbb{E}\big[r^o_{ijt}\big|\mathcal{J}_{ijt}\big] - f_{ijt} - (1 - d_{ijt-1})s_{ijt} \\ &+ \rho\mathbb{E}\big[d(0_t)_{ijt+1}\big(\eta_j^{-1}r_{ijt+1} - f_{ijt+1}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0\big] \\ &+ \sum_{t'=t+2} \rho^{t'-t}\mathbb{E}\big[d(0_t)_{ijt'}\big(\eta_j^{-1}r_{ijt'} - f_{ijt'} - (1 - d(0_t)_{ijt'-1})s_{ijt'}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0\big]. \end{aligned} \tag{F.3}$$

Fourth, we compute the expected discounted sum of profits of firm i in market j conditional on choosing not to export to j at t, $a_{ijt} = 0$, and choosing in every period $t' > t$ the path that would have been optimal if firm i had exported to j at period t as

$$\begin{aligned} W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, a_{ijt} = 0) &= \rho\mathbb{E}\big[d(1_t)_{ijt+1}\big(\eta_j^{-1}r_{ijt+1} - f_{ijt+1} - s_{ijt+1}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1\big] \\ &+ \sum_{t'=t+2} \rho^{t'-t}\mathbb{E}\big[d(1_t)_{ijt'}\big(\eta_j^{-1}r_{ijt'} - f_{ijt'} - (1 - d(1_t)_{ijt'-1})s_{ijt'}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1\big]. \end{aligned} \tag{F.4}$$

By definition, if the independence condition in equation (25) holds, then

$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) \geq 0, \tag{F.5a}$$
$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0. \tag{F.5b}$$

Given the functional form assumptions in equations (5) and (24), the assumption that ν_{ijt} is independent over time, and the definition of the variable $d(1_t)_{ijt'}$ for $t' > t$ as the optimal choice of firm i in country j at period t' conditional on exporting to j at t, it holds that, for any period t' larger than t,

$$\mathbb{E}[d(0_t)_{ijt'}\eta_j^{-1}r_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0] = \eta_j^{-1}\mathbb{E}[d(0_t)_{ijt'}r_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{F.6a}$$
$$\mathbb{E}[d(0_t)_{ijt'}f_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0] = (\beta_0 + \beta_1 dist_j)\mathbb{E}[d(0_t)_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{F.6b}$$
$$\mathbb{E}[d(0_t)_{ijt'}s_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0] = (\gamma_0 + \gamma_1 dist_j)\mathbb{E}[d(0_t)_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{F.6c}$$

and, analogously,

$$\mathbb{E}[d(1_t)_{ijt'}\eta_j^{-1}r_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1] = \eta_j^{-1}\mathbb{E}[d(1_t)_{ijt'}r_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{F.7a}$$
$$\mathbb{E}[d(1_t)_{ijt'}f_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1] = (\beta_0 + \beta_1 dist_j)\mathbb{E}[d(1_t)_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{F.7b}$$
$$\mathbb{E}[d(1_t)_{ijt'}s_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1] = (\gamma_0 + \gamma_1 dist_j)\mathbb{E}[d(1_t)_{ijt'}|\mathcal{J}_{ijt}, dist_j]. \tag{F.7c}$$

Equations (F.6) and (F.7) will allow us to re-express equations (F.1), (F.2), (F.3), (F.4) as a function of the parameter vector of interest $(\beta_0, \beta_1, \gamma_0, \gamma_1, \sigma)$.

Finally, using our notation, we can define the value function $V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1})$ for every firm i, country j and period t as

$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}) \equiv \max\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1), V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0)\}, \tag{F.8}$$

and we can then rewrite equation (26) as

$$d_{ijt} = \mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\}. \tag{F.9}$$

In Section F.2, we show how to combine equations (F.1) to (F.9) to derive odds-based moment inequalities in the dynamic context. In Section F.3, we derive revealed-preference inequalities consistent with the dynamic model.

F.2 Odds-Based Moment Inequalities for a Dynamic Model

The definition of the random variable d_{ijt} in equation (F.9) implies that, for every firm i, country j and period t, we can write the following two inequalities

$$d_{ijt} - \mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\} \geq 0, \tag{F.10a}$$
$$\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\} - d_{ijt} \geq 0. \tag{F.10b}$$

Equation (F.10a) exploits the fact that

$$\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\}, \tag{F.11}$$

is a sufficient condition for $d_{ijt} = 1$. Equation (F.10b) exploits the fact that the inequality inside the indicator function in equation (F.11) is a necessary condition for $d_{ijt} = 1$. We will derive an odds-based moment inequality from each of the inequalities in equation (F.10). We show first how to derive an inequality from equation (F.10a) and do the same for equation (F.10b) below.

Combining equations (F.5b) and (F.10a), we obtain:

$$d_{ijt} - \mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\} \geq 0, \tag{F.12}$$

and, from equations (5), (24), (F.1), (F.4), and (F.7c), we can write the variable inside the indicator function as:

$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) =$$

$$\eta_j^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - f_{ijt} - (1-d_{ijt-1})s_{ijt} + \rho\mathbb{E}\big[d(1_t)_{ijt+1}s_{ijt+1}\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1\big] =$$
$$\eta_j^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - (\beta_0+\beta_1 dist_j+\nu_{ijt}) - (1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j) + (\gamma_0+\gamma_1 dist_j)\rho\mathbb{E}\big[d(1_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big] =$$
$$\eta_j^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - (\beta_0+\beta_1 dist_j+\nu_{ijt}) - (1-d_{ijt-1}-\rho\mathbb{E}\big[d(1_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big])(\gamma_0+\gamma_1 dist_j). \quad \text{(F.13)}$$

For simplicity in the notation, we denote this expression as:

$$\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt},$$

with $\theta^*_D \equiv (\beta_0, \beta_1, \sigma, \gamma_0, \gamma_1)$ and, therefore,

$$\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) \equiv$$
$$\eta_j^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho\mathbb{E}\big[d(1_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big])(\gamma_0+\gamma_1 dist_j). \quad \text{(F.14)}$$

Using this expression, we can rewrite (F.12) as

$$d_{ijt} - \mathbb{1}\{\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt} \geq 0\} \geq 0, \quad \text{(F.15)}$$

or, equivalently,

$$\mathbb{1}\{\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt} \leq 0\} - (1-d_{ijt}) \geq 0. \quad \text{(F.16)}$$

Given that this inequality must hold for every firm i, country j and period t, it must also hold in expectation

$$\mathbb{E}[\mathbb{1}\{\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt} \leq 0\} - (1-d_{ijt}) \geq 0\}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \quad \text{(F.17)}$$

and, given the distributional assumption in equation (6) and the assumption that ν_{ijt} is independent over time, we can rewrite this expression as

$$\mathbb{E}(1-\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D)) - (1-d_{ijt})|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \quad \text{(F.18)}$$

Following analogous steps as those described in the proof to Lemma C.5, we can rewrite this inequality as

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D))}{\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D))} - (1-d_{ijt})\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right] \geq 0. \quad \text{(F.19)}$$

This inequality is analogous to that in equation (C.5) with two differences: (1) the lagged export status d_{ijt-1} is included in the conditioning set and (2) the term inside the function $\Phi(\cdot)$ accounts for how the export decision at t affects export profits at t and $t+1$.

The term $\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D)$ in equation (F.19) depends on the unobserved expectations

$$\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] \qquad \text{and} \qquad \mathbb{E}[d(1_t)_{ijt+1}|\mathcal{J}_{ijt}, dist_j]$$

and, therefore, the researcher would not know it even if she knew the true parameter vector θ^*_D. We define an analogous expression $\Delta_1^{obs}(r^o_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta^*_D)$ that depends only on covariates that the researcher observes *ex post*:

$$\Delta_1^{obs}(r^o_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta^*_D) \equiv$$
$$\eta_j^{-1}r^o_{ijt} - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho d(1_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j). \quad \text{(F.20)}$$

By definition,

$$\mathbb{E}[\Delta_1^{obs}(r^o_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta^*_D) - \Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D)|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] = 0. \quad \text{(F.21)}$$

Therefore, exploiting the convexity of the function $(1-\Phi(\cdot))/\Phi(\cdot)$, we can apply a reasoning similar to that in Lemma C.2 and conclude that:

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D))}{\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D))} - (1-d_{ijt})\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right] \leq$$
$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}\Delta_1^{obs}(r^o_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta^*_D))}{\Phi(\sigma^{-1}\Delta_1^{obs}(r^o_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta^*_D))} - (1-d_{ijt})\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right]. \quad \text{(F.22)}$$

Combining equations (F.19), (F.21) and (F.22), and given a vector Z_{ijt} whose distribution conditional on the vector $(\mathcal{J}_{ijt}, dist_j, d_{ijt-1})$ is degenerate, we can therefore derive the weaker inequality:

$$\mathbb{E}\Big[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta_j^{-1}r^o_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1}-\rho d(1_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j)))}{\Phi(\sigma^{-1}(\eta_j^{-1}r^o_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1}-\rho d(1_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j)))}-(1-d_{ijt})\Big|Z_{ijt}\Big]\geq 0.$$

Furthermore, $d(1_t)_{ijt+1}$ denotes the actual export behavior of firm i in country j at period $t+1$ conditional on this firm having exported to j at t; therefore,

$$d(1_t)_{ijt+1} = d_{ijt+1} \qquad \text{if } d_{ijt}=1,$$

and, therefore, we can write our first odds-based moment inequality as

$$\mathbb{E}\Big[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta_j^{-1}r^o_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1}-\rho d_{ijt+1})(\gamma_0+\gamma_1 dist_j)))}{\Phi(\sigma^{-1}(\eta_j^{-1}r^o_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1}-\rho d_{ijt+1})(\gamma_0+\gamma_1 dist_j)))}-(1-d_{ijt})\Big|Z_{ijt}\Big]\geq 0, \tag{F.23}$$

where d_{ijt+1} takes value 1 if firm i is observed to export to country j in year t. This is the first of the two odds-based conditional moment inequalities we will use for identification of the parameter vector θ^*_D

Starting from the inequality in equation (F.10b), we derive here a second odds-based moment inequality that allows us to identify the parameter vector θ^*_D. Adding and subtracting 1 to equation (F.10b), we obtain:

$$\begin{aligned}
\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1)-V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0)\geq 0\}-1+(1-d_{ijt})\geq 0,\\
(1-d_{ijt})+(\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1)-V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0)\geq 0\}-1)\geq 0,\\
(1-d_{ijt})-(\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1)-V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0)\leq 0\})\geq 0,\\
(1-d_{ijt})-\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0)-V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1)\geq 0\}\geq 0.
\end{aligned}$$

Combining the last inequality with equation (F.5a), we obtain

$$(1-d_{ijt})-\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0)-W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1)\geq 0\}\geq 0, \tag{F.24}$$

and, from equations (5), (24), (F.2), (F.3), and (F.6c), we can write the variable inside the indicator function as:

$$\begin{aligned}
V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0)-W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1)=\\
-\eta_j^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]+f_{ijt}+(1-d_{ijt-1})s_{ijt}-\rho\mathbb{E}\big[d(0_t)_{ijt+1}s_{ijt+1}\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0\big]=\\
-\eta_j^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]+(\beta_0+\beta_1 dist_j+\nu_{ijt})+(1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)-(\gamma_0+\gamma_1 dist_j)\rho\mathbb{E}\big[d(0_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big]=\\
-\eta_j^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]+(\beta_0+\beta_1 dist_j+\nu_{ijt})+(1-d_{ijt-1}-\rho\mathbb{E}\big[d(0_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big])(\gamma_0+\gamma_1 dist_j).
\end{aligned} \tag{F.25}$$

For simplicity in the notation, we denote this expression as:

$$\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1};\theta^*_D)+\nu_{ijt},$$

with $\theta^*_D \equiv (\beta_0, \beta_1, \sigma, \gamma_0, \gamma_1)$ and, therefore,

$$\begin{aligned}
\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1};\theta^*_D)\equiv\\
-\eta_j^{-1}\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}]+(\beta_0+\beta_1 dist_j)+(1-d_{ijt-1}-\rho\mathbb{E}\big[d(0_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big])(\gamma_0+\gamma_1 dist_j).
\end{aligned} \tag{F.26}$$

Using this expression, we can rewrite (F.24) as

$$(1-d_{ijt})-\mathbb{1}\{\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1};\theta^*_D)+\nu_{ijt}\geq 0\}\geq 0, \tag{F.27}$$

or, equivalently,

$$\begin{aligned}
-d_{ijt}+1-\mathbb{1}\{\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1};\theta^*_D)+\nu_{ijt}\geq 0\}\geq 0,\\
\mathbb{1}\{\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1};\theta^*_D)+\nu_{ijt}\leq 0\}-d_{ijt}\geq 0,
\end{aligned}$$

$$\mathbb{1}\{-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt} \geq 0\} - d_{ijt} \geq 0. \tag{F.28}$$

Following analogous steps to those described in the proof to Lemma C.6, we can rewrite this inequality as

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}{1-\Phi(\sigma^{-1}(-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))} - d_{ijt}\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right] \geq 0. \tag{F.29}$$

This inequality is analogous to that in equation (C.6) with two differences: (1) the lagged export status d_{ijt-1} is included in the conditioning set and (2) the term inside the function $\Phi(\cdot)$ accounts for how the export decision at t affects export profits at t and $t+1$.

The term $\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)$ in equation (F.29) depends on the unobserved expectations

$$\mathbb{E}[r^o_{ijt}|\mathcal{J}_{ijt}] \qquad \text{and} \qquad \mathbb{E}[d(0_t)_{ijt+1}|\mathcal{J}_{ijt}, dist_j]$$

and, therefore, the researcher would not know it even if she knew the true parameter vector θ_D^*. We define an analogous expression $\Delta_1^{obs}(r^o_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$ that depends only on covariates that the researcher observes *ex post*:

$$\begin{gathered}\Delta_0^{obs}(r^o_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \equiv \\ -\eta_j^{-1} r^o_{ijt} + (\beta_0 + \beta_1 dist_j) + (1 - d_{ijt-1} - \rho d(0_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j).\end{gathered} \tag{F.30}$$

By definition,

$$\mathbb{E}[\Delta_0^{obs}(r^o_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) - \Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] = 0. \tag{F.31}$$

Therefore, exploiting the convexity of the function $\Phi(\cdot)/(1-\Phi(\cdot))$, we can apply a reasoning similar to that in Lemma C.4 and conclude that:

$$\begin{gathered}\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}{1-\Phi(\sigma^{-1}(-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right] \leq \\ \mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(-\Delta_0^{obs}(r^o_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)))}{1-\Phi(\sigma^{-1}(-\Delta_0^{obs}(r^o_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)))}\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right].\end{gathered} \tag{F.32}$$

Combining equations (F.29), (F.31) and (F.32), and given a vector Z_{ijt} whose distribution conditional on the vector $(\mathcal{J}_{ijt}, dist_j, d_{ijt-1})$ is degenerate, we can therefore derive the weaker inequality:

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta_j^{-1}r^o_{ijt} - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho d(0_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j)))}{1-\Phi(\sigma^{-1}(\eta_j^{-1}r^o_{ijt} - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho d(0_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j)))} - d_{ijt}\middle|Z_{ijt}\right] \geq 0.$$

Furthermore, note that $d(0_t)_{ijt+1}$ denotes the actual export behavior of firm i in country j at period $t+1$ conditional on this firm not exporting to j at t; therefore,

$$d(0_t)_{ijt+1} = d_{ijt+1} \qquad \text{if } 1 - d_{ijt} = 1,$$

and, therefore, we can write our second odds-based moment inequality as

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta_j^{-1}r^o_{ijt} - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho d_{ijt+1})(\gamma_0+\gamma_1 dist_j)))}{1-\Phi(\sigma^{-1}(\eta_j^{-1}r^o_{ijt} - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho d_{ijt+1})(\gamma_0+\gamma_1 dist_j)))} - d_{ijt}\middle|Z_{ijt}\right] \geq 0, \tag{F.33}$$

where d_{ijt+1} takes value 1 if firm i is observed to export to country j in year t.

Equations (F.23) and (F.33) denote the two conditional odds-based moment inequalities we may use for identification. As in the static case, we derive a finite set of unconditional moment inequalities consistent with equations (F.23) and (F.33). Specifically, we use twice as many unconditional moment inequalities as in the static case, as we interact each of the instrument functions described in Appendix A.5 both with the dummy variable $\mathbb{1}\{d_{ijt-1} = 0\}$ and with the dummy variable $\mathbb{1}\{d_{ijt-1} = 1\}$. Adding the dummy variable d_{ijt-1} to the vector Z_{ijt} that we employ to form unconditional moment inequalities allows us to separately identify the average fixed costs parameters, (β_0, β_1), and the average sunk costs parameters, (γ_0, γ_1).

F.3 Revealed-Preference Moment Inequalities for Dynamic Model

The definition of the random variable d_{ijt} in equation (F.9) implies that, for every firm i, country j and period t, we can write the following two inequalities

$$d_{ijt}(V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0) \geq 0, \tag{F.34a}$$

$$(1-d_{ijt})(V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1) \geq 0. \tag{F.34b}$$

Equation (F.34a) exploits the fact that

$$\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0) \geq 0\}, \tag{F.35}$$

is a necessary condition for $d_{ijt}=1$. Equation (F.34b) exploits the fact that the inequality inside the indicator function in equation (F.35) is a sufficient condition for $d_{ijt}=1$. We will derive an revealed-preference moment inequality from each of the inequalities in equation (F.34). We show first how to derive an inequality from equation (F.34a) and do the same for equation (F.34b) below.

Combining equations (F.5b) and (F.34a), we obtain:

$$d_{ijt}(V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0)) \geq 0. \tag{F.36}$$

As above, we can denote the expression in parenthesis as $\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt}$, with $\theta^*_D \equiv (\beta_0, \beta_1, \sigma, \gamma_0, \gamma_1)$ and $\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D)$ defined in equation (F.14). Using this expression, we can rewrite (F.36) as

$$d_{ijt}(\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt}) \geq 0. \tag{F.37}$$

Given that this inequality must hold for every firm i, country j and period t, it must also hold in expectation

$$\mathbb{E}[d_{ijt}(\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt})|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \tag{F.38}$$

or, equivalently,

$$\mathbb{E}[d_{ijt}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D)|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] - \mathbb{E}[d_{ijt}\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{F.39}$$

Focusing on the second term, we note that

$$\begin{aligned}\mathbb{E}[d_{ijt}\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] &= -\mathbb{E}[(1-d_{ijt})\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}]\\ &= -\mathbb{E}[\mathbb{E}[(1-d_{ijt})\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt}=0]|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}]\\ &= -\mathbb{E}[(1-d_{ijt})\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt}=0]|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}].\end{aligned} \tag{F.40}$$

Focusing further on the conditional expectation $\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt}=0]$, we note that

$$\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt}=0] =$$
$$\mathbb{E}[\nu_{ijt}|V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0) \leq 0, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}];$$

from equations (F.1) and (F.2), we can rewrite

$$\begin{aligned}&V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0) =\\ &\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt},\end{aligned} \tag{F.41}$$

and, therefore,

$$\begin{aligned}&\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt}=0] =\\ &\mathbb{E}[\nu_{ijt}|\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) - \nu_{ijt} \leq 0, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] =\\ &\mathbb{E}[\nu_{ijt}|\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta^*_D) \leq \nu_{ijt}, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}].\end{aligned}$$

Given the distributional assumption in equation (6) and the assumption that ν_{ijt} is independent over time, we can rewrite this expression as

$$\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 0] = \sigma \frac{\phi(\sigma^{-1}\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}. \tag{F.42}$$

Combining equations (F.39), (F.40) and (F.42), we obtain the following moment inequality

$$\mathbb{E}[d_{ijt}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{F.43}$$

As equations (F.1) and (F.2) show, the term $\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)$ will depend on the difference in the expected discounted sum of future profits depending on whether firm i exports to country j at t and, therefore, cannot be computed without specifying precisely the content of $\mathcal{J}_{ijt}$. However, from equations (F.5b), (F.13), and (F.41), one can conclude that

$$\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) \geq \tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*),$$

and, therefore, in combination with equation (F.43), one can derive the weaker inequality

$$\mathbb{E}[d_{ijt}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{F.44}$$

Exploiting the mean independence restriction in equation (F.21) and the convexity of the function $\phi(\cdot)/(1 - \Phi(\cdot))$, we can apply a reasoning similar to that in lemmas C.9 and C.10 and conclude that:

$$\begin{aligned}\mathbb{E}[&d_{ijt}\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \\ &+ (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0.\end{aligned} \tag{F.45}$$

where the term $\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$ is defined in equation (F.20). This inequality cannot be used for identification directly because for all observations i, j and t such that $d_{ijt} = 0$, we will not observe the random variable $d(1_t)_{ijt+1}$. We therefore cannot compute the term

$$(1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}, \tag{F.46}$$

as a function of data and the parameter vector θ_D^*. However, as equation (F.20) shows, the function

$$\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$$

is increasing in the value of the dummy variable $d(1_t)_{ijt+1}$. As equation (F.45) is also increasing in the term $\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$, we can therefore derive a weaker inequality by substituting the unobserved dummy variable $d(1_t)_{ijt+1}$ by the largest value in its support:

$$\begin{aligned}\mathbb{E}[&d_{ijt}\Delta_1^{obs}(r_{ijt}^o, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \\ &+ (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}^o, 1, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}^o, 1, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0.\end{aligned} \tag{F.47}$$

From equation (F.20), we therefore obtain the following inequality

$$\begin{aligned}\mathbb{E}[&d_{ijt}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d(1_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j)) \\ &+ (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho)(\gamma_0 + \gamma_1 dist_j)))}{1 - \Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho)(\gamma_0 + \gamma_1 dist_j)))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0,\end{aligned} \tag{F.48}$$

which implies that, for any random vector Z_{ijt} whose distribution conditional on $(\mathcal{J}_{ijt}, dist_j, d_{ijt-1})$ is degenerate, the following inequality holds

$$\mathbb{E}[d_{ijt}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d(1_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j)) \\ + (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho)(\gamma_0 + \gamma_1 dist_j)))}{1 - \Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho)(\gamma_0 + \gamma_1 dist_j)))} | Z_{ijt}] \geq 0, \tag{F.49}$$

This is the first of the revealed-preference inequalities we will use for identification of the parameter θ_D^*.

Combining equations (F.5a) and (F.34b), we obtain:

$$(1 - d_{ijt})(V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) \geq 0. \tag{F.50}$$

As above, we can denote the expression in parenthesis as $\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt}$, with $\theta_D^* \equiv (\beta_0, \beta_1, \sigma, \gamma_0, \gamma_1)$ and $\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)$ defined in equation (F.26). Using this expression, we can rewrite (F.50) as

$$(1 - d_{ijt})(\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt}) \geq 0. \tag{F.51}$$

Given that this inequality must hold for every firm i, country j and period t, it must also hold in expectation

$$\mathbb{E}[(1 - d_{ijt})(\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt}) | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \tag{F.52}$$

or, equivalently,

$$\mathbb{E}[(1 - d_{ijt})\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] + \mathbb{E}[(1 - d_{ijt})\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{F.53}$$

Focusing on the second term, we note that

$$\begin{aligned} \mathbb{E}[(1 - d_{ijt})\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] &= -\mathbb{E}[d_{ijt}\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \\ &= -\mathbb{E}[\mathbb{E}[d_{ijt}\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1] | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \\ &= -\mathbb{E}[d_{ijt}\mathbb{E}[\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1] | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}]. \end{aligned} \tag{F.54}$$

Focusing further on the conditional expectation $\mathbb{E}[\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1]$, we note that

$$\mathbb{E}[\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1] = \\ \mathbb{E}[\nu_{ijt} | V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] = \\ \mathbb{E}[\nu_{ijt} | V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) \leq 0, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}];$$

from equations (F.1) and (F.2), we can rewrite

$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) = \\ \tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt}, \tag{F.55}$$

and, therefore,

$$\mathbb{E}[\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1] = \\ \mathbb{E}[\nu_{ijt} | \tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt} \leq 0, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] = \\ \mathbb{E}[\nu_{ijt} | -\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) \geq \nu_{ijt}, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}].$$

Given the distributional assumption in equation (6) and the assumption that ν_{ijt} is independent over time, we can rewrite this expression as

$$\mathbb{E}[\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1] = -\sigma \frac{\phi(\sigma^{-1}(-\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}{\Phi(\sigma^{-1}(-\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}. \tag{F.56}$$

Combining equations (F.53), (F.54) and (F.56), we obtain the following moment inequality

$$\mathbb{E}[(1-d_{ijt})\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + d_{ijt}\sigma\frac{\phi(\sigma^{-1}(-\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}{\Phi(\sigma^{-1}(-\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0.$$

or, equivalently,

$$\mathbb{E}[(1-d_{ijt})\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + d_{ijt}\sigma\frac{\phi(\sigma^{-1}\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1-\Phi(\sigma^{-1}\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{F.57}$$

As equations (F.1) and (F.2) show, the term $\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)$ will depend on the difference in the expected discounted sum of future profits depending on whether firm i exports to country j at t and, therefore, cannot be computed without specifying precisely the content of $\mathcal{J}_{ijt}$. However, from equations (F.5a), (F.26), and (F.55), one can conclude that

$$\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) \geq \tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*),$$

and, therefore, in combination with equation (F.57), one can derive the weaker inequality

$$\mathbb{E}[(1-d_{ijt})\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + d_{ijt}\sigma\frac{\phi(\sigma^{-1}\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1-\Phi(\sigma^{-1}\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{F.58}$$

Exploiting the mean independence restriction in equation (F.31) and the convexity of the function $\phi(\cdot)/(1-\Phi(\cdot))$, we can apply a reasoning similar to that in lemmas C.9 and C.10 and conclude that:

$$\begin{aligned}\mathbb{E}[(1-d_{ijt})&\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)\\ &+ d_{ijt}\sigma\frac{\phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}{1-\Phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0,\end{aligned} \tag{F.59}$$

where the term $\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$ is defined in equation (F.30). This inequality cannot be used for identification directly because for observations i, j and t such that $d_{ijt} = 1$, we do not observe the random variable $d(0_t)_{ijt+1}$. We therefore cannot compute the term

$$d_{ijt}\sigma\frac{\phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}{1-\Phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}. \tag{F.60}$$

However, as equation (F.30) shows, the function

$$\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$$

is decreasing in the value of the dummy variable $d(0_t)_{ijt+1}$. As equation (F.59) is increasing in the term $\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$, we can therefore derive a weaker inequality by substituting the unobserved dummy variable $d(0_t)_{ijt+1}$ by the smallest value in its support:

$$\begin{aligned}\mathbb{E}[(1-d_{ijt})&\Delta_0^{obs}(r_{ijt}^o, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)\\ &+ d_{ijt}\sigma\frac{\phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}^o, 0, dist_j, d_{ijt-1}; \theta_D^*))}{1-\Phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}^o, 0, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0,\end{aligned} \tag{F.61}$$

From equation (F.30), we therefore obtain the following inequality

$$\begin{aligned}\mathbb{E}[(1-d_{ijt})&(-\eta_j^{-1}r_{ijt}^o + (\beta_0+\beta_1 dist_j) + (1-d_{ijt-1}-\rho d(0_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j))\\ &+ d_{ijt}\sigma\frac{\phi(\sigma^{-1}(-\eta_j^{-1}r_{ijt}^o + (\beta_0+\beta_1 dist_j) + (1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)))}{1-\Phi(\sigma^{-1}(-\eta_j^{-1}r_{ijt}^o + (\beta_0+\beta_1 dist_j) + (1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0,\end{aligned} \tag{F.62}$$

or, equivalently,

$$\mathbb{E}[-(1-d_{ijt})(\eta_j^{-1}r_{ijt}^o - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho d(0_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j))$$

$$+ d_{ijt}\sigma\frac{\phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1})(\gamma_0 + \gamma_1 dist_j)))}{\Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1})(\gamma_0 + \gamma_1 dist_j)))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \quad \text{(F.63)}$$

which implies that, for any random vector Z_{ijt} whose distribution conditional on $(\mathcal{J}_{ijt}, dist_j, d_{ijt-1})$ is degenerate, the following inequality holds

$$\begin{aligned}\mathbb{E}[-(1 - d_{ijt})(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d(0_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j))\\ + d_{ijt}\sigma\frac{\phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1})(\gamma_0 + \gamma_1 dist_j)))}{\Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}^o - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1})(\gamma_0 + \gamma_1 dist_j)))}|Z_{ijt}] \geq 0,\end{aligned} \quad \text{(F.64)}$$

This is the second of the revealed-preference inequalities we will use for identification of the parameter θ_D^*.

Equations (F.49) and (F.64) denote the two conditional revealed-preference moment inequalities we may use for identification. As in the static case, we derive a finite set of unconditional moment inequalities consistent with equations (F.49) and (F.64). Specifically, as discussed above for the case of the odds-based moment inequalities, we use twice as many unconditional moment inequalities as in the static case, as we interact each of the instrument functions described in Appendix A.5 both with the dummy variable $\mathbb{1}\{d_{ijt-1} = 0\}$ and with the dummy variable $\mathbb{1}\{d_{ijt-1} = 1\}$.

G Firm-Country Export Revenue Shocks: Details

G.1 Known to Firms When Deciding on Export Entry

We model export revenues as

$$r_{ijt} = \alpha_{jt} r_{iht} + \omega_{ijt} + e_{ijt}, \qquad \mathbb{E}[e_{ijt}|\mathcal{J}_{ijt}, r_{iht}, f_{ijt}] = 0, \text{ and } \mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}] = \omega_{ijt}; \tag{G.1}$$

export participation decisions are determined by

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\}; \tag{G.2}$$

and the distribution of unobserved determinants of export revenues and fixed costs is

$$\left(\begin{array}{c} \omega_{ijt} \\ \nu_{ijt} \end{array} \right) \Bigg| (\mathcal{W}_{ijt}, dist_j) \sim \mathbb{N}\left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma^2_\omega & \sigma_{\omega\nu} \\ \sigma_{\omega\nu} & \sigma^2 \end{array} \right) \right), \tag{G.3}$$

where $\mathcal{W}_{ijt}$ is a subset of $\mathcal{J}_{ijt}$ such that

$$\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] = \mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{W}_{ijt}]. \tag{G.4}$$

Relative to the model in Section 2, this model adds, for every firm, country and year, a revenue shock ω_{ijt} that is unobserved to the researcher but known to the firm when it decides whether to export (see equation (G.1)) and is jointly normally distributed with the firm-country-year fixed costs shock, ν_{ijt}. Combining equations (G.1), (G.2), and (G.4), we can rewrite the export participation dummy d_{ijt} as in equation (28); i.e.

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j - (\nu_{ijt} - \eta^{-1}\omega_{ijt}) \geq 0\}. \tag{G.5}$$

Taking into account that we assume both ν_{ijt} and ω_{ijt} are jointly normally distributed, it holds that the unobserved (to the researcher) term in equation (28) will also be normally distributed:

$$(\nu_{ijt} - \eta^{-1}\omega_{ijt})|(\mathcal{W}_{ijt}, dist_j) \sim \mathbb{N}(0, \sigma^2 + \eta^{-2}\sigma^2_\omega - 2\eta^{-1}\sigma_{\omega\nu}). \tag{G.6}$$

The parameters of this model are $(\{\alpha_{jt}; \forall(j,t)\}, \beta_0, \beta_1, \eta, \sigma, \sigma_{\omega\nu}, \sigma_\omega)$. These parameters are not identified, unless a normalization is imposed on them. As a normalization, we follow the literature and initially impose that

$$\sigma_{\omega\nu} = 0. \tag{G.7}$$

The key difference between the estimation procedures described in Section 4 and the estimation approaches that must be followed to estimate the parameters of this model is that the parameter vector $\{\alpha_{jt}; \forall j \text{ and } t\}$ can no longer be identified separately from the parameter vector $(\beta_0, \beta_1, \eta, \sigma)$. Given that our sample period covers 10 years and 22 countries, the parameter vector $\{\alpha_{jt}; \forall j \text{ and } t\}$ incorporates over 200 elements. Thus, unless some restriction is imposed on these parameters, our moment inequality estimation approach would require estimating jointly a confidence set for over 200 parameters. While this is theoretically possible, as far as we know, it is infeasible given current computing power and our inference approach. Therefore, we simplify the estimation by assuming

$$\alpha_{jt} = \alpha, \qquad \text{for all } j \text{ and } t. \tag{G.8}$$

The resulting parameter vector of the model described above is $\theta_S \equiv (\alpha, \beta_0, \beta_1, \sigma_\omega, \sigma, \eta)$.

G.1.1 Maximum Likelihood Estimation: Perfect Foresight

Deriving a maximum likelihood estimator for θ_S requires either assuming that $e_{ijt} = 0$ for all i, j and t, or imposing parametric assumptions on the joint distribution of e_{ijt} across firms and countries. For the sake of simplicity, we opt for the former assumption and, thus, assume that all unobserved (to the researcher) year t revenue shocks are known to the firm when it decides whether to export to j at t.

Given equations (G.1) to (G.7), the assumption that $e_{ijt} = 0$ for all i, j and t, and an assumed information set $\mathcal{W}^a_{ijt}$ such that $\mathbb{E}[\alpha r_{iht}|\mathcal{W}^a_{ijt}] = \alpha r_{iht}$, we can estimate the export revenue parameters (α, σ_ω) and the fixed export costs parameters $(\eta, \beta_0, \beta_1, \sigma)$ in a single step using a maximum likelihood procedure. Following the same steps as in Das et al. (2007), we can write the contribution of each firm-country-year triplet to the likelihood function as

$$f(d_{ijt}, d_{ijt}r_{ijt}|r_{iht}, dist_j) = f(d_{ijt}, \omega_{ijt}^{obs}|r_{iht}, dist_j) = f(d_{ijt}|\omega_{ijt}^{obs}, r_{iht}, dist_j)f(\omega_{ijt}^{obs}|r_{iht}, dist_j)$$

where $\omega_{ijt}^{obs} \equiv \{r_{ijt} - \alpha r_{iht}; d_{ijt} = 1\}$ and $f(\cdot)$ denotes the corresponding density function. This individual log-likelihood can be further rewritten as

$$\begin{gathered} f(d_{ijt}, \omega_{ijt}^{obs}|r_{iht}, dist_j) = \\ \Big(\int_{\omega_{it}} \mathcal{P}(d_{ijt} = 0|\omega_{ijt}, r_{iht}, dist_j)f(\omega_{ijt})d\omega_{ijt}\Big)^{1-d_{ijt}} \times \\ \Big(\mathcal{P}(d_{ijt} = 1|\omega_{ijt}^{obs}, r_{iht}, dist_j)f(\omega_{ijt}^{obs})\Big)^{d_{ijt}}, \end{gathered}$$

where $f(\omega_{ijt}^{obs})$ denotes the marginal density function of ω_{ijt} evaluated at the value $\omega_{ijt}^{obs} = r_{ijt} - \alpha r_{iht}$ for all i, j and t such that $d_{ijt} = 1$. Using the expression for d_{ijt} in equation (G.5), we can rewrite

$$\begin{gathered} f(d_{ijt}, \omega_{ijt}^{obs}|r_{iht}, dist_j) = \\ \Big(\int_{\omega_{it}}\int_{\nu_{it}} \mathbb{1}\{\eta^{-1}\alpha r_{iht} - \beta_0 - \beta_1 dist_j - (\nu_{ijt} - \eta^{-1}\omega_{ijt}) < 0\}f(\nu_{ijt}|\omega_{ijt})f(\omega_{ijt})d\nu_{ijt}d\omega_{ijt}\Big)^{1-d_{ijt}} \times \\ \Big(\int_{\nu_{it}} \mathbb{1}\{\eta^{-1}\alpha r_{iht} - \beta_0 - \beta_1 dist_j - \nu_{ijt} + \eta^{-1}\omega_{ijt}^{obs} \geq 0\}f(\nu_{ijt}|\omega_{ijt}^{obs})f(\omega_{ijt}^{obs})d\nu_{ijt}\Big)^{d_{ijt}} = \\ \Big(\int_{\omega_{it}}\int_{\nu_{it}} \mathbb{1}\{\eta^{-1}\alpha r_{iht} - \beta_0 - \beta_1 dist_j - (\nu_{ijt} - \eta^{-1}\omega_{ijt}) < 0\}f(\nu_{ijt} - \eta^{-1}\omega_{ijt})d\omega_{ijt}\Big)^{1-d_{ijt}} \times \\ \Big(\int_{\nu_{it}} \mathbb{1}\{\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\}f(\nu_{ijt}|\omega_{ijt}^{obs})f(\omega_{ijt}^{obs})\Big)^{d_{ijt}}, \end{gathered}$$

where the last equality uses the definition of ω_{ijt}^{obs}. Given the normalization in equation (G.7), we can rewrite

$$\begin{gathered} f(d_{ijt}, \omega_{ijt}^{obs}|r_{iht}, dist_j) = \\ \Big(1 - \Phi\big((1/\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2})(\eta^{-1}\alpha r_{iht} - \beta_0 - \beta_1 dist_j)\big)\Big)^{1-d_{ijt}} \times \\ \Big(\Phi\big((1/\sigma)(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)\phi\big((1/\sigma_\omega)(r_{ijt} - \alpha r_{iht})\big)(1/\sigma_\omega)\Big)^{d_{ijt}}, \end{gathered}$$

Taking logs and computing its expectation conditional on $(\mathcal{J}_{ijt}, dist_j)$, we can thus write the log-likelihood function as

$$\begin{aligned} \mathcal{L}(d_{ijt}, d_{ijt}r_{ijt}|\mathcal{J}_{ijt}, dist_j; \theta_S) = \mathbb{E}\Big[& d_{ijt}\ln\Big(\Phi\big((1/\sigma)(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)\Big) + \\ &(1 - d_{ijt})\ln\Big(1 - \Phi\big((1/\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2})(\eta^{-1}\alpha r_{iht} - \beta_0 - \beta_1 dist_j)\big)\Big) + \\ &d_{ijt}\ln\Big(\phi\big((1/\sigma_\omega)(r_{ijt} - \alpha r_{iht})\big)(1/\sigma_\omega)\Big)\Big|\mathcal{J}_{ijt}, dist_j; \theta_S\Big], \end{aligned} \tag{G.9}$$

and its sample analogue as

$$\begin{aligned} \sum_i\sum_j\sum_t\Big\{& d_{ijt}\ln\Big(\Phi\big((1/\sigma)(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)\Big) + \\ &(1 - d_{ijt})\ln\Big(1 - \Phi\big((1/\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2})(\eta^{-1}\alpha r_{iht} - \beta_0 - \beta_1 dist_j)\big)\Big) + \\ &d_{ijt}\ln\Big(\phi\big((1/\sigma_\omega)(r_{ijt} - \alpha r_{iht})\big)(1/\sigma_\omega)\Big)\Big\}. \end{aligned} \tag{G.10}$$

This likelihood function corresponds to that described in footnote 17 of Das et al. (2007). The estimator that maximizes this likelihood function yields consistent estimates of the parameter vector $\theta_S \equiv (\alpha, \beta_0, \beta_1, \sigma_\omega, \sigma, \eta)$. Once we have recovered these estimates of θ_S, we re-normalize them (and the value of $\sigma_{\omega\nu}$) to match the assumption that $\eta = 5$ (see Section 2.4).

G.1.2 Moment Inequality Estimation: Defining Moment Inequalities

To compute bounds on the parameter vector θ_S, we impose five conditional moment inequalities. Four of these inequalities are slight variations of the moment inequalities described in Sections 4.2. The additional inequality arises from equation (G.1) when the only assumption on $\mathcal{W}_{ijt}$ that we want to impose is that we observe a vector Z_{ijt} such that $Z_{ijt} \subseteq \mathcal{W}_{ijt}$. We describe here in detail how we derive each of these five inequalities. Importantly, all these inequalities are valid in the presence of the unobserved (to the researcher) export revenue shock e_{ijt} as long as it satisfies the mean independence restriction in equation (G.1).

Odds-based inequalities. Equations (G.5) to (G.7) imply that

$$\mathcal{P}(d_{ijt}=1|\mathcal{J}_{ijt},dist_j) = \Phi\big((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}\big(\eta^{-1}\alpha\mathbb{E}[r_{iht}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\big). \tag{G.11}$$

Given this expression, following the same steps as in Appendix C, one can derive the following two odds-based inequalities

$$\mathcal{M}^{ob}(Z_{ijt};\beta_0,\beta_1,\alpha,\beta_0,\beta_1,\sigma_\omega,\sigma,\eta) = \tag{G.12}$$
$$\mathbb{E}\left[\begin{array}{c|c} d_{ijt}\frac{1-\Phi\left((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}\left(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j\right)\right)}{\Phi\left((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}\left(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j\right)\right)}-(1-d_{ijt}) & \\ (1-d_{ijt})\frac{\Phi\left((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}\left(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j\right)\right)}{1-\Phi\left((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}\left(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j\right)\right)}-d_{ijt} & \end{array} Z_{ijt}\right] \geq 0.$$

These inequalities are analogous to those in equation (14).

Revealed-preference inequalities. Following the same steps as in Appendix C, one can derive the following two revealed-preference inequalities

$$\mathcal{M}^{r}(Z_{ijt};\alpha,\beta_0,\beta_1,\sigma_\omega,\sigma,\eta) = \tag{G.13}$$
$$\mathbb{E}\left[\begin{array}{c|c} -(1-d_{ijt})\big(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j\big)+d_{ijt}(\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})\frac{\phi\left((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j)\right)}{\Phi\left((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j)\right)} & \\ d_{ijt}\big(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j\big)+(1-d_{ijt})(\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})\frac{\phi\left((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j)\right)}{1-\Phi\left((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\alpha r_{iht}-\beta_0-\beta_1 dist_j)\right)} & \end{array} Z_{ijt}\right] \geq 0.$$

These inequalities are analogous to those in equation (17) for the case in which the revenue shock ω_{ijt}.

Revenue-equation inequality. From equation (G.1) and the fact that $\mathcal{W}_{ijt} \subset \mathcal{J}_{ijt}$, we know that

$$\mathbb{E}[r_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt},dist_j] = \alpha\mathbb{E}[r_{iht}|d_{ijt}=1,\mathcal{W}_{ijt},dist_j]+\mathbb{E}[\omega_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt},dist_j]$$

and equations (G.3), (G.5) and (G.7) imply that

$$\begin{aligned}
&\mathbb{E}[\omega_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt},dist_j] = \\
&\quad = \mathbb{E}[\omega_{ijt}|\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j-(\nu_{ijt}-\eta^{-1}\omega_{ijt})\geq 0,\mathcal{W}_{ijt},dist_j] \\
&\quad = \mathbb{E}[\omega_{ijt}|\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j \geq \nu_{ijt}-\eta^{-1}\omega_{ijt},\mathcal{W}_{ijt},dist_j], \\
&\quad = \mathbb{E}[\mathbb{E}[\omega_{ijt}|\nu_{ijt}-\eta^{-1}\omega_{ijt}]|\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j \geq \nu_{ijt}-\eta^{-1}\omega_{ijt},\mathcal{W}_{ijt},dist_j], \\
&\quad = [(-\eta^{-1}\sigma_\omega^2)/(\sigma^2+\eta^{-2}\sigma_\omega^2)]\mathbb{E}[\nu_{ijt}-\eta^{-1}\omega_{ijt}|\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j \geq \nu_{ijt}-\eta^{-1}\omega_{ijt},\mathcal{W}_{ijt},dist_j], \\
&\quad = \frac{\eta^{-1}\sigma_\omega^2}{\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2}}\frac{\phi((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}.
\end{aligned}$$

Therefore, we can write

$$\mathbb{E}\Big[r_{ijt}-\alpha r_{iht}-\frac{\eta^{-1}\sigma_\omega^2}{\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2}}\frac{\phi((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi((\sqrt{\sigma^2+\eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\Big|d_{ijt}=1,\mathcal{W}_{ijt},dist_j\Big]=0.$$

and, by the Law of Iterated Expectations, this equation plus the assumption that $Z_{ijt} \subseteq \mathcal{J}_{ijt}$ implies that

$$\mathbb{E}\Big[r_{ijt} - \alpha r_{iht} - \frac{\eta^{-1}\sigma_\omega^2}{\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2}} \frac{\phi((\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))}{\Phi((\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))}\Big| d_{ijt} = 1, Z_{ijt}, dist_j\Big] = 0.$$

This equation cannot be used directly for identification, as it depends on the unobserved expectation $\mathbb{E}[\alpha r_{iht}|\mathcal{W}_{ijt}]$. However, given that $\phi(\cdot)/\Phi(\cdot)$ is convex, we can use Jensen's inequality to derive the following inequality

$$\mathbb{E}\Big[r_{ijt} - \alpha r_{iht} - \frac{\eta^{-1}\sigma_\omega^2}{\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2}} \frac{\phi((\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\alpha r_{iht} - \beta_0 - \beta_1 dist_j))}{\Phi((\sqrt{\sigma^2 + \eta^{-2}\sigma_\omega^2})^{-1}(\eta^{-1}\alpha r_{iht} - \beta_0 - \beta_1 dist_j))}\Big| d_{ijt} = 1, Z_{ijt}, dist_j\Big] \leq 0. \tag{G.14}$$

Combining inequalities for estimation. The conditional inequalities in equations (G.12), (G.13), and (G.14) may be used to compute bounds on the parameter vector $\theta_S \equiv (\alpha, \beta_0, \beta_1, \sigma_\omega, \sigma, \eta)$. However, as discussed in Section 4.2.3, exploiting all the information contained in these conditional moment inequalities can be computationally challenging and, consequently, we base our inference on a fixed number of unconditional moment inequalities implied by the conditional moment inequalities in equations (G.12), (G.13), and (G.14).

Alternative inequalities. Note that neither the odds-based moment inequalities in equation (G.12) nor the revealed-preference moment inequalities in equation (G.13) exploit the available data on realized export revenues, r_{ijt}, for those firm-country-year triplets for which we observe positive exports; i.e. those i, j, and t for which we observe $d_{ijt} = 1$. This is the same approach we have followed in the main draft, where r^o_{ijt} is substituted by $\alpha_{jt} r_{iht}$ (instead of being substituted by r_{ijt}) everywhere in equations (14) and (17).

One could think of alternative odds-based and revealed-preference inequalities that exploit the information on $d_{ijt} r_{ijt}$ whenever possible. We leave this possibility for future work.

G.1.3 Moment Inequality Estimation: Computing Confidence Sets

We describe here the procedure we follow to compute the confidence set for the true parameter vector θ^*_S. The procedure is very similar to that described in Appendix A.7.1, but here θ_S is not a function of preliminary estimates of the export revenue parameters $\{\alpha_{jt}; \forall j, t\}$; all parameters are estimated jointly in a single step. We describe here a procedure that implements the asymptotic version of the Generalized Moment Selection (GMS) test described in page 135 of Andrews and Soares (2010). 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

$$\bar{m}_k(\theta_S) \geq 0, \qquad k = 1, \ldots, K,$$

where, for every $k = 1, \ldots, K$,

$$\bar{m}_k(\theta_S) \equiv \frac{1}{n} \sum_i \sum_j \sum_t m_k(X_{ijt}, Z_{ijt}, \theta_S),$$

with n denoting the sample size (i.e. sum of distinct ijt triplets included in our sample). The MMM statistic is therefore defined as

$$Q(\theta_S) = \sum_{k=1}^{K} (\min\{\frac{\bar{m}_k(\theta_S)}{\hat{\sigma}_k(\theta_S)}, 0\})^2, \tag{G.15}$$

where $\hat{\sigma}_k(\theta_S) = \sqrt{\hat{\sigma}^2_k(\theta_S)}$ and

$$\hat{\sigma}^2_k(\theta_S) = \frac{1}{n} \sum_i \sum_j \sum_t (m_k(X_{ijt}, Z_{ijt}, \theta_S) - \bar{m}_k(X_{ijt}, Z_{ijt}; \theta_S))^2.$$

In the notation introduced in sections 4.2.1 and 4.2.2, $X_{ijt} \equiv (d_{ijt}, d_{ijt} r_{ijt}, r_{iht}, dist_j)$ and $m_k(\cdot)$ may be an odds-based, a revealed-preference moment function or the revenue-equation inequality introduced in Appendix

G.1.2. 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 moment inequalities described in Appendix G.1.2. Given the set of unconditional moment inequalities $k = 1, \ldots, K$ and the test statistic in equation (A.27), we compute confidence sets for the true parameter value θ_S^* using the following steps:

Step 1: define a grid $\Theta_{S,g}$ 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 θ_S. The procedure is analogous to that in step 1 in Appendix A.7.1

Step 2: choose a point $\theta_{S,p} \in \Theta_{S,g}$. The following steps will test the null hypothesis that the vector $\theta_{S,p}$ equals the true value of θ_S:

$$H_0 : \theta_S^* = \theta_{S,p} \qquad \text{vs.} \qquad H_0 : \theta_S^* \neq \theta_{S,p}.$$

Step 3: evaluate the MMM test statistic at $\theta_{S,p}$:

$$Q(\theta_{S,p}) = \sum_{k=1}^{K} (\min\{\frac{\bar{m}_k(\theta_{S,p})}{\hat{\sigma}_k(\theta_{S,p})}, 0\})^2, \tag{G.16}$$

Step 4: compute the correlation matrix of moments evaluated at $\theta_{S,p}$:

$$\hat{\Omega}(\theta_{S,p}) = Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_{S,p}))\hat{\Sigma}(\theta_{S,p})Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_{S,p})),$$

where $Diag(\hat{\Sigma}(\theta_{S,p}))$ is the $K \times K$ diagonal matrix whose diagonal elements are equal to those of $\hat{\Sigma}(\theta_{S,p})$, $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_{S,p}))$ is a matrix such that $Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_{S,p}))Diag^{-\frac{1}{2}}(\hat{\Sigma}(\theta_{S,p})) = Diag^{-1}(\hat{\Sigma}(\theta_{S,p}))$ and

$$\hat{\Sigma}(\theta_{S,p}) = \frac{1}{n} \sum_i \sum_j \sum_t (m(X_{ijt}, Z_{ijt}, \theta_{S,p}) - \bar{m}(\theta_{S,p}))(m(X_{ijt}, Z_{ijt}, \theta_{S,p}) - \bar{m}(\theta_{S,p}))',$$

where $\bar{m}(\theta_{S,p}) = (\bar{m}_1(\theta_{S,p}), \ldots, \bar{m}_K(\theta_{S,p}))$.

Step 5: simulate the asymptotic distribution of $Q(\theta_{S,p})$. Take 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 criterion function $Q^{AA}_{n,r}(\theta_{S,p})$ as

$$Q^{AA}_{n,r}(\theta_{S,p}) = \sum_{k=1}^{K} \Big\{ (\min\{[\hat{\Omega}^{\frac{1}{2}}(\theta_{S,p})\zeta_r]_k, 0\})^2 \times \mathbb{1}\{\sqrt{n}\frac{\bar{m}_k(\theta_{S,p})}{\hat{\sigma}_k(\theta_{S,p})} \leq \sqrt{\ln n}\} \Big\}$$

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

Step 6: compute the critical value. The critical $\hat{c}^{AA}_n(\theta_{S,p}, 1-\delta)$ is the $(1-\delta)$-quantile of the distribution of $Q^{AA}_{n,r}(\theta_{S,p})$ across the R draws taken in the previous step.

Step 7: accept/reject $\theta_{S,p}$. Include $\theta_{S,p}$ in the estimated $(1-\delta)$% confidence set, $\hat{\Theta}^{1-\delta}$, if $Q(\theta_{S,p}) \leq \hat{c}^{AA}_n(\theta_{S,p}, 1-\delta)$.

Step 8: repeat steps 2 to 7 for every $\theta_{S,p}$ in the grid $\Theta_{S,g}$.

Step 9: compare the points included in the set $\hat{\Theta}_S^{1-\alpha}$ to those in the set $\Theta_{S,g}$. If (a) some of the points included in the set $\hat{\Theta}_S^{1-\alpha}$ are at the boundary of the set $\Theta_{S,g}$, expand the limits of $\Theta_{S,g}$ and repeat steps 2 to 9. If (b) the set of points included in $\hat{\Theta}_S^{1-\alpha}$ is only a small fraction of those included in $\Theta_{S,g}$, redefine a set $\Theta_{S,g}$ that is again a 3-dimensional orthotope whose limits are the result of adding a small number to the corresponding limits of the set $\hat{\Theta}_S^{1-\alpha}$ and repeat steps 2 to 9. If neither (a) nor (b) applies, define $\hat{\Theta}_S^{1-\alpha}$ as the 95% confidence set for θ_S^*.

G.1.4 Re-normalization

The estimation in appendices G.1.1 and G.1.2 is performed under the normalization in equation (G.7). However, as described in Section 2.4, the baseline estimates reported in Section 5 are computed under the normalization that $\eta = 5$. To facilitate the comparison of the estimates computed under the assumption that $\omega_{ijt} \subseteq \mathcal{J}_{ijt}$ to those computed under the assumption that $\mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}] = 0$, we re-normalize the estimates computed according to the procedures in appendices G.1.1 and G.1.2 to make them consistent with the normalization $\eta = 5$.

For the case of the maximum likelihood estimation, we denote the estimates we obtain from maximizing the log-likelihood function in equation (G.10) as

$$\hat{\theta}_S \equiv (\hat{\alpha}, \hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_\omega, \hat{\sigma}, \hat{\eta}).$$

These estimates are obtained under the normalization that $\sigma_{\nu\omega} = 0$, as imposed by equation (G.7). If we want instead to impose a normalization that fixes the value of η^{-1} to 0.2, then we need to renormalize our other estimates of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}$ as

$$\frac{0.2}{\widehat{\eta^{-1}}}\hat{\beta}_0, \qquad \frac{0.2}{\widehat{\eta^{-1}}}\hat{\beta}_1, \qquad \frac{0.2}{\widehat{\eta^{-1}}}\hat{\sigma}.$$

Furthermore, the value of $\sigma_{\nu\omega}$ consistent with this normalization will not be zero anymore. Rather, it is the value of $\sigma_{\nu\omega}$ such that

$$\frac{0.2}{\widehat{\eta^{-1}}}\sqrt{\hat{\sigma}^2 + (\widehat{\eta^{-1}})^2\hat{\sigma}_\omega^2} = \sqrt{\left(\frac{0.2}{\widehat{\eta^{-1}}}\right)^2 \hat{\sigma}^2 + \left(\frac{0.2}{\widehat{\eta^{-1}}}\right)^2 (\widehat{\eta^{-1}})^2\hat{\sigma}_\omega^2 - 2\frac{0.2}{\widehat{\eta^{-1}}}\widehat{\eta^{-1}}\sigma_{\nu\omega}}.$$

Notice that this re-normalization does not affect the maximum likelihood estimates of α_{jt} and σ_ω.

For the case of the moment inequality estimation, let's denote a generic value of θ_S included in the estimated confidence set computed using inequalities (G.12), (G.13), and (G.14) as

$$\theta'_S \equiv (\alpha', \beta'_0, \beta'_1, \sigma'_\omega, \sigma', \eta').$$

As in the case of the maximum likelihood estimates, the confidence set that includes this value of θ_S is computed under the normalization imposed in equation (G.7). If we want instead to assume that η^{-1} to 0.2, then we need to renormalize the values of β'_0 and β'_1 of every point in the estimated confidence set as

$$\frac{0.2}{\widehat{\eta^{-1}}}\beta'_0, \qquad \frac{0.2}{\widehat{\eta^{-1}}}\beta'_1,$$

and the value of σ' must change to a new value σ'' such that

$$\frac{(\eta')^{-1}}{\sqrt{(\sigma')^2 + ((\eta')^{-1})^2(\sigma'_\omega)^2}} = \frac{0.2}{\sqrt{(\sigma'')^2 + 0.2^2(\sigma'_\omega)^2}}.$$