What do Exporters Know?*

Michael J. Dickstein

New York University and NBER

Eduardo Morales

Princeton University and NBER

October 22, 2016

Abstract

Much of the variation in international trade volume is driven by firms' extensive margin decisions to participate in export markets. To understand these decisions and predict the sensitivity of export flows to changes in trade costs, we estimate a standard model of firms' export participation. In choosing whether to export, firms weigh the fixed costs of exporting against the forecasted profits from serving a foreign market. We show that the estimated parameters and counterfactual predictions from the model depend heavily on how the researcher specifies firms' expectations over these profits. In response, we adopt a moment inequality approach, placing weaker assumptions on firms' expectations. We use data from Chilean exporters to show that, relative to methods that require specifying firms' information sets, our approach finds fixed export costs that are 70-90% smaller, leading to distinct predictions under counterfactual export promotion policies. Finally, we 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.

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

*We thank Tim Bresnahan, Lorenzo Caliendo, Jan De Loecker, Dave Donaldson, Jonathan Eaton, Liran Einav, Alon Eizenberg, Guido Imbens, Ariel Pakes, Esteban Rossi-Hansberg, James Tybout and seminar participants at the CEPR-JIE conference on Applied Industrial Organization, Columbia University, Dartmouth College, LMU, MIT, McGill University, New York University, the NBER ITI meeting, 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. All errors are our own. Email: michael.dickstein@nyu.edu, ecmorale@princeton.edu.

1 Introduction

In 2014, approximately 300,000 US firms chose to export to foreign markets.[1] The decision of these firms to sell abroad drives much of the variation in trade volume from the US.[2] Thus, to predict how aggregate exports may change with lower trade costs, exchange rate movements, or other policy or market fluctuations, researchers need to understand firms' extensive margin decisions to participate in export markets.

A large literature in international trade focuses on modeling firms' export decisions.[3] Empirical analyses of these decisions, however, face a serious data obstacle: the decision to export depends on a firm's expectations 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 importantly influence the overall measurement, as Manski (1993, 2004), and Cunha and Heckman (2007) show in the context of evaluating the returns to schooling. In the 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; Helpman et al., 2008) under two scenarios: the "perfect foresight" case, under which firms perfectly predict their profits when exporting, and a minimal information specification in which firms only use a specific observed set of variables to predict their export profits. Under each assumption on firms' information, we recover values for the fixed costs of exporting and predict counterfactual changes in exports across markets under a policy that reduces these fixed costs by 40%. 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 approach, firms may gather different signals about their productivity relative to competitors, or about the evolution of exchange rates, trade policy, and foreign demand. Crucially, we do not require the researcher to have full knowledge of each 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

[1] Department of Commerce (2016).

[2] According to Bernard et al. (2010), approximately 70% of the cross-sectional variation in exports comes from firms entering a market rather than changing their export volume.

[3] See for example Das et al. (2007), Eaton et al. (2008), Arkolakis (2010), Moxnes (2010), Eaton et al. (2011), Arkolakis et al. (2014), Eaton et al. (2014), Cherkashin et al. (2015) and Ruhl and Willis (2016). The literature has also recently focused on the decisions of importers (Antràs et al., 2016; Blaum et al., 2016), and on how exporters and importers match (Bernard et al., 2016; Eaton et al., 2016).

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 true parameters of interest. To do so, we develop a new type of moment inequality, which we label the odds-based inequality, and combine it with inequalities based on revealed preference.[4] Using these inequalities, our empirical burden is twofold. First, we must show that placing fewer assumptions on expectations matters both for the estimates of the parameters of the exporter's problem and for the predictions of export flows under counterfactual trade policy. Second, our approach must generate bounds on the model's parameters and on predicted exports that are small enough to be informative.

We perform our empirical analysis in the context of a standard partial equilibrium, two period model of export participation. In our model, firms may obtain different export profits in a country due to differences both in their productivity and in fixed export costs. We later extend this baseline model to account for dynamic incentives, as in Das et al. (2007), and to allow firms to react to firm-country specific export revenue shocks, as in Eaton et al. (2011). We estimate our model using data on Chilean exporters in two manufacturing sectors, the manufacture of chemicals and food products.

We have three main contributions. First, we demonstrate the sensitivity of both the estimated fixed costs and model-based counterfactual predictions to assumptions the researcher imposes on firms' profit forecasts. Specifically, using maximum likelihood, we estimate a perfect foresight model under which firms predict perfectly the revenues they will earn upon entry. Under this assumption, we find export costs in the chemicals sector from Chile to Argentina, Japan, and 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 in which we specify the variables firms use to form their expectations. We specify that firms know: distance to the export market, aggregate exports from Chile to that market in the prior year, and the firm's own productivity from the prior year. We estimate fixed costs of exporting under this approach that are approximately 40-60% smaller than those found under the perfect foresight assumption, in both the chemicals and food sector.

That the fixed cost estimates differ under the two approaches reflects a bias in the estimation. Both require the researcher to specify precisely the content of the agent's information set. If firms actually employ a different set of variables—either more information or less— to

[4] A growing empirical literature employs moment inequalities derived from revealed preference arguments, including Ho (2009), Crawford and Yurukoglu (2012), Ho and Pakes (2014), Eizenberg (2014), Morales et al. (2015), and Wollman (2016). This work generally follows the methodology developed in Pakes (2010) and Pakes et al. (2015); our revealed preference inequalities apply this methodology in a new setting with a distinct error structure. We also combine these inequalities with our odds-based inequalities for additional identification power.

predict their potential export profits, the estimates of the model parameters will generally be biased. Thus, our second contribution is to employ moments 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 productivity 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 approach, we find much lower fixed costs, representing only 10-15% of the perfect foresight values.

Finally, as a third contribution, we use our model's estimates to address the substantive question of "what do exporters know?", and to illustrate the implications for trade policy design. Our empirical approach is well suited for this question. Under rational expectations, our model requires us to specify a set of variables we assume firms use to predict their export revenue. We can therefore 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) for moment inequalities, we can look for evidence against the null that firms use a particular set of variables in their revenue forecasts.

We have several findings. First, we test 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 country specific shifters of potential export revenue. According to our model, these shifters are sufficient statistics for the effect of foreign market characteristics (i.e. size, price index, trade costs and demand shifters) on firms' export revenue. 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 doesn't appear to be driven by prior export experience. That is, even when we focus only on large firms that chose 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.

Finally, we take our new knowledge of exporter's information to examine the implications for trade policy. In particular, we provide bounds that indicate how firms would respond to a policy that reduces the fixed costs of exporting by 40%. Comparing predictions from the perfect foresight model to those computed using our moment inequalities, in the latter we predict very different patterns of export participation. For the United States, for example, we predict 30-40% lower participation growth relative to the perfect foresight estimate. Crucially, when the researcher presumes firms have perfect foresight, the model predicts that those firms

that would earn the most from exporting are the first to participate under the policy change. This assumption will tend to overestimate the export volume response to trade policy, relative to an approach that imposes more realistic assumptions on the firm's information set.

We demonstrate our contributions using the exporter's problem. However, our estimation approach can apply more broadly to many discrete choices in economics that depend on agents' forecasts of key variables. For example, to determine whether to invest in research and development projects, the firm 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 the likely future demand (Bernard et al., 2010; Bilbiie et al., 2012; Arkolakis et al., 2014). Firms deciding whether to enter health insurance markets must also form expectations about the type of health risks that will enroll in their plans (Dickstein et al., 2015) and consumers choosing among insurance plans must form expectations about their future health and financial risks (Handel and Kolstad, 2015). In household finance, a retiree's decision to purchase a private annuity (Ameriks et al., 2015) depends on her expectations about life expectancy and, in education, the decision to attend college crucially depends on potential students' expectations about the difference in lifetime 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 agent's preferences (Manski, 2004), our approach can recover bounds on the economic primitives of the agent's problem without imposing strong assumptions on agents' expectations.

We proceed in this paper by first describing our model of firm exports in Section 2, building up to an expression for firms' export participation decisions. 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 predict the effect on export participation and export volume from a reduction in fixed export costs. In Section 8 we demonstrate how to proceed under various extensions of our baseline model and Section 9 concludes.

2 Export Model

We begin with a model of firms' export decisions. All firms located in country h may choose to sell in every export market j. We index the firms located in h and active at period t by $i = 1, \ldots, N_t$.[5] 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.[6] When choosing among export destinations, firms may differ in

[5] For ease of notation, we will eliminate the subindex for the country of origin h.

[6] In Section 8.1, we consider a fully dynamic export participation model in which forward-looking firms must

their degree of uncertainty about the profits they will obtain upon exporting. In the second period, conditional on entering a foreign market, all firms set their prices optimally.

2.1 Demand, Supply, Market Structure, and Information

Firms face an isoelastic demand in every country: $x_{ijt} = p_{ijt}^{-\eta} P_{jt}^{\eta-1} Y_{jt}$. Here, the quantity demanded x_{ijt} depends on the price firm i sets in destination j at t, denoted p_{ijt}, on the total expenditure in the sector in which i operates, denoted Y_{jt}, and on a price index, P_{jt}, which captures the competition that firm i faces in market j from all other firms selling in this market. Conditional on entering market j in year t, firms are assumed to set p_{ijt} optimally taking P_{jt} as given; this specification implies that every firm faces a constant demand elasticity equal to η in every destination.

Firm i produces one unit of output with a constant marginal cost c_{it}. Here, c_{it} is a function of both the cost of a bundle of inputs to firm i at time t and of the number of bundles of inputs that firm i uses to produce one unit of output. We impose no restriction on the joint distribution of c_{it} across firms or time periods. When i chooses to sell in a foreign market j, it must pay two additional export costs: a variable cost, τ_{jt}, and a fixed cost, f_{ijt}. We adopt the "iceberg" specification of variable export costs and thus assume that firm i must ship τ_{jt} units of output to country j for one unit to arrive. The total marginal cost for firm i of exporting one unit to country j at period t is therefore $\tau_{jt} c_{it}$. Fixed export costs f_{ijt} are paid by firms exporting a positive amount to j at t and are independent of the quantity exported.[7]

We denote the firm's potential export 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 export revenue. Thus, $\mathcal{J}_{ijt}$ includes all variables that firm i uses to predict r_{ijt}. We also assume that, at the time it chooses its export destinations for period t, the firm i knows the determinants of fixed costs of exporting f_{ijt}, $(dist_j, \nu_{ijt})$, for every country j. Therefore, if relevant to predict r_{ijt}, these determinants of fixed export costs will also belong to $\mathcal{J}_{ijt}$.

2.2 Export Revenue

When entering a destination market, every seller observes both η and his marginal cost of exporting and sets his price p_{ijt} optimally; $p_{ijt} = (\eta/(\eta-1))\tau_{jt} c_{it}$. The demand and supply assumptions together imply that the optimal revenue firm i would obtain if it were to export to $j$in year t is

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

also pay a sunk export entry cost à la Das et al. (2007).

[7] In Section 8.2.1, we introduce variable trade costs τ_{jt} that vary by firm, destination, and year.

Potential export revenue for entrants is thus a function of (1) market size in the destination market, Y_{jt}; (2) competition by other suppliers, as captured by the price index, P_{jt}; (3) marginal production costs, c_{it} and, (4) variable export costs, τ_{jt}. The set $\mathcal{J}_{ijt}$ thus includes all variables firm i knows when deciding whether to export and uses to predict any of these four variables. We can rewrite this revenue in more simplified form as:[8]

$$r_{ijt} = \alpha_{jt} r_{iht}, \qquad \text{with} \qquad \alpha_{jt} = \left(\frac{\tau_{jt}}{\tau_{ht}} \frac{P_{ht}}{P_{jt}}\right)^{1-\eta} \frac{Y_{jt}}{Y_{ht}}. \tag{2}$$

In this form, α_{jt} is a destination-year specific shifter of export revenues that accounts for the impact of variable trade costs, market size and the price index. In a later extension to the model, we allow for firm-country-year export revenue shocks, such that $r_{ijt} = \alpha_{jt} r_{iht} + \omega_{ijt}$.[9]

Our model does not restrict the relationship between the information set firm i uses to predict r_{ijt} and the firm's marginal cost, c_{it}, the price index, P_{jt}, market size, Y_{jt}, or trade costs, τ_{jt}. Our framework therefore permits firms to face different degrees of uncertainty in each market. For example, we can allow more productive firms to be systematically better informed than less productive firms about their export profitability in foreign markets, or for all firms to have more information about markets that are closer to the market of origin. Similarly, firms' uncertainty may also vary freely across time periods.

2.3 Export Profits

Given that every potential exporter i is monopolistically competitive and faces constant marginal export costs in every destination market j and period t, the export profits that i would obtain in j if it were to export at t are

$$\pi_{ijt} = \eta^{-1} r_{ijt} - f_{ijt}. \tag{3}$$

We model fixed export costs as:

$$f_{ijt} = \beta_0 + \beta_1 dist_j + \nu_{ijt}, \tag{4}$$

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.[10] As discussed in Section 2.1, we assume that firms know f_{ijt}.

[8] We show the full derivation in Appendix A.1.

[9] In sections 8.2.1 and 8.2.2, we discuss two cases, respectively: (1) the firm does not have any information on firm-country-year specific export revenue shocks when deciding whether to export, so that $\mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}] = 0$, and (2) the firm anticipates these export revenue shocks, such that $\mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}] = \omega_{ijt}$.

[10] In Appendix B.2, we generalize the specification in equation (4) and present estimates for a model in which we assume fixed export costs to be $f_{ijt} = \beta_j + \nu_{ijt}$, where β_j is a country j specific fixed effect.

The estimation procedure introduced in Section 4.2 requires ν_{ijt} to be independently distributed of $\mathcal{J}_{ijt}$ and that its distribution is known up to a scale parameter and is log-concave.[11] To match the typical discrete choice model, we assume that ν_{ijt} is distributed normally and independently of other determinants of the export participation decision:

$$\nu_{ijt}|(\mathcal{J}_{ijt}, dist_j) \sim \mathbb{N}(0, \sigma^2). \tag{5}$$

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

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 variable in the information set of firm i used to predict the potential export profit in country j. Combining equations (3) and (4), we can write

$$\mathbb{E}[\pi_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt}, \tag{6}$$

Here, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ by our definition of $\mathcal{J}_{ijt}$ and ν_{ijt}.

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 (6), we can write

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\}, \tag{7}$$

and, given equations (5) and (7), we can write the probability that i exports to j at t conditional on $\mathcal{J}_{ijt}$ and $dist_j$ as

$$\begin{aligned}\mathcal{P}_{jt}(d_{ijt} = 1|\mathcal{J}_{ijt}, dist_j) &= \int_\nu \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu \geq 0\}\phi(\nu)d\nu \\ &= \Phi\big(\sigma^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big),\end{aligned} \tag{8}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the standard normal probability density function and cumulative distribution function. Equation (8) 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 expectations of

[11] More precisely, while for the odds-based moment inequalities introduced in Section 4.2.1 we require the distribution of ν_{ijt} to be log-concave; for the revealed-preference moment inequalities introduced in Section 4.2.2 we need the left and right-truncated expectations of ν_{ijt} to be convex. Both the normal and the logistic distribution satisfy these two conditions. Other log-concave distributions include the uniform, exponential, type I extreme value and laplace (or double exponential) distributions. Pratt (1981), Heckman and Honoré (1990), and Bagnoli and Bergstrom (2005) provide more information on the properties of log-concave distributions.

the revenue it will earn in j at t upon entry. The key complication to estimation, which we discuss in Section 4, is that researcher rarely observe these expectations.

From equation (8), we have four parameters to estimate: $(\sigma, \eta, \beta_0, \beta_1)$. However, the form of equation (8) implies that, even were we to observe firms' actual expectations, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, data on export choices alone do not allow us to identify the scale of this parameter vector. That is, if we multiply these four parameters by the same positive number, the probability in equation (8) remains constant. To normalize for scale in export entry models, researchers typically use additional data to estimate or calibrate the demand elasticity η (e.g. Das et al., 2007) . In our estimation, we set $\eta = 5$.[12] For simplicity of notation, we use θ to denote the remaining parameter vector and θ^* to denote its true value; i.e.

$$\theta^* \equiv (\beta_0, \beta_1, \sigma). \tag{9}$$

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 collect the annual survey data for the same years observed in the customs data. We merge these two datasets using firm identifiers, allowing us to examine the export participation and export volume of each firm along with their domestic activity.[13]

The firms in our dataset operate in one of two sectors: the manufacture of chemicals and the food products sector.[14] These are the two largest Chilean export manufacturing sectors by volume. 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. Across the time period used in our empirical analysis, 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

[12] This elasticity of substitution is within the range of values estimated in the literature. See, for example, Simonovska and Waugh (2014) and Head and Mayer (2014) and the references cited therein.

[13] We aggregate the information from ENIA across plants in order to obtain firm-level information that matches the customs data. There are some cases in which firms are identified as exporters in ENIA but do not have any exports listed with customs. In these cases, we assume that the customs database is more accurate and thus identify 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. Nevertheless, the remaining firms account for roughly 80% of total export flows.

[14] 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 on average $2.18 million in 1996 and grow to $3.58 million in 2005, with a dip in 2001 and 2002.[15] The median level of exports is much lower, at around $120,000 to $200,000. In the food sector, we observe 372 unique firms, 30% of which export in a typical year. The mean exporter in this sector is much larger, with an average revenue across years of $7.7 million per exporter. The median exporter across all years sells approximately $2.24 million abroad. Relative to the chemicals sector, firms in the food sector also typically export to a greater number of destination markets, typically 5-6 markets on average. In the chemicals sector, the average exporter serves 3-4 countries.

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 USD in year 2000 terms.

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.[16]

[15] The revenue values we report are in year 2000 US dollars.

[16] Available at http://www.cepii.fr/anglaisgraph/bdd/distances.htm. Mayer and Zignago (2006) provide a detailed explanation of the content of this database.

4 Empirical Approach

In the model we describe in Section 2, firm i's potential export revenue to destination market j at time t, r_{ijt}, is a function of its own marginal production costs, and of country j's market size, price index, and trade barriers. Firms may know only some of these variables when deciding whether to export; they therefore base their export participation decisions on expectations of potential export revenues conditional on their information set, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$. In the theoretical model, we did not impose assumptions on the content of this information set. However, as Manski (1993) illustrates, identifying the parameter vector θ and performing counterfactuals requires placing restrictions on $\mathcal{J}_{ijt}$.[17]

We discuss three alternative empirical approaches to recover the parameters of the firm's export decision when these decisions depend on unobserved expectations. First, we consider a perfect foresight model. With this model, researchers assume an information set $\mathcal{J}^a_{ijt}$ for potential exporters such that $\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}] = r_{ijt}$.[18] That is, firms predict r_{ijt} perfectly.

In a second model, we allow firms to face some uncertainty—for example, they may lack perfect knowledge of the size of the market or the degree of competition they will face. In this empirical 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) aggregate exports to the destination country 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; specifically, $\mathcal{J}^a_{ijt} = (r_{iht-1}, R_{jt-1}, dist_j)$. These three variables are easily accessible by every firm in any year t. However, this information set is likely to be strictly smaller than the actual information set firms possess when deciding on the set of export destination countries.[19] Furthermore, specifying $\mathcal{J}^a_{ijt}$ as above also 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.[20]

[17] Specifically, Manski (1993) shows that different assumptions on $\mathcal{J}_{ijt}$ might generate identical likelihood functions for a given set of reduced form parameters. In these cases, one cannot use goodness-of-fit measures to discriminate among these different assumptions on agents' information sets. However, each reduced form parameter has a different structural interpretation under each of these assumptions. Therefore, even in this cases, different assumptions on $\mathcal{J}_{ijt}$ will imply different counterfactual predictions.

[18] The assumption of perfect foresight is common in static general equilibrium models of export and import participation. E.g. Arkolakis (2010), Eaton et al. (2011), Arkolakis et al. (2014), Antràs et al. (2016). The model described in Section 2 is partial equilibrium. Extending our flexible treatment of firms' information sets to general equilibrium models is not trivial and we therefore leave it for future research.

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

[20] The assumption that the true information set of potential exporters $\mathcal{J}_{ijt}$ is identical to a vector of observed covariates is common in firm-level empirical analysis of export participation. E.g. Roberts and Tybout (1997), Bernard and Jensen (2004). Some other studies allow $\mathcal{J}_{ijt}$ to be unobserved by the econometrician, but then need to specify the exact parametric form of its distribution; see for example Das et al. (2007). Our moment inequality approach in Section 4.2 allows $\mathcal{J}_{ijt}$ to be unobserved and imposes no distributional assumption on it.

Finally, third, we discuss how to identify the model parameters and perform counterfactuals in a discrete choice context without imposing strong assumptions on firms' information sets. 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 $Z_{ijt} \subseteq \mathcal{J}_{ijt}$.[21] 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.

To proceed in estimation under any of the three approaches described above, we need an observed measure of ex post revenues. We need this measure of revenue both for firms that choose to export in the data and those who do not. Since customs data does not include measures of observed revenue for non-exporters, we use our model to generate an appropriate measure. According to our model, $r_{ijt} = \alpha_{jt} r_{iht}$, where r_{iht} reflects the firm's revenues in the home market of Chile. Since we observe r_{iht}, to obtain a measure of r_{ijt} for every firm, country and year we need a measure of α_{jt} for every country and year. We show how to estimate α_{jt} consistently in Appendix A.2. In brief, if we observe export revenues, r_{ijt}^{obs} for exporters. Allowing for error, e_{ijt}, in our observed measure of revenue, we can write

$$r_{ijt}^{obs} = d_{ijt}(r_{ijt} + e_{ijt}), \tag{10}$$

where again $d_{ijt} = 1$ if firm i exports to country j in period t. Our model therefore predicts that $r_{ijt}^{obs} = d_{ijt}(\alpha_{jt} r_{iht} + e_{ijt})$. As long as the mean of e_{ijt}, the measurement error in export revenues, is independent of a firm's domestic revenue and export decision and is equal to zero, the following moment then point identifies α_{jt} for every j and t:

$$\mathbb{E}_{jt}[r_{ijt}^{obs} - \alpha_{jt} r_{iht} | r_{iht}, d_{ijt} = 1] = 0, \tag{11}$$

where $\mathbb{E}_{jt}[\cdot]$ denotes the expectation across firms in a given country-year pair jt.

4.1 Perfect Knowledge of Exporters' Information Sets

Under the assumption that the econometrician's specified information set, $\mathcal{J}_{ijt}^{a}$, equals the firm's true information set, $\mathcal{J}_{ijt}$, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}^{a}]$ is a perfect proxy for $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and one can identify θ^* as the value of the unknown parameter $\theta = (\theta_0, \theta_1, \theta_2)$ that maximizes a standard

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

log-likelihood function

$$\mathcal{L}(\theta|d, \mathcal{J}^a_{ijt}, dist) =$$
$$\mathbb{E}\Big[\sum_{j,t} d_{ijt} \log\big(\mathcal{P}(d_{jt}=1|\mathcal{J}^a_{ijt}, dist_j;\theta)\big) + (1-d_{ijt})\log\big(\mathcal{P}(d_{jt}=0|\mathcal{J}^a_{ijt}, dist_j;\theta)\big)\Big], \quad (12)$$

where the expectation is taken over firms, and[22]

$$\mathcal{P}(d_{jt}=1|\mathcal{J}^a_{ijt}, dist_j;\theta) = \Phi\big(\theta_2^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}] - \theta_0 - \theta_1 dist_j\big)\big). \quad (13)$$

The key assumption underlying this procedure is that the researcher correctly specifies the agent's information set. We denote the difference between firms' true expectations and the researcher's proxy as ξ_{ijt}:

$$\xi_{ijt} \equiv \mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}] - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]. \quad (14)$$

Whenever ξ_{ijt} differs from zero, the estimator of θ under this procedure will be biased. Biased estimates of θ will imply biased estimates of fixed export costs and incorrect model predictions in our counterfactual exercise. In Appendix D, we present simulation results that illustrate the direction of the bias in the θ estimates in three cases: when the researcher assumes perfect foresight, when the researcher's information set is smaller than the firm's true information set, and when the researcher specifies an information set that is larger than the firm's information. Here, we provide intuition for the perfect foresight case.

To discuss the bias under perfect foresight, we first define $\varepsilon_{ijt} \equiv r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ as the true expectational error that firm i makes when predicting its export revenue upon entry. If a researcher assumes perfect foresight, then the researcher's error, defined in equation 14, coincides with the firm's error–i.e. $\xi_{ijt} = \varepsilon_{ijt}$. Bias in the estimates of θ therefore arises in instances in which $\varepsilon_{ijt} \neq 0$. Yatchew and Griliches (1985) analyzed bias of this form under particular distributional assumptions. Specifically, if firms' true expectations are normally distributed, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] \sim \mathbb{N}(0, \sigma_e^2)$, and the expectational error is also normally distributed, $\varepsilon_{ijt}|(\mathcal{J}_{ijt}, \nu_{ij}) \sim \mathbb{N}(0, \sigma_\varepsilon^2)$, there is an upward bias in the estimates of the fixed costs parameters β_0, β_1 and σ. Furthermore, the upward bias increases in the variance of the expectational error, σ_ε^2, relative to the variance of the true unobserved expectations, σ_e^2. That is, the worse the researcher's proxy for the true expectations, the greater the bias. When either firms' true expectations or the expectational error are not normally distributed, there is no analytic

[22] To use the expressions in equations (12) and (13) to estimate θ, one first needs to compute $\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$. When the researcher assumes perfect foresight, $\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}] = \alpha_{jt} r_{iht}$, where α_{jt} is identified according to equation (11). When the researcher assumes $\mathcal{J}^a_{ijt}$ is equal to a set of observed covariates, $\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$ is the outcome of projecting $\alpha_{jt} r_{iht}$ on the observed vector $\mathcal{J}^a_{ijt}$. See Manski (1991) and Ahn and Manski (1993) for additional details on this two-step estimation approach.

expression for the bias of the maximum likelihood estimator of θ. However, our simulations in Appendix D illustrate that the upward bias in the estimates of β_0, β_1 and σ from wrongly assuming perfect foresight appears fairly general.[23]

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 smaller vector of covariates in their data that represent 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}; one can find the former in firms' accounting statements, while the latter appears in publicly available trade data. Similarly, firms can also easily obtain information on the distance to each destination country, $dist_j$, which might potentially affect variable trade costs. Thus, while $(r_{iht-1}, R_{it-1}, dist_j)$ might not reflect firms' complete information sets, firms 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 represents a subset of the information firms use to forecast export revenues, i.e. $Z_{ijt} \subseteq \mathcal{J}_{ijt}$. We form two types of moment inequalities that partially identify the parameters of the firm's entry decision, θ.[24]

[23] The intuition for upward bias in the maximum likelihood estimates of β_0, β_1, and σ caused by wrongly assuming perfect foresight shares the same basis as the well-known attenuation bias affecting OLS estimates in linear models when a covariate is affected by classical measurement error (see page 73 in Wooldridge (2002)). Rational expectations implies that firms' expectational errors are mean independent of their unobserved true expectation and, therefore, correlated with the ex-post realization of the variable whose expectation affects firms' decisions; i.e. rational expectations implies that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}] = 0$ and $cov(\varepsilon_{ijt}, r_{ijt}) \neq 0$. Thus, if we were in a linear regression setting, wrongly assuming perfect foresight and using the ex-post realized revenue, r_{ijt}, as a regressor instead of the unobserved expectation, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, would generate a downward bias on the coefficient on r_{ijt}. The probit model in equation (13) differs from this linear setting in two dimensions. First, we normalize the scale of the unknown parameter vector by setting the coefficient on the covariate measured with error, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, to a given value. This implies that the bias generated by the correlation between the expectational error, ε_{ijt}, and the realized export revenue, r_{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_{ijt} but also on the functional form of the distribution of unobserved expectations and expectational error.

[24] As shown in Appendix A.3, 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 the parameter vector θ. Whether the bounds defined by the inequalities in Sections 4.2.1 and 4.2.2 are sharp is left for future research. However, as the results in Section 5 show, in our empirical application, they generate bounds that are small enough to be informative.

4.2.1 Odds-based Moment Inequalities

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

$$\mathcal{M}^{ob}(Z_{ijt};\theta) = \mathbb{E}\left[\begin{array}{c} m_l^{ob}(d_{ijt}, r_{ijt}, dist_j;\theta) \\ m_u^{ob}(d_{ijt}, r_{ijt}, dist_j;\theta) \end{array}\middle| Z_{ijt}\right] \geq 0, \tag{15a}$$

where the two moment functions are defined as

$$m_l^{ob}(\cdot) = d_{ijt}\frac{1-\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}-\theta_0-\theta_1 dist_j\big)\big)}{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}-\theta_0-\theta_1 dist_j\big)\big)} - (1-d_{ijt}), \tag{15b}$$

$$m_u^{ob}(\cdot) = (1-d_{ijt})\frac{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}-\theta_0-\theta_1 dist_j\big)\big)}{1-\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}-\theta_0-\theta_1 dist_j\big)\big)} - d_{ijt}. \tag{15c}$$

We denote the set of all possible values of the parameter vector θ as Θ and the subset of those that are consistent with the conditional moment inequalities described in equation (15) as Θ_0^{ob}. As in earlier sections, we denote the true parameter vector as $\theta^* = \big(\beta_0, \beta_1, \sigma\big)$. The following theorem contains the main property of the inequalities defined in equation (15):

Theorem 1 *Let θ^* be the parameter defined by equation* (9)*. Then $\theta^* \in \Theta_0^{ob}$.*

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

We focus on intuition behind the moment function in equation (15c); the intuition for (15b) is analogous. In brief, if firm i decides to export to j in period t, by revealed preference we know that it must have been profitable to do so: $\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0$ for firms that chose to export. We can use the definition of the export dummy d_{ijt} in equation (7) to write:

$$\mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\} - d_{ijt} \geq 0, \tag{16}$$

We then take the expectation of this inequality conditional on $(d_{ijt}, \mathcal{J}_{ijt}, dist_j)$.[25] Given the distributional assumption in equation (5), we can use simple algebraic transformations to rewrite the resulting inequality as

$$\mathbb{E}[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))} - d_{ijt}|\mathcal{J}_{ijt}] \geq 0. \tag{17}$$

If we were to write this inequality as a function of the unknown parameter θ, it would only hold at its true value θ^*. However, it cannot be used directly for identification because it depends on the unknown true information set, $\mathcal{J}_{ijt}$, and the unobserved expectation, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$.

[25] Pakes and Porter (2014) exploit the expectation of the inequality in equation (16) in settings with more than two choices. In their set-up, agents have perfect foresight on all payoff relevant variables.

We show in Appendix C.1 that, given the rational expectations assumption and the convexity of $\Phi(\cdot)/(1-\Phi(\cdot))$, we can use Jensen's inequality to derive weaker moments than those in equation (17) by plugging r_{ijt} in lieu of $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and by conditioning both moments on Z_{ijt} instead of $\mathcal{J}_{ijt}$. Consequently, the moment inequalities in equation (15) will generally only partially identify the parameter vector of interest.[26]

There is one case in which the inequalities equation (15) will point-identify the parameter of interest: when agents have perfect foresight, $r_{ijt} = \mathbb{E}[r_{it}|\mathcal{J}_{ijt}]$, and the vector of instruments Z_{ijt} the researcher chooses includes all variables that agents use to predict either the ex post revenues or the fixed export costs, i.e. $Z_{ijt} = (\mathcal{J}_{ijt}, dist_j)$, the set Θ_0^{ob} is a singleton and identical to the true value of the parameter vector, θ^*. As we deviate from this special case, the size of the set Θ_0^{ob} increases monotonically in the variance of the expectational error, $\varepsilon_{ijt} \equiv r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$.

The moment functions in equations (15b) and (15c) are not redundant. Take for example, the identification of the parameter θ_0. Given observed values of d_{ijt}, r_{ijt}, and $dist_j$, and given any arbitrary value of the parameters θ_1 and θ_2, the moment function $m_l^{ob}(\cdot)$ in equation (15b) is increasing in θ_0 and, therefore, will identify a lower bound on θ_0. With the same observed values, $m_u^{ob}(\cdot)$ in equation (15c) is decreasing in θ_0 and will thus identify an upper bound on θ_0. The same intuition applies for identifying upper and lower bounds for θ_1 and θ_2.

4.2.2 Revealed Preference Moment Inequalities

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

$$\mathcal{M}^r(Z_{ijt};\theta) = \mathbb{E}\left[\begin{array}{c} m_l^r(d_{ijt}, r_{ijt}, dist_j;\theta) \\ m_u^r(d_{ijt}, r_{ijt}, dist_j;\theta) \end{array}\middle| Z_{ijt}\right] \geq 0, \tag{18a}$$

where the two moment functions are defined as

$$m_l^r(\cdot) = -(1-d_{ijt})\big(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j\big) + d_{ijt}\theta_2 \frac{\phi\big(\theta_2^{-1}(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j)\big)}{\Phi\big(\theta_2^{-1}(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j)\big)}, \tag{18b}$$

$$m_u^r(\cdot) = d_{ijt}\big(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j\big) + (1-d_{ijt})\theta_2 \frac{\phi\big(\theta_2^{-1}(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j)\big)}{1-\Phi\big(\theta_2^{-1}(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j)\big)}. \tag{18c}$$

We denote subset of values in the parameter space for θ that are consistent with the conditional moment inequalities in equation (18) as Θ_0^r. The following theorem contains the main property of the set Θ_0^r:

Theorem 2 *Let θ^* be the parameter defined by equation* (9)*. Then $\theta^* \in \Theta_0^r$.*

[26] The key assumption needed for this result is that both $\Phi(\cdot)/(1-\Phi(\cdot))$ and $(1-\Phi(\cdot))/\Phi(\cdot)$ are globally convex. The assumption of normality of the distribution of ν is therefore sufficient but not necessary to derive the inequalities in equation (15): for any distribution of ν with cumulative distribution function $F_\nu(\cdot)$, the odds-ratios $F_\nu(\cdot)/(1-F_\nu(\cdot))$ and $(1-F_\nu(\cdot))/F_\nu(\cdot)$ are globally convex as long as the distribution of ν is log-concave.

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 (18b) and (18c) are derived using standard revealed preference arguments. We focus our discussion on moment function (18c); the intuition behind the derivation of moment (18b) is analogous. If firm i decides to export to 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 $(d_{ijt}, \mathcal{J}_{ijt}, dist_j)$, we obtain

$$d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + S_{ijt} \geq 0, \tag{19}$$

where $S_{ijt} = \mathbb{E}[-d_{ijt}\nu_{ijt}|d_{ijt}, \mathcal{J}_{ijt}, dist_j]$. The term S_{ijt} is a selection correction and accounts for the fact that firms might decide whether to export to j at t based partly on determinants of profits that are accounted for by the unobserved (to the researcher) term ν_{ijt}.[27] We cannot directly use the inequality in equation (19) because it depends on the unobserved agents' expectations, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, both directly and through the term S_{ijt}. However, similar to the odds-based inequalities, the inequality in equation (19) becomes weaker if we introduce the observed proxy, r_{ijt}, in the place of the unobserved expectations $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and take the expectation of the resulting expression conditional on an observed vector $Z_{ijt} \subseteq \mathcal{J}_{ijt}$. Consequently, if the inequality in equation (19) holds at the true value of the parameter vector, that in equation (18c) will also hold at $\theta = \theta^*$.[28] As Appendix C.2 shows, the key condition to be able to derive the inequality in equation (18) from that in equation (19) is that $\phi(\cdot)/\Phi(\cdot)$ and $\phi(\cdot)/(1 - \Phi(\cdot))$ are globally convex.[29]

The moment functions in equations (18b) and (18c) follow the revealed preference inequalities introduced in Pakes (2010) and Pakes et al. (2015), and previously applied in Eizenberg (2014) and Morales 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.[30]

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

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

[28] We can similarly derive the inequality in equation (18b) if we start from the revealed preference inequality $-(1 - d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt}\big) \geq 0$.

[29] The assumption of normality of the structural error term is sufficient but not necessary for equation (18) to hold. As long as the distribution of the structural error ν is such that both $f_\nu(\cdot)/F_\nu(\cdot)$ and $f_\nu(\cdot)/(1 - F_\nu(\cdot))$ are globally convex, we may write inequalities that also satisfy Theorem 2.

[30] In our empirical application, we find the standard deviation of ν_{ijt}, σ, 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 (18).

4.2.3 Combining Inequalities for Estimation

For our estimation approach, we combine the odds-based and revealed preference moment inequalities described in equations (15) and (18). As indicated in Section 4.2.1, the set defined by the odds-based inequalities is a singleton 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 $\mathcal{J}_{ijt}$. 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 beyond that provided by the odds-based inequalities.[31]

The set of inequalities we define in equations (15) and (18) condition on particular values of the instrument vector, Z. Exploiting all the information contained in these conditional moment inequalities is computationally challenging.[32] In this paper, we base our inference on a fixed number of unconditional moment inequalities implied by the conditional moment inequalities in equations (15) and (18).[33] We denote the set of values of $\theta \in \Theta$ consistent with our set of 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 parameter estimates, on counterfactual choice probabilities, and also allow us in some cases to reject hypotheses about the specific information firms use to forecast export revenue.

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 to estimate the exporter's fixed costs when we assume perfect foresight. Second, we again use maximum likelihood methods, but under the two-step procedure described in Manski (1991) and Ahn and Manski (1993) in which we project realized revenues on a set of observable covariates that we assume form a firm's information set. Specifically, we assume $\mathcal{J}^a_{ijt} = (r_{iht-1}, R_{jt-1}, dist_j)$. Finally, third, we carry out our moment inequality approach under

[31] How the identified sets defined by each type of inequality compare in size is difficult to characterize generally. As the results in Appendix A.6 show, in our empirical application, the 95% confidence set for the true parameter θ^* jointly defined by the revealed preference and the odds-based inequalities is smaller than the analogous confidence sets that arise if only revealed preference or only odds-based inequalities are used for estimation.

[32] Examples of theoretical papers that exploit all information contained in moment inequality models are: Andrews and Shi (2013), Chernozhukov et al. (2013), Chetverikov (2013), Amstrong (2014), Armstrong (2015), and Amstrong and Chan (2016).

[33] We describe in Appendix A.4 the unconditional moments we use to compute the estimates in Section 5.

the assumption that the firm knows the same three variables as in the two-step approach. However, unlike the two-step approach, the inequalities allow additional unobserved variables to enter the firm's true information set, $\mathcal{J}_{ijt}$, and these variables may vary idiosyncratically by firm, market, and time period.

Before implementing any of these three procedures to estimate the key parameter vector θ, we first need to compute our proxy for export revenue. We describe in Section 4 and Appendix A.2 how to obtain this proxy for revenue, which requires estimating demand shifters, α_{jt}, for each market j and time period t. We report the resulting estimates, $\{\hat{\alpha}_{jt}; \forall j, t\}$ in Appendix B.1 (see Figure B.1 for the estimates corresponding to the chemicals sector and Figure B.2 for the food sector).

5.1 Average Fixed Export Costs

In Table 2, we report the estimates and confidence regions for the parameters of our entry model, normalizing the demand elasticity, η, at a value of five. The first coefficient, σ, represents the standard deviation of the structural error ν_{ijt} affecting the fixed export costs. It therefore 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.

The estimates in Table 2 reveal that the models that require full knowledge of the exporter's information sets 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, the set of parameter values that satisfy the moments imply an added cost of \$142,600 to \$197,100 when the export destination is 10,000 kilometers farther in distance. Under the two maximum likelihood procedures, the estimates of the added cost equal \$1,087,800 and \$447,100 for the same added distance.

Table 2: Parameter estimates

	Chemicals			Food		
Estimator	σ	β_0	β_1	σ	β_0	β_1
Perfect Foresight	1,038.6	745.2	1,087.8	1,578.1	2,025.1	214.5
(MLE)	(11.7)	(8.9)	(12.9)	(16.9)	(3.7)	(23.6)
Minimal Information	395.5	298.3	447.1	959.9	1,259.3	129.4
(MLE)	(2.6)	(2.2)	(6.1)	(8.1)	(2.2)	(18.1)
Moment Inequality	[85.1, 117.6]	[62.8, 82.4]	[142.6, 197.1]	[114.9, 160.0]	[167.1, 264.0]	[36.4, 81.3]

Notes: All parameters are reported in thousands of year 2000 USD and are conditional on the assumption that $\eta = 5$. 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. These confidence sets are projections of a confidence set for $(\beta_0, \beta_1, \sigma)$ computed according to the procedure described in Appendix A.5.

The moment inequality bounds on each of the elements of the parameter vector θ re-

ported in Table 2 arise from projecting a three-dimensional 95% confidence set for the vector $(\beta_0, \beta_1, \sigma)$, computed following the procedure in Andrews and Soares (2010).[34] In Appendix A.5 we describe our implementation in detail.[35] In Appendix A.6, we show 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 in our specification that combines both types of inequalities.

We translate the coefficients reported in Table 2 into estimates of the average fixed costs of exporting by country. To start, we report the results in Table 3 for three countries (Argentina, Japan, and the United States) out of the 22 destinations in the chemicals sector and 34 countries in the food sector used in our estimation; we show the results for all countries in graphical form in Figure 1 below. 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 fixed export costs.

Under perfect foresight, we estimate the 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 fixed cost estimates under perfect foresight in these three countries equal \$2.05 million, \$2.40 million, and \$2.20 million, respectively. As Table 4 shows, comparing the estimates under perfect foresight to the estimates that assume an information set that is likely to be "too small", the latter produces entry cost estimates that are about 60% smaller in the chemicals sector and 38% smaller in the food sector. This is consistent with the discussion of the relative bias of the maximum likelihood estimates in Section 4.1.

Under our moment inequality estimator, we find 95% confidence sets for the fixed costs of exporting in the chemicals sector between \$79,100 and \$104,100 for Argentina, \$309,200 and \$420,500 for Japan, and \$181,300 and \$243,600 for the United States.[36] Across Argentina, Japan, and the US, in both the chemicals and food sector, the estimated bounds we find from the inequalities equal only a fraction of the perfect foresight estimates, with a level between 85% and 91% smaller than the perfect foresight values. Comparing the bounds of the fixed costs from the inequalities to the estimates from the two-step approach, reported

[34] Formally, denoting $\hat{\Theta}^{95\%}$ as the 95% confidence set for the vector $(\beta_0, \beta_1, \sigma)$, the confidence set for β_0 in Table 2, 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 $\hat{\Theta}^{95\%}$.

[35] Our reported confidence sets for β_0, β_1 and σ are confidence sets for a subvector of θ^*. Bugni et al. (2016) introduce a new inference procedure that dominates our projection-based inference of each of the parameters β_0, β_1 and σ in terms of power. We report here confidence sets based on the projection of the confidence set $\hat{\Theta}^{95\%}$ 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 $\hat{\Theta}^{95\%}$.

[36] We compute the confidence sets for the average fixed costs for country j, $\bar{f}_j = \beta_0 + \beta_1 dist_j$, by projecting the confidence set for θ^*, $\hat{\Theta}^{95\%}$. Specifically, we compute the lower bound on $\bar{f}_j$ for each country j as $\min_{\theta \in \hat{\Theta}^{95\%}} \theta_0 + \theta_1 dist_j$ and the upper bound as $\max_{\theta \in \hat{\Theta}^{95\%}} \theta_0 + \theta_1 dist_j$.

Table 3: Average fixed export costs

Estimator	Chemicals Argentina	Chemicals Japan	Chemicals United States	Food Argentina	Food Japan	Food United States
Perfect Foresight (MLE)	868.0 (51.7)	2,621.4 (159.4)	1,645.0 (97.6)	2,049.3 (87.2)	2,395.1 (103.9)	2,202.5 (93.5)
Minimal Information (MLE)	348.7 (12.9)	1,069.4 (40.9)	668.1 (24.2)	1,273.9 (43.1)	1,482.4 (50.3)	1,366.3 (45.5)
Moment Inequality	[79.1, 104.1]	[309.2, 420.5]	[181.3, 243.6]	[175.6, 270.1]	[269.1, 361.0]	[227.3, 308.9]

Notes: All parameters are reported in thousands of year 2000 USD and are conditional on the assumption that $\eta = 5$. 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. These confidence sets are projections of a confidence set for $(\beta_0, \beta_1, \sigma)$ computed according to the procedure described in Appendix A.5.

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

Estimator	Chemicals Argentina	Chemicals Japan	Chemicals United States	Food Argentina	Food Japan	Food United States
Minimal Info.	40.2%	40.8%	40.6%	62.1%	61.8%	62.0%
Moment Ineq.	[9.1%, 11.9%]	[11.0%, 14.8%]	[11.8%, 16.3%]	[8.6%, 13.1%]	[10.3%, 14.0%]	[11.2%, 15.0%]

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

in Table 3, again the bounds are much smaller. The results are in line with the discussion in Section 4.1 and Appendix D.3 of the bias that arises if the researcher incorrectly assumes firms have perfect foresight. Here, we observe that assuming the specific minimal information set $\mathcal{J}^a_{ijt} = (r_{iht-1}, R_{jt-1}, dist_j)$ also appears to generate an upward bias in the estimates of the fixed costs.[37]

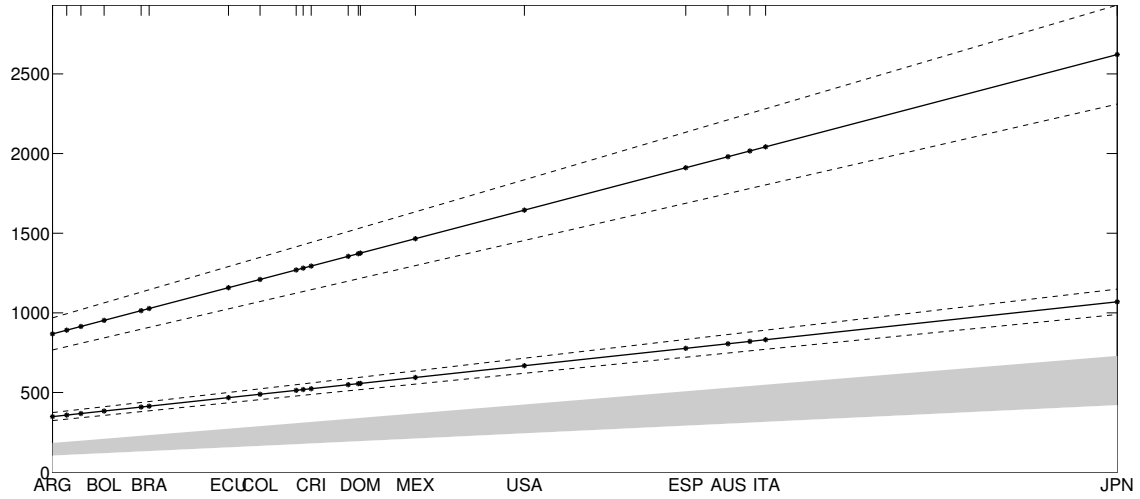
It may seem counterintuitive that the maximum likelihood estimates obtained under the assumption $\mathcal{J}^a_{ijt} = (r_{iht-1}, R_{it-1}, dist_j)$ are not contained in the confidence set computed under the assumption that $(r_{iht-1}, R_{it-1}, dist_j) \subseteq \mathcal{J}_{ijt}$. However, as we discuss in Section 4.2.3, the key property of the identified set Θ_0 is that it is guaranteed to contain the maximum likelihood estimate of θ only if that likelihood function uses the correct information set. There is no guarantee that every information set $\mathcal{J}^a_{ijt}$ consistent with our assumption that $(r_{iht-1}, R_{jt-1}, dist_j) \subseteq \mathcal{J}_{ijt}$ must generate a likelihood function whose maximand is contained in Θ_0.

Returning to the full sample of countries, Figure 1 illustrates that the results we show in Tables 3 and 4 hold for all countries in our sample. In both figures, the vertical axis indicates average fixed export costs in thousands of USD and the horizontal axis indicates the distance between Chile and each export destination. In the figures, the maximum likelihood estimates

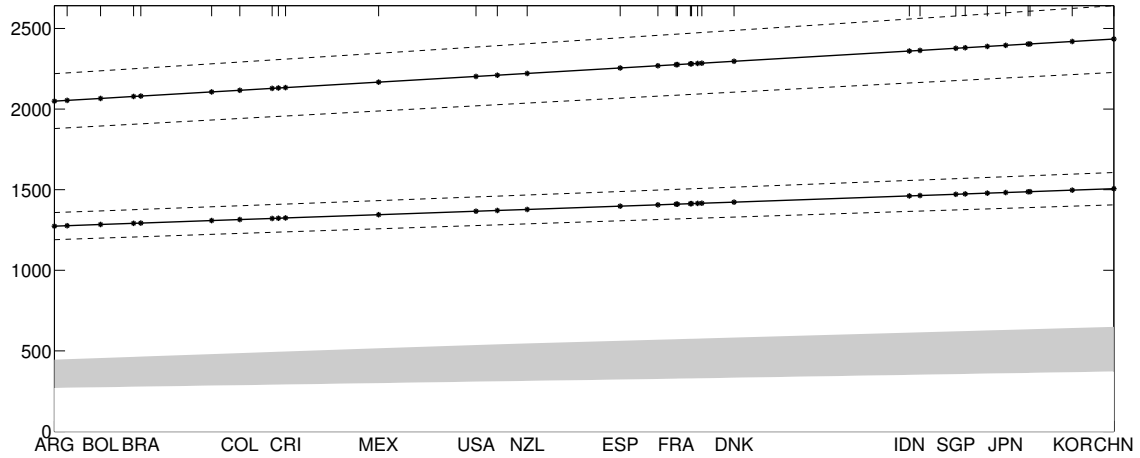
[37] This upward bias is consistent with the simulation in Appendix D.5 in which the distribution of the difference between the true expectation, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, and the one implied by the minimal information set, $\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$, is not symmetric.

Figure 1: Country-specific fixed costs estimates

(a) Chemicals



(b) Food



In both figures, the light-grey shaded area denotes the 95% confidence set generated by our moment inequalities. In panels (a) and (b), the continuous black lines correspond to the ML point estimates under the perfect foresight (upper line) and the minimal information assumption (lower line). The dotted black lines denote the bounds of the corresponding ML 95% confidence intervals.

are always larger than the upper bound of the confidence set.[38]

[38] In Appendix B.2, we generalize the parametric assumptions imposed in equation (4) on how average fixed export costs vary across countries and estimate instead average fixed costs for each country j as country fixed effects. Confidence sets and confidence intervals are larger in this case, reflecting the larger number of parameters to estimate. The qualitative results are similar.

5.2 Distribution of Fixed Export Costs

Table 3 and Figure 1 focus on estimates of the average fixed export costs. However, given the distributional assumptions in equations (4) and (5), we can also compute both maximum likelihood estimates and moment inequality bounds on the fixed costs of exporting for any quantile of the distribution of these costs. In Figure 2, we illustrate the distribution of these fixed export costs in Argentina, Japan and the United States for both the chemical and food sectors. The figures illustrate that both (a) the distance between the two maximum likelihood estimates and (b) 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.

6 Testing Content of Exporters' Information Sets

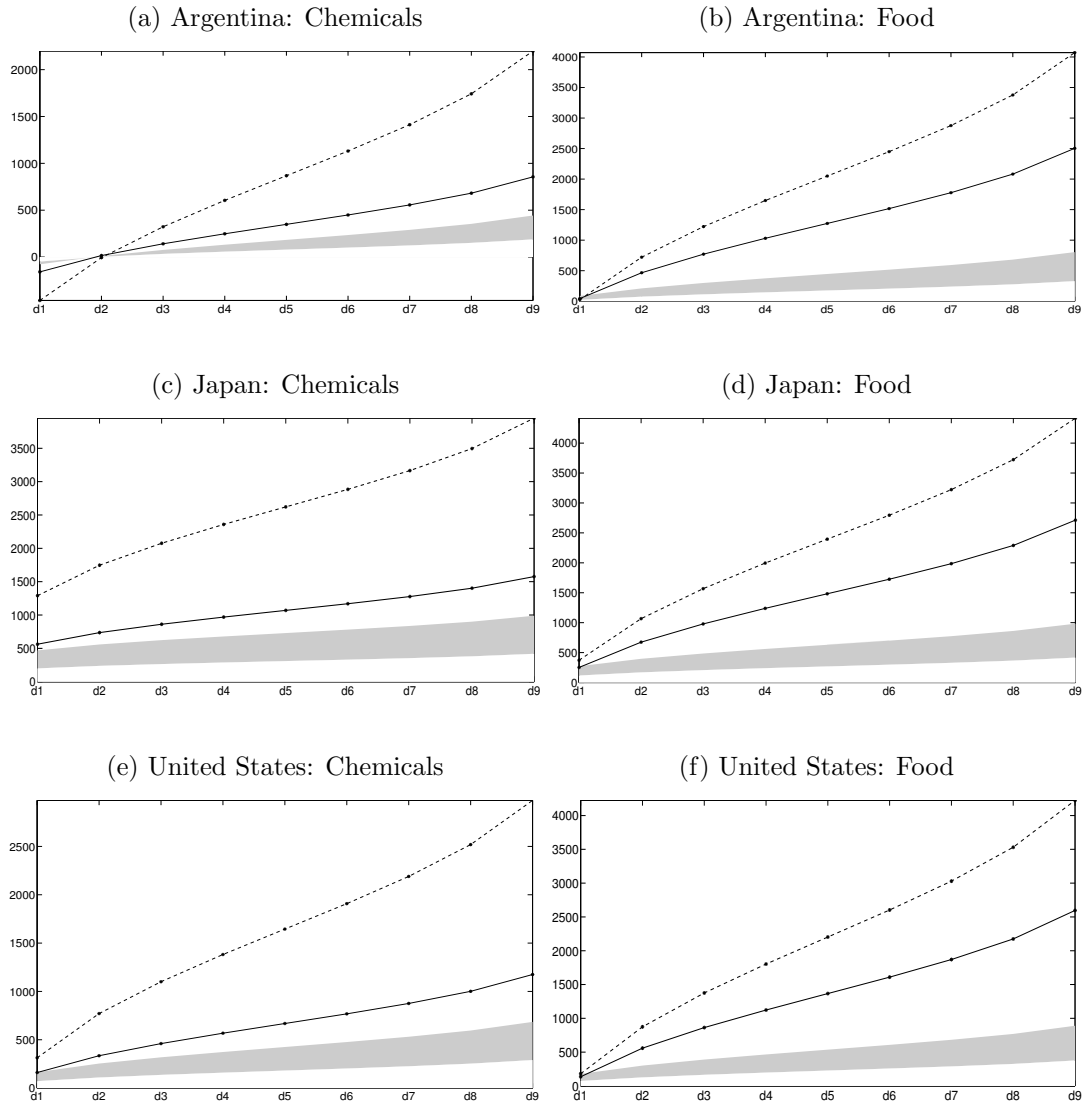
With our estimated parameters, we now can address the question "what do exporters know?" and to illustrate the implications of exporters' information for the design of trade policy. To examine exporters' information sets, we exploit an implication of our empirical model: under rational expectations, any variable in the information set the firm uses to predict revenue serves as an instrument in our empirical moments. Thus, we can define alternative sets of observed variables, labeled Z_{ijt}, as being in the firm's information set, and then use the specification test in Bugni et al. (2015) to test the null hypothesis that there exists a value of the parameter vector that rationalizes our 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 in equation (8) does not hold in the data or that (b) the assumption that the set of observed variables Z_{ijt} we specified are not all elements of the firm's information set. To distinguish between these two conclusions, we repeat our test with the same underlying model but different observables Z_{ijt}.[39]

The p-values for the different vectors Z_{ijt} that we test appear in Table 5. First, we test our main specification of the moment inequalities, in which we include three covariates in the vector Z_{ijt}: 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 the null that the model is correctly specified at conventional

[39] Formally, our setting involves the simultaneous testing of more than one hypothesis. We could approach this test similar to the problem of selecting the valid and relevant moment inequalities among many candidate inequalities. Andrews (1999) and Cheng and Liao (2015), among others, describe two different procedures to perform this moment selection for generalized method of moments (GMM) estimation. As far as we know, no equivalent procedure exists in the literature for moment inequality estimation.

Figure 2: Distribution of Fixed Export Costs



In all the six 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 for each decile 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.4 in Appendix B.3.

significance levels. That is, we fail to reject the hypothesis that potential exporters know at least these three covariates when predicting the revenue from exporting r_{ijt}.[40]

[40] While our goal is to test whether a set of variables is contained in the firm's information set, in practice, we would fail to reject our null hypothesis when the set of variables we specify are either (a) irrelevant or (b) relevant and in the agent's information set. Specifically, such null hypothesis will be rejected only if the expectational error in the firm's revenue forecast, $\varepsilon_{ijt} \equiv r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, does not satisfy the condition,

Table 5: Testing Content of Information Sets

Set of Firms	Set of Export Destinations	Variable Tested	Chemicals Reject at 5%	Chemicals p-value RC	Food Reject at 5%	Food p-value RC
All	All	$(dist_j, r_{iht-1}, R_{jt-1})$	No	0.140	No	0.975
All	All	$(\alpha_{jt} r_{iht})$	Yes	0.005	Yes	0.005
Large	Popular	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	No	0.110	No	0.940
Large	Unpopular	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	No	0.110	No	0.970
Small	Popular	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	Yes	0.005	Yes	0.005
Small	Unpopular	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	Yes	0.020	Yes	0.005
Small & Exporter$_{t-1}$	All	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	Yes	0.005	Yes	0.005
Large & Non-exporter$_{t-1}$	All	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	No	0.145	No	0.990
Small & Non-Exporter$_{t-1}$	All	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	Yes	0.005	Yes	0.005
Large & Exporter$_{t-1}$	All	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	No	0.105	No	0.985

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. *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*$_{t-1}$ with respect to a country j if $d_{ijt-1} = 1$ and as a *Non-exporter*$_{t-1}$ if $d_{ijt-1} = 0$. For details on how to compute these p-values, see Bugni et al. (2015). All numbers reported in this table are independent of the value of η chosen as the normalizing constant.

In a second test, we run our moment inequality procedure under the assumption of perfect foresight. Here, we presume the firm knows r_{ijt} when it chooses whether to export. We can reject, at conventional significance levels, that firms know their exact future revenue when deciding whether to export. The p-value is less than 1% for both the chemicals and food sectors.

In all the remaining tests, we re-run the same empirical model as in our main specification, but we add an additional variable to the vector of instruments: we assume firms also know the lagged value of the country-year shifter of revenue, α_{jt-1}. From the model in Section 2, this shifter accounts for how supply factors in a market and aggregate demand affect export revenues at each time period. We test our new specification of Z_{ijt} first for all firms, and then for large firms vs. small firms, for previous exporters vs. non-exporters, for popular vs. unpopular destinations, and for various combinations of these characteristics. More precisely, we split firms in two equal groups ("large" or "small") depending on whether they are above or below the median domestic sales in $t-1$ and split countries also in two groups ("popular" or "unpopular") depending on whether the number of Chilean exporters in the corresponding sector to each destination country is above or below the median for the year $t-1$.

For both the chemical and food sector, the results in rows 3 to 6 in Table 5 suggest two broad conclusions: (a) large firms have more information than small firms; and (b) the information a firm has about a destination country does not depend on the popularity of the

$\mathbb{E}[\varepsilon_{ijt}|Z_{ijt}] = 0$. This condition will hold when Z_{ijt} is irrelevant to predict r_{ijt} or when Z_{ijt} is relevant and 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 we test, to check if such variable has predictive power for r_{ijt}. The results from this pre-test are included in Table B.7 in Appendix B.5. We confirm the relevancy of the variables Z_{it} we test in this section.

market. The results show that, at the 5% significance level, we can reject that α_{jt-1} is in the information set of small firms, both for popular and unpopular markets, and we cannot reject that α_{jt-1} is in the information set of large firms, for any market type. Put differently, we find little evidence that firms in popular markets learn from other exporters, but do find that firms that are either more productive or sell higher quality products also tend to have an informational advantage when forecasting market conditions in foreign countries.[41]

Finally, we test a further possible source of heterogeneity in information across firms: do firms with export experience in the prior year have better information on the prior year's revenue shocks α_{jt-1} than those who did not export? Given that large firms are more likely to have exported in the past to any given destination country, the tests in rows 7 to 10 in Table 5 also represent our attempt to disentangle whether the extra information large firms possess appears due to their productivity or simply their previous export experience. The results from these four test clearly suggest that learning from previous export experience is not an important phenomenon in explaining information acquisition about the revenue shifter. We cannot reject that large firms know α_{jt-1} even if they did not export to j in the prior year and, conversely, we can reject that small firms have this information, including those that exported to j in the prior period.[42]

Our identification of differences in information across firms is very different from alternative identification approaches in the literature that identify firms' learning and, therefore, information acquisition, from patterns of correlation in either export entry decisions, export prices or export quantities— see, for example, work by Albornoz et al. (2012), Fernandes and Tang (2014), Berman et al. (2015), and Fitzgerald et al. (2016). We view our approach as complementary: we also use data on the export participation of firms to learn about their information sets but exploit the implication of the rational expectations assumption directly.

We do not directly investigate why large firms may have better information. We can however, offer some insight from the Chilean case study. The literature offers two hypotheses. First, large firms may acquire information about foreign markets through affiliates located abroad. Second, large firms may be more likely to participate in international trade fairs and may better exploit the information resources of governmental trade promotion agencies ((Weiss, 2008)). In Chile in our sample period, none of the firms is a multinational firm, limiting the information-via-affiliates channel. However, Alvarez and Crespi (2000) notes that firms with larger domestic sales were more likely to participate in programs sponsored by the Chilean National Agency for Export Promotion.[43]

[41] In our sample, popular destination countries are geographically closer to Chile. Therefore, one could also interpret the findings in rows 3 to 6 in Table 5 as suggesting that potential exporters do not have systematically more information about geographically close countries than about further away ones.

[42] Our tests here are of learning about a destination-year aggregate shifter of either foreign demand or foreign trade costs; we do not test whether there is learning about a firm-specific taste or supply shock (as in Albornoz et al., 2012) or about the demand shifter in a particular buyer-seller relationship (as in Eaton et al., 2014).

[43] As Alvarez and Crespi (2000) describe, the Chilean National Agency for Export Promotion "manages a

7 Counterfactuals

Finally, we use our model and the insights on firms' information to explore the effect of trade policy on exporting. Here, we conduct a counterfactual exercise in which we imagine government programs reduce the exporters' fixed costs of participation by 40%.

With our counterfactual policy, we aim to capture in a stylized way the effect of export promotion programs on the fixed costs of exporting and ultimately on export participation. Such programs are common. Van Biesebroeck et al. (2015) discuss how the Canadian Trade Commissioner Service lower entry barriers to increase export participation. Volpe Martincus and Carballo (2008) and Volpe Martincus et al. (2010) document similar measures in Peru and Uruguay. According to Lederman et al. (2009), typical programs include country image building (advertising, promotional events, advocacy), export support services (exporter training, technical assistance on logistics, customs, and packaging), and marketing (trade fairs, follow-up services offered by representatives abroad). It is hard to quantify the precise savings in fixed export costs that these services imply; our choice of a 40% reduction illustrates one possible level.

Predicting counterfactual export participation and export volume requires more care in our setting, both because we identify and estimate bounds on our parameter of interest, and because we do not want to impose at this stage any additional assumptions on the exact set of covariates firms use to predict their potential export revenue. In Appendix A.7, we show how to derive bounds on the effect that counterfactual changes in the parameter vector θ have on both export participation and export volume. Broadly, the procedure involves four steps. First, we derive bounds on choice probabilities conditional on a value of θ and on an observed subset, Z_{ijt}, of exporters' true information sets. Second, we take the value of θ^* as known and use the bounds derived in the first step to further bound the effect of changing the parameter vector θ from θ^* to a counterfactual value θ'. Third, we account for the fact that θ^* is only partially identified. Finally, fourth, we aggregate our bounds across firms with different values of Z_{ijt}.

We leave much of the algorithm for Appendix A.7, but show here the key theorem that allows us to bound the probability of participation under counterfactual parameter values. The hurdle we face in our setting is that even when we condition on a particular value of θ and Z_{ijt}, choice probabilities are not point identified because we only observe a subset Z_{ijt} of the variables in the true information set $\mathcal{J}_{ijt}$ firms use to predict export revenues. Thus, we cannot compute firms' unobserved expectations, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, exactly and therefore cannot compute the export probabilities in equation (8). However, we may still derive bounds on the expected probability that firm i exports to j at t, conditional on Z_{ijt}.

system that provides information to firms. It is used by companies interested in obtaining specific information about international markets, for example: external prices, transport costs, entrance regulations and trade barriers."

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

$$\mathcal{P}_{ijt}(\theta) = \mathcal{P}_{jt}(d_{ijt} = 1|\mathcal{J}_{ijt};\theta) = \Phi\big(\theta_2^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \theta_0 - \theta_1 dist_j\big)\big). \tag{20}$$

Then,

$$\mathcal{P}^l_{jt}(Z_{ijt};\theta) \leq \mathcal{P}_{jt}(Z_{ijt};\theta) \leq \mathcal{P}^u_{jt}(Z_{ijt};\theta), \tag{21}$$

where

$$\mathcal{P}^l_{jt}(Z_{ijt};\theta) = \frac{1}{1 + B^l_{jt}(Z_{ijt};\theta)}, \tag{22a}$$

$$\mathcal{P}^u_{jt}(Z_{ijt};\theta) = \frac{B^u_{jt}(Z_{ijt};\theta)}{1 + B^u_{jt}(Z_{ijt};\theta)}, \tag{22b}$$

and

$$B^l_{jt}(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1 - \Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)}{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)}\middle| Z_{ijt}\right], \tag{23a}$$

$$B^u_{jt}(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)}{1 - \Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)}\middle| Z_{ijt}\right]. \tag{23b}$$

The proof of Theorem 3 is in Appendix A.7.1. Equation (21) defines bounds on export probabilities conditional on a particular value of the instrument vector Z_{ijt} and a particular value of the parameter vector θ.

To compute the relative change in $\mathcal{P}_{jt}(Z_{ijt};\theta)$ due to a change in the parameter vector from its true value θ^* to a particular counterfactual value θ', we can use equation (21) and compute bounds on this relative change as

$$\frac{\mathcal{P}^l_{jt}(Z_{ijt};\theta')}{\mathcal{P}^u_{jt}(Z_{ijt};\theta^*)} \leq \frac{\mathcal{P}_{jt}(Z_{ijt};\theta')}{\mathcal{P}_{jt}(Z_{ijt};\theta^*)} \leq \frac{\mathcal{P}^u_{jt}(Z_{ijt};\theta')}{\mathcal{P}^l_{jt}(Z_{ijt};\theta^*)}. \tag{24}$$

Here, the lower bound on the relative change is defined as the ratio of the lower bound on the export probability evaluated at the new value of θ, θ', and the upper bound evaluated at the initial true value of θ, θ^*. The reverse case is used to compute the upper bound on the relative change.

We conduct the counterfactuals using only data from the year 2005, and compare the predictions from both our moment inequality approach and from the models that require the researcher to specify the exact set of covariates included in firms' information sets. We report

our counterfactual predictions in Table 6.

Table 6: Impact of 40% Reduction in Fixed Costs

Estimator	Chemicals Argentina	Chemicals Japan	Chemicals United States	Food Argentina	Food Japan	Food United States
			% Change in Number of Exporters			
Perfect Foresight	51.6	632.7	201.9	123.4	141.6	90.9
(MLE)	(0.2)	(102.4)	(0.01)	(0.2)	(8.6)	(6.9)
Minimal Info.	53.5	755.1	135.8	126.2	97.6	79.2
(MLE)	(0.9)	(180.0)	(13.2)	(0.7)	(7.6)	(6.2)
Moment Inequality	[43.1, 47.1]	[83.3, 111.9]	[443.5, 533.5]	[127.4, 204.2]	[57.6, 71.8]	[35.5, 41.6]
			% Change in Volume of Exports			
Perfect Foresight	45.9	214.9	201.5	117.6	61.3	24.9
(MLE)	(1.7)	(108.0)	(0.01)	(1.4)	(17.6)	(6.9)
Minimal Info.	36.1	226.3	38.5	106.8	24.7	19.4
(MLE)	(5.5)	(194.0)	(17.3)	(5.3)	(6.3)	(4.8)
Moment Inequality	[15.0, 17.4]	[20.5, 29.8]	[435.2, 520.8]	[52.5, 92.3]	[8.4, 12.1]	[3.5, 4.5]

Notes: All variables are reported in percentages. For the two ML estimators, standard errors are reported in parentheses. For the moment inequality estimates, all points in the identified set and confidence sets are used to compute the counterfactual changes. The corresponding minimum and maximum predicted values obtained using all parameter values contained in the identified set are reported in square brackets; the minimum and maximum values obtained using all parameter values contained in the confidence set are reported in parentheses. All numbers reported in this table are independent of the value of η chosen as normalizing constant.

Comparing first the changes predicted by the two maximum likelihood estimators, we find that, relative to the predictions from perfect foresight, the predicted export participation under the two-step approach is lower in three out of the six cases considered: for the US in both sectors and for Japan in the food sector. For export volume, the perfect foresight estimate predicts larger effects in five out of the six cases. Specifically, in the chemicals sector in Argentina, the two-step approach predicts a slightly larger impact in terms of number of exporters (53.5% vs. 51.6%) but a much smaller impact in terms of volume (36.1% vs. 45.9%). This reflects a general principle: for the same change in the number of firms expected to switch to exporting, we will predict the largest change in total export volume under the perfect foresight assumption. Intuitively, for a given level of fixed export costs for each firm, the perfect foresight assumption implies that the first set of firms that will switch to exporting are those with the highest realized revenues. That is, the firms that in the end earn the most upon exporting are precisely those the model predicts will decide to participate. Conversely, when firms have imperfect information about ex post export sales, some large firms might not enter (because their expectations are too low) while some small firms will (because their expectations are too high). Therefore, conditional on a predicted number of new exporters, if the perfect foresight assumption is wrong, a model that imposes this informational assumption will always overestimate the effect of a policy change.[44]

[44]The overall effect of the policy change predicted under the perfect foresight assumption might not be larger

The moment inequality approach, which imposes weaker assumptions on the content of firms' information sets, produces predictions that may be larger or smaller than those from the maximum likelihood approach. For example, the moment inequality estimator predicts growth in export participation in the United States that is more than twice as large as the perfect foresight prediction in the chemicals sector and less than half as large in the food sector. Similarly differences appear for export volume predictions. These differences are large and likely to be important for the evaluation of any export promotion policy .

The estimates reveal substantive economic effects from a hypothetical export promotion measure that reduces fixed export costs. As Table 6 shows, decreasing export costs by 40% may lead to a large increases in export flows. As a percentage of the baseline level, the policy that causes fixed costs to fall 40% leads to a 15.0 to 17.4% increase in export volume to Argentina in the chemicals sector. In the food sector, the effect on trade flows between Chile and Argentina is somewhat larger: the reduction in fixed costs produces an increase in volume of between 52.5 and 92.3%.

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. (2015) to allow for sunk export entry costs and forward-looking exporters.

Second, we relax the assumption, captured in equation (2), that potential export revenues in a country-year pair are proportional to domestic sales. Instead, in Section 8.2, we allow these export revenues to depend on firm-country-year specific export revenue shocks. We consider two cases. In Section 8.2.1, we assume firms do no know the export revenue shocks when they decide whether to export–that is, the shocks are mean independent of exporters' information sets, $\mathcal{J}_{ijt}$. We show that our benchmark model can accommodate shocks of this form with only minor changes to notation; our empirical results remain unchanged. In Section 8.2.2, we show how to derive moment inequalities when the firm's participation decision depends on revenue shocks the firm anticipates but the econometrician cannot observe.

The extensions in sections 8.1 and 8.2.2 require much more computing time to estimate than our benchmark model and, therefore, we restrict our empirical analysis below to the chemical sector.[45]

than the actual one even if this informational assumption is wrong, because the perfect foresight assumption might predict a relatively small increase in the total number of exporters. This is the true for Japan in the chemical sector. As Table 6 shows, the predicted impact on aggregate export volume is slightly smaller for the perfect foresight case than for the minimal information case because the policy's effect on the number of exporters is much smaller.

[45]Both extensions involve larger dimensional parameter vectors than our benchmark specification. Ho and Rosen (2016) comment on how the computation required for inference increases with the dimensionality of the

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 our dynamic model, exporting firms must still pay fixed costs f_{ijt} in every period in which they choose to export, but they must also now 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 profits to firm i of exporting to j at period t net of fixed and sunk costs are

$$\pi_{ijt} = \eta_j^{-1} r_{ijt} - f_{ijt} - (1 - d_{ijt-1}) s_{ijt}. \tag{25}$$

We maintain the assumptions in equations (4) and (5) on the distribution of fixed export costs, f_{ijt}. We model sunk export costs as:

$$s_{ijt} = \gamma_0 + \gamma_1 dist_j \tag{26}$$

We assume information sets evolve independently of past export decisions:[46]

$$(\mathcal{J}_{ijt+1}, f_{ijt+1}, s_{ijt+1}) | (\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt}) \sim (\mathcal{J}_{ijt+1}, f_{ijt+1}, s_{ijt+1}) | (\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}). \tag{27}$$

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

$$\begin{aligned} d_{ijt} = \mathbb{1}\{&\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - (1 - d_{ijt-1})(\gamma_0 + \gamma_1 dist_j) - \nu_{ijt} \\ &+ \rho\mathbb{E}[V(\mathcal{J}_{ijt+1}, f_{ijt+1}, s_{ijt+1}, d_{ijt})|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt} = 1] \\ &- \mathbb{E}[V(\mathcal{J}_{ijt+1}, f_{ijt+1}, s_{ijt+1}, d_{ijt})|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt} = 0] \geq 0\}, \end{aligned} \tag{28}$$

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

The firm's export decision now depends on the firm's expectations of both static revenues r_{ijt} and the difference in the value function depending on whether firm i exported to j in period t. We can follow the approach from the static case to find a measure of r_{ijt}, but finding a measure of the difference in value functions $V(\cdot)$ 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 take into account profits at periods t and $t+1$ when making a decision at t, we can only find a measure of either $V(\cdot, d_{ijt} = 1)$ or $V(\cdot, d_{ijt} = 0)$. To solve this lack of measurement, we adjust the Euler approach in Morales et al.(2015) to find bounds on θ_D^* without a measure of the difference in value functions. This approach

parameter vector.

[46] This assumption rules out learning by exporting.

follows the methodology developed in Hansen and Singleton (1982) and Luttmer (1999) for continuous controls, but adapted for our partially identified model with discrete controls.[47] 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 7 we report the results from projecting the 95% confidence set for θ_D^* to compute analogous confidence sets for the fixed and sunk costs of exporting. The estimates show that sunk entry costs are significantly larger than fixed export costs, consistent with previous evidence in Das et al. (2007). Not surprisingly, fixed and sunk export costs are clearly increasing in distance. Furthermore, the sensitivity of these cost parameters 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 fixed costs bounds in the benchmark model in Section 2 to the ones arising from this dynamic model, we note two key differences. First, the bounds are wider; this is a consequence of having to estimate fixed and sunk costs simultaneously and the difficulties of separately identifying both types of costs. Second, while the fixed export costs for Argentina are similar in the static and dynamic models, those for the United States and Japan are larger in the dynamic model than in the static one, because we estimate the effect of distance on fixed export costs, β_1, to be larger when accounting for the forward-looking behavior of firms.[48]

Table 7: Export fixed and sunk costs: firm average

			Chemicals	
Estimator	Cost	Argentina	Japan	United States
Benchmark	Fixed	[79.1, 104.1]	[309.2, 420.5]	[181.3, 243.6]
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]
Selection	Fixed	[67.7, 135.1]	[1,033.9, 2,064.3]	[495.8, 989.9]

Notes: All variables are reported in thousands of year 2000 USD. Extreme points of the confidence set are reported in parentheses. Confidence sets are computed using the procedure described in Andrews and Soares (2010).

[47] Appendix E shows how to adapt the Euler approach in Morales et al. (2015) to the specific characteristics of the model described in Section 2 and in equations (25) to (28). Specifically, Morales et al. (2015) consider moment inequality models in which the authors assume the unobserved component ν_{ijt} is constant across groups of countries for each firm-year specific pair. The Euler approach to the construction of moment inequalities 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 but also on the number of periods ahead that each firm takes into account when making its export participation decision; see Appendix E for more details.

[48] It may seem counterintuitive that accounting for sunk export entry costs increases the estimates of fixed export costs. This pattern would not arise if exporters were simply 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 $V(\cdot)$ in equation (28) 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.

8.2 Firm-Country Export Revenue Shocks

In equation (2) in our benchmark model, potential export revenues r_{ijt} in a given country-year pair jt are proportional to each firm i's domestic sales, r_{iht}. Here we relax this assumption and impose instead that

$$r_{ijt} = \alpha_{jt} r_{iht} + \omega_{ijt}, \tag{29}$$

where ω_{ijt} is an unobserved (to the researcher) firm-country-year export revenue shock. We consider below the case in which the export revenue shock ω_{ijt} is mean independent of the firm i's information set, $\mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}] = 0$, and the case in which the firm knows this revenue shock, $\mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}] = \omega_{ijt}$.

8.2.1 Unknown to Firms When Deciding on Export Entry

If we assume that the export revenue shocks ω_{ijt} are mean independent of the information set $\mathcal{J}_{ijt}$ and the firms' domestic sales, r_{iht}, and that these shocks are mean zero in every country-year pair—that is, $\mathbb{E}_{jt}[\omega_{ijt}|r_{iht}, \mathcal{J}_{ijt}] = 0$—then the random variable ω_{ijt} has the same statistical properties as the measurement error in export revenue e_{ijt} considered in equation (10). Therefore, even in the presence of the export revenue shocks ω_{ijt}, one may still exploit the moment condition in equation (11) to estimate the export revenue coefficients $\{\alpha_{jt}, \forall j, t\}$.

Furthermore, if one assumes that export revenue upon entry is captured by equation (29) and that $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}]$, then one can write the probability that firm i exports to country j in period t as

$$\mathcal{P}_{jt}(d_{ijt} = 1|\mathcal{J}_{ijt}, dist_j) = \Phi\big(\sigma^{-1}\big(\eta^{-1}\mathbb{E}[\alpha_{ijt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big), \tag{30}$$

which is identical the corresponding expression in the benchmark model. Therefore, conditional on first-stage estimates of α_{jt} and the corresponding assumptions imposed in the information set $\mathcal{J}_{ijt}$, the presence of the revenue shock ω_{ijt} does not affect the consistency of the maximum likelihood estimators described in Section 4.1 nor the properties of the moment inequalities described in Section 4.2. One possible microfoundation for the export revenue shock ω_{ijt} is to assume that, contrary to what is assumed in the benchmark model in Section 2, variable trade costs τ vary across firms within a single country-year pair; i.e. $\tau_{ijt} \neq \tau_{i'jt}$ for $i \neq i'$.

8.2.2 Known When Deciding on Export Entry

In this section, we generalize the benchmark model described in Section 2 and assume instead that

$$\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] + \omega_{ijt}, \tag{31}$$

where ω_{ijt} is a component known to the firm but not the econometrician. We generalize the distributional assumption in equation (5) to also account for the distribution of ω_{ijt} and assume that

$$\begin{pmatrix} \omega_{ijt} \\ \nu_{ijt} \end{pmatrix} \Bigg| (\mathcal{J}_{jt}, dist_j) \sim \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\omega^2 & \sigma_{\omega\nu} \\ \sigma_{\omega\nu} & \sigma^2 \end{pmatrix} \right), \tag{32}$$

where, as in the main model, ν_{ijt} is an unobserved component element of fixed export costs (see equation (4)). The export dummy d_{ijt} therefore becomes

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - (\nu_{ijt} + \eta^{-1}\omega_{ijt}) \geq 0\}. \tag{33}$$

If we assume that potential exporters have perfect foresight with respect to the component $\alpha_{jt} r_{iht}$ of their potential export revenues, $\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] = \alpha_{jt} r_{iht}$, then we would be able to identify 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). Appendix F.2 shows how to estimate this parameter vector when we only impose the assumption that we observe a vector $Z_{ijt} \subseteq \mathcal{J}_{ijt}$.

Before discussing the parameter estimates from this approach, we address one additional complication in estimation from allowing firms to account for ω_{ijt} in their decision. Allowing this shock, we can no longer use linear regression to estimate the export revenue coefficients $\{\alpha_{jt}\}_{j,t}$; our assumptions now imply that the export revenue coefficients $\{\alpha_{jt}\}_{j,t}$ are only partially identified and must be estimated jointly with the fixed costs parameters $(\beta_0, \beta_1, \sigma_\omega, \sigma_{\omega\nu}, \sigma_\nu)$. Given that our sample period covers 10 years and 22 countries, this implies 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. Therefore, we simplify the problem by assuming $\alpha_{jt} = \alpha_0 + \alpha_1 R_{jt}$ and estimate the parameter vector $\theta_S \equiv (\alpha_0, \alpha_1, \beta_0, \beta_1, \sigma_\omega, \sigma_{\omega\nu}, \sigma_\nu)$.[49]

We report the results in Table 7. Our selection model generates larger bounds on the fixed cost parameters than our benchmark model, and these costs increase faster with distance from Chile than in the benchmark. That we find the fixed costs from the selection model to be

[49] When estimating the model introduced in Heckman (1979), it is typical to fix one of the components of the variance matrix in equation (32) as a normalization. In our case, to facilitate the comparison of the estimates with those computed for the benchmark model in Section 2, we opt to maintain the normalization $\eta = 5$.

larger than those in the benchmark model is not surprising. Intuitively, firms with higher domestic sales, r_{iht}, are *ceteris paribus* more likely to export and, therefore, for the subset of firms, countries and years with positive exports, we should expect a negative correlation between r_{iht} and ω_{ijt}. This implies that the estimates of α_{jt} that ignore ω_{ijt} are likely to be downward biased. Given a fixed value of η^{-1}, a downward bias in α_{jt} will generate a downward bias in the estimates of the variance of the fixed export cost shock, σ, and this will translate into a downward bias in the fixed export cost parameters β_0 and β_1.

9 Conclusion

Policymakers seeking to affect the balance of trade often pursue export promotion policies, including country image building and export support services. The success of these programs depends on what information firms use to predict revenue, and on how their participation decisions react to changes in export costs. In this paper, we develop a model not only to measure the fixed costs of exporting, but also to test exactly what information firms use to predict their export revenue and to quantify how firms will react to export promotion polices.

The estimated fixed costs from our inequality model are between one third and one half the size of the costs found using the approaches common in the earlier international trade literature. The predictions in a counterfactual economic environment in which fixed export costs fall 40% also differ substantially across alternative methods. The bounds we estimate for the effect of this counterfactual on export participation and export volume are sufficiently tight to inform policy.

Finally, when we test alternative assumptions on the content of the information sets firms use in their export decision—i.e. when we test what exporters know—we reject that firms can perfectly predict the revenue they'll earn upon entering a market. Further, we find important heterogeneity by firm size: large firms have better information on demand conditions in foreign markets than small firms. This effect appears driven by more than simply past export experience, as even large firms without export experience in the prior year appear to possess better information on characteristics of the prior year's export market. We leave for future work a careful analysis of what drives this information acquisition and learning.

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, 12, 17]

Albornoz, Facundo, Héctor F. Calvo Pardo, Gregory Corcos, and Emanuel Ornelas, "Sequential Exporting," *Journal of International Economics*, 2012, *88* (1), 1–24. [25]

Alvarez, Roberto and Gustavo Crespi, "Exporter Performance and Promotion Instruments: Chilean Empirical Evidence," *Estudios de Economía*, 2000, *27* (2), 225–241. [25]

Ameriks, John, Joseph Briggs, Andrew Caplin, Matthew D. Shapiro, and Christopher Tonetti, "Long-Term Care Insurance, Annuities, and the Under-Insurance Puzzle," *mimeo*, February 2015. [4]

Amstrong, Tim B., "Weighted KS Statistic for Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 2014, *182* (2), 92–116. [17]

_ **and Hock Peng Chan**, "Multiscale Adaptive Inference on Conditional Moment Inequalities," *Journal of Econometrics*, 2016, *194* (1), 24–43. [17]

Andrews, Donald W. K., "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 1999, *67* (3), 543–564. [22]

Andrews, Donald W.K. and Gustavo Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78*, 119–157. [19, 31, 4, 24, 25, 28, 29]

_ **and Patrik Guggenberger**, "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 2009, *25*, 669–709. [29]

_ **and Xiaoxia Shi**, "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, *81* (2), 609–666. [17]

Antràs, Pol, Teresa Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S.," *mimeo*, December 2016. [1, 10]

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151–1199. [1, 10]

_ **, Sharat Ganapati, and Marc-Andreas Muendler**, "The Extensive Margin of Exporting Goods: A Firm-level Analysis," *mimeo*, September 2014. [1, 4, 10]

Armstrong, Tim B., "Asymptotically Exact Inference in Conditional Moment Inequality Models," *Journal of Econometrics*, 2015, *186* (1), 51–65. [17]

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. [7]

Berman, Nicolas, Vincent Rebeyrol, and Vincent Vicard, "Demand Learning and Firm Dynamics: Evidence from Exporters," *mimeo*, July 2015. [25]

Bernard, Andrew B. and J. Bradford Jensen, "Why Some Firms Export," *Review of Economics and Statistics*, 2004, *86* (2), 561–569. [10]

_ , Andreas Moxnes, and Karen Helen Ulltveit-Moe, "Two-Sided Heterogeneity and Trade," *mimeo*, January 2016. [1]

_ , 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, *forthcoming.* [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, "Global Firms in the Innovation Economy," *mimeo*, September 2016. [4]

Blaum, Joaquín, Claire Lelarge, and Michael Peters, "The Gains from Input Trade with Heterogeneous Importers," *mimeo*, August 2016. [1]

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, 22, 24, 25, 29]

_ , _ , and _ , "Inference for Functions of Partially Identified Parameters in Moment Inequality Models," *Quantitative Economics*, 2016, *forthcoming.* [19]

Cheng, Xu and Zhipeng Liao, "Select the Valid and Relevant Moments: An Information-based LASSO for GMM with Many Moments," *Journal of Econometrics*, 2015, *186* (2), 443–464. [22]

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. [17]

Chetverikov, Denis, "Adaptive Test of Conditional Moment Inequalities," *mimeo*, 2013. [17]

Coşar, Kerem A., Paul L. E. Grieco, Shengyu Li, and Felix Tintelnot, "What Drives Home Market Advantage?," *mimeo*, 2016. [20]

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

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, 2, 5, 8, 10, 29, 31]

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. [1]

_ , Marcela Eslava, C.J. Krizan, Maurice Kugler, and James R. Tybout, "A Search and Learning Model of Export Dynamics," *mimeo*, February 2014. [1, 25]

_ , _ , 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, 2, 10]

Eizenberg, Alon, "Upstream Innovation and Product Variety in the U.S. Home PC Market," *Review of Economic Studies*, 2014, *81*, 1003–1045. [2, 16]

Fernandes, Ana and Heiwei Tang, "Learning to Export from Neighbors," *Journal of International Economics*, 2014, *94* (1), 87–94. [25]

Fitzgerald, Doireann, Stefanie Haller, and Yaniv Yedid-Levi, "How Exporters Grow," *mimeo*, January 2016. [25]

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

Handel, Benjamin R. and Jonathan T. Kolstad, "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare," *American Economic Review*, 2015, *105* (8), 2449–2500. [4]

Hansen, Lars Peter and Kenneth J. Singleton, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 1982, *50* (5), 1269–1286. [31]

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. [33, 66]

_ **and Bo E. Honoré**, "The Empirical Content of the Roy Model," *Econometrica*, 1990, *58* (5), 1121–1149. [7]

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 2008, *123* (2), 441–487. [1]

Ho, Katherine, "Insurer-Provider Networks in the Medical Care Market," *American Economic Review*, 2009, *99* (1), 393–430. [2]

_ **and Adam Rosen**, "Partial Identification in Applied Research: Benefits and Challenges," *mimeo*, 2016. [29]

_ **and Ariel Pakes**, "Hospital Choices, Hospital Prices, and Financial Incentives to Physicians," *American Economic Review*, 2014, *104* (12), 3841–3884. [2]

Imbens, Guido and Charles F. Manski, "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 2004, *72*, 1845–1857. [16, 19]

Lederman, Daniel, Marcelo Olarreaga, and Lucy Payton, "Export Promotion Agencies Revisited," *World Bank Policy Research Working Paper*, 2009, *5125.* [26]

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. [31]

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, 12, 17]

_ , "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*, May 2006. [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*, January 2015. [2, 16, 29, 31]

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, 16]

_ and Jack R. Porter, "Moment Inequalities for Multinomial Choice with Fixed Effects," *mimeo*, August 2014. [14]

_ , _ , Katherine Ho, and Joy Ishii, "Moment Inequalities and their Application," *Econometrica*, 2015, *83* (1), 315–334. [2, 16]

Pratt, John W., "Concavity of the the Log Likelihood," *Journal of the American Statistical Association*, 1981, *76*, 103–106. [7]

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. [10]

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. [29]

Ruhl, Kim J. and Jonathan L. Willis, "New Exporter Dynamics," *International Economic Review*, forthcoming 2016. [1]

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

Weiss, Kenneth D., *Building an Import-Export Business*, Hoboken NJ: John Wiley & Sons, Inc., 2008. [25]

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," *mimeo*, May 2016. [2]

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA: MIT Press, 2002. [13]

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

October 2016

A Model and Estimation Strategy: Details

A.1 Expected Export Revenue: Details

We describe here how we can exploit the structure introduced in Section 2.1 to derive the expression for the export revenue conditional on entry in equation (1).

In Section 2.1, we assume potential exporters face: (a) a constant elasticity of substitution demand function; (b) a constant marginal cost; and (c) a monopolistically competitive market. These three assumptions imply that we can write the potential revenue that firm i would obtain in market j at period t as indicated in equation (1); i.e.

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

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

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

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

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

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

$$\alpha_{jt} \equiv \left[\frac{\tau_{jt}}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right]^{1-\eta}\frac{Y_{jt}}{Y_{ht}}, \tag{A.3}$$

we obtain equation (1).

A.2 Estimation of Export Revenue Shifters

In this section, 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^{obs}_{ijt}, r_{iht}) for every exporting firm–i.e. where $d_{ijt} = 1$. With this data, we use OLS to estimate the parameter α_{jt} in the following linear regression:

$$r^{obs}_{ijt} = \alpha_{jt} r_{iht} + e_{ijt}, \tag{A.4}$$

If $\mathbb{E}_{jt}[e_{ijt}|r_{iht}, d_{ijt} = 1] = 0$, where $\mathbb{E}_{jt}[\cdot]$ denotes an expectation conditional on a given country-year pair jt, then standard results for OLS estimators guarantee that

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

The mean independence condition in equation (**??**) is likely to hold given the definition of e_{ijt} as measurement error in reported trade flows.

A.3 Partial Identification: Example

Here we prove that the model described in Section 2, combined with the assumption that the econometrician only observes a vector $(d_{ijt}, Z_{ijt}, r_{ijt})$ such that $Z_{ijt} \subseteq \mathcal{J}_{ijt}$, is not enough to point identify the parameter vector of interest θ^*.

The data are informative about the joint distribution of $(d_{ijt}, Z_{ijt}, r_{ijt})$ across i, j, and t. To simplify notation, we define Z_{ijt} such that $dist_j \in Z_{ijt}$. We denote the joint distribution of the vector $(d_{ijt}, Z_{ijt}, r_{ijt})$ as $\mathbb{P}(d_{ijt}, Z_{ijt}, r_{ijt})$. In this section, we use $\mathbb{P}(\cdot)$ to denote distributions that may be directly estimated given the available data on $(d_{ijt}, Z_{ijt}, r_{ijt})$. For the sake of simplicity in the notation, we use r^e_{ijt} to denote $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$

in this section. Without loss of generality, we can write

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

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

$$\mathbb{P}(d_{ijt}, Z_{ijt}, r_{ijt}) = \int f^y(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}) f^y(r_{ijt}|r^e_{ijt}, Z_{ijt}) f^y(r^e_{ijt}|Z_{ijt}) \mathbb{P}(Z_{ijt}) dr^e_{ijt}, \tag{A.5}$$

where we use $\mathbb{P}(Z_{ijt})$ to denote that the marginal distribution of Z_{ijt} directly observable in the data. Any structure $S^y \equiv \{f^y(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}),\ f^y(r_{ijt}|r^e_{ijt}, Z_{ijt}),\ f^y(r^e_{ijt}|Z_{ijt})\}$ is admissible as long as it verifies the restrictions imposed in Section 2 and equation (A.5). The model in Section 2 imposes the following restriction on the elements of equation (A.5):

$$f^y(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}) = f(d_{ijt}|r^e_{ijt}, Z_{ijt}; \gamma^y) =$$
$$\Big(\Phi\big((\theta_2)^{-1}\big(\eta^{-1} r^e_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)\Big)^{d_{ijt}} \Big(1 - \Phi\big((\theta_2)^{-1}\big(\eta^{-1} r^e_{ijt} - \theta_0 - \theta_1 dist_j\big)\big)\Big)^{1-d_{ijt}}. \tag{A.6}$$

Here, we show that θ 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.5)

$$\theta_1 \text{ is known and equal to } 0, \tag{A.7a}$$
$$Z_{ijt} = r^e_{ijt} + \xi_{ijt} \qquad \xi_{ijt}|r^e_{ijt} \sim \mathbb{N}((\sigma_\xi/\sigma_{r^e})\rho_{\xi r^e}(r^e_{ijt} - \mu_{r^e}), (1 - \rho^2_{\xi r^e})\sigma^2_\xi) \tag{A.7b}$$
$$r_{ijt} = r^e_{ijt} + \varepsilon_{ijt}, \qquad \varepsilon_{ijt}|(r^e_{ijt}, \xi_{ijt}) \sim \mathbb{N}(0, \sigma^2_\varepsilon) \tag{A.7c}$$
$$r^e_{ijt} \sim \mathbb{N}(\mu_{r^e}, \sigma^2_{r^e}). \tag{A.7d}$$

Equation (A.7a) restricts the model in Section 2 by assuming that distance does not affect fixed export costs. Equation (A.7b) 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.7c) assumes that firms' expectational error is normally distributed and independent of both firms' unobserved expectations and the difference between the instrument 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.7d) imposes that firms' unobserved expectations are normally distributed; a distributional assumption that is not imposed in the model in the main text. Therefore, it is clear that equation (A.7) 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.7), we can still find at least two structures

$$S^{y_1} \equiv \{(\theta^{y_1}_0, \theta^{y_1}_2), f^{y_1}(r_{ijt}|r^e_{ijt}, Z_{ijt}), f^{y_1}(r^e_{ijt}|Z_{ijt})\},$$
$$S^{y_2} \equiv \{(\theta^{y_2}_0, \theta^{y_2}_2), f^{y_2}(r_{ijt}|r^e_{ijt}, Z_{ijt}), f^{y_2}(r^e_{ijt}|Z_{ijt})\},$$

that verify: (1) equations (A.6) and (A.7); (2) equation (A.5); and (3) $\theta^{y_1} \neq \theta^{y_2}$. 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.7a) simplifies the identification exercise discussed here because the only parameters that are left to identify are (θ_0, θ_2); i.e. we can set $\theta_1 = 0$ in equation (A.6). Equation (A.7c) assumes that the expectational error not only has mean zero and finite variance but is also normally distributed. This implies that the conditional density $f(r_{ijt}|r^e_{ijt}, Z_{ijt})$ is normal:

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

By applying Bayes' rule, both equations (A.7b) and (A.7d) jointly determine the conditional density $f(r^e_{ijt}|Z_{ijt})$ entering equation (A.5).

Result A.3.1 *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.5), (A.6), *and* (A.7)*;*
2. $\theta^{y_1} \neq \theta^{y_2}$.

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

Lemma A.3.1 *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.6) 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.5) 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.3.2 *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.7c), (A.7b) and (A.7d), we can conclude that r_{ijt} and Z_{ijt} are jointly normal. Therefore, all the information arising from observing their joint distribution is summarized in three moments:

$$\begin{aligned}
\sigma_r^2 &= \sigma_{r^e}^2 + \sigma_\varepsilon^2, \\
\sigma_z^2 &= \sigma_{r^e}^2 + \sigma_\xi^2 + 2\rho_{\xi r^e}\sigma_{r^e}\sigma_\xi, \\
\sigma_{rz} &= \sigma_{r^e}^2 + \rho_{\xi r^e}\sigma_{r^e}\sigma_\xi
\end{aligned} \tag{A.8}$$

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 $\sigma_{rz} = \sigma_{r^e}^2$ and, therefore, from Lemma A.3.1, 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.8) only allows us to define bounds on $\sigma_{r^e}^2$. We can rewrite the system of equations in equation (A.8) as

$$\begin{aligned}
\sigma_r^2 &= \sigma_{r^e}^2 + \sigma_\varepsilon^2, \\
\sigma_z^2 &= \sigma_{r^e}^2 + \sigma_\xi^2 + 2\sigma_{\xi r^e} \\
\sigma_{rz} &= \sigma_{r^e}^2 + \sigma_{\xi r^e}.
\end{aligned} \tag{A.9}$$

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

A.4 Deriving Unconditional Moments

The moment inequalities described in equations (15) and (18) 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 an instrument function $g(\cdot)$, we derive unconditional moments that are consistent with our conditional moments:

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

where $m_l^{ob}(\cdot)$, $m_u^{ob}(\cdot)$, $m_l^{r}(\cdot)$, $m_u^{r}(\cdot)$, and Z_{ijt} are defined in equations (15) and (18).

In Section 5, we present results based on a set of positive-valued 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 the instrument Z_{kijt} and the median value of this instrument. 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

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

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 (15) and (18)), 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 $\hat{\Theta}^{95\%}$.[50]

A.5 Confidence Sets for True Parameter: Details

A.5.1 Computation

We describe here the procedure we follow to compute the confidence set for the true parameter vector θ^*. This procedure 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, indexing the finite set of inequalities that we use for estimation by $k = 1, \ldots, K$ and denoting them as

$$\mathbb{E}[m_k(X_{ijt}, Z_{ijt}, \theta)] \geq 0, \qquad k = 1, \ldots, K,$$

the MMM statistic corresponds to

$$Q(\theta) = \sum_{k=1}^{K} (\min\{\frac{\mathbb{E}[m_k(X_{ijt}, Z_{ijt}; \theta)]}{\sqrt{var(m_k(\theta))}}, 0\})^2, \tag{A.10}$$

where, in terms of the notation introduced in sections 4.2.1 and 4.2.2, $X_{ijt} \equiv (d_{ijt}, r_{ijt}, dist_j)$ 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.4 contains additional details on the unconditional moments that we employ. Given the set of unconditional moment inequalities $k = 1, \ldots, K$ and the test statistic in equation (A.10), 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. For the case of the confidence set for the parameter vector $\theta^* \equiv (\beta_0, \beta_1, \sigma)$, Θ_g is a 3-dimensional orthotope. In order to define the limits of this 3-dimensional orthotope, we solve the following nonlinear optimization

$$\begin{aligned} & min_\theta \ \mathbb{d} \cdot \theta \\ & \text{subject to} \\ & \frac{1}{n} \sum_i \sum_j \sum_t m(X_{ijt}, Z_{ijt}, \theta) + \ln n \geq 0, \end{aligned} \tag{A.11}$$

[50] 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 both identified and confidence sets depend on the instrument functions, the main conclusions are robust. The exact results are available upon request.

where n denotes the sample size (i.e. sum of distinct ijt triplets included in our sample), $m(X_{ijt}, Z_{ijt}, \theta_p) \equiv (m_1(X_{ijt}, Z_{ijt}, \theta_p), m_2(X_{ijt}, Z_{ijt}, \theta_p), \ldots, m_K(X_{ijt}, Z_{ijt}, \theta_p))$, 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 that $\mathbb{D}$ has 6 elements, we will therefore solve six nonlinear optimizations like that in equation (A.11). Denote the six 3-dimensional vectors θ that solve each of these optimizations as $(\theta_{1+}, \theta_{1-}, \theta_{2+}, \theta_{2-}, \theta_{3+}, \theta_{3-})'$ and 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 the 3-dimensional orthotope Θ_g we fill it up with 64,000 equidistant points.

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

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

Step 3: evaluate the MMM test statistic at θ_p:

$$Q(\theta_p) = \sum_{k=1}^{K} (\min\{\frac{\mathbb{E}[m_k(X_{ijt}, Z_{ijt}; \theta_p)]}{\sqrt{var(m_k(\theta_p))}}, 0\})^2. \tag{A.12}$$

Step 4: compute 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

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

where

$$\bar{m}(\theta_p) \equiv \frac{1}{n} \sum_i \sum_j \sum_t m(X_{ijt}, Z_{ijt}, \theta_p).$$

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

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

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

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

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

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 the 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.5.2 Figures

The previous section describes in detail the steps that we follow to compute a confidence set for θ^* conditioning on a given value of η^{-1}. In practice, however, we compute such a confidence set in two steps. We first compute a confidence set for the vector $(\beta_0, \beta_1, \eta^{-1})$ conditioning on the normalization $\sigma = 1$. Let's denote such confidence set as $\hat{\Theta}_\sigma^{95\%}$. Then, for each element $\theta_\sigma \in \hat{\Theta}_\sigma^{95\%}$, we compute the corresponding element of $\hat{\Theta}^{1-\alpha}$ by renormalizing it so that it is consistent with our assumed value of η, $\eta = 5$. Specifically, for each element $\theta_\sigma = (\theta_{\sigma 0}, \theta_{\sigma 1}, \theta_{\sigma 2})$ in the confidence set for $(\beta_0, \beta_1, \eta^{-1})$ conditional on the normalization $\sigma = 1$, we compute the corresponding element $\theta = (\theta_0, \theta_1, \theta_2)$ in the confidence set for $(\beta_0, \beta_1, \sigma)$ conditional on the normalization $\eta^{-1} = 0.2$ in the following way:

$$\theta_0 = \theta_{\sigma 0} \times \frac{0.2}{\theta_{\sigma 2}},$$
$$\theta_1 = \theta_{\sigma 1} \times \frac{0.2}{\theta_{\sigma 2}},$$
$$\theta_2 = 1 \times \frac{0.2}{\theta_{\sigma 2}}.$$

For the specific case of the chemicals sector, Figure A.1a plots the resulting confidence set $\hat{\Theta}_\sigma^{95\%}$ in the $(\theta_0, \theta_1, \theta_2)$ dimension. Figure A.1b contains all pairs (θ_0, θ_1) such that there exists a value of θ_2 for which the corresponding triplet $(\theta_0, \theta_1, \theta_2)$ is included in $\hat{\Theta}_\sigma^{95\%}$; it is therefore the outcome of projecting $\hat{\Theta}_\sigma^{95\%}$ in the (θ_0, θ_1) dimension. Similarly, Figure A.1c contains all pairs (θ_1, θ_2) such that there exists a value of θ_0 for which the corresponding triplet $(\theta_0, \theta_1, \theta_2)$ is included in $\hat{\Theta}_\sigma^{95\%}$.

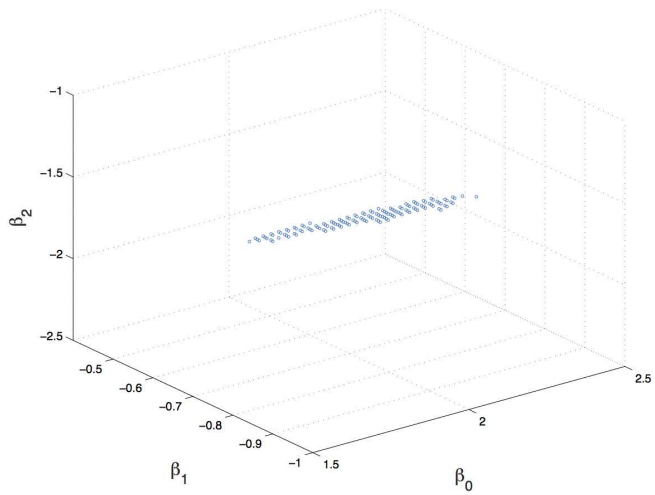
A.6 Moment Inequality Estimates Using Subsets of Inequalities

Here we discuss estimates that are based only on revealed-preference or only on odds-based inequalities. As Table B.5 shows, the bounds on export fixed costs conditional on $\eta = 5$ that arise if we use only odds-based moment inequalities or only revealed-preference inequalities are much wider than those that arise if we combine both our odds-based and revealed-preference inequalities in our estimation.

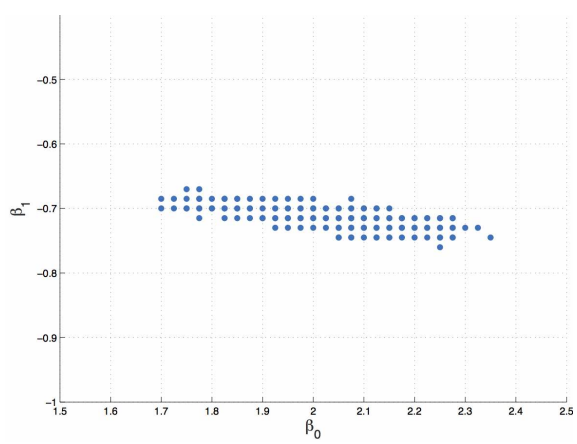
We illustrate the large difference in confidence sets depending on whether we use only odds-based, only revealed-preference or both types of inequalities in figures A.1 and A.2. Specifically, Figure A.2 computes the same three plots shown in Figure A.1 for the case in which we only use revealed-preference inequalities (plots (a), (c), and (e)) and for the case in which we use only odds-based inequalities (plots (b), (d), and (f)). It is immediately apparent from the comparison of figures A.1 and A.2 that information is lost if we exclusively employ odds-based moment inequalities or revealed-preference inequalities. Furthermore, the confidence sets reported in Figure A.2 touch the boundaries of the parameter space employed to compute the plots in such a figure. Therefore, the true size of the confidence sets that employ only revealed-preference or only odds-base inequalities is actually larger than what Figure A.2 reflects. Conversely, the confidence set shown in Figure A.1 is clearly within the bounds of the grid Θ_g employed for its computation.

Figure A.1: Confidence Sets

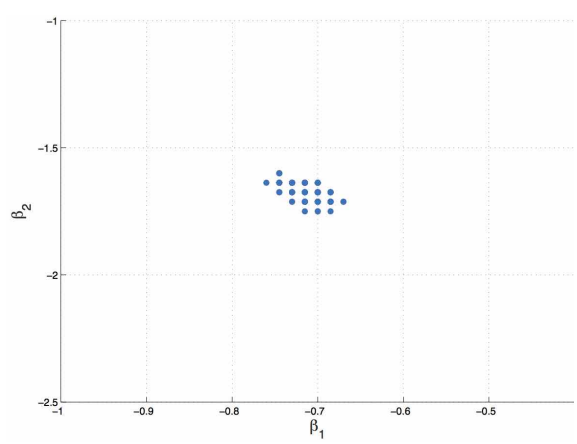
(a) 3-dimensional confidence set



(b) 2-dimensional (β_0, β_1) projection



(c) 2-dimensional (β_1, β_2) projection



In all three figures, the different axis denote the parameter space in which we have performed the estimation and the the blue dots denote the points in the grid expanding this parameter space for which we cannot reject at the 95% confidence level the null hypothesis that they correspond to the true value of the parameter vector.

A.7 Confidence Set for Counterfactual Predictions: Details

A.7.1 Proof of Theorem 3

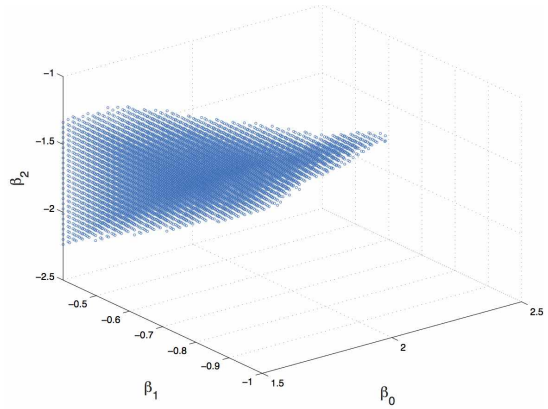
Lemma 1 *Suppose the assumptions in equations* (5) *and* (8) *hold. Then*

$$\mathbb{E}\left[\frac{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}\middle|\mathcal{J}_{ijt}\right] \geq \mathbb{E}\left[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\middle|\mathcal{J}_{ijt}\right]. \tag{A.13}$$

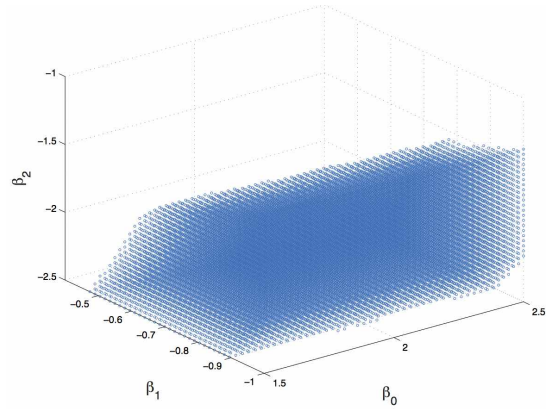
Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the assumption in equation (5) that

Figure A.2: Confidence Sets Using Exclusively Revealed-preference or Odds-based Inequalities

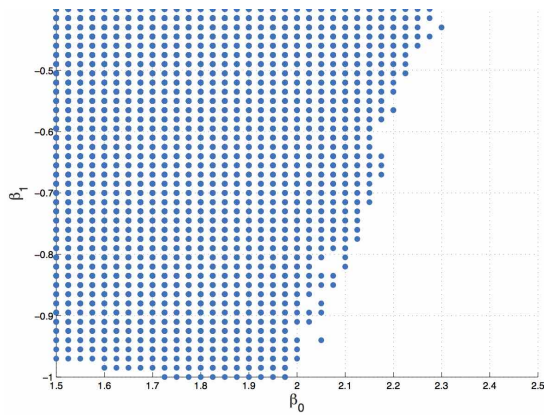
(a) Revealed-preference, 3-dimensional



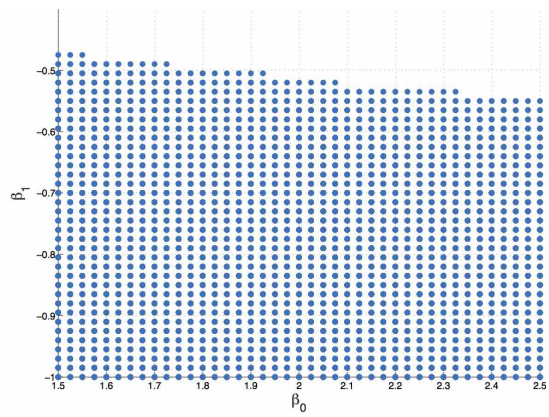
(b) Odds-based, 3-dimensional



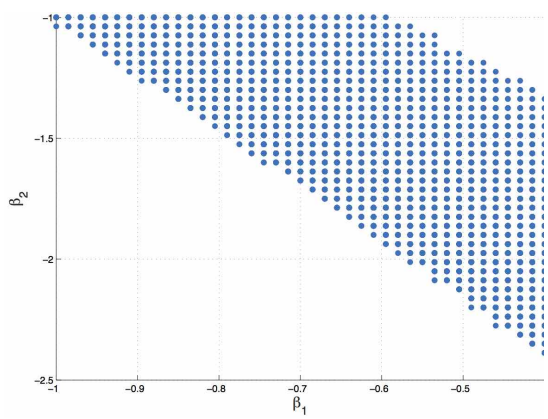
(c) Revealed-preference, 2-dimensional (β_0, β_1)



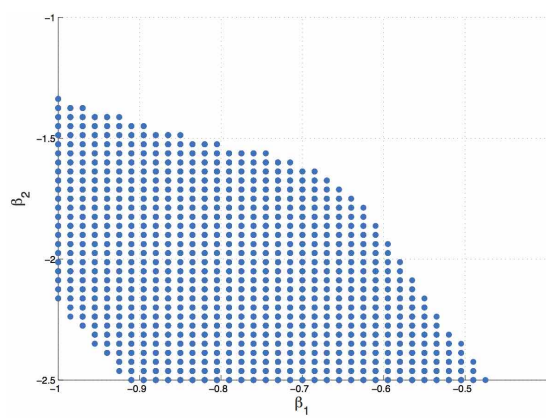
(d) Odds-based, 2-dimensional (β_0, β_1)



(e) Revealed-preference, 2-dimensional (β_1, β_2)



(f) Odds-based, 2-dimensional (β_1, β_2)



In all three figures, the different axes denote the parameter space in which we have performed the estimation and the the blue dots denote the points in the grid for which we cannot reject the null hypothesis that they correspond to the true value of the parameter (at a 95% confidence level).

$\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. Since

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

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

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

Equation (A.13) follows from the equality $\eta^{-1}r_{ijt} = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$ and the definition of $\mathcal{P}_{ijt}$ in equation (8). ■

Lemma 2 *Suppose the assumptions in equations* (5) *and* (8) *hold. Then*

$$\mathbb{E}\left[\frac{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j))}\middle|\mathcal{J}_{ijt}\right] \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\middle|\mathcal{J}_{ijt}\right]. \tag{A.14}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the assumption in equation (5) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. Since

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

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

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

Equation (A.13) follows from the equality $\eta^{-1}r_{ijt} = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$ and the definition of $\mathcal{P}_{ijt}$ in equation (8). ■

Lemma 3 *Suppose the distribution of Z_{ijt} conditional on $\mathcal{J}_{ijt}$ is degenerate, then*

$$\mathbb{E}\left[\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}\middle|\mathcal{J}_{ijt}\right] \geq \mathbb{E}\left[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\middle|Z_{ijt}\right], \tag{A.15}$$

and

$$\mathbb{E}\left[\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + \eta^{-1}\varepsilon_{ijt}))}\middle|\mathcal{J}_{ijt}\right] \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\middle|Z_{ijt}\right]. \tag{A.16}$$

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

Lemma 4 *Suppose Y is a variable with support in* $(0,1)$*, then*

$$\mathbb{E}\Big[\frac{1-Y}{Y}\Big] \geq \frac{1-\mathbb{E}[Y]}{\mathbb{E}[Y]}, \tag{A.17}$$

and

$$\mathbb{E}\Big[\frac{Y}{1-Y}\Big] \geq \frac{\mathbb{E}[Y]}{1-\mathbb{E}[Y]}. \tag{A.18}$$

Proof: We can rewrite the left hand side of equation (A.17) as

$$\mathbb{E}\Big[\frac{1-Y}{Y}\Big] = \mathbb{E}\Big[\frac{1}{Y} - 1\Big] = \mathbb{E}\Big[\frac{1}{Y}\Big] - 1, \tag{A.19}$$

and the right hand side of equation (A.17) as

$$\frac{1-\mathbb{E}[Y]}{\mathbb{E}[Y]} = \frac{1}{\mathbb{E}[Y]} - 1. \tag{A.20}$$

As Y takes values in the interval $(0,1)$, Jensen's inequality implies

$$\mathbb{E}\Big[\frac{1}{Y}\Big] \geq \frac{1}{\mathbb{E}[Y]}. \tag{A.21}$$

Equations (A.19), (A.20), and (A.21) imply that equation (A.17) holds.

Define a random variable $X = 1 - Y$ and rewrite the left hand side of equation (A.18) as

$$\mathbb{E}\Big[\frac{1-X}{X}\Big].$$

As the support of Y is $(0,1)$, the support of X is also $(0,1)$. Equations (A.19), (A.20), and (A.21) only depend on the property that the support of Y is $(0,1)$. Therefore, from these equations, it must also be true that

$$\mathbb{E}\Big[\frac{1-X}{X}\Big] \geq \frac{1-\mathbb{E}[X]}{\mathbb{E}[X]},$$

and, applying the inequality $X = 1 - Y$, we can conclude that equation (A.18) holds. ■

Corollary 1 *Suppose $\mathcal{P}_{ijt}$ is defined as in equation* (8)*, then*

$$\mathbb{E}\Big[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big] \geq \frac{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}, \tag{A.22}$$

and

$$\mathbb{E}\Big[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big] \geq \frac{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}. \tag{A.23}$$

Proof: Equation (8) implies that the support of $\mathcal{P}_{ijt}$ is the interval $(0,1)$. Therefore, Lemma 4 implies that equations (A.22) and (A.23) hold. ■

Lemma 5 *Suppose the distribution of Z_{ijt} conditional on $\mathcal{J}_{ijt}$ is degenerate and define $\mathcal{P}(Z_{ijt}) = \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]$, with $\mathcal{P}_{ijt}$ defined in equation* (8)*. Then,*

$$\frac{1}{1+B_l(Z_{ijt};\theta)} \leq \mathcal{P}(Z_{ijt}) \leq \frac{B_u(Z_{ijt};\theta)}{1+B_u(Z_{ijt};\theta)}, \tag{A.24}$$

where

$$B_l(Z_{ijt};\theta) = \mathbb{E}\Bigg[\frac{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}\Bigg|Z_{ijt}\Bigg]. \tag{A.25}$$

$$B_u(Z_{ijt};\theta) = \mathbb{E}\Bigg[\frac{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}\Bigg|Z_{ijt}\Bigg], \tag{A.26}$$

Proof: Combining equations (A.15) and (A.22),

$$B_l(Z_{ijt};\theta) \geq \mathbb{E}\Bigg[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\Bigg|Z_{ijt}\Bigg] \geq \frac{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]},$$

and, reordering terms, we obtain the inequality

$$\frac{1}{1+B_l(Z_{ijt};\theta)} \leq \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]. \tag{A.27}$$

Combining equations (A.16) and (A.23),

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

and, reordering terms, we obtain the inequality

$$\frac{B_u(Z_{ijt};\theta)}{1+B_u(Z_{ijt};\theta)} \geq \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]. \tag{A.28}$$

Combining the inequalities in equations (A.27) and (A.28) we obtain equation (A.24). ■

A.7.2 Confidence Set for Change in Export Probability

We describe the procedure we follow to compute the confidence set for the relative change in the average export probability due to a change in the parameter vector θ from its true value θ^* to a counterfactual value $\theta' = \lambda\theta^*$, for some given vector λ. In our empirical application, $\lambda = (0.6, 0.6, 0)$ and θ' thus implies a 40% reduction in the average fixed costs parameters β_0 and β_1. Formally, we describe here how to compute a confidence set for

$$\frac{\overline{\mathcal{P}}_{jt}(\lambda\theta^*)}{\overline{\mathcal{P}}_{jt}(\theta^*)},$$

where

$$\overline{\mathcal{P}}_{jt}(\lambda\theta^*) \equiv N_t^{-1}\sum_{i=1}^{N_t}\mathcal{P}_{jt}(Z_{ijt};\lambda\theta^*),$$
$$\overline{\mathcal{P}}_{jt}(\theta^*) \equiv N_t^{-1}\sum_{i=1}^{N_t}\mathcal{P}_{jt}(Z_{ijt};\theta^*),$$

where N_t is the total number of firms active in period t, both exporters and non-exporters. We base our confidence set on the result from Theorem 3, which states that, for any value of the parameter vector θ and any value of the instrument vector Z_{ijt}, we can compute a lower bound on $\mathcal{P}_{jt}(Z_{ijt};\theta)$ as

$$\mathcal{P}^l_{jt}(Z_{ijt};\theta) \equiv \frac{1}{1+B^l_{jt}(Z_{ijt};\theta)}, \quad \text{with} \quad B^l_{jt}(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1-\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}-\theta_0-\theta_1 dist_j\big)\big)}{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}-\theta_0-\theta_1 dist_j\big)\big)}\middle| Z_{ijt}\right],$$

and an upper bound on $\mathcal{P}_{jt}(Z_{ijt};\theta)$ as

$$\mathcal{P}^u_{jt}(Z_{ijt};\theta) \equiv \frac{B^u_{jt}(Z_{ijt};\theta)}{1+B^u_{jt}(Z_{ijt};\theta)}, \quad \text{with} \quad B^u_{jt}(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}-\theta_0-\theta_1 dist_j\big)\big)}{1-\Phi\big(\theta_2^{-1}\big(\eta^{-1}r_{ijt}-\theta_0-\theta_1 dist_j\big)\big)}\middle| Z_{ijt}\right].$$

Using the result in Theorem 3 and conditioning on a given value of θ, it is immediate to compute a lower bound $\overline{\mathcal{P}}^l_{jt}(\theta)$ and an upper bound $\overline{\mathcal{P}}^u_{jt}(\theta)$ on the true average export probability $\overline{\mathcal{P}}_{jt}(\theta)$ as

$$\overline{\mathcal{P}}^l_{jt}(\theta) \equiv N_t^{-1}\sum_{i=1}^{N_t}\mathcal{P}^l_{jt}(Z_{ijt};\theta),$$
$$\overline{\mathcal{P}}^u_{jt}(\theta) \equiv N_t^{-1}\sum_{i=1}^{N_t}\mathcal{P}^l_{jt}(Z_{ijt};\theta).$$

Given that $\overline{\mathcal{P}}^l_{jt}(\theta) \leq \overline{\mathcal{P}}_{jt}(\theta) \leq \overline{\mathcal{P}}^u_{jt}(\theta)$, we can compute bounds for the the relative change in $\overline{\mathcal{P}}_{jt}(\theta)$ due to a change in the parameter vector of interest from its true value θ^* to a particular counterfactual value $\lambda\theta^*$ as

$$\frac{\overline{\mathcal{P}}^l_{jt}(\lambda\theta^*)}{\overline{\mathcal{P}}^u_{jt}(\theta^*)} \leq \frac{\overline{\mathcal{P}}_{jt}(\lambda\theta^*)}{\overline{\mathcal{P}}_{jt}(\theta^*)} \leq \frac{\overline{\mathcal{P}}^u_{jt}(\lambda\theta^*)}{\overline{\mathcal{P}}^l_{jt}(\theta^*)}. \tag{A.29}$$

The lower bound on the relative change is defined as the ratio of the lower bound on the export probability evaluated at the new value of θ, $\lambda\theta^*$, and the upper bound evaluated at the initial true value of θ, θ^*. The

reverse case is used to compute the upper bound on the relative change.

The bounds in equation (A.29) cannot be directly used for estimation because the true value of the parameter vector θ^* is not identified. Taking into account that our moment inequalities only restrict the true parameter vector θ^* to be contained in a subset of the parameter space Θ defined by the identified set Θ_0, we can derive the following bounds:

$$\min_{\theta\in\Theta_0}\left\{\frac{\overline{\mathcal{P}}^l_{jt}(\lambda\theta)}{\overline{\mathcal{P}}^u_{jt}(\theta)}\right\} \leq \frac{\overline{\mathcal{P}}_{jt}(\lambda\theta^*)}{\overline{\mathcal{P}}_{jt}(\theta^*)} \leq \max_{\theta\in\Theta_0}\left\{\frac{\overline{\mathcal{P}}^u_{jt}(\lambda\theta)}{\overline{\mathcal{P}}^l_{jt}(\theta)}\right\}. \tag{A.30}$$

We compute separate confidence sets for the parameter $\overline{\mathcal{P}}_{jt}(\lambda\theta^*)/\overline{\mathcal{P}}_{jt}(\theta^*)$ for each country-year pair. When computing a confidence set for the parameter $\overline{\mathcal{P}}_{jt}(\lambda\theta^*)/\overline{\mathcal{P}}_{jt}(\theta^*)$, we take into account that both upper and lower bound are a function of two sets of unobserved parameters: (1) a parameter vector that includes all conditional expectations $B^l_{jt}(Z_{ijt};\theta_p)$, $B^u_{jt}(Z_{ijt};\theta_p)$, $B^l_{jt}(Z_{ijt};\lambda\theta_p)$, and $B^u_{jt}(Z_{ijt};\lambda\theta_p)$ for every Z_{ijt} observed in the sample; and, (2) the identified set Θ_0. We therefore compute estimates of the upper and lower bound in equation (A.30) that rely on: (a) a non-parametric estimator of $B^l_{jt}(Z_{ijt};\theta_p)$, $B^u_{jt}(Z_{ijt};\theta_p)$, $B^l_{jt}(Z_{ijt};\lambda\theta_p)$, and $B^u_{jt}(Z_{ijt};\lambda\theta_p)$ for every Z_{ijt}; and, (b) the confidence set for the true parameter vector $\hat{\Theta}^{95\%}$ computed according to the procedure described in Appendix A.5.

Specifically, our procedure to compute confidence intervals for the parameter $\overline{\mathcal{P}}_{jt}(\lambda\theta^*)/\overline{\mathcal{P}}_{jt}(\theta^*)$ for a specific pair jt has seven steps. Steps two to six condition on a particular value of the parameter vector θ in the confidence set $\hat{\Theta}^{95\%}$, θ_p, and show how we estimate the four conditional expectations $B^l_{jt}(Z_{ijt};\theta_p)$, $B^u_{jt}(Z_{ijt};\theta_p)$, $B^l_{jt}(Z_{ijt};\lambda\theta_p)$, and $B^u_{jt}(Z_{ijt};\lambda\theta_p)$ at each observed value of Z_{ijt}. In these steps, we account for the standard errors of these estimates in order to compute a confidence interval for $\overline{\mathcal{P}}_{jt}(\lambda\theta_p)/\overline{\mathcal{P}}_{jt}(\theta_p)$ at the particular value of the parameter vector θ_p. Steps one and seven show how we account the fact that we do not know the value of the true parameter vector θ^* but just a confidence interval $\hat{\Theta}^{95\%}$.

Step 1: Choose a point $\theta_p \in \hat{\Theta}^{95\%}$.

Step 2: For every Z_{ijt} observed in the data, we compute a non-parametric estimate of

$$\{B^l_{jt}(Z_{ijt};\theta_p), B^u_{jt}(Z_{ijt};\theta_p), B^l_{jt}(Z_{ijt};\lambda\theta_p), B^u_{jt}(Z_{ijt};\lambda\theta_p)\}. \tag{A.31}$$

To do so, we run four different non-parametric regressions using a Nadaraya-Watson estimator. Given that the only covariate in the vector Z_{ijt} that varies across firms i within a given country-year pair jt is lagged domestic sales, r_{iht-1}, we use a univariate kernel in each of these regressions. Specifically, we employ the Epanechnikov kernel in all four regressions and use two separate bandwidth parameters $h^l_{jt}(\theta_p)$ and $h^u_{jt}(\theta_p)$, where $h^l_{jt}(\theta_p)$ is employed in the nonparametric regressions of $B^l_{jt}(Z_{ijt};\theta_p)$ and $B^l_{jt}(Z_{ijt};\lambda\theta_p)$ and $h^u_{jt}(\theta_p)$ is employed in the nonparametric regressions of $B^u_{jt}(Z_{ijt};\theta_p)$ and $B^u_{jt}(Z_{ijt};\lambda\theta_p)$. Both $h^l_{jt}(\theta_p)$ and $h^u_{jt}(\theta_p)$ are computed using a cross-validation approach applied to the nonparametric regressions for $B^l_{jt}(Z_{ijt};\theta_p)$ and $B^u_{jt}(Z_{ijt};\theta_p)$, respectively.[51] We denote the predicted values generated by these non-parametric regressions as

$$\{\hat{B}^l_{jt}(Z_{ijt};\theta_p), \hat{B}^u_{jt}(Z_{ijt};\theta_p), \hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p), \hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p)\}, \tag{A.32}$$

and, for every firm i, we compute the residuals from these non-parametric regressions as

$$\{\hat{e}^l_{ijt}(\theta_p), \hat{e}^u_{ijt}(\theta_p), \hat{e}^l_{ijt}(\lambda\theta_p), \hat{e}^u_{ijt}(\lambda\theta_p)\}. \tag{A.33}$$

Using these residuals, we compute an estimate of the asymptotic variance of each our four non-parametric

[51] We have also experimented with four different bandwidths, one for each of the four variables in equation (A.48). The difference in the optimal bandwidths of the non-parametric regression for $B^l_{jt}(Z_{ijt};\theta_p)$ and $B^l_{jt}(Z_{ijt};\lambda\theta_p)$ is negligible, as it is also the case for $B^u_{jt}(Z_{ijt};\theta_p)$ and $B^u_{jt}(Z_{ijt};\lambda\theta_p)$.

estimates as

$$\hat{\mathbb{V}}(\hat{B}^l_{jt}(Z_{ijt};\theta_p)) = \frac{R_{ep}\hat{\mathbb{E}}[(\hat{e}^l_{ijt}(\theta_p))^2|Z_{ijt}]}{N_t h^l_{jt}(\theta_p)\hat{f}_{jt}(Z_{ijt})}$$
$$\hat{\mathbb{V}}(\hat{B}^u_{jt}(Z_{ijt};\theta_p)) = \frac{R_{ep}\hat{\mathbb{E}}[(\hat{e}^u_{ijt}(\theta_p))^2|Z_{ijt}]}{N_t h^u_{jt}(\theta_p)\hat{f}_{jt}(Z_{ijt})}$$
$$\hat{\mathbb{V}}(\hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p)) = \frac{R_{ep}\hat{\mathbb{E}}[(\hat{e}^l_{ijt}(\lambda\theta_p))^2|Z_{ijt}]}{N_t h^l_{jt}(\theta_p)\hat{f}_{jt}(Z_{ijt})}$$
$$\hat{\mathbb{V}}(\hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p)) = \frac{R_{ep}\hat{\mathbb{E}}[(\hat{e}^u_{ijt}(\lambda\theta_p))^2|Z_{ijt}]}{N_t h^u_{jt}(\theta_p)\hat{f}_{jt}(Z_{ijt})}, \tag{A.34}$$

and an estimate of the following two covariances

$$\hat{\mathbb{C}}(\hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p), \hat{B}^u_{jt}(Z_{ijt};\theta_p)) = \frac{R_{ep}\hat{\mathbb{E}}[\hat{e}^l_{ijt}(\lambda\theta_p)\hat{e}^u_{ijt}(\theta_p)|Z_{ijt}]}{N_t h^l_{jt}(\theta_p)\hat{f}_{jt}(Z_{ijt})},$$
$$\hat{\mathbb{C}}(\hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p), \hat{B}^l_{jt}(Z_{ijt};\theta_p)) = \frac{R_{ep}\hat{\mathbb{E}}[\hat{e}^u_{ijt}(\lambda\theta_p)\hat{e}^l_{ijt}(\theta_p)|Z_{ijt}]}{N_t h^u_{jt}(\theta_p)\hat{f}_{jt}(Z_{ijt})}, \tag{A.35}$$

where N_t is the total number of firms active in Chile at period t, $R_{ep} = 3/5$, $\hat{f}_{jt}(Z_{ijt})$ denotes the kernel estimate of the density function of Z in the country-year pair jt evaluated at Z_{ijt}, and $\hat{\mathbb{E}}[\cdot|Z_{ijt}]$ denotes a Nadaraya-Watson estimate of the corresponding function of the residuals listed in equation (A.33). To set a bandwidth for these non-parametric estimates, we again use cross-validation.

Furthermore, an implication of the Nadaraya-Watson estimator is that the random variables in equation (A.32) are asymptotically jointly normally distributed.[52]

Step 3: Transform estimates of the parameters of the distribution of the random variables in equation (A.32) **into estimates of the parameters of the distribution of the random variables**

$$\{\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\theta_p), \hat{\mathcal{P}}^u_{jt}(Z_{ijt};\theta_p), \hat{\mathcal{P}}^l_{jt}(Z_{ijt};\lambda\theta_p), \hat{\mathcal{P}}^u_{jt}(Z_{ijt};\lambda\theta_p)\}.$$

For every Z_{ijt}, we use the estimates in equation (A.32) to compute the point estimates

$$\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\theta_p) = \frac{1}{1+\hat{B}^l_{jt}(Z_{ijt};\theta_p)},$$
$$\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\theta_p) = \frac{\hat{B}^u_{jt}(Z_{ijt};\theta_p)}{1+\hat{B}^u_{jt}(Z_{ijt};\theta_p)},$$
$$\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\lambda\theta_p) = \frac{1}{1+\hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p)},$$
$$\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\lambda\theta_p) = \frac{\hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p)}{1+\hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p)}, \tag{A.36}$$

[52] In an abuse of notation, we use $\{\hat{B}^l_{jt}(Z_{ijt};\theta_p), \hat{B}^u_{jt}(Z_{ijt};\theta_p), \hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p), \hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p)\}$ here to denote both the random variable and the particular value that we find in our sample.

apply the Delta method and use the estimates in equation (A.34) to compute the variances estimates

$$\begin{aligned}
\hat{\mathbb{V}}(\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\theta_p)) &= \frac{1}{(1+\hat{B}^l_{jt}(Z_{ijt};\theta_p))^4}\hat{\mathbb{V}}(\hat{B}^l_{jt}(Z_{ijt};\theta_p)),\\
\hat{\mathbb{V}}(\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\theta_p)) &= \frac{1}{(1+\hat{B}^u_{jt}(Z_{ijt};\theta_p))^4}\hat{\mathbb{V}}(\hat{B}^u_{jt}(Z_{ijt};\theta_p)),\\
\hat{\mathbb{V}}(\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\lambda\theta_p)) &= \frac{1}{(1+\hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p))^4}\hat{\mathbb{V}}(\hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p)),\\
\hat{\mathbb{V}}(\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\lambda\theta_p)) &= \frac{1}{(1+\hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p))^4}\hat{\mathbb{V}}(\hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p)),
\end{aligned} \tag{A.37}$$

and apply again the Delta method and use the estimates in equations (A.34) and (A.35) to compute the covariance estimates

$$\begin{aligned}
\hat{\mathbb{C}}(\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\lambda\theta_p),\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\theta_p)) &= \frac{\hat{\mathbb{C}}(\hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p),\hat{B}^u_{jt}(Z_{ijt};\theta_p))}{(1+\hat{B}^l_{jt}(Z_{ijt};\lambda\theta_p))^2(1+\hat{B}^u_{jt}(Z_{ijt};\theta_p))^2},\\
\hat{\mathbb{C}}(\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\lambda\theta_p),\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\theta_p)) &= \frac{\hat{\mathbb{C}}(\hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p),\hat{B}^l_{jt}(Z_{ijt};\theta_p))}{(1+\hat{B}^u_{jt}(Z_{ijt};\lambda\theta_p))^2(1+\hat{B}^l_{jt}(Z_{ijt};\theta_p))^2}.
\end{aligned} \tag{A.38}$$

Furthermore, an implication of the Delta method and Step 2 is that the random variables

$$\{\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\theta_p),\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\theta_p),\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\lambda\theta_p),\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\lambda\theta_p)\} \tag{A.39}$$

are asymptotically jointly normally distributed for every Z_{ijt} and every θ_p.

Step 4: Transform estimates of the parameters of the distribution of $\mathcal{P}^l_{jt}(Z_{ijt};\cdot)$ and $\mathcal{P}^u_{jt}(Z_{ijt};\cdot)$ into estimates of the parameters of the distribution of

$$\{\hat{\overline{\mathcal{P}}}^l_{jt}(\theta_p),\hat{\overline{\mathcal{P}}}^u_{jt}(\theta_p),\hat{\overline{\mathcal{P}}}^l_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{P}}}^u_{jt}(\lambda\theta_p)\}. \tag{A.40}$$

Using the estimates computed in equation (A.36), we compute

$$\begin{aligned}
\hat{\overline{\mathcal{P}}}^l_{jt}(\theta_p) &= N_t^{-1}\sum_{i=1}^{N_t}\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\theta_p),\\
\hat{\overline{\mathcal{P}}}^u_{jt}(\theta_p) &= N_t^{-1}\sum_{i=1}^{N_t}\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\theta_p),\\
\hat{\overline{\mathcal{P}}}^l_{jt}(\lambda\theta_p) &= N_t^{-1}\sum_{i=1}^{N_t}\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\lambda\theta_p),\\
\hat{\overline{\mathcal{P}}}^u_{jt}(\lambda\theta_p) &= N_t^{-1}\sum_{i=1}^{N_t}\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\lambda\theta_p).
\end{aligned} \tag{A.41}$$

These four expressions are functions of non-parametric estimates and, therefore, random variables. We estimate

their variances using the Delta method and the estimates in equation (A.37) as

$$\hat{\mathbb{V}}(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}\hat{\mathbb{V}}(\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\theta_p)),$$
$$\hat{\mathbb{V}}(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}\hat{\mathbb{V}}(\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\theta_p)),$$
$$\hat{\mathbb{V}}(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}\hat{\mathbb{V}}(\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\lambda\theta_p)),$$
$$\hat{\mathbb{V}}(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}\hat{\mathbb{V}}(\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\lambda\theta_p)), \tag{A.42}$$

and two covariances that will be relevant below can be consistently estimated as

$$\hat{\mathbb{C}}(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}\hat{\mathbb{C}}(\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\lambda\theta_p),\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\theta_p)),$$
$$\hat{\mathbb{C}}(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}\hat{\mathbb{C}}(\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\lambda\theta_p),\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\theta_p)). \tag{A.43}$$

Furthermore, an implication of the Delta method and Step 3 is that the random variables

$$\{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p),\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p),\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)\} \tag{A.44}$$

are asymptotically jointly normally distributed for every Z_{ijt} and every θ_p.

Step 5: Transform estimates of the parameters of the distribution of the random variables in equation (A.53) **into estimates of the parameters of the distribution of**

$$\left\{\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)},\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)}\right\}. \tag{A.45}$$

Using the estimates in equation (A.50), we compute

$$\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)} = \frac{\sum_{i=1}^{N_t}\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\lambda\theta_p)}{\sum_{i=1}^{N_t}\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\theta_p)},$$
$$\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)} = \frac{\sum_{i=1}^{N_t}\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\lambda\theta_p)}{\sum_{i=1}^{N_t}\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\theta_p)},$$

and using the Delta method and the estimates in equations (A.51) and (A.52), we compute the following variance estimates:

$$\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)}\right) = \begin{pmatrix}\frac{1}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)}\\ -\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p))^2}\end{pmatrix}' \begin{pmatrix}\hat{\mathbb{V}}(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)) & \hat{\mathbb{C}}(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p))\\ \hat{\mathbb{C}}(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)) & \hat{\mathbb{V}}(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p))\end{pmatrix}\begin{pmatrix}\frac{1}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)}\\ -\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p))^2}\end{pmatrix},$$

$$\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)}\right) = \begin{pmatrix}\frac{1}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)}\\ -\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p))^2}\end{pmatrix}' \begin{pmatrix}\hat{\mathbb{V}}(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)) & \hat{\mathbb{C}}(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p))\\ \hat{\mathbb{C}}(\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)) & \hat{\mathbb{V}}(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p))\end{pmatrix}\begin{pmatrix}\frac{1}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)}\\ -\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{(\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p))^2}\end{pmatrix}.$$

Furthermore, an implication of the Delta method and Step 3 is that the random variables

$$\left\{ \frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)}, \frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)} \right\}.$$

are asymptotically jointly normally distributed for every Z_{ijt} and every θ_p.

Step 6: compute a confidence set for the parameter

$$\frac{\overline{\mathcal{P}}_{jt}(\lambda\theta_p)}{\overline{\mathcal{P}}_{jt}(\theta_p)},$$

using the information on the asymptotic distribution of (A.54) **derived in step 5.** In order to compute this confidence set, we apply the results in Imbens and Manski (2004) and compute such a confidence interval as

$$\left[\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)} - \overline{C}_{N_t}(\theta_p)\sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)}\right)}, \frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)} + \overline{C}_{N_t}(\theta_p)\sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)}\right)} \right]$$

where $\overline{C}_{N_t}(\theta_p)$ satisfies

$$\Phi\left(\overline{C}_{N_t}(\theta_p) + \sqrt{N_t}\frac{\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)} - \frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)}}{\max\left\{ \sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)}\right)}, \sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)}\right)} \right\}} \right) - \Phi(-\overline{C}_{N_t}(\theta_p)) = 0.95.$$

Step 7: account for the uncertainty in the true value of the parameter vector θ, θ^*. Repeat steps 2 to 6 for every θ_p in the confidence set $\hat{\Theta}^{95\%}$ and compute the resulting confidence set as

$$\left[\min_{\theta_p \in \hat{\Theta}^{95\%}} \left\{ \frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)} - \overline{C}_{N_t}(\theta_p)\sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\theta_p)}\right)} \right\}, \max_{\theta_p \in \hat{\Theta}^{95\%}} \left\{ \frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)} + \overline{C}_{N_t}(\theta_p)\sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{P}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{P}}}^{l}_{jt}(\theta_p)}\right)} \right\} \right]$$

A.7.3 Confidence Set for Change in Aggregate Export Revenues

We describe the procedure we follow to compute the confidence set for the relative change in aggregate export revenues due to a change in the parameter vector θ from its true value θ^* to a counterfactual value $\theta' = \lambda\theta^*$, for some given vector λ. In our empirical application, $\lambda = (0.6, 0.6, 0)$ and θ' thus implies a 40% reduction in the average fixed costs parameters β_0 and β_1. Formally, we describe here how to compute a confidence set for

$$\frac{\overline{\mathcal{R}}_{jt}(\lambda\theta^*)}{\overline{\mathcal{R}}_{jt}(\theta^*)},$$

where

$$\overline{\mathcal{R}}_{jt}(\lambda\theta^*) \equiv N_t^{-1}\sum_{i=1}^{N_t}\mathcal{R}_{jt}(Z_{ijt};\lambda\theta^*),$$

$$\overline{\mathcal{R}}_{jt}(\theta^*) \equiv N_t^{-1}\sum_{i=1}^{N_t}\mathcal{R}_{jt}(Z_{ijt};\theta^*),$$

where N_t is the total number of firms active in period t—including both exporters and non-exporters—and, for every value of the parameter vector θ,

$$\mathcal{R}_{jt}(Z_{ijt};\theta) = r_{ijt}\mathbb{E}[d_{ijt}|Z_{ijt};\theta] = r_{ijt}\mathcal{P}(Z_{ijt};\theta). \tag{A.46}$$

Following the same steps discussed in Appendix A.7.2, we can define bounds similar to those in equation (A.30) as

$$\min_{\theta\in\Theta_0}\left\{\frac{\overline{\mathcal{R}}^l_{jt}(\lambda\theta)}{\overline{\mathcal{R}}^u_{jt}(\theta)}\right\} \leq \frac{\overline{\mathcal{R}}_{jt}(\lambda\theta^*)}{\overline{\mathcal{R}}_{jt}(\theta^*)} \leq \max_{\theta\in\Theta_0}\left\{\frac{\overline{\mathcal{R}}^u_{jt}(\lambda\theta)}{\overline{\mathcal{R}}^l_{jt}(\theta)}\right\}. \tag{A.47}$$

where, for every θ,

$$\overline{\mathcal{R}}^l_{jt}(\theta) = N_t^{-1}\sum_{i=1}^{N_t} r_{ijt}\mathcal{P}^l_{jt}(Z_{ijt};\theta),$$
$$\overline{\mathcal{R}}^u_{jt}(\theta) = N_t^{-1}\sum_{i=1}^{N_t} r_{ijt}\mathcal{P}^u_{jt}(Z_{ijt};\theta),$$

where $\mathcal{P}^l_{jt}(Z_{ijt};\theta)$ and $\mathcal{P}^u_{jt}(Z_{ijt};\theta)$ are defined in Theorem 3.

In order to compute confidence intervals for the parameter $\overline{\mathcal{R}}_{jt}(\lambda\theta^*)/\overline{\mathcal{R}}_{jt}(\theta^*)$ for a specific pair jt we follow seven steps analogous to those described in Appendix A.7.2.

Step 1: Choose a point $\theta_p \in \hat{\Theta}^{95\%}$.

Step 2: For every Z_{ijt} observed in the data, we compute a non-parametric estimate of

$$\{B^l_{jt}(Z_{ijt};\theta_p), B^u_{jt}(Z_{ijt};\theta_p), B^l_{jt}(Z_{ijt};\lambda\theta_p), B^u_{jt}(Z_{ijt};\lambda\theta_p)\}, \tag{A.48}$$

and of its variance-covariance matrix. To do so, we follow exactly the same procedure described in step 2 of the process described in Appendix A.7.2.

Step 3: Transform estimates of the parameters of the distribution of the random variables in equation (A.32) **into estimates of the parameters of the distribution of the random variables**

$$\{\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\theta_p), \hat{\mathcal{P}}^u_{jt}(Z_{ijt};\theta_p), \hat{\mathcal{P}}^l_{jt}(Z_{ijt};\lambda\theta_p), \hat{\mathcal{P}}^u_{jt}(Z_{ijt};\lambda\theta_p)\}.$$

To do so, we follow exactly the same procedure described in step 2 of the process described in Appendix A.7.2.

Step 4: Transform estimates of the parameters of the distribution of $\mathcal{P}^l_{jt}(Z_{ijt};\cdot)$ and $\mathcal{P}^u_{jt}(Z_{ijt};\cdot)$ into estimates of the parameters of the distribution of

$$\{\hat{\overline{\mathcal{R}}}^l_{jt}(\theta_p), \hat{\overline{\mathcal{R}}}^u_{jt}(\theta_p), \hat{\overline{\mathcal{R}}}^l_{jt}(\lambda\theta_p), \hat{\overline{\mathcal{R}}}^u_{jt}(\lambda\theta_p)\}. \tag{A.49}$$

Using the estimates computed in equation (A.36), we compute

$$\hat{\overline{\mathcal{R}}}^l_{jt}(\theta_p) = N_t^{-1}\sum_{i=1}^{N_t} r_{ijt}\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\theta_p),$$
$$\hat{\overline{\mathcal{R}}}^u_{jt}(\theta_p) = N_t^{-1}\sum_{i=1}^{N_t} r_{ijt}\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\theta_p),$$
$$\hat{\overline{\mathcal{R}}}^l_{jt}(\lambda\theta_p) = N_t^{-1}\sum_{i=1}^{N_t} r_{ijt}\hat{\mathcal{P}}^l_{jt}(Z_{ijt};\lambda\theta_p),$$
$$\hat{\overline{\mathcal{R}}}^u_{jt}(\lambda\theta_p) = N_t^{-1}\sum_{i=1}^{N_t} r_{ijt}\hat{\mathcal{P}}^u_{jt}(Z_{ijt};\lambda\theta_p). \tag{A.50}$$

These four expressions are functions of non-parametric estimates and, therefore, random variables. We estimate

their variances using the Delta method and the estimates in equation (A.37) as

$$\hat{\mathbb{V}}(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}(r_{ijt})^2\hat{\mathbb{V}}(\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\theta_p)),$$
$$\hat{\mathbb{V}}(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}(r_{ijt})^2\hat{\mathbb{V}}(\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\theta_p)),$$
$$\hat{\mathbb{V}}(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}(r_{ijt})^2\hat{\mathbb{V}}(\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\lambda\theta_p)),$$
$$\hat{\mathbb{V}}(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}(r_{ijt})^2\hat{\mathbb{V}}(\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\lambda\theta_p)), \tag{A.51}$$

and two covariances that will be relevant below can be consistently estimated as

$$\hat{\mathbb{C}}(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}(r_{ijt})^2\hat{\mathbb{C}}(\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\lambda\theta_p),\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\theta_p)),$$
$$\hat{\mathbb{C}}(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)) = N_t^{-2}\sum_{i=1}^{N_t}(r_{ijt})^2\hat{\mathbb{C}}(\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\lambda\theta_p),\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\theta_p)). \tag{A.52}$$

Furthermore, an implication of the Delta method and Step 3 is that the random variables

$$\{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p),\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p),\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)\} \tag{A.53}$$

are asymptotically jointly normally distributed for every Z_{ijt} and every θ_p.

Step 5: Transform estimates of the parameters of the distribution of the random variables in equation (A.53) **into estimates of the parameters of the distribution of**

$$\left\{\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)},\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)}\right\}. \tag{A.54}$$

Using the estimates in equation (A.50), we compute

$$\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)} = \frac{\sum_{i=1}^{N_t} r_{ijt}\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\lambda\theta_p)}{\sum_{i=1}^{N_t} r_{ijt}\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\theta_p)},$$
$$\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)} = \frac{\sum_{i=1}^{N_t} r_{ijt}\hat{\mathcal{P}}^{u}_{jt}(Z_{ijt};\lambda\theta_p)}{\sum_{i=1}^{N_t} r_{ijt}\hat{\mathcal{P}}^{l}_{jt}(Z_{ijt};\theta_p)},$$

and using the Delta method and the estimates in equations (A.51) and (A.52), we compute the following variance estimates:

$$\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)}\right) = \begin{pmatrix} \frac{1}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)} \\ -\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p))^2} \end{pmatrix}' \begin{pmatrix} \hat{\mathbb{V}}(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)) & \hat{\mathbb{C}}(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)) \\ \hat{\mathbb{C}}(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)) & \hat{\mathbb{V}}(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)) \end{pmatrix} \begin{pmatrix} \frac{1}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)} \\ -\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p))^2} \end{pmatrix},$$

$$\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)}\right) = \begin{pmatrix} \frac{1}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)} \\ -\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p))^2} \end{pmatrix}' \begin{pmatrix} \hat{\mathbb{V}}(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)) & \hat{\mathbb{C}}(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)) \\ \hat{\mathbb{C}}(\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p),\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)) & \hat{\mathbb{V}}(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)) \end{pmatrix} \begin{pmatrix} \frac{1}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)} \\ -\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{(\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p))^2} \end{pmatrix}.$$

Furthermore, an implication of the Delta method and Step 3 is that the random variables

$$\left\{ \frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)}, \frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)} \right\}.$$

are asymptotically jointly normally distributed for every Z_{ijt} and every θ_p.

Step 6: Compute a confidence set for the parameter

$$\frac{\overline{\mathcal{R}}_{jt}(\lambda\theta_p)}{\overline{\mathcal{R}}_{jt}(\theta_p)},$$

using the information on the asymptotic distribution of (A.54) **derived in step 5.** In order to compute this confidence set, we apply the results in Imbens and Manski (2004) and compute such a confidence interval as

$$\left[\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)} - \overline{C}_{N_t}(\theta_p)\sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)}\right)}, \frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)} + \overline{C}_{N_t}(\theta_p)\sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)}\right)} \right]$$

where $\overline{C}_{N_t}(\theta_p)$ satisfies

$$\Phi\left(\overline{C}_{N_t}(\theta_p) + \sqrt{N_t}\frac{\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)} - \frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)}}{\max\left\{ \sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)}\right)}, \sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)}\right)} \right\}} \right) - \Phi(-\overline{C}_{N_t}(\theta_p)) = 0.95.$$

Step 7: Account for the uncertainty in the true value of the parameter vector θ, θ^*. Repeat steps 2 to 6 for every θ_p in the confidence set $\hat{\Theta}^{95\%}$ and compute the resulting confidence set as

$$\left[\min_{\theta_p \in \hat{\Theta}^{95\%}} \left\{ \frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)} - \overline{C}_{N_t}(\theta_p)\sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\theta_p)}\right)} \right\}, \max_{\theta_p \in \hat{\Theta}^{95\%}} \left\{ \frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)} + \overline{C}_{N_t}(\theta_p)\sqrt{\hat{\mathbb{V}}\left(\frac{\hat{\overline{\mathcal{R}}}^{u}_{jt}(\lambda\theta_p)}{\hat{\overline{\mathcal{R}}}^{l}_{jt}(\theta_p)}\right)} \right\} \right]$$

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 period t. We describe the estimation procedure to compute these estimates in Appendix A.2. 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, price index, and market size in destination market j at period t relative to the same variables in the home market in the same time period:

$$\alpha_{jt} \equiv \left[\frac{\tau_{jt}}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right]^{1-\eta}\frac{Y_{jt}}{Y_{ht}}.$$

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 time period t. This is consistent with τ_{jt} being significantly smaller 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.[53]

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 boxe plots for the former two countries appearing higher up in figures B.1 and B.2. Table B.1 contains one additional piece of information on the distribution of $\hat{\alpha}_{jt}$ that is not captured by the figures B.1 and B.2: for some of the countries, α_{jt} is serially correlated and, therefore, for these countries, knowledge of α at any period t will help predict its value in subsequent periods; i.e. $\mathbb{E}[\alpha_{jt+1}|\alpha_{jt}] \neq \mathbb{E}[\alpha_{jt+1}]$.

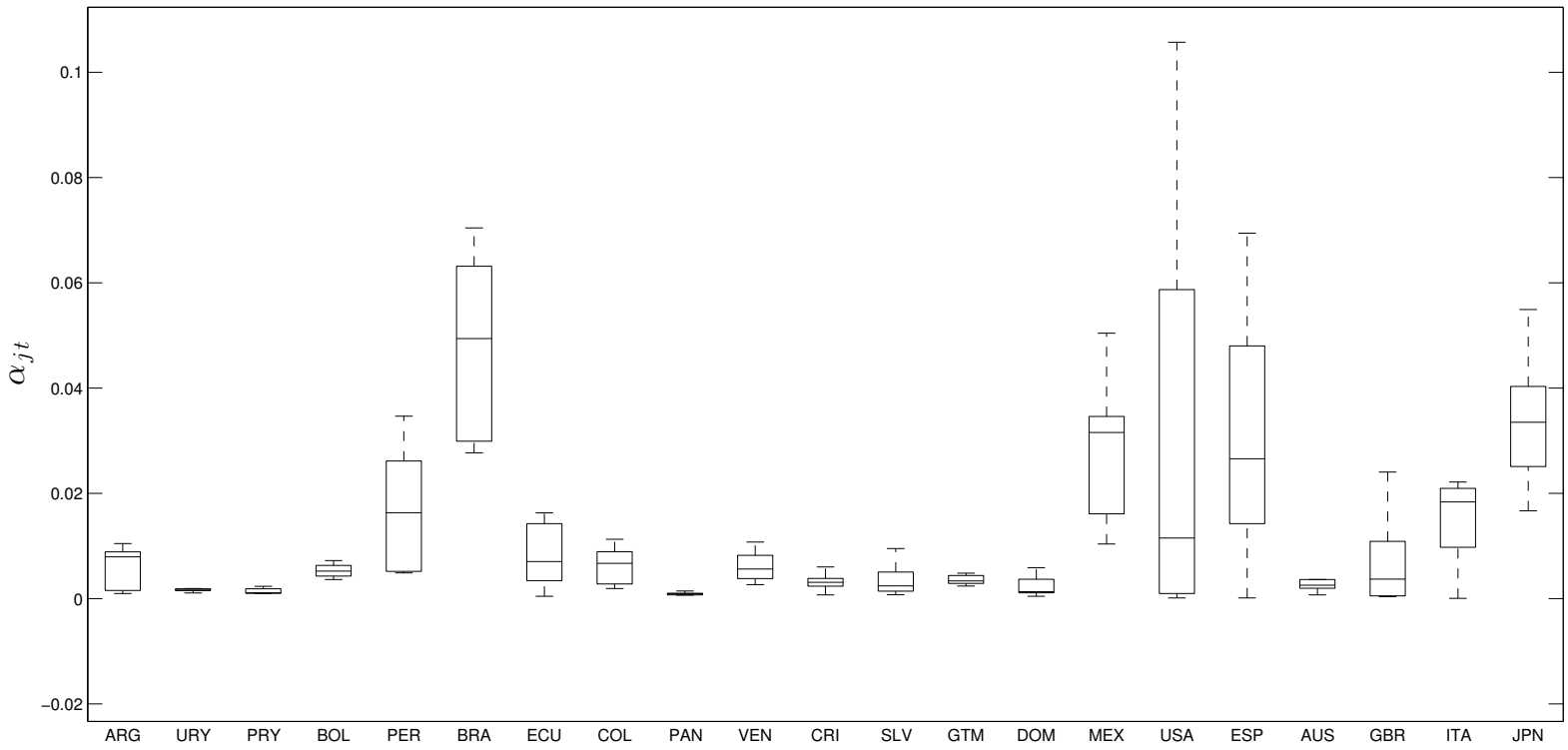
Table B.1: Moments of the distribution of α_{jt}

	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%
Autocorrelation Coef.	0.68	0.36	0.18	-0.17	-0.08	0.24

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

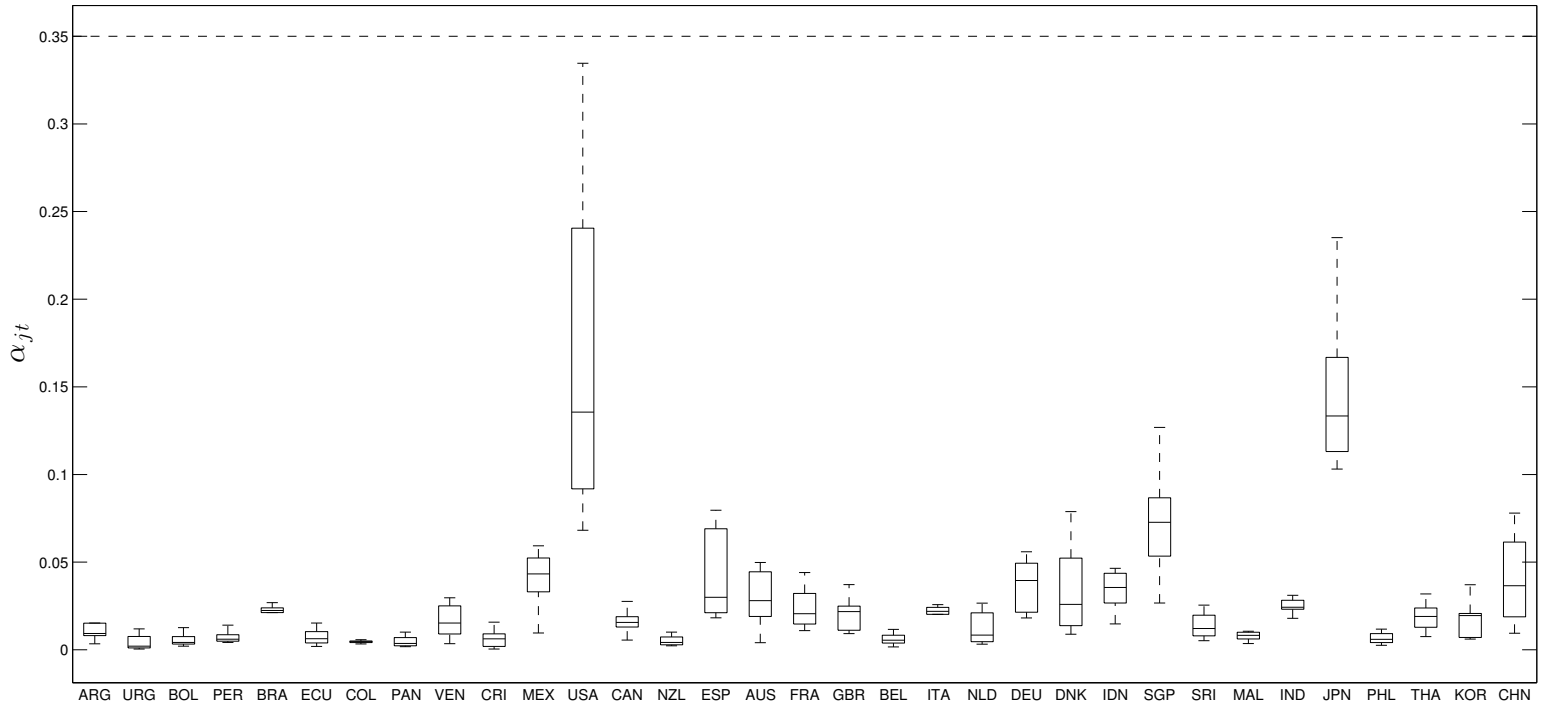
[53] Coşar et al. (2016) provide empirical evidence on the importance of home bias in consumption as a determinant of a firm's home market advantage.

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 period t, the parameter α_{jt} captures the expected potential export revenue a firm might obtain if it exports to j at 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 considers to be not outliers. Points are drawn as 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 period t, the parameter α_{jt} captures the expected potential export revenue a firm might obtain if it exports to j at 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 considers to be not outliers. Points are drawn as 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 Country-Specific Fixed Export Costs

In this section, we report both maximum likelihood and moment inequality estimates of country-specific fixed costs. Specifically, we generalize the model describe in Section 2 in two dimensions: (a) we substitute the specification of fixed export costs in equation (4) by the alternative specification $f_{ijt} = \beta_j + \nu_{ijt}$; (b) we allow the dispersion in fixed export costs to be country-specific; i.e. σ_j may be different from $\sigma_{j'}$ for two different countries j and j'. In order to estimate the parameter vector (β_j, σ_j) for each country j, we divide the samples of both the chemical and the food sector into country-specific subsamples and use each of them separately to compute both the maximum likelihood and moment inequality estimates of this two-parameter vector. For each of these subsamples, we compute the moment inequality confidence sets using both the odds-based and revealed-preference moment inequalities described in Appendix A.4, with the only difference that we exclusively use instrument functions $g_a(Z_{ijt})$ that depend either on firms' lagged domestic sales, r_{iht-1}, or lagged aggregate exports to each destination market j, R_{jt-1}.

Table B.2 and Figure B.3 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, in 16 of them both the perfect foresight and the minimal information maximum likelihood estimates are larger than the upper bound of the moment inequality confidence set. Of the three countries with empty 95% confidence intervals, only two of them have p-values below 0.1%.

Table B.3 and Figure B.4 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 27 of them have perfect foresight maximum likelihood estimates larger than the upper bound of the MI 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 of them have p-values below 0.1%.

As Figures B.3 and B.4 show, it is true for both sectors that these flexibly estimated fixed export costs generally grow with distance, consistent with the benchmark estimates shown in Figure 1. Furthermore, the slope of these fixed costs with respect to distance is significantly larger in the chemicals sector than in the food sector.

Table B.2: Country-specific Fixed Export Costs: Chemicals

	Num. Export Obs.	Perfect Foresight	Minimal Information	Moment Inequality
Argentina	511	794.7 (77.7)	535.5 (34.6)	[125.0, 464.1] (0.102)
Australia	86	4,009.9 (989.6)	2,077.5 (308.6)	[265.4, 1,532.2] (0.146)
Bolivia	617	463.1 (37.6)	388.1 (26.3)	[264.6, 476.1] (0.054)
Brazil	269	18,638.6 (2,673.8)	13,870.4 (1,608.6)	[3,145.3, 8,175.7] (0.081)
Colombia	274	889.4 (63.5)	835.5 (54.9)	[424.6, 492.5] (0.012)
Costa Rica	163	1,940.1 (356.3)	1,429.8 (198.8)	[345.5, 679.7] (0.056)
Dominican Republic	103	2,368.2 (501.1)	1,231.3 (165.2)	[202.1, 497.7] (0.042)
Ecuador	393	1,861.1 (223.5)	1,219.4 (103.4)	[-, -] (0.003)
Spain	88	36,312.5 (7,303.4)	16,879.1 (2,033.8)	[2,156.7, 16,744.3] (0.120)
Great Britain	66	32,455.5 (29,157.9)	10,454.8 (3,303.6)	[-, -] (0.001)

Table B.2: Country-specific Fixed Export Costs: Chemicals (cont.)

	Num. Export Obs.	Perfect Foresight	Minimal Information	Moment Inequality
Guatemala	126	1,600.5 (213.4)	1,351.3 (159.8)	[265.6, 865.7] (0.104)
Italy	58	18,660.6 (4,759.2)	17,797.6 (4,401.5)	[2,274.0, 9,184.3] (0.124)
Japan	59	28,616.2 (6,294.2)	25,346.2 (5,295.0)	[8,500.4, 40,028.0] (0.199)
Mexico	173	12,607.8 (1,701.6)	8,028.7 (813.4)	[2,005.7, 6,903.3] (0.067)
Panama	153	379.1 (43.4)	351.3 (38.5)	[132.8, 191.7] (0.030)
Peru	651	850.5 (52.3)	883.2 (50.8)	[458.4, 616.5] (0.023)
Paraguay	324	294.8 (29.7)	237.6 (20.1)	[101.0, 205.9] (0.092)
El Salvador	83	2,529.0 (429.7)	1,749.6 (222.6)	[310.7, 1,327.5] (0.184)
Uruguay	314	315.1 (29.9)	271.7 (22.6)	[106.4, 200.1] (0.043)
United States	201	61,239.5 (21,847.0)	45,624.9 (16,171.7)	[6,658.6, 11.543.7] (0.057)
Venezuela	215	1,492.8 (198.6)	1,780.1 (195.6)	[-, -] (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.3: Country-specific Fixed Export Costs: Food

	Num. Export Obs.	Perfect Foresight	Minimal Information	Moment Inequality
Argentina	363	3,220,4 (424.3)	2,116,1 (195.8)	[-, -] (0.001)
Australia	149	19,304.6 (524.7)	12,689.0 (308.6)	[2,200.4, 3,281.1] (0.020)
Belgium	123	4,806.3 (1,294.2)	2,277.0 (347.3)	[498.0, 2,289.5] (0.171)
Bolivia	149	3,223.6 (629.5)	1,879.0 (246.7)	[425.4, 1,053.9] (0.104)
Brazil	368	8,643.3 (1,942.5)	7,421.3 (1,477.5)	[2,430.6, 2,430.6] (0.010)
Canada	263	8,579.8 (1,801.1)	5,668.0 (875.6)	[-, -] (0.007)
China	265	9,913.8 (1,036.4)	6,146.0 (449.8)	[2,744.7, 5,326.4] (0.053)
Colombia	301	3,858.6 (821.6)	1,356.3 (126.5)	[327.7, 540.2] (0.066)
Costa Rica	105	5,559.0 (1,958.7)	2,544.9 (468.6)	[595.4, 1,563.3] (0.046)

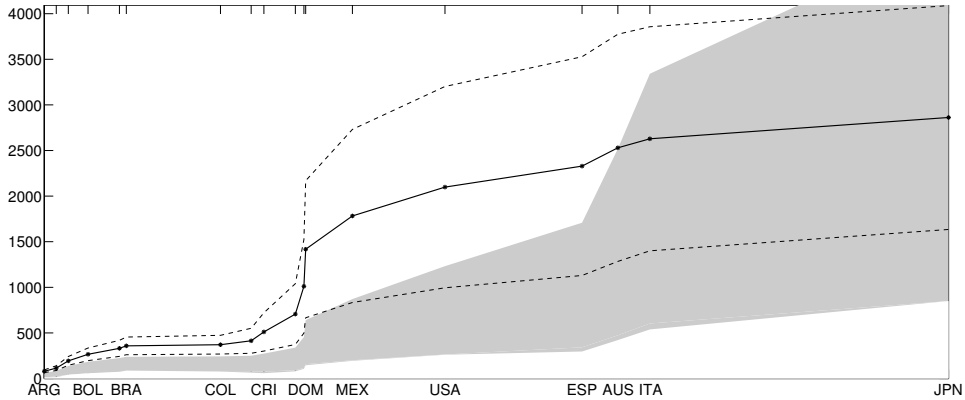
Table B.3: Country-specific Fixed Export Costs: Food (cont.)

	Num. Export Obs.	Perfect Foresight	Minimal Information	Moment Inequality
Germany	319	8,866.4 (999.2)	7,216.1 (711.4)	[2,821.3, 6,204.7] (0.188)
Denmark	111	26,272.3 (1,028.0)	16,075.6 (3,675.5)	[2,766.0, 16,164.4] (0.098)
Ecuador	185	2,103.6 (249.4)	1,308.8 (106.6)	[410.4, 838.5] (0.066)
Spain	258	36,456.3 (16,958.1)	22,561.78 (7,227.1)	[4,772.8, 6,770.2] (0.022)
France	257	11,537.5 (2,226.2)	8,387.0 (1,299.6)	[2,302.6, 5,694.1] (0.009)
Great Britain	214	5,980.3 (687.6)	4,979.7 (501.7)	[1,983.5, 2,223.9] (0.001)
Indonesia	122	12,892.5 (1,751.1)	11,067.4 (1,395.6)	[3,017.0, 4,767.0] (0.036)
India	72	13,556.5 (2,413.9)	8,383.7 (997.8)	[2,167.9, 4,941.6] (0.103)
Italy	167	11,956.8 (2,115.5)	6,840.9 (773,5)	[1,665.5, 4,202.3] (0.109)
Japan	636	22,051.8 (2,675.1)	18,856.4 (1,949.1)	[9,935.7, 13,390.2] (0.016)
South Korea	207	6,486.9 (921.7)	5,212.0 (662.4)	[1,430.9, 3,538.5] (0.161)
Mexico	321	8,091.8 (807.1)	6,329.3 (529.5)	[2,391.8, 4,440.9] (0.072)
Malaysia	98	3,277.1 (496.9)	2,885.4 (410.1)	[839.0, 1,112.1] (0.017)
Netherlands	185	8,407.6 (1,802.6)	4,275.4 (595.3)	[777.0, 3,243.9] (0.174)
New Zealand	102	3,003.2 (583.8)	2,368.2 (396.8)	[612.3, 712.3] (0.011)
Panama	109	3,402.3 (896.8)	1,821.8 (305.6)	[358.1, 1,821.8] (0.138)
Peru	282	1,664.6 (173.5)	1,222.1 (99.8)	[384.6, 1,212.3] (0.191)
Philippines	116	3,554.8 (557.4)	2,251.7 (250.4)	[600.8, 1,442.6] (0.098)
Singapore	117	7,048.5 (1,309.2)	5,868.7 (1,003.8)	[1,218.3, 2,471.2] (0.016)
Thailand	155	17,448.7 (3,587.2)	17,466.5 (3,909.3)	[2,866.5, 2,924.4] (0.013)
Uruguay	184	5,859.8 (2,826.3)	1,731.8 (293.1)	[-, -] (0.006)
United States	595	76,548.9 (21,436.6)	52,183.9 (11,211.1)	[-, -] (0.001)
Venezuela	231	8,495.8 (1,719.7)	6,821.7 (1,220.7)	[-, -] (0.007)

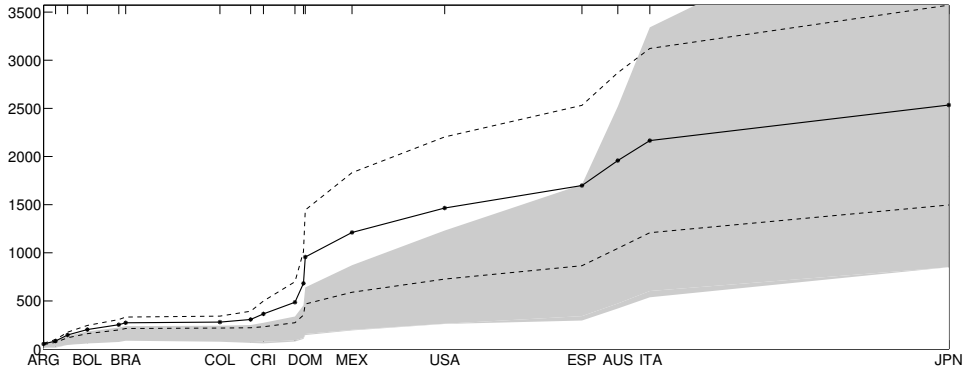
Notes: All variables are reported in thousands of year 2000 USD. For the two ML estimators, standard errors are reported in parentheses. For the moment inequality estimates, extreme points of the 95% confidence set are reported in square brackets and p-values are reported in parenthesis. The MI confidence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

Figure B.3: Country-specific fixed costs estimates: Chemicals

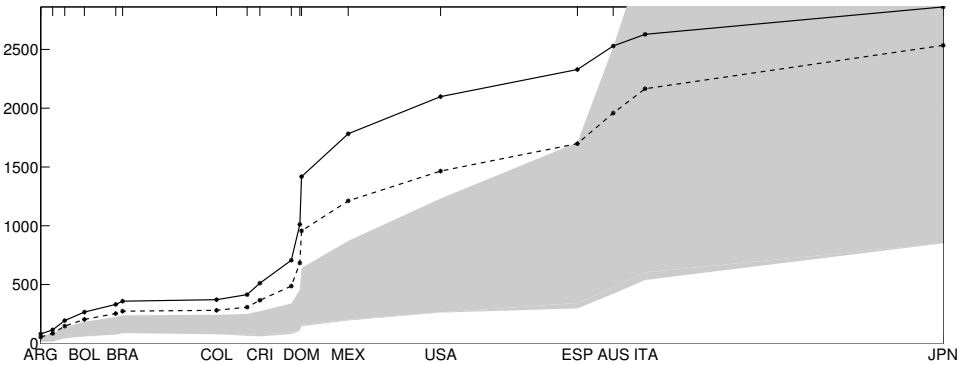
(a) MI & Perfect Foresight ML Confidence Sets



(b) MI & Minimal Information ML Confidence Sets



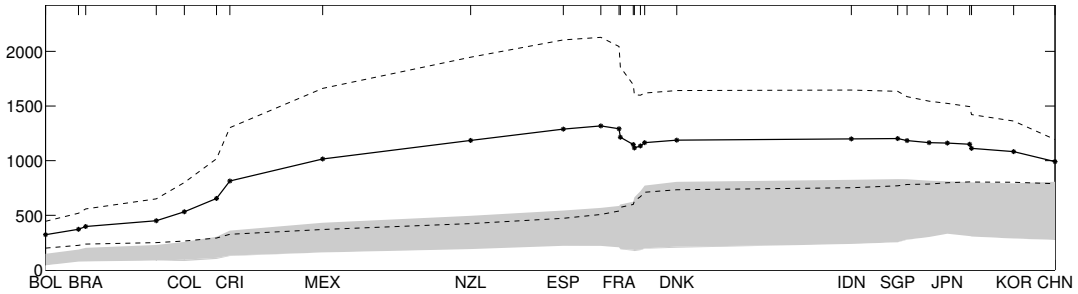
(c) MI Confidence Set & ML Point Estimates



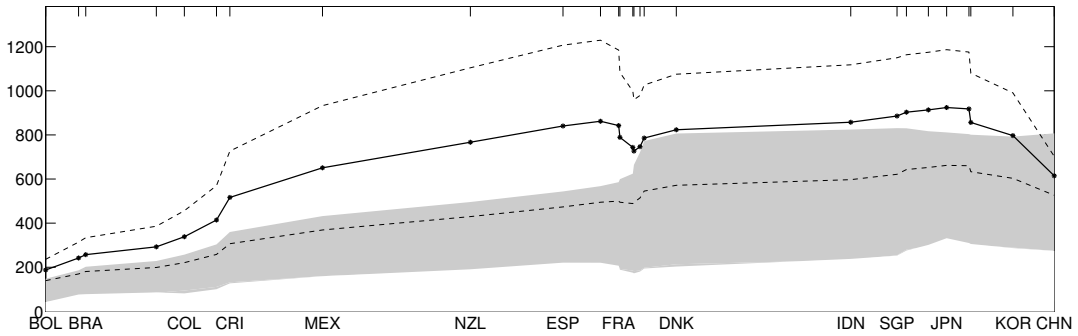
In all the three figures, the vertical axis indicates fixed export costs in thousands of year 2000 USD and the horizontal axis indicates different export destinations. The countries are placed along the horizontal axis according to their distance to Chile and we have limited the labeling to only a few countries for aesthetic purposes. 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.

Figure B.4: Country-specific fixed costs estimates: Food

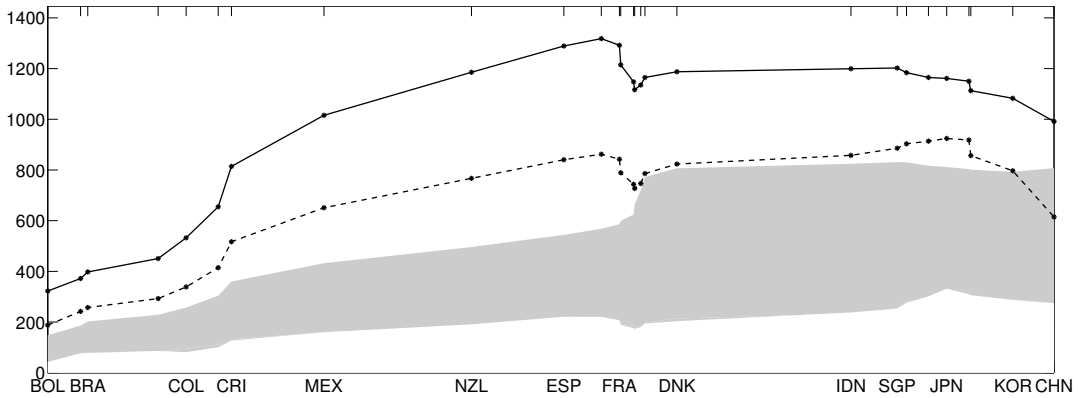
(a) MI & Perfect Foresight ML Confidence Sets



(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 aesthetic purposes. 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.

B.3 Quantiles of Distribution of Fixed Export Costs across Firms

Given equations (4) and (5),

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

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 firm and time periods for a given country j. Given equation (5) 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 $(\beta_0, \beta_1, \sigma)$, 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.4: 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 parentheses. Confidence sets are computed using the procedure described in Andrews and Soares (2010). All the information in this table is reflected in Figure 2

B.4 Confidence Set Computed Using Subsets of Moment Inequalities

For both the chemicals and food sector, Table B.5 reports extreme points of the 95% confidence set of fixed export costs for Argentina, Japan and the United States using three different moment inequality estimators. For the sake of facilitating the comparison, the first row displays again our benchmark estimates, reported also in Table 3. The second row displays equivalent confidence sets computed using exclusively the odds-based inequalities described in Section 4.2.1. The third row reports analogous confidence sets computed using exclusively the revealed-preference inequalities described in Section 4.2.2. All three confidence sets in Table B.5 are computed using a finite number of unconditional moment inequalities; specifically, in all three cases we maintain the set of instrument functions introduced in Appendix A.4. The results show that the confidence sets that exploit only the odds-based or only the revealed-preference inequalities are always strictly larger than the confidence sets that simultaneously exploit both sets of moment inequalities.

Table B.5: Fixed export costs: different moment inequality estimators

	Chemicals			Food		
	Argentina	Japan	United States	Argentina	Japan	United States
Both	[79.1, 104.1]	[309.2, 420.5]	[181.3, 243.6]	[175.6, 270.1]	[269.1, 361.0]	[227.3, 308.9]
Only Odds-Based	[66.6, 164.7]	[269.7, 694.2]	[165.4, 395.1]	[75.6, 690.4]	[141.3, 1,693.7]	[130.2, 1,049.1]
Only Rev. Pref.	[58.8, 144.4]	[274.4, 567.8]	[181.3, 300.1]	[130.1, 1319.7]	[244.4, 1,636.5]	[224.6, 1,381.3]

Notes: Extreme points of the confidence set are reported in parentheses. Confidence sets are computed using the procedure described in Andrews and Soares (2010).

B.5 What Do Exporters Know? Additional Details

B.5.1 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 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.6 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, these are never large enough to modify the conclusion of our 5% significance level tests; i.e. either both the BP and RC values are above 5% or both are below 5%.

B.5.2 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 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, $\varepsilon_{ijt} \equiv r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, 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_{ijt}, even if Z_{ijt} belongs to the information set of exporters $\mathcal{J}_{ijt}$. Second, the mean independence condition will hold if the set of covariates Z_{ijt} is relevant to predict r_{ijt} and the distribution of Z_{ijt} conditional on the information set $\mathcal{J}_{ijt}$ is degenerate. To rule out the

Table B.6: Testing Content of Information Sets

Set of Firms	Set of Export Destinations	Variable Tested	Chemicals Reject at 5%	Chemicals p-value BP	Food Reject at 5%	Food p-value BP
All	All	$(dist_j, r_{iht-1}, R_{jt-1})$	No	0.140	No	0.975
All	All	α_{jt}	Yes	0.020	Yes	0.005
Large	Popular	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	No	0.135	No	0.940
Large	Unpopular	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	No	0.110	No	0.985
Small	Popular	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	Yes	0.005	Yes	0.005
Small	Unpopular	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	Yes	0.030	Yes	0.005
Small & Exporter$_{t-1}$	All	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	Yes	0.005	Yes	0.005
Large & Non-exporter$_{t-1}$	All	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	No	0.145	No	0.995
Small & Non-Exporter$_{t-1}$	All	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	Yes	0.005	Yes	0.005
Large & Exporter$_{t-1}$	All	$(dist_j, r_{iht-1}, R_{jt-1}, \alpha_{jt-1})$	No	0.110	No	0.985

Notes: Firm i at period t is defined as *Large* if $\mathbb{1}\{domsales_{jt-1} \geq median(domsales_{jt-1})\} = 1$ and as *Small* if $\mathbb{1}\{domsales_{jt-1} < median(domsales_{jt-1})\} = 1$. Country j at period t is defined as *Popular* if $\mathbb{1}\{N_{jt-1} \geq median(N_{jt-1})\} = 1$, where N_{jt-1} denotes the total number of Chilean firms in the corresponding sector (chemicals or food) to export to j at period t, and as *Unpopular* if $\mathbb{1}\{N_{jt-1} < median(N_{jt-1})\} = 1$. We define a firm i at period t as *Exporter*$_{t-1}$ with respect to a country j if $d_{ijt-1} = 1$ and as a *Non-exporter*$_{t-1}$ if $d_{ijt-1} = 0$.

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 are contained in Table 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.

Each of the ten columns in Table B.7 correspond to each of the ten rows in tables 5 and B.6. The results show that, for all subsets of firms and countries considered in tables 5 and B.6, each of the covariates included in the vector Z_{ijt} is relevant as a predictor of the potential export revenues $r_{ijt} \equiv \alpha_{jt} r_{iht}$. 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 $R_{jt} \equiv \alpha_{jt} \sum_{i=1}^{N_t} r_{iht}$ are impacted by 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_{ijt} across destinations.

We note here that our model in Section 2 does not impose assumptions on the functional form of the relationship between the predicted revenue r_{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_{ijt}, the researcher need only find at least one functional form that relates Z_{ijt} to r_{ijt}. In Table B.7, we assume a linear relationship between r_{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.7: Instrument Relevance

Panel A: Chemicals

Covariates	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
R_{jt-1}	0.010^a	0.007^a	0.010^a	0.002	0.001^a	0.000	0.001^a	0.013^a	0.001^a	0.017^a
	(20.7)	(11.1)	(6.58)	(0.56)	(11.8)	(0.71)	(4.09)	(8.59)	(19.32)	(5.69)
r_{iht-1}	0.012^a	0.012^a	0.013^a	0.010^a	0.016^a	0.013^a	0.014^a	0.014^a	0.015^a	0.009^a
	(16.0)	(16.1)	(9.95)	(11.5)	(32.6)	(29.0)	(6.32)	(10.7)	(41.77)	(13.09)
$dist_j$	0.201^a	0.159^a	0.339^a	0.442^a	0.016^a	0.022^a	0.050^a	0.253^a	0.014^a	0.579^a
	(23.9)	(12.8)	(5.11)	(21.9)	(8.19)	(26.1)	(2.85)	(11.3)	(21.16)	(5.12)
α_{jt-1}		2.817^a	12.87^a	3.896^b	0.585^a	0.195^a	0.22	4.14^a	0.238^a	13.18^a
		(5.14)	(7.00)	(2.27)	(12.3)	(3.50)	(0.68)	(3.78)	(7.63)	(3.82)
Firms	All	All	Large	Large	Small	Small	Small & Exporter$_{t-1}$	Large & Non-Exporter$_{t-1}$	Small & Non Exporter$_{t-1}$	Large & Exporter$_{t-1}$
Countries	All	All	Popular	Unpopular	Popular	Unpopular	All	All	All	All
Num. Obs.	44,037	44,037	11,047	10,982	11,196	10,812	664	17,887	21,344	4,142
R^2	0.293	0.295	0.343	0.283	0.391	0.195	0.230	0.324	0.314	0.305

Panel B: Food

Covariates	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
R_{jt-1}	0.005^a	0.003^a	0.007^a	-0.004^a	0.001^a	0.000	0.001^a	0.006^a	0.001^a	0.007^a
	(28.2)	(15.7)	(13.9)	(-2.71)	(19.7)	(-1.00)	(5.26)	(11.3)	(21.96)	(10.49)
r_{iht-1}	0.028^a	0.028^a	0.036^a	0.018^a	0.041^a	0.019^a	0.075^a	0.027^a	0.028^a	0.032^a
	(33.2)	(33.3)	(22.2)	(26.4)	(30.8)	(43.9)	(6.55)	(29.0)	(48.01)	(10.76)
$dist_j$	0.055^a	0.035^a	-0.035	0.091^a	-0.002^a	0.004^a	-0.023^b	0.042^a	0.004^a	0.065
	(10.5)	(5.40)	(-1.58)	(14.0)	(-2.84)	(14.4)	(-2.14)	(3.03)	(9.95)	(1.65)
α_{jt-1}		2.947^a	5.876^a	9.983^a	0.171^a	0.526^a	0.375^a	7.359^a	0.251^a	5.381^a
		(8.86)	(6.76)	(18.4)	(7.29)	(23.2)	(2.72)	(9.48)	(15.05)	(5.29)
Firms	All	All	Large	Large	Small	Small	Small & Exporter$_{t-1}$	Large & Non-Exporter$_{t-1}$	Small & Non Exporter$_{t-1}$	Large & Exporter$_{t-1}$
Countries	All	All	Popular	Unpopular	Popular	Unpopular	All	All	All	All
Num. Obs.	84,975	84,975	21,341	21,163	21,519	20,952	1,250	36,736	41,221	5,768
R^2	0.315	0.322	0.388	0.391	0.294	0.244	0.205	0.372	0.334	0.359

Notes: ^{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. Conversely, *Unpopular* denotes $\mathbb{1}\{N_{jt-1} < median(N_{jt-1})\} = 1$. *Exporter*$_{t-1}$ denotes $d_{ijt-1} = 1$ and *Non-exporter*$_{t-1}$ is $d_{ijt-1} = 0$.

B.6 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, price indices and aggregate market size in both country j and in 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_{jt} and τ_{ht} as exogenous parameters and, therefore, within the theoretical framework they may be invariant to average fixed export costs. The invariance of the price index and market size in destination country j, P_{jt} and Y_{jt}, 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%
(continuation on next page)					

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

	1999	2000	2001	2002	2003
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%
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 Section C.1.1 and the second one in Section C.1.2. The first proof makes use of the score function corresponding to the model described in Section 2. The heuristic derivation of the odds-based inequalities in Section 4.2.1 follows the proof in Section C.1.2.

C.1.1 First Proof of Theorem 1

Lemma 6 *Let $L(d_{ijt}|\mathcal{J}_{ijt}, dist_j; \theta)$ denote the log-likelihood conditional on $\mathcal{J}_{ijt}$. Suppose equation (8) holds. Then:*

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

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

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

The score function is given by

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

Reordering terms

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

Given that

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

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

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

we can simplify:

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

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

Lemma 7 *Suppose the equations* (5) *and* (8) *hold. Then*

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

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}] = 0$. From equations (4), (7) and the assumption that $dist_j \subseteq \mathcal{J}_{ijt}$ it follows that d_{ijt} may be written as a function of the vector $(\mathcal{J}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, d_{ijt}] = 0$. Since

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

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

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

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

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

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

and

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

Proof: The result follow from Lemmas 6 and 7 and the application of the Law of Iterated Expectations. ■

Lemma 8 *Let $L(d_{ijt}|\mathcal{J}_{ijt}, dist_j; \theta)$ denote the log-likelihood conditional on $\mathcal{J}_{ijt}$. Suppose equation* (8) *holds. Then:*

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

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

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

Given that

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

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

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

we can simplify:

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

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

Lemma 9 *Suppose the equations* (5) *and* (8) *hold. Then*

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

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt}=r_{ijt}-\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}]=0$ where, as reminder, the set $\mathcal{J}_{ijt}$ includes every covariate known by the firm at the time of deciding on export destinations. From equations (4) and (7), it follows that d_{ijt} may be written as a function of the vector $(\mathcal{J}_{ijt},,dist_j,\nu_{ijt})$; i.e. $d_{ijt}=d(\mathcal{J}_{ijt},dist_j,\nu_{ijt})$. 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 $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt},d_{ijt}]=0$, by Jensen's Inequality

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

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

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

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

and

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

Proof: The results follow from Lemmas 8 and 9 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 (15) and (15b). Combining equations (C.9) and (C.10), we obtain the inequality defined by equations (15) and (15c). ■

C.1.2 Second Proof of Theorem 1

Lemma 10 *Suppose equations* (5) *and* (7) *hold. Then:*

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

Proof: Suppose equation (7) holds. Then:

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

or, equivalently,

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

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 $(\mathcal{J}_{ijt}, dist_j)$:

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

Imposing the distributional assumption in equation (5),

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

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

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

Adding and subtracting $1-d_{ijt}$

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

and, doing some simple algebra, we obtain

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

and, regrouping terms,

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

Lemma 11 *Suppose equations* (5) *and* (7) *hold. Then:*

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

Proof: Suppose equation (7) holds. Then:

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

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 $(\mathcal{J}_{ijt}, dist_j)$:

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

or, equivalently,

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

Imposing the distributional assumption in equation (5),

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

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

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

Adding and subtracting d_{ijt}

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

and, doing some simple algebra, we obtain

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

Second Proof of Theorem 1 Combining equations (C.4), (C.5), (C.6), and (C.11), we obtain the inequality defined by equations (15) and (15b). Combining equations (C.8), (C.9), (C.10), and (C.12), we obtain the inequality defined by equations (15) and (15c)

C.2 Proof of Theorem 2

Lemma 12 *Suppose equations* (4) *and* (7) *hold. Then,*

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

Proof: From equations (4) and (7),

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

This implies

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

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

Lemma 13 *Suppose equations* (4), (5), *and* (7) *hold. Then*

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

Proof: From equation (C.13),

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

Since the assumption in equation (5) implies that $\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}]=0$, 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

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

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

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

Using the definition of d_{ijt} in equation (7), 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 (5), we can rewrite

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

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

Lemma 14 *Suppose the equation* (5) *holds. Then*

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

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

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

From equation (5) 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 15 *Suppose equation* (5) *holds. Then*

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

Proof: From the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, the assumption in equation (5), 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 equations (4) and (7), 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

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

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

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

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

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

and

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

Proof: The results follow from Lemmas 13, 14 and 15 and the application of the Law of Iterated Expectations. ■

Lemma 16 *Suppose equations* (4) *and* (7) *hold. Then,*

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

Proof: From equations (4) and (7),

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

This implies

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

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

Lemma 17 *Suppose equations* (4), (5), *and* (7). *Then*

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

Proof: From equation (C.24),

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

Since the assumption in equation (5) implies that $\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}]=0$, 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

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

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

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

Using the definition of d_{ijt} in equation (7), it follows

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

and, following equation (5), we can rewrite

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

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

Lemma 18 *Suppose equation* (5) *holds. Then*

$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}, dist_j\Big]$$
$$= \mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}, dist_j\Big] \tag{C.29}$$

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

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

From equation (5) and the definition of $\mathcal{J}_{ijt}$ as the information set firm i uses to predict revenue when it decides whether to export, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = 0$. From equations (4), (7), 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,

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

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

Lemma 19 *Suppose equation* (5) *holds. Then*

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

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the definition of $\mathcal{J}_{ijt}$ as the information set firm i uses to predict revenue when it decides whether to export, that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, dist_j, \nu_{ijt}] = 0$. From equations (4), (7), 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

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

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

Corollary 5 *Suppose the distribution of Z_{ijt} conditional on $(\mathcal{J}_{ijt}, dist_j)$ is degenerate, then*

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

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

and

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

Proof of Theorem 2 Combining equations (C.21), (C.22), and (C.23) we obtain the inequality defined by equations (18) and (18b). Combining equations (C.32), (C.33), and (C.34) we obtain the inequality defined by equations (18) and (18c). ■

D Bias in Maximum Likelihood Estimates

We provide here additional details on the content of Section 4.1. 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 potential export revenue. In Section D.2, we report results from several 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 its potential revenue, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, using observable data. If the researcher assumes perfect foresight, she sets the proxy equal to the observed export revenues r_{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 $\mathbb{J}^a_{ijt}$, and uses the projection as her proxy.

As in equation (14), we use ξ_{ijt} to denote the difference between the researcher's assumed proxy $\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$ and the firm's true expectation $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$. Here, if $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ is the true unobserved covariate entering the firm's export decision (see equation (7)) and $\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$ is the researcher's proxy for it, then ξ_{ijt} represents measurement error in the definition of the proxy.

We now evaluate the bias when the researcher assumes $\mathbb{E}[r_{ijt}|\mathcal{J}^a_{ijt}]$ is a perfect proxy for firms' unobserved expectations $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$. Under this assumption, and using the decision rule in equation (7) and the distributional assumption in equation (5), the researcher will assume that the random variable d_{ijt} follows:

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

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

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

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

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

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

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 is

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

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). This difference arises because the conditional density function $f_\chi(\chi|\mathcal{J}^a_{ijt}, dist_j; \theta_2)$ in equation (D.4) is different from the conditional density function $f_\nu(\nu|\mathcal{J}^a_{ijt}, dist_j; \theta_2)$ in equation (D.2). These density functions, and the resulting log likelihood maximand, may differ for three reasons.

First, statistical independence may fail. While equation (5) 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_{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 (5) 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, differences in the variance parameter θ_2 can generate differences in the estimates. 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 likelihood function in equation (D.4) and the researcher's assumed 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 $var(\chi_{ijt}) = var(\nu_{ijt} + \eta^{-1}\xi_{ijt}) = var(\nu_{ijt}) + \eta^{-2}var(\xi_{ijt}) = \sigma^2 + \eta^{-2}var(\xi_{ijt})$. In this case, the parameter vector (θ_0, θ_1) that maximizes the log-likelihood function in equation (D.2) will be identical to the true parameter vector (β_0, β_1). Conversely, 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 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 in the description of the model. Each firm i decides whether to export according to the decision rule

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

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 $\sigma^2 = 1$:

$$\nu_i \sim \mathbb{N}(0, \sigma^2). \tag{D.6}$$

For the actual revenue from exporting in our simulation, we set:

$$r_i = x_{1i} + x_{2i} + x_{3i}, \tag{D.7}$$

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 most empirical settings, 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 $\{\mathcal{J}_i^a; i = 1, \ldots, N\}$).[54]

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,

$$\mathbb{E}[r_i|\mathcal{J}_i] = x_{1i} \qquad \text{and} \qquad \mathbb{E}[r_i|\mathcal{J}_i^a] = r_i = x_{1i} + x_{2i} + x_{3i}, \tag{D.8}$$

and, from equation (14), the measurement error introduced by the misspecification of agents' expectations is thus

$$\xi_i = \mathbb{E}[r_i|\mathcal{J}_i^a] - \mathbb{E}[r_i|\mathcal{J}_i] = r_i - x_{1i} = x_{2i} + x_{3i}, \tag{D.9}$$

which is identical to the expectational error that firm i makes; i.e. $\xi_i = r_i - \mathbb{E}[r_i|\mathcal{J}_i]$. 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

$$L_a(\theta|d, \mathcal{J}^a) = \sum_{i=1}^{N} \Big\{ d_i \Phi(\theta_2^{-1}(\eta^{-1} r_i - \theta_0)) + (1 - d_i)(1 - \Phi(\theta_2^{-1}(\eta^{-1} r_i - \theta_0))) \Big\}. \tag{D.10}$$

However, the correct log-likelihood function is:

$$L(\theta|d, \mathcal{J}) =$$
$$\sum_{i=1}^{N} \Big\{ d_{ijt} \int_{\nu + \eta^{-1}(x_{2i} + x_{3i})} \mathbb{1}\{\eta^{-1} r_i - \theta_0 - (\nu_i + \eta^{-1}(x_{2i} + x_{3i})) \geq 0\} f(\nu_i + \eta^{-1}(x_{2i} + x_{3i})|r_i; \theta_2)$$
$$+ (1 - d_{ijt}) \int_{\nu + \eta^{-1}(x_{2i} + x_{3i})} \mathbb{1}\{\eta^{-1} r_i - \theta_0 - (\nu_i + \eta^{-1}(x_{2i} + x_{3i})) < 0\} f(\nu_i + \eta^{-1}(x_{2i} + x_{3i})|r_i; \theta_2) \Big\}. \tag{D.11}$$

The conditional densities $f(\nu_i|r_i; \theta_2)$ and $f(\nu_i + \eta^{-1}(x_{2i} + x_{3i})|r_i; \theta_2)$ differ in at least two dimensions. First, while ν_i is independent of r_i, $\nu_i + \eta^{-1}(x_{2i} + x_{3i})$ 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, $\mathbb{E}[r_i|\mathcal{J}_i^a] = r_i$:

$$\begin{aligned} cov(\xi_i, \mathbb{E}[r_i|\mathcal{J}_i^a]) &= cov(\xi_i, r_i) \\ &= cov(x_{2i} + x_{3i}, x_{1i} + x_{2i} + x_{3i}) = (var(x_{2i}) + var(x_{3i})). \end{aligned} \tag{D.12}$$

Therefore, the aggregate error term, $\chi_i = \nu_i + \eta^{-1}\xi_i$, is also correlated with $\mathbb{E}[r_i|\mathcal{J}_i^a]$. Second, the variance of

[54] 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 $\sigma = 1$.

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 $f(\nu_i|r_i;\theta_2)$ will also differ from that of $f(\nu_i + \eta^{-1}(x_{2i} + x_{3i})|r_i;\theta_2)$.

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 expectations $\mathbb{E}[r_i|\mathcal{J}_i]$. 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 expectations term, r_i. 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	$log\text{-}normal(0,1)$	$log\text{-}normal(0,1)$	1.8737 (0.0005)	3.4698 (0.0014)
9	$-log\text{-}normal(0,1)$	$-log\text{-}normal(0,1)$	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 $(\eta^{-1}, \beta_0) = (0.5, 0.5)$ 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 model in which both the true exporters' expectation, $\mathbb{E}[r_i|\mathcal{J}_i]$, and the exporters' expectational error, ξ_i, are normally distributed. The results show that the researcher's ML estimates converge to values of the unknown parameter vector, (θ_0, θ_2), that are larger than the true value of the parameter vector $(\beta_0, \sigma) = (0.5, 0.5)$ 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 departures from the setting 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 the student t distribution that has fatter tails than the normal.

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. Specifically, in model 8 we use distributions that are positively skewed, and in model 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 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,

$$\mathbb{E}[r_i|\mathcal{J}_i] = x_{1i} \qquad \text{and} \qquad \mathbb{E}[r_i|\mathcal{J}_i^a] = x_{1i} + x_{2i}. \tag{D.14}$$

and, from equation (14), the measurement error introduced by the misspecification of agents' expectations is thus

$$\xi_i = \mathbb{E}[r_i|\mathcal{J}_i^a] - \mathbb{E}[r_i|\mathcal{J}_i] = (x_{1i} + x_{2i}) - x_{1i} = x_{2i}. \tag{D.15}$$

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

$$L_a(\theta|d, \mathcal{J}^a) = \sum_{i=1}^{N} \Big\{ d_i \Phi(\theta_2^{-1}(\eta^{-1}(x_{1i} + x_{2i}) - \theta_0)) + (1 - d_i)(1 - \Phi(\theta_2^{-1}(\eta^{-1}(x_{1i} + x_{2i}) - \theta_0))) \Big\}. \tag{D.16}$$

However, the correct log-likelihood function is:

$$L(\theta|d, \mathcal{J}) =$$
$$\sum_{i=1}^{N} \Big\{ d_{ijt} \int_{\nu + \eta^{-1} x_{2i}} \mathbb{1}\{\eta^{-1}(x_{1i} + x_{2i}) - \theta_0 - (\nu_i + \eta^{-1} x_{2i}) \geq 0\} f(\nu_i + \eta^{-1} x_{2i} | x_{1i} + x_{2i}; \theta_2)$$
$$+ (1 - d_{ijt}) \int_{\nu + \eta^{-1} x_{2i}} \mathbb{1}\{\eta^{-1}(x_{1i} + x_{2i}) - \theta_0 - (\nu_i + \eta^{-1} x_{2i}) < 0\} f(\nu_i + \eta^{-1} x_{2i} | x_{1i} + x_{2i}; \theta_2) \Big\}. \tag{D.17}$$

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. Here, both $cov(\chi_i, \mathbb{E}[r_i|\mathcal{J}_i^a])$ 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 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 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 $\mathbb{E}[r_i|\mathcal{J}_i]$. 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, $\mathbb{E}[r_i|\mathcal{J}_i^a]$.

Comparing results in Tables D.1 and D.2, the biases have the same sign but are smaller in absolute value. 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 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. Specifically, the researcher assumes that only x_{1i} is in the exporters'

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	$\textit{log-normal}(0,1)$	$\textit{log-normal}(0,1)$	1.8189 (0.0005)	3.3602 (0.0014)
9	$-\textit{log-normal}(0,1)$	$-\textit{log-normal}(0,1)$	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 $(\eta^{-1}, \beta_0) = (0.5, 0.5)$ and those reported in columns 4 and 5 show the asymptotic bias generated of the corresponding ML estimates.

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

$$\mathbb{E}[r_i|\mathcal{J}_i] = x_{1i} + x_{2i} \qquad \text{and} \qquad \mathbb{E}[r_i|\mathcal{J}_i^a] = x_{1i}, \tag{D.19}$$

and, therefore, the measurement error introduced by the misspecification of agents' expectations is

$$\xi_i = \mathbb{E}[r_i|\mathcal{J}_i^a] - \mathbb{E}[r_i|\mathcal{J}_i] = x_{1i} - (x_{1i} + x_{2i}) = -x_{2i}. \tag{D.20}$$

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

$$L_a(\theta|d, \mathcal{J}^a) = \sum_{i=1}^{N} \Big\{ d_i \Phi(\theta_2^{-1}(\eta^{-1}x_{1i} - \theta_0)) + (1-d_i)(1 - \Phi(\theta_2^{-1}(\eta^{-1}x_{1i} - \theta_0))) \Big\}. \tag{D.21}$$

However, the correct log-likelihood function is:

$$L(\theta|d,\mathcal{J}) =$$

$$\sum_{i=1}^{N}\Big\{d_{ijt}\int_{\nu-\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}x_{1i}-\theta_0-(\nu_i-\eta^{-1}x_{2i})\geq 0\}f(\nu_i-\eta^{-1}x_{2i}|x_{1i};\theta_2)$$

$$+(1-d_{ijt})\int_{\nu-\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}x_{1i}-\theta_0-(\nu_i-\eta^{-1}x_{2i})< 0\}f(\nu_i-\eta^{-1}x_{2i}|x_{1i};\theta_2)\Big\} =$$

$$\sum_{i=1}^{N}\Big\{d_{ijt}\int_{\nu-\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}x_{1i}-\theta_0-(\nu_i-\eta^{-1}x_{2i})\geq 0\}f(\nu_i-\eta^{-1}x_{2i};\theta_2)$$

$$+(1-d_{ijt})\int_{\nu-\eta^{-1}x_{2i}} \mathbb{1}\{\eta^{-1}x_{1i}-\theta_0-(\nu_i-\eta^{-1}x_{2i})< 0\}f(\nu_i-\eta^{-1}x_{2i};\theta_2)\Big\}, \tag{D.22}$$

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 different 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, $\mathbb{E}[r_i|\mathcal{J}_i^a]$. Specifically,

$$cov(\xi_i, \mathbb{E}[r_i|\mathcal{J}_i^a]) = cov(\xi_i, x_{1i}) = cov(-x_{2i}, x_{1i}) = 0. \tag{D.23}$$

This represents a very special case in which ξ_i is also fully independent of $\mathbb{E}[r_i|\mathcal{J}_i^a]$. In this case, both χ_i and ν_i would be independent of $\mathcal{J}_i^a$, meaning 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 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 particular special case, if ξ_i is only mean independent of $\mathbb{E}[r_i|\mathcal{J}_i^a]$ or if the distribution of ξ_i is such that 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.

The results reported in rows 1 to 7 of Tables D.2 and D.3 are very different: in our simple simulation, the bias is much larger when the information set assumed by the researcher is too large than when it is too small. The crucial difference is that when the assumed information set is too large, the measurement error in firms' expectations is more likely correlated with the assumed proxy for these expectations, 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. We may find the measure of the agents' true expectations in this case to be exogenous and, therefore, the bias in the parameter estimates becomes less severe.

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	$\textit{log-normal}(0,1)$	$\textit{log-normal}(0,1)$	0.6092 (0.0011)	1.2370 (0.0014)
9	$-\textit{log-normal}(0,1)$	$-\textit{log-normal}(0,1)$	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 $(\eta^{-1}, \beta_0) = (0.5, 0.5)$ and those reported in columns 4 and 5 show the asymptotic bias of the corresponding ML estimates.

E Sunk Costs of Exporting and Forward-Looking Firms

E.1 Introduction to Dynamic Model

We show here 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 model introduced in Section 8.1. Under this extension of the benchmark model in Section 2, we now allow firms to take into account how the decision to export to destination j at time t, d_{ijt}, 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.

For the discussion presented 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 are akin to '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 $t' > t$,

$$\begin{aligned} V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) &= \eta_j^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - f_{ijt} - (1 - d_{ijt-1})s_{ijt} \\ &+ \rho\mathbb{E}\big[d(1_t)_{ijt+1}\big(\eta_j^{-1}r_{ijt+1} - f_{ijt+1}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1\big] \\ &+ \sum_{t'=t+2} \rho^{t'-t}\mathbb{E}\big[d(1_t)_{ijt'}\big(\eta_j^{-1}r_{ijt'} - f_{ijt'} - (1 - d(1_t)_{ijt'-1})s_{ijt'}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1\big]; \end{aligned} \tag{E.1}$$

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

$$\begin{aligned} V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) &= \rho\mathbb{E}\big[d(0_t)_{ijt+1}\big(\eta_j^{-1}r_{ijt+1} - f_{ijt+1} - s_{ijt+1}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0\big] \\ &+ \sum_{t'=t+2} \rho^{t'-t}\mathbb{E}\big[d(0_t)_{ijt'}\big(\eta_j^{-1}r_{ijt'} - f_{ijt'} - (1 - d(0_t)_{ijt'-1})s_{ijt'}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0\big]. \end{aligned} \tag{E.2}$$

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

$$\begin{aligned} W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, a_{ijt} = 1) &= \eta_j^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - f_{ijt} - (1 - d_{ijt-1})s_{ijt} \\ &+ \rho\mathbb{E}\big[d(0_t)_{ijt+1}\big(\eta_j^{-1}r_{ijt+1} - f_{ijt+1}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0\big] \\ &+ \sum_{t'=t+2} \rho^{t'-t}\mathbb{E}\big[d(0_t)_{ijt'}\big(\eta_j^{-1}r_{ijt'} - f_{ijt'} - (1 - d(0_t)_{ijt'-1})s_{ijt'}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0\big]; \end{aligned} \tag{E.3}$$

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

$$\begin{aligned} W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, a_{ijt} = 0) &= \rho\mathbb{E}\big[d(1_t)_{ijt+1}\big(\eta_j^{-1}r_{ijt+1} - f_{ijt+1} - s_{ijt+1}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1\big] \\ &+ \sum_{t'=t+2} \rho^{t'-t}\mathbb{E}\big[d(1_t)_{ijt'}\big(\eta_j^{-1}r_{ijt'} - f_{ijt'} - (1 - d(1_t)_{ijt'-1})s_{ijt'}\big)\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1\big]. \end{aligned} \tag{E.4}$$

By definition, if the independence condition in equation (27) holds, then

$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) \geq 0 \tag{E.5a}$$
$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0. \tag{E.5b}$$

Given the functional form assumptions in equations (4) and (26), 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,

$$\mathbb{E}[d(0_t)_{ijt'}\eta_j^{-1}r_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0] = \eta_j^{-1}\mathbb{E}[d(0_t)_{ijt'}r_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{E.6a}$$
$$\mathbb{E}[d(0_t)_{ijt'}f_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0] = (\beta_0 + \beta_1 dist_j)\mathbb{E}[d(0_t)_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{E.6b}$$
$$\mathbb{E}[d(0_t)_{ijt'}s_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0] = (\gamma_0 + \gamma_1 dist_j)\mathbb{E}[d(0_t)_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{E.6c}$$

and, analogously,

$$\mathbb{E}[d(1_t)_{ijt'}\eta_j^{-1}r_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1] = \eta_j^{-1}\mathbb{E}[d(1_t)_{ijt'}r_{ijt'}|\mathcal{J}_{ijt}, a_{ijt} = 1], \tag{E.7a}$$
$$\mathbb{E}[d(1_t)_{ijt'}f_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1] = (\beta_0 + \beta_1 dist_j)\mathbb{E}[d(1_t)_{ijt'}|\mathcal{J}_{ijt}, dist_j], \tag{E.7b}$$
$$\mathbb{E}[d(1_t)_{ijt'}s_{ijt'}|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1] = (\gamma_0 + \gamma_1 dist_j)\mathbb{E}[d(1_t)_{ijt'}|\mathcal{J}_{ijt}, dist_j]. \tag{E.7c}$$

Equations (E.6) and (E.7) will allow us to re-express equations (E.1), (E.2), (E.3), (E.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

$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}) \equiv \max\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1), V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0)\}, \tag{E.8}$$

We can then rewrite equation (28) as

$$d_{ijt} = \mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\}. \tag{E.9}$$

In Section E.2, we show how to combine equations (E.1) to (E.7) to derive odds-based moment inequalities in the dynamic context. In Section E.3, we derive revealed-preference inequalities consistent with the dynamic model.

E.2 Odds-Based Moment Inequalities for a Dynamic Model

The definition of the random variable d_{ijt} in equation (E.9) implies that, for every firm i, country j and period t, we can write the following two inequalities

$$d_{ijt} - \mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\} \geq 0, \tag{E.10a}$$
$$\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\} - d_{ijt} \geq 0. \tag{E.10b}$$

Equation (E.10a) exploits the fact that

$$\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\}, \tag{E.11}$$

is a sufficient condition for $d_{ijt} = 1$. Equation (E.10b) exploits the fact that the inequality inside the indicator function in equation (E.35) is a necessary condition for $d_{ijt} = 1$. We will derive an odds-based moment inequality from each of the inequalities in equation (E.10). We show first how to derive an inequality from equation (E.10a) and do the same for equation (E.10b) below.

Combining equations (E.5b) and (E.10a), we obtain:

$$d_{ijt} - \mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\} \geq 0, \tag{E.12}$$

and, from equations (4), (26), (E.1), (E.4), and (E.7c), we can write the variable inside the indicator function

as:

$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=0) =$$
$$\eta_j^{-1}\mathbb{E}[r_{ijt}\big|\mathcal{J}_{ijt}] - f_{ijt} - (1-d_{ijt-1})s_{ijt} + \rho\mathbb{E}\big[d(1_t)_{ijt+1}s_{ijt+1}\big|\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt}=1\big] =$$
$$\eta_j^{-1}\mathbb{E}[r_{ijt}\big|\mathcal{J}_{ijt}] - (\beta_0+\beta_1 dist_j+\nu_{ijt}) - (1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j) + (\gamma_0+\gamma_1 dist_j)\rho\mathbb{E}\big[d(1_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big] =$$
$$\eta_j^{-1}\mathbb{E}[r_{ijt}\big|\mathcal{J}_{ijt}] - (\beta_0+\beta_1 dist_j+\nu_{ijt}) - (1-d_{ijt-1}-\rho\mathbb{E}\big[d(1_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big])(\gamma_0+\gamma_1 dist_j). \tag{E.13}$$

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,

$$\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) \equiv$$
$$\eta_j^{-1}\mathbb{E}[r_{ijt}\big|\mathcal{J}_{ijt}] - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho\mathbb{E}\big[d(1_t)_{ijt+1}\big|\mathcal{J}_{ijt}, dist_j\big])(\gamma_0+\gamma_1 dist_j). \tag{E.14}$$

Using this expression, we can rewrite (E.12) as

$$d_{ijt} - \mathbb{1}\{\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt} \geq 0\} \geq 0, \tag{E.15}$$

or, equivalently,[55]

$$\mathbb{1}\{\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt} \leq 0\} - (1-d_{ijt}) \geq 0. \tag{E.16}$$

Given that this inequality must hold for every firm i, country j and period t, it must also hold in expectation

$$\mathbb{E}[\mathbb{1}\{\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt} \leq 0\} - (1-d_{ijt}) \geq 0]|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \tag{E.17}$$

and, given the distributional assumption in equation (5) and the assumption that ν_{ijt} is independent over time, we can rewrite this expression as

$$\mathbb{E}(1-\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)) - (1-d_{ijt})|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{E.18}$$

Following analogous steps as those described in the proof to Lemma 10, we can rewrite this inequality as

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))} - (1-d_{ijt})\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right] \geq 0. \tag{E.19}$$

This inequality is analogous to that in equation (10) 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 (E.19) depends on the unobserved expectations

$$\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] \qquad \text{and} \qquad \mathbb{E}[d(1_t)_{ijt+1}|\mathcal{J}_{ijt}, dist_j]$$

and, therefore, the researcher would not know it even if she knew the true parameter vector θ_D^*. We define an analogous expression $\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$ that depends only on covariates that the researcher observes *ex post*:

$$\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \equiv$$
$$\eta_j^{-1}r_{ijt} - (\beta_0+\beta_1 dist_j) - (1-d_{ijt-1}-\rho d(1_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j). \tag{E.20}$$

By definition,

$$\mathbb{E}[\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) - \Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] = 0. \tag{E.21}$$

Therefore, exploiting the convexity of the function $(1-\Phi(\cdot))/\Phi(\cdot)$, we can apply a reasoning similar to that in

[55] See the proof to Lemma 10 for a step-by-step derivation of an equation analogous to (E.28) from an equation analogous to (E.51).

Lemma 7 and conclude that:

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt},dist_j,d_{ijt-1};\theta_D^*))}{\Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt},dist_j,d_{ijt-1};\theta_D^*))}-(1-d_{ijt})\middle|\mathcal{J}_{ijt},dist_j,d_{ijt-1}\right]\leq$$
$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt},d(1_t)_{ijt+1},dist_j,d_{ijt-1};\theta_D^*))}{\Phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt},d(1_t)_{ijt+1},dist_j,d_{ijt-1};\theta_D^*))}-(1-d_{ijt})\middle|\mathcal{J}_{ijt},dist_j,d_{ijt-1}\right]. \tag{E.22}$$

Combining equations (E.19), (E.21) and (E.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:

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1dist_j)-(1-d_{ijt-1}-\rho d(1_t)_{ijt+1})(\gamma_0+\gamma_1dist_j)))}{\Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1dist_j)-(1-d_{ijt-1}-\rho d(1_t)_{ijt+1})(\gamma_0+\gamma_1dist_j)))}-(1-d_{ijt})\middle|Z_{ijt}\right]\geq 0.$$

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

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1dist_j)-(1-d_{ijt-1}-\rho d_{ijt+1})(\gamma_0+\gamma_1dist_j)))}{\Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1dist_j)-(1-d_{ijt-1}-\rho d_{ijt+1})(\gamma_0+\gamma_1dist_j)))}-(1-d_{ijt})\middle|Z_{ijt}\right]\geq 0, \tag{E.23}$$

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 (E.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 (E.10b), we obtain:

$$\mathbb{1}\{V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=1)-V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=0)\geq 0\}-1+(1-d_{ijt})\geq 0,$$
$$(1-d_{ijt})+(\mathbb{1}\{V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=1)-V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=0)\geq 0\}-1)\geq 0,$$
$$(1-d_{ijt})-(\mathbb{1}\{V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=1)-V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=0)\leq 0\})\geq 0,$$
$$(1-d_{ijt})-\mathbb{1}\{V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=0)-V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=1)\geq 0\}\geq 0.$$

Combining the last inequality with equation (E.5a), we obtain

$$(1-d_{ijt})-\mathbb{1}\{V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=0)-W(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=1)\geq 0\}\geq 0, \tag{E.24}$$

and, from equations (4), (26), (E.2), (E.3), and (E.6c), we can write the variable inside the indicator function as:

$$V(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=0)-W(\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=1)=$$
$$-\eta_j^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]+f_{ijt}+(1-d_{ijt-1})s_{ijt}-\rho\mathbb{E}\big[d(0_t)_{ijt+1}s_{ijt+1}\big|\mathcal{J}_{ijt},f_{ijt},s_{ijt},d_{ijt-1},a_{ijt}=0\big]=$$
$$-\eta_j^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]+(\beta_0+\beta_1dist_j+\nu_{ijt})+(1-d_{ijt-1})(\gamma_0+\gamma_1dist_j)-(\gamma_0+\gamma_1dist_j)\rho\mathbb{E}\big[d(0_t)_{ijt+1}\big|\mathcal{J}_{ijt},dist_j\big]=$$
$$-\eta_j^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]+(\beta_0+\beta_1dist_j+\nu_{ijt})+(1-d_{ijt-1}-\rho\mathbb{E}\big[d(0_t)_{ijt+1}\big|\mathcal{J}_{ijt},dist_j\big])(\gamma_0+\gamma_1dist_j). \tag{E.25}$$

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,

$$\Delta_0(\mathcal{J}_{ijt},dist_j,d_{ijt-1};\theta_D^*)\equiv$$
$$-\eta_j^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]+(\beta_0+\beta_1dist_j)+(1-d_{ijt-1}-\rho\mathbb{E}\big[d(0_t)_{ijt+1}\big|\mathcal{J}_{ijt},dist_j\big])(\gamma_0+\gamma_1dist_j). \tag{E.26}$$

Using this expression, we can rewrite (E.24) as

$$(1 - d_{ijt}) - \mathbb{1}\{\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt} \geq 0\} \geq 0, \tag{E.27}$$

or, equivalently,

$$\begin{aligned} -d_{ijt} + 1 - \mathbb{1}\{\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt} \geq 0\} \geq 0, \\ \mathbb{1}\{\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt} \leq 0\} - d_{ijt} \geq 0, \\ \mathbb{1}\{-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt} \geq 0\} - d_{ijt} \geq 0. \end{aligned} \tag{E.28}$$

Following analogous steps to those described in the proof to Lemma 11, we can rewrite this inequality as

$$\mathbb{E}\left[(1 - d_{ijt})\frac{\Phi(\sigma^{-1}(-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}{1 - \Phi(\sigma^{-1}(-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))} - d_{ijt}\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right] \geq 0. \tag{E.29}$$

This inequality is analogous to that in equation (11) 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 (E.29) depends on the unobserved expectations

$$\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] \qquad \text{and} \qquad \mathbb{E}[d(0_t)_{ijt+1}|\mathcal{J}_{ijt}, dist_j]$$

and, therefore, the researcher would not know it even if she knew the true parameter vector θ_D^*. We define an analogous expression $\Delta_1^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$ that depends only on covariates that the researcher observes *ex post*:

$$\begin{aligned} \Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \equiv \\ -\eta_j^{-1} r_{ijt} + (\beta_0 + \beta_1 dist_j) + (1 - d_{ijt-1} - \rho d(0_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j). \end{aligned} \tag{E.30}$$

By definition,

$$\mathbb{E}[\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) - \Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] = 0. \tag{E.31}$$

Therefore, exploiting the convexity of the function $\Phi(\cdot)/(1 - \Phi(\cdot))$, we can apply a reasoning similar to that in Lemma 9 and conclude that:

$$\begin{aligned} \mathbb{E}\left[(1 - d_{ijt})\frac{\Phi(\sigma^{-1}(-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}{1 - \Phi(\sigma^{-1}(-\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right] \leq \\ \mathbb{E}\left[(1 - d_{ijt})\frac{\Phi(\sigma^{-1}(-\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)))}{1 - \Phi(\sigma^{-1}(-\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)))}\middle|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}\right]. \end{aligned} \tag{E.32}$$

Combining equations (E.29), (E.31) and (E.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:

$$\mathbb{E}\left[(1 - d_{ijt})\frac{\Phi(\sigma^{-1}(\eta_j^{-1} r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d(0_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j)))}{1 - \Phi(\sigma^{-1}(\eta_j^{-1} r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d(0_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j)))} - d_{ijt}\middle|Z_{ijt}\right] \geq 0.$$

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

$$\mathbb{E}\left[(1 - d_{ijt})\frac{\Phi(\sigma^{-1}(\eta_j^{-1} r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d_{ijt+1})(\gamma_0 + \gamma_1 dist_j)))}{1 - \Phi(\sigma^{-1}(\eta_j^{-1} r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d_{ijt+1})(\gamma_0 + \gamma_1 dist_j)))} - d_{ijt}\middle|Z_{ijt}\right] \geq 0, \tag{E.33}$$

where d_{ijt+1} takes value 1 if firm i is observed to export to country j in year t.

Equations (E.23) and (E.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 (E.23) and (E.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.4 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).

E.3 Revealed-Preference Moment Inequalities for Dynamic Model

The definition of the random variable d_{ijt} in equation (E.9) implies that, for every firm i, country j and period t, we can write the following two inequalities

$$d_{ijt}(V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0, \tag{E.34a}$$

$$(1 - d_{ijt})(V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) \geq 0. \tag{E.34b}$$

Equation (E.34a) exploits the fact that

$$\mathbb{1}\{V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \geq 0\}, \tag{E.35}$$

is a necessary condition for $d_{ijt} = 1$. Equation (E.34b) exploits the fact that the inequality inside the indicator function in equation (E.35) is a sufficient condition for $d_{ijt} = 1$. We will derive an revealed-preference moment inequality from each of the inequalities in equation (E.34). We show first how to derive an inequality from equation (E.34a) and do the same for equation (E.34b) below.

Combining equations (E.5b) and (E.34a), we obtain:

$$d_{ijt}(V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0)) \geq 0. \tag{E.36}$$

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 (E.14). Using this expression, we can rewrite (E.36) as

$$d_{ijt}(\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt}) \geq 0. \tag{E.37}$$

Given that this inequality must hold for every firm i, country j and period t, it must also hold in expectation

$$\mathbb{E}[d_{ijt}(\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt})|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \tag{E.38}$$

or, equivalently,

$$\mathbb{E}[d_{ijt}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] - \mathbb{E}[d_{ijt}\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{E.39}$$

Focusing on the second term, we note that

$$\begin{aligned}
\mathbb{E}[d_{ijt}\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] &= -\mathbb{E}[(1 - d_{ijt})\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \\
&= -\mathbb{E}[\mathbb{E}[(1 - d_{ijt})\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 0]|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \\
&= -\mathbb{E}[(1 - d_{ijt})\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 0]|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}].
\end{aligned} \tag{E.40}$$

Focusing further on the conditional expectation $\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 0]$, we note that

$$\begin{gathered}
\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 0] = \\
\mathbb{E}[\nu_{ijt}|V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) \leq 0, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}];
\end{gathered}$$

from equations (E.1) and (E.2), we can rewrite

$$\begin{gathered}
V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) = \\
\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt},
\end{gathered} \tag{E.41}$$

and, therefore,

$$\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 0] =$$
$$\mathbb{E}[\nu_{ijt}|\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) - \nu_{ijt} \leq 0, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] =$$
$$\mathbb{E}[\nu_{ijt}|\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) \leq \nu_{ijt}, \mathcal{J}_{ijt}, dist_j, d_{ijt-1}].$$

Given the distributional assumption in equation (5) and the assumption that ν_{ijt} is independent over time, we can rewrite this expression as

$$\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 0] = \sigma \frac{\phi(\sigma^{-1}\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}. \tag{E.42}$$

Combining equations (E.39), (E.40) and (E.42), we obtain the following moment inequality

$$\mathbb{E}[d_{ijt}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\tilde{\Delta}_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{E.43}$$

As equations (E.1) and (E.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 (E.5b), (E.13), and (E.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 (E.43), one can derive the weaker inequality

$$\mathbb{E}[d_{ijt}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_1(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{E.44}$$

Exploiting the mean independence restriction in equation (E.21) and the convexity of the function $\phi(\cdot)/(1 - \Phi(\cdot))$, we can apply a reasoning similar to that in lemmas 14 and 15 and conclude that:

$$\begin{aligned}\mathbb{E}[&d_{ijt}\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \\ &+ (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0.\end{aligned} \tag{E.45}$$

where the term $\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$ is defined in equation (E.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

$$(1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}, \tag{E.46}$$

as a function of data and the parameter vector θ_D^*. However, as equation (E.20) shows, the function

$$\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$$

is increasing in the value of the dummy variable $d(1_t)_{ijt+1}$. As equation (E.45) is also increasing in the term $\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$, 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:

$$\begin{aligned}\mathbb{E}[&d_{ijt}\Delta_1^{obs}(r_{ijt}, d(1_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \\ &+ (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}, 1, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_1^{obs}(r_{ijt}, 1, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0.\end{aligned} \tag{E.47}$$

From equation (E.20), we therefore obtain the following inequality

$$\mathbb{E}[d_{ijt}(\eta_j^{-1}r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d(1_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j)) \\ + (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}(\eta_j^{-1}r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho)(\gamma_0 + \gamma_1 dist_j)))}{1 - \Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho)(\gamma_0 + \gamma_1 dist_j)))} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \tag{E.48}$$

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

$$\mathbb{E}[d_{ijt}(\eta_j^{-1}r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho d(1_t)_{ijt+1})(\gamma_0 + \gamma_1 dist_j)) \\ + (1 - d_{ijt})\sigma \frac{\phi(\sigma^{-1}(\eta_j^{-1}r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho)(\gamma_0 + \gamma_1 dist_j)))}{1 - \Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt} - (\beta_0 + \beta_1 dist_j) - (1 - d_{ijt-1} - \rho)(\gamma_0 + \gamma_1 dist_j)))} | Z_{ijt}] \geq 0, \tag{E.49}$$

This is the first of the revealed-preference inequalities we will use for identification of the parameter θ_D^*.

Combining equations (E.5a) and (E.34b), we obtain:

$$(1 - d_{ijt})(V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) - W(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) \geq 0. \tag{E.50}$$

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 (E.26). Using this expression, we can rewrite (E.50) as

$$(1 - d_{ijt})(\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt}) \geq 0. \tag{E.51}$$

Given that this inequality must hold for every firm i, country j and period t, it must also hold in expectation

$$\mathbb{E}[(1 - d_{ijt})(\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt}) | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \tag{E.52}$$

or, equivalently,

$$\mathbb{E}[(1 - d_{ijt})\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] + \mathbb{E}[(1 - d_{ijt})\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{E.53}$$

Focusing on the second term, we note that

$$\begin{aligned} \mathbb{E}[(1 - d_{ijt})\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] &= -\mathbb{E}[d_{ijt}\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \\ &= -\mathbb{E}[\mathbb{E}[d_{ijt}\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1] | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \\ &= -\mathbb{E}[d_{ijt}\mathbb{E}[\nu_{ijt} | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1] | \mathcal{J}_{ijt}, dist_j, d_{ijt-1}]. \end{aligned} \tag{E.54}$$

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 (E.1) and (E.2), we can rewrite

$$V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 0) - V(\mathcal{J}_{ijt}, f_{ijt}, s_{ijt}, d_{ijt-1}, a_{ijt} = 1) = \\ \tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + \nu_{ijt}, \tag{E.55}$$

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 (5) and the assumption that ν_{ijt} is independent over time, we can rewrite this expression as

$$\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}, d_{ijt} = 1] = -\sigma \frac{\phi(\sigma^{-1}(-\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}{\Phi(\sigma^{-1}(-\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*)))}. \tag{E.56}$$

Combining equations (E.53), (E.54) and (E.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,

$$\mathbb{E}[(1 - d_{ijt})\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + d_{ijt}\sigma \frac{\phi(\sigma^{-1}\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\tilde{\Delta}_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{E.57}$$

As equations (E.1) and (E.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 (E.5a), (E.26), and (E.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 (E.57), one can derive the weaker inequality

$$\mathbb{E}[(1 - d_{ijt})\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*) + d_{ijt}\sigma \frac{\phi(\sigma^{-1}\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_0(\mathcal{J}_{ijt}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0. \tag{E.58}$$

Exploiting the mean independence restriction in equation (E.31) and the convexity of the function $\phi(\cdot)/(1 - \Phi(\cdot))$, we can apply a reasoning similar to that in lemmas 14 and 15 and conclude that:

$$\begin{aligned}\mathbb{E}[&(1 - d_{ijt})\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \\ &+ d_{ijt}\sigma \frac{\phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0,\end{aligned} \tag{E.59}$$

where the term $\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$ is defined in equation (E.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

$$d_{ijt}\sigma \frac{\phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*))}. \tag{E.60}$$

However, as equation (E.30) shows, the function

$$\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$$

is decreasing in the value of the dummy variable $d(0_t)_{ijt+1}$. As equation (E.59) is increasing in the term $\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*)$, 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:

$$\begin{aligned}\mathbb{E}[&(1 - d_{ijt})\Delta_0^{obs}(r_{ijt}, d(0_t)_{ijt+1}, dist_j, d_{ijt-1}; \theta_D^*) \\ &+ d_{ijt}\sigma \frac{\phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}, 0, dist_j, d_{ijt-1}; \theta_D^*))}{1 - \Phi(\sigma^{-1}\Delta_0^{obs}(r_{ijt}, 0, dist_j, d_{ijt-1}; \theta_D^*))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0,\end{aligned} \tag{E.61}$$

From equation (E.30), we therefore obtain the following inequality

$$\mathbb{E}[(1-d_{ijt})(-\eta_j^{-1}r_{ijt}+(\beta_0+\beta_1 dist_j)+(1-d_{ijt-1}-\rho d(0_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j)) \\ + d_{ijt}\sigma\frac{\phi(\sigma^{-1}(-\eta_j^{-1}r_{ijt}+(\beta_0+\beta_1 dist_j)+(1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)))}{1-\Phi(\sigma^{-1}(-\eta_j^{-1}r_{ijt}+(\beta_0+\beta_1 dist_j)+(1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \quad \text{(E.62)}$$

or, equivalently,

$$\mathbb{E}[-(1-d_{ijt})(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1}-\rho d(0_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j)) \\ + d_{ijt}\sigma\frac{\phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)))}{\Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)))}|\mathcal{J}_{ijt}, dist_j, d_{ijt-1}] \geq 0, \quad \text{(E.63)}$$

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

$$\mathbb{E}[-(1-d_{ijt})(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1}-\rho d(0_t)_{ijt+1})(\gamma_0+\gamma_1 dist_j)) \\ + d_{ijt}\sigma\frac{\phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)))}{\Phi(\sigma^{-1}(\eta_j^{-1}r_{ijt}-(\beta_0+\beta_1 dist_j)-(1-d_{ijt-1})(\gamma_0+\gamma_1 dist_j)))}|Z_{ijt}] \geq 0, \quad \text{(E.64)}$$

This is the second of the revealed-preference inequalities we will use for identification of the parameter θ_D^*.

Equations (E.49) and (E.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 (E.49) and (E.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.4 both with the dummy variable $\mathbb{1}\{d_{ijt-1}=0\}$ and with the dummy variable $\mathbb{1}\{d_{ijt-1}=1\}$.

F Firm-Country Export Revenue Shocks: Details

F.1 Unknown to Firms When Deciding on Export Entry

In this section, we show that the estimation procedures we describe in sections 4.1 and 4.2 as well as the results we present in sections 5, 6 and 7 are valid under a generalization of the model described in Section 2. Here, we allow variable trade costs, τ_{ijt}, to vary freely across firms for a single country-year pair jt. To add firm heterogeneity to our baseline model, we need to impose one restriction on the variability of τ_{ijt} across firms in a single country and year:

$$\mathbb{E}_{jt}[\tau_{ijt}^{1-\eta}|\mathcal{J}_{ijt}, c_{it}] = \tau_{jt}^{1-\eta}, \tag{F.1}$$

where $\mathbb{E}_{jt}[\cdot]$ denotes the expectation across firms conditional on a single country-year pair. In words, equation (F.1) implies that firms cannot predict the firm-specific component of variable trade costs in j and t when they decide whether to export to that destination. The firm-specific component must also be mean independent of firms' variable production costs, c_{it}.[56]

The model introduced in Section 2 has two key estimating equations. The first is the expression for the probability of exporting conditional on the exporters' information set $(\mathcal{J}_{ijt}, dist_j)$, reported in equation (8). This equation is used in combination with additional assumptions on the content of exporters' information sets to justify both the maximum likelihood and the moment inequality estimators introduced in sections 4.1 and 4.2. The second key estimating equation that our model generates is the expression that allows us to consistently estimate the export revenue parameters $\{\alpha_{jt}; \forall j, t\}$ using information on observed export revenues and domestic sales; i.e. equations (10) and (11). We show here that our two estimating equations remain unchanged when we drop the assumption that $\tau_{ijt} = \tau_{jt}$ for every i and impose instead the weaker assumption in equation (F.1).

Export probability equation. Given equations (2) and (8), the benchmark model described in Section 2 implies that we can write the probability that firm i exports to j at t conditional on $(\mathcal{J}_{ijt}, dist_j)$ as:

$$\mathcal{P}_{jt}(d_{ijt} = 1|\mathcal{J}_{ijt}, dist_j) = \Phi\big(\sigma^{-1}\big(\eta^{-1}\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big). \tag{F.2}$$

This equation will also hold exactly in the model with firm-varying variable trade costs if the firm heterogeneity in trade costs verifies equation (F.1).

Specifically, under assumption (F.1), we can write potential export revenues r_{ijt} as in equation (29), where α_{jt} is exactly as indicated in equation (2), and

$$\omega_{ijt} = (\tau_{ijt}^{1-\eta} - \tau_{jt}^{1-\eta})\left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}} r_{iht}. \tag{F.3}$$

Therefore, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ will be equal to $\mathbb{E}[\alpha_{jt} r_{iht}|\mathcal{J}_{ijt}]$ and equation (F.2) will therefore hold in the model with firm-varying trade costs if

$$\mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}] = 0. \tag{F.4}$$

[56] For example, $\tau_{ijt}^{1-\eta} = \tau_{jt}^{1-\eta} e_{1,ijt}^{\tau} + e_{2,ijt}^{\tau}$ would satisfy this restriction, where both $\mathbb{E}[e_{1,ijt}^{\tau}|\mathcal{J}_{ijt}, c_{it}] = 1$ and $\mathbb{E}[e_{2,ijt}^{\tau}|\mathcal{J}_{ijt}, c_{it}] = 0$.

As the following derivation shows, equation (F.1) is sufficient for equation (F.4) to hold:

$$\begin{aligned}
\mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}] &= \mathbb{E}[\mathbb{E}_{jt}[\omega_{ijt}|\mathcal{J}_{ijt}, c_{it}]|\mathcal{J}_{ijt}] \\
&= \mathbb{E}\Big[\mathbb{E}_{jt}\Big[(\tau_{ijt}^{1-\eta} - \tau_{jt}^{1-\eta})\Big(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\Big)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht}\Big|\mathcal{J}_{ijt}, c_{it}\Big]\Big|\mathcal{J}_{ijt}\Big] \\
&= \mathbb{E}\Big[\mathbb{E}_{jt}\Big[(\tau_{ijt}^{1-\eta} - \tau_{jt}^{1-\eta})r_{iht}\Big|\mathcal{J}_{ijt}, c_{it}\Big]\Big(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\Big)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}\Big|\mathcal{J}_{ijt}\Big] \\
&= \mathbb{E}\Big[\mathbb{E}_{jt}\Big[(\tau_{ijt}^{1-\eta} - \tau_{jt}^{1-\eta})\Big[\frac{\eta}{\eta-1}\frac{\tau_{ht}c_{it}}{P_{ht}}\Big]^{1-\eta}Y_{ht}\Big|\mathcal{J}_{ijt}, c_{it}\Big]\Big(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\Big)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}\Big|\mathcal{J}_{ijt}\Big] \\
&= \mathbb{E}\Big[\mathbb{E}_{jt}\Big[(\tau_{ijt}^{1-\eta} - \tau_{jt}^{1-\eta})\Big|\mathcal{J}_{ijt}, c_{it}\Big]c_{it}^{1-\eta}\Big[\frac{\eta}{\eta-1}\frac{\tau_{ht}}{P_{ht}}\Big]^{1-\eta}Y_{ht}\Big(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\Big)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}\Big|\mathcal{J}_{ijt}\Big] \\
&= \mathbb{E}\Big[0 \times c_{it}^{1-\eta}\Big[\frac{\eta}{\eta-1}\frac{\tau_{ht}}{P_{ht}}\Big]^{1-\eta}Y_{ht}\Big(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\Big)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}\Big|\mathcal{J}_{ijt}\Big] = 0,
\end{aligned}$$

where the first equality uses the Law of Iterated Expectations; the second equality replaces ω_{ijt} with its expression in equation (F.3); the third equality takes into account that price indices and market sizes only vary at the country-year level; the fourth equality replaces r_{ijt} with its expression in equation (1) for the specific case of $j = h$; the fifth equality takes into account that some of the determinants of r_{iht} only vary at the year level; and the sixth equality applies assumption (F.1).

Therefore, equation (F.2) is consistent with a model in which variable trade costs vary across firms within country-year pairs if equation (F.1) holds.

Export revenue equation. Given equation (2) and allowing for measurement error e_{ijt} in observed export revenues, the benchmark model described in Section 2 implies that we can use the moment condition in equation (11) to consistently estimate the parameter vector $\{\alpha_{jt}; \forall j \text{ and } t\}$. As shown in the following lines, the moment condition in equation (11) also identifies the parameter vector $\{\alpha_{jt}; \forall j \text{ and } t\}$ in the model with firm-varying variable trade costs if the firm heterogeneity in trade costs verifies equation (F.1).

Specifically, under assumption (F.1), we can write potential export revenues r_{ijt} as in equation (29), where α_{jt} is exactly as indicated in equation (2), and ω_{ijt} is as indicated in equation (F.3). Therefore, allowing again for measurement error e_{ijt} in observed export revenues as in equation (10), we can write observed export revenues as

$$r_{ijt}^{obs} = d_{ijt}(\alpha_{jt}r_{iht} + \omega_{ijt} + e_{ijt}). \tag{F.5}$$

Given that $\mathbb{E}_{jt}[e_{ijt}|r_{iht}, d_{ijt} = 1] = 0$ by assumption imposed on the properties of the measurement error variable e_{ijt}, the moment condition in equation (11) is consistent with equation (F.5) if $\mathbb{E}[\omega_{ijt}|r_{iht}, d_{ijt} = 1] = 0$. The following derivation shows that this mean independence condition on ω_{ijt} is a direct consequence of the

mean independence condition in equation (F.1):

$$
\begin{aligned}
\mathbb{E}_{jt}[\omega_{ijt}|r_{iht}, d_{ijt}=1] &= \mathbb{E}_{jt}[(\tau_{ijt}^{1-\eta}-\tau_{jt}^{1-\eta})\left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht}|r_{iht}, d_{ijt}=1] \\
&= \mathbb{E}_{jt}[(\tau_{ijt}^{1-\eta}-\tau_{jt}^{1-\eta})|r_{iht}, d_{ijt}=1] \times \left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= \mathbb{E}_{jt}[\mathbb{E}_{jt}[(\tau_{ijt}^{1-\eta}-\tau_{jt}^{1-\eta})|\mathcal{J}_{ijt}, c_{it}, \nu_{ijt}, r_{iht}, d_{ijt}=1]|r_{iht}, d_{ijt}=1] \times \left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= \mathbb{E}_{jt}[\mathbb{E}_{jt}[(\tau_{ijt}^{1-\eta}-\tau_{jt}^{1-\eta})|\mathcal{J}_{ijt}, c_{it}, \nu_{ijt}, d_{ijt}=1]|r_{iht}, d_{ijt}=1] \times \left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= \mathbb{E}_{jt}[\mathbb{E}_{jt}[(\tau_{ijt}^{1-\eta}-\tau_{jt}^{1-\eta})|\mathcal{J}_{ijt}, c_{it}, \nu_{ijt}]|r_{iht}, d_{ijt}=1] \times \left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= \mathbb{E}_{jt}[\mathbb{E}_{jt}[(\tau_{ijt}^{1-\eta}-\tau_{jt}^{1-\eta})|\mathcal{J}_{ijt}, c_{it}]|r_{iht}, d_{ijt}=1] \times \left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= \mathbb{E}_{jt}[0|r_{iht}, d_{ijt}=1] \times \left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} \\
&= 0 \times \left(\frac{1}{\tau_{ht}}\frac{P_{ht}}{P_{jt}}\right)^{1-\eta}\frac{Y_{jt}}{Y_{ht}}r_{iht} = 0,
\end{aligned}
$$

where the first equality replaces ω_{ijt} with its expression in equation (F.3); the second equality takes out of the expectation those covariates whose distribution either on r_{iht} or on a set of country-year fixed effects are degenerate; the third equality applies the Law of Iterated Expectations; the fourth equality takes into account that r_{iht} is a function of c_{it} and country-year covariates and, therefore, redundant in the conditioning set; the fifth equality takes into account that d_{ijt} is a function of $\mathcal{J}_{ijt}$, country-year covariates (specifically, $dist_j$) and ν_{ijt} and, therefore, also redundant in the conditioning set; the sixth equality, with $\mathcal{J}_{ijt}$ defined as the set of variables the firm uses to predict r_{ijt}, accounts for the fact that ν_{ijt} must be either included in $\mathcal{J}_{ijt}$ or irrelevant to predict τ_{ijt}; the seventh equality applies equation (F.1); and all remaining equalities follow trivially from the previous expression.

Therefore, equation (11) is consistent with a model in which variable trade costs vary across firms within country-year pairs if equation (F.1) holds.

F.2 Known to Firms When Deciding on Export Entry

Assume a setting characterized by the the following three equations

$$
r_{ijt}^{obs} = d_{ijt}(\alpha_{jt}r_{iht} + \omega_{ijt} + e_{ijt}), \qquad \mathbb{E}[e_{ijt}|d_{ijt}=1, \mathcal{J}_{ijt}, dist_j] = 0, \text{ and } \mathbb{E}[\omega_{ijt}|\mathcal{J}_{ijt}, dist_j] = \omega_{ijt}, \tag{F.6}
$$

$$
\begin{pmatrix} \omega_{ijt} \\ \nu_{ijt} \end{pmatrix} \Bigg| (\mathcal{J}_{ijt}, dist_j) \sim \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\omega^2 & \sigma_{\nu\omega} \\ \sigma_{\omega\nu} & \sigma^2 \end{pmatrix}\right), \tag{F.7}
$$

$$
d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\}. \tag{F.8}
$$

Relative to the model in Section 2, this model adds a firm-country-year specific revenue shock ω_{ijt} that we assume the firm knows when deciding on export destinations (see equation (F.6)) and that is jointly normally distributed with the firm-country-year fixed costs shock, ν_{ijt}. Combining equations (F.6) and (F.8), we can rewrite the export participation dummy d_{ijt} as in equation (33). Taking into account the both ν_{ijt} and ω_{ijt} are assumed to be jointly normally distributed, it holds that the unobserved (to the researcher) term in equation (33) will also be normally distributed:

$$
\eta^{-1}\omega_{ijt} - \nu_{ijt}|(\mathcal{J}_{ijt}, dist_j) \sim \mathbb{N}(0, \eta^{-2}\sigma_\omega^2 + \sigma^2 - 2\eta^{-1}\sigma_{\omega\nu}). \tag{F.9}
$$

For simplicity in the notation, we henceforth use

$$
\tilde{\sigma}^2 \equiv \eta^{-2}\sigma_\omega^2 + \sigma^2 - 2\eta^{-1}\sigma_{\omega\nu}. \tag{F.10}
$$

Therefore, from equations (F.6), (F.8), and (F.9), we can conclude that the probability that firm i exports to market j at period t conditional on the vector $(\mathcal{J}_{ijt}, dist_j)$ becomes

$$\mathcal{P}_{jt}(d_{ijt} = 1|\mathcal{J}_{ijt}, dist_j) = \Phi\big(\tilde{\sigma}^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big). \tag{F.11}$$

This equation has the same functional form as the analogous probability in our benchmark model; see equation (F.11). The equations differ in how we interpret structurally the variance of the probit shock. Here, it is equal to the variance of a weighted sum of fixed costs shocks and revenue shocks known to the the firm when deciding on export destinations, $\tilde{\sigma}^2$.

The key difference between our benchmark model and this extension appears in the estimating equation used to identify the parameter vector $\{\alpha_{jt}; \forall j \text{ and } t\}$. While these export revenue parameters were point-identified in our benchmark model, they will be only partially identified in the presence of the known export revenue shocks ω_{ijt}. Given that our sample period covers 10 years and 22 countries, this implies that we would need to estimate 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. Therefore, we simplify the problem by assuming $\alpha_{jt} = \alpha_0 + \alpha_1 R_{jt}$ and estimate the parameter vector $\theta_S \equiv (\alpha_0, \alpha_1, \beta_0, \beta_1, \sigma_\xi, \sigma_{\xi\nu}, \sigma_\nu)$. Given this parametric restriction on α_{jt}, equation (F.11) becomes

$$\mathcal{P}_{jt}(d_{ijt} = 1|\mathcal{J}_{ijt}, dist_j) = \Phi\big(\tilde{\sigma}^{-1}\big(\eta^{-1}\mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big). \tag{F.12}$$

The additional moment inequality that arises from equation (F.6) can be derived as follows:

$$\begin{aligned}
\mathbb{E}[r^{obs}_{ijt}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] &= \mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] + \mathbb{E}[\omega_{ijt}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] \\
&+ \mathbb{E}[e_{ijt}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] \\
&= \mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] + \mathbb{E}[\omega_{ijt}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] \\
&= \mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] \\
&+ \mathbb{E}[\omega_{ijt}|\eta^{-1}\mathbb{E}[(\alpha_0 + \alpha_1 R_{jt}) r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + \eta^{-1}\omega_{ijt} - \nu_{ijt} \geq 0, \mathcal{J}_{ijt}, dist_j] \\
&= \mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] \\
&+ \mathbb{E}[\omega_{ijt}|\eta^{-1}\omega_{ijt} - \nu_{ijt} \geq -\eta^{-1}\mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|\mathcal{J}_{ijt}] + \beta_0 + \beta_1 dist_j, \mathcal{J}_{ijt}, dist_j]
\end{aligned}$$

Imposing the distributional assumption in equation (F.7), we can further rewrite this expression as

$$\begin{aligned}
\mathbb{E}_{jt}[r^{obs}_{ijt}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] &= \mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] \\
&+ \frac{cov(\omega_{ijt}, \eta^{-1}\omega_{ijt} - \nu_{ijt})}{\sqrt{var(\eta^{-1}\omega_{ijt} - \nu_{ijt})}} \frac{\phi(\tilde{\sigma}^{-1}(-\eta^{-1}\mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|\mathcal{J}_{ijt}] + \beta_0 + \beta_1 dist_j))}{1 - \Phi(\tilde{\sigma}^{-1}(-\eta^{-1}\mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|\mathcal{J}_{ijt}] + \beta_0 + \beta_1 dist_j))} \\
&= \mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] \\
&+ \frac{\tilde{\sigma}_{\omega\nu}}{\tilde{\sigma}} \frac{\phi(\tilde{\sigma}^{-1}(\eta^{-1}\mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}{\Phi(\tilde{\sigma}^{-1}(\eta^{-1}\mathbb{E}[\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}.
\end{aligned}$$

where $\tilde{\sigma}_{\omega\nu} = cov(\omega_{ijt}, \eta^{-1}\omega_{ijt} - \nu_{ijt}) = \eta^{-1}\sigma^2_\omega + \sigma_{\nu\omega}$, and, therefore one can obtain the following moment condition

$$\begin{aligned}
&\mathbb{E}[r^{obs}_{ijt} - (\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \\
&\frac{\tilde{\sigma}_{\omega\nu}}{\tilde{\sigma}} \frac{\phi(\tilde{\sigma}^{-1}(\eta^{-1}\mathbb{E}[(\alpha_0 + \alpha_1 R_{jt}) r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}{\Phi(\tilde{\sigma}^{-1}(\eta^{-1}\mathbb{E}[(\alpha_0 + \alpha_1 R_{jt}) r_{iht}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}|d_{ijt} = 1, \mathcal{J}_{ijt}, dist_j] = 0.
\end{aligned} \tag{F.13}$$

Equations (F.12) and (F.13) are the key equations to identify the parameter vector of the model discussed in Section 8.2. Given the normalization that $\eta = 5$, the parameters that appear either in equation (F.12) or in equation (F.13) are: $(\alpha_0, \alpha_1, \beta_0, \beta_1, \tilde{\sigma}, \tilde{\sigma}_{\nu\omega})$. However, in order to be able to exploit equations (F.12) and (F.13) for identification, the researcher must still impose some assumption on the content of the true information set of firms $\mathcal{J}_{ijt}$. As in the main model described in Section 2, the researcher can opt between three alternatives.

First, if the researcher is willing to assume that firms have perfect foresight, then the researcher can use

the following two equations for identification:

$$\mathbb{E}[r_{ijt}^{obs} - (\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \frac{\tilde{\sigma}_{\omega\nu}}{\tilde{\sigma}} \frac{\phi(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j))}{\Phi(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j))} | d_{ijt} = 1, R_{jt}, r_{iht}, dist_j] = 0, \tag{F.14a}$$

$$\mathcal{P}_{jt}(d_{ijt} = 1 | r_{iht}, R_{jt}, dist_j) = \Phi\big(\tilde{\sigma}^{-1}\big(\eta^{-1}(\alpha_0 r_{iht} + \alpha_1 R_{jt} r_{iht}) - \beta_0 - \beta_1 dist_j\big)\big). \tag{F.14b}$$

These two equations transform the model in Section 8.2 into a sample selection model à la Heckman (1979). The parameter vector can then be estimated either the two-step procedure introduced in Heckman (1979) or a maximum likelihood estimator.

Second, if the researcher is willing to assume that the information set of exporters, $\mathcal{J}_{ijt}$, is identical to a vector of observed covariates, $\mathcal{J}_{ijt}^{obs}$, then the researcher can use the following two equations for identification:

$$\mathbb{E}[r_{ijt}^{obs} - (\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \frac{\tilde{\sigma}_{\omega\nu}}{\tilde{\sigma}} \frac{\phi(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0 \mathbb{E}[r_{iht}|\mathcal{J}_{ijt}^{obs}] + \alpha_1 \mathbb{E}[R_{jt} r_{iht}|\mathcal{J}_{ijt}^{obs}]) - \beta_0 - \beta_1 dist_j))}{\Phi(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0 \mathbb{E}[r_{iht}|\mathcal{J}_{ijt}^{obs}] + \alpha_1 \mathbb{E}[R_{jt} r_{iht}|\mathcal{J}_{ijt}^{obs}]) - \beta_0 - \beta_1 dist_j))} | d_{ijt} = 1, \mathcal{J}_{ijt}^{obs}, dist_j] = 0, \tag{F.15a}$$

$$\mathcal{P}_{jt}(d_{ijt} = 1 | \mathcal{J}_{ijt}^{obs}, dist_j) = \Phi\big(\tilde{\sigma}^{-1}\big(\eta^{-1}(\alpha_0 \mathbb{E}[r_{iht}|\mathcal{J}_{ijt}^{obs}] + \alpha_1 \mathbb{E}[R_{jt} r_{iht}|\mathcal{J}_{ijt}^{obs}]) - \beta_0 - \beta_1 dist_j\big)\big). \tag{F.15b}$$

To estimate this model, the researcher must first project r_{iht} and $R_{jt} r_{iht}$ on the vector $\mathcal{J}_{ijt}^{obs}$, in order to recover a consistent estimate of $\mathbb{E}[r_{iht}|\mathcal{J}_{ijt}^{obs}]$ and $\mathbb{E}[R_{jt} r_{iht}|\mathcal{J}_{ijt}^{obs}]$, respectively. Once these expectations are known, these two equations again transform the model in Section 8.2 into a model à la Heckman (1979).

Third, if the researcher is only willing to assume that some observed vector Z_{ijt} has a distribution conditional on $(\mathcal{J}_{ijt}, dist_j)$ that is degenerate, then she cannot directly use equations (F.12) and (F.13). Instead of equation (F.12), the researcher can use odds-based and revealed-preference inequalities analogous to those introduced in Section 4.2. Specifically, following the same steps as in the proof in Appendix C, one can derive the following two odds-based inequalities

$$\mathcal{M}^{ob}(Z_{ijt}; \beta_0, \beta_1, \alpha_0, \alpha_1, \tilde{\sigma}_1) = \tag{F.16}$$
$$\mathbb{E}\left[\begin{array}{c} d_{ijt} \frac{1-\Phi\left(\tilde{\sigma}^{-1}\left(\eta^{-1}(\alpha_0+\alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j\right)\right)}{\Phi\left(\tilde{\sigma}^{-1}\left(\eta^{-1}(\alpha_0+\alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j\right)\right)} - (1 - d_{ijt}) \\ (1 - d_{ijt}) \frac{\Phi\left(\tilde{\sigma}^{-1}\left(\eta^{-1}(\alpha_0+\alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j\right)\right)}{1-\Phi\left(\tilde{\sigma}^{-1}\left(\eta^{-1}(\alpha_0+\alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j\right)\right)} - d_{ijt} \end{array} \middle| Z_{ijt} \right] \geq 0,$$

and the following revealed-preference moment inequalities

$$\mathcal{M}^{r}(Z_{ijt}; \beta_0, \beta_1, \alpha_0, \alpha_1, \tilde{\sigma}_1) = \tag{F.17}$$
$$\mathbb{E}\left[\begin{array}{c} -(1 - d_{ijt})\big(\eta^{-1}(\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j\big) + d_{ijt} \tilde{\sigma} \frac{\phi\left(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0+\alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j)\right)}{\Phi\left(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0+\alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j)\right)} \\ d_{ijt}\big(\eta^{-1}(\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j\big) + (1 - d_{ijt}) \tilde{\sigma} \frac{\phi\left(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0+\alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j)\right)}{1-\Phi\left(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0+\alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j)\right)} \end{array} \middle| Z_{ijt} \right] \geq 0.$$

Similarly, the researcher that is willing to assume only that a vector Z_{ijt} is included in firms' informations sets $\mathcal{J}_{ijt}$ cannot use the moment equality in equation (F.13) directly for identification, as it depends on the unobserved expectation $\mathbb{E}[(\alpha_0 + \alpha_1 R_{jt}) r_{iht}|\mathcal{J}_{ijt}]$. However, given that

$$\mathbb{E}[(\alpha_0 + \alpha_1 R_{jt}) r_{iht} - (\alpha_0 + \alpha_1 R_{jt}) r_{iht}|\mathcal{J}_{ijt}] = 0,$$

and $\phi(\cdot)/\Phi(\cdot)$ is convex, we can use Jensen's inequality to derive the following inequality

$$\mathbb{E}[r_{ijt}^{obs} - (\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \frac{\eta^{-1}\sigma_\omega^2 + \sigma_{\nu\omega}}{\tilde{\sigma}} \frac{\phi(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j))}{\Phi(\tilde{\sigma}^{-1}(\eta^{-1}(\alpha_0 + \alpha_1 R_{jt}) r_{iht} - \beta_0 - \beta_1 dist_j))} | d_{ijt} = 1, Z_{ijt}] \leq 0. \tag{F.18}$$

Equations (F.17) to (F.18) may be used to identify the parameter vector of interest.