What do Exporters Know?*

Michael J. Dickstein

Stanford University and NBER

Eduardo Morales

Princeton University and NBER

May 17, 2015

Abstract

The decision of firms to participate in export markets drives much of the variation in the volume of trade. To understand these decisions and predict how export flows react to changes in trade policy, we estimate a model of firms' export participation. Crucially, in choosing to export, firms must forecast the likely profits from serving a foreign market, to trade off against the fixed costs of exporting. We show that the estimated parameters and predicted exports from the model depend heavily on how the researcher specifies firms' expectations over these profits. In response, we develop a novel moment inequality approach that allows us to recover fixed costs and predict exports placing weaker assumptions on firms' expectations. Our approach introduces a new set of moment inequalities, odds-based inequalities, and generalizes the revealed-preference inequalities introduced in Pakes (2010) to a new setting. We use data from Chilean exporters to show that, relative to methods that must specify firms' information sets, our approach generates estimates of fixed export costs that are approximately 65-85% smaller. We also predict gains in export participation from counterfactual reductions in fixed costs that are 30% smaller on average than those predicted by existing approaches, depending upon the destination market and industrial sector.

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

*We thank Tim Bresnahan, Lorenzo Caliendo, Jan De Loecker, Dave Donaldson, Liran Einav, Guido Imbens, Ariel Pakes, Esteban Rossi-Hansberg, James Tybout and seminar participants at Dartmouth College, NBER ITI meeting, Pennsylvania State University, Princeton University, the Stanford/Berkeley IO Fest, UCLA, University of Virginia, and Wharton School for helpful suggestions. All errors are our own. Email: mjd@stanford.edu, ecmorale@princeton.edu.

1 Introduction

In 2012, approximately 300,000 US firms chose to export to foreign markets. The decision of these firms to sell abroad drives the volume of trade from the US—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. Thus, to predict how aggregate exports may change with trade policy, exchange rate movements, or other policy or market fluctuations, researchers need to quantify the margin on which individual firms will respond.

A large literature in international trade focuses on the role of the exporting firm in aggregate export flows.[1] Empirical analyses of firms' export participation 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. The exact assumptions the researcher imposes on agents' information often matter, 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 fixed costs incurred when exporting and predictions of how firms will respond to counterfactual changes in trade costs.

In this paper, we first document the sensitivity of the standard estimates of the parameters driving firms' export decisions to assumptions on how firms form expectations. We compare the predictions of a standard model in the international trade literature (Melitz, 2003; Helpman et al., 2008) under two different set of assumptions on how exporters form their expectations: the "perfect foresight" case, under which firms perfectly predict their profits when exporting, and a limited information specification in which firms only use a specific observed set of variables recorded either on firms' own balance sheets or on governments' official statistics to predict their own export revenue were the firm to export. Under each assumption on firms' information, we recover values for the fixed costs of exporting and predict changes in exports across markets in reaction to a policy that reduces fixed costs of exporting by 40%. Finding important differences in the predictions from the two models, we then develop a new empirical model of export participation that places few restrictions on firms' expectations.

Under our new approach, firms may gather different signals about their productivity relative to competitors, about the evolution of exchange rates, trade policy, or political stability abroad, or about their foreign demand; 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 the subset included in the specification, but the remaining variables that affect the firm's expectations

[1] See for example Das et al. (2007), Arkolakis (2010), Moxnes (2010), Eaton et al. (2011), Ruhl and Willis (2014), Arkolakis et al. (2014a), and Cherkashin et al. (2015). A recent literature also focuses on the decisions of importers; e.g. Antràs et al. (2014).

need not be observed and 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 new types of moments inequalities, which we label odds-based and generalized revealed preference inequalities.[2] Using these inequalities, our empirical burden is twofold. We must show that placing fewer assumptions on expectations affects both the estimates of the parameters of the exporter's problem and the predictions of export flows under counterfactual trade policy, and that our robust approach generates bounds on these 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.[3]. We estimate this model using data on Chilean exporters in two industrial sectors, the manufacture of chemicals and food products.

We proceed in three steps. First, we demonstrate the sensitivity of both the estimated fixed costs of exporting and the predictions of firms' export participation to assumptions the researcher imposes on how firms forecast revenues. Specifically, under the assumption that firms' expectations are rational and using maximum likelihood methods, we estimate a perfect foresight model under which firms predict perfectly the revenues they will earn upon entry. Under this assumption, for example, we find export costs in the chemicals sector from Chile to Argentina, Japan, and United States to equal \$894,000, \$2.8 million, and \$1.7 million, respectively. We compare these estimates to an alternative model, from Willis and Rosen (1979), Manski (1991) and Ahn and Manski (1993), in which, in a first stage, we specify the exact set of variables firms use to form their unobserved expectations. The estimated fixed costs of exporting under this two-step approach are approximately 20-30% lower than those found under the perfect foresight assumption, in both the chemicals and food sector.

That the fixed cost estimates differ under perfect foresight and the two-step approach reflects a possible bias in the estimation. Both the two-step procedure and the perfect foresight approach 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—the expected revenue term will contain error. Thus, our next step is to employ our new types of moments inequalities to partially identify the exporter's fixed costs under weaker assumptions. Here, we assume that firms know the distance to the export destination, the aggregate exports to that market in the prior year, and their own domestic sales from the prior year. Importantly,

[2] 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), Wollman (2014), and Morales et al. (2015). This work generally follows the methodology developed in Pakes (2010) and Pakes et al. (2015); our generalized revealed preference inequalities extend this methodology to a new setting with a distinct error structure.

[3] By combining the insights in this paper with the Euler's perturbation method introduced in Morales et al. (2015), we could similarly perform our analysis in the context of a fully dynamic export participation problem à la Das et al. (2007).

the inequalities we define do not restrict firms to use only these three variables, but require that the firm know at least these variables. We can use the specification test suggested in Andrews and Soares (2010) to test the null hypothesis that the content of potential exporters' information sets satisfies this minimal requirement. That is, conditional on the model, we can test our specification that presumes these three variables are in the firm's information set.

Under the traditional maximum likelihood methods, the estimated fixed costs for exports from Chile to Argentina, for example, ranged from \$594,000 to \$894,000 under traditional methods in the chemicals sector. Using our inequalities approach, we find a much lower range of fixed costs, between approximately \$270,000 and \$298,000 in the chemicals sector. This range is small enough to be informative for policy. In addition, in model specification tests using data from both the chemicals and food sectors, we cannot reject the null hypothesis that exporters know distance, lagged domestic sales, and lagged aggregate exports when making their export decisions. To address further the question of "what do exporters know?", we repeat this test under the same model, but place one additional variable in the firm's information set. In this alternative, we assume the firm also knows the productivity of other firms that exported in the prior year to each destination country. Repeating the test, we now reject the null that firms knew this information when making their export decision at the 4% level in the chemicals sector and the 1% level in the food sector.

Finally, as a third key step, we conduct counterfactuals using our inequalities, imposing the same minimal requirements on firms' information sets as we imposed in estimation. Our counterfactual predictions are also set-identified. We provide bounds that indicate how firms would respond to a counterfactual policy that reduces the fixed costs of exporting by 40%. Starting with the approach that requires explicit assumptions on the firm's information set, we find that the results differ substantially with these assumptions. For example, relative to predictions under perfect foresight, the predicted export participation under the two step procedure in the chemicals sector is 1% and 11% lower for Argentina and Japan and 13% higher for the United States. That is, we would predict the relative effect of the policy on exports to the United States to be different depending on our specification. Comparing the predictions from these two models to those computed using the moment inequalities, in the latter we predict gains in export participation from counterfactual reductions in fixed costs that are 30% smaller on average, depending upon the destination market and industrial sector.

We illustrate our contribution using the exporter's problem. Our approach, however, provides a robust methodology to estimate the parameters of many decisions in economics that depend on agents' forecasts of key variables. For example, 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., 2014b). 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). On the consumer side, Greenstone et al. (2014) examine the enlistment of soldiers

in the US Army; the decision to reenlist depends on the soldiers' expectations about the riskiness of the task assigned. Similarly, a retiree's decision to purchase a private annuity (Ameriks et al., 2014) depends on her expectations about life expectancy. 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, as in Arcidiacano et al. (2014), 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. In Sections 3 and 4 we describe our data, empirical setting, and three alternative empirical models. We first outline the maximum likelihood procedures that require full knowledge by the researcher of agents' information sets. We then introduce our moment inequality estimator and discuss how to build these inequalities as well as conduct counterfactuals with possibly set-identified parameters. In Section 5, we compare the parameter estimates resulting from the alternative empirical models. In Section 6, we use our inequality approach to predict the effect on export participation and export volume from changes to the economic environment. Section 7 concludes.[4]

2 Export Model

We begin with a model of a firm's 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 $i = 1, \dots, N_t$.[5] We index the potential destination countries by $j = 1, \dots, J$.

We model firms' export decisions using a two-period model. In the first period, firms choose the set of countries they wish to enter. To enter a market, firms must pay a fixed export cost. When choosing to export, firms may differ in their degree of uncertainty about the profits they will obtain upon exporting. In the second period, conditional on entering a foreign market, all firms observe supply and demand conditions and set their prices optimally.

2.1 Demand

Every firm i faces an isoelastic demand in country j in year t:

$$x_{ijt} = \frac{p_{ijt}^{-\eta} Y_{jt}}{P_{jt}^{1-\eta}}, \tag{1}$$

[4] Proofs and additional details are contained in the Appendix.

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

where p is the price firm i sets in destination country j at time t, Y is the total expenditure in country j at time t in the sector in which firm i operates, and P is the ideal price index:

$$P_{jt} = \left[\int_{i \in A_{jt}} p_{ijt}^{1-\eta} di \right]^{\frac{1}{1-\eta}},$$

where A_{jt} denotes the set of all firms in the world selling in j. This specification implies that every firm faces a constant demand elasticity equal to η in every destination country.

2.2 Supply

Firm i produces one unit of output with a cost-minimizing combination of inputs that costs $a_{it}c_t$, where c represents the cost of this bundle in country h and a_{it} is the number of bundles of inputs that firm i uses to produce one unit of output. The inverse of a_{it} denotes firm i's productivity level in t. A cumulative distribution function $G_t(a)$ describes the distribution of a across firms located in h in year t. This distribution function may vary freely across time periods. We also allow firms' productivity to be correlated over time.

When i wants to sell in a foreign market j, it must pay production costs and two additional costs: a transport cost, τ_{jt}, and a fixed cost, f_{ijt}. We adopt the "iceberg" specification of transport costs and assume that firm i must ship τ_{jt} units of a product from country h for one unit to arrive to j. The fixed export costs are

$$f_{ijt} = \beta_0 + \beta_1 dist_j + \nu_{ijt}, \tag{2}$$

where $dist_j$ denotes the distance from country h to country j, and ν_{ijt} is an aggregate of all remaining determinants of f_{ijt} that the researcher does not observe.[6] We assume that

$$\nu_{ijt} \sim \mathbb{N}(0, \sigma_\nu^2), \tag{3}$$

where σ_ν^2 measures the heterogeneity in fixed export costs across firms, countries and time periods.[7]

[6] We assume that the fixed export costs, f_{ijt}, are independent of the previous export experience of firm i in country j. However, given that the time process of the term a_{it} is unrestricted, our model can match any observed persistence in export status. One can also allow fixed exports costs to depend on previous export experience and apply the moment inequalities introduced in Section 4.2 to the corresponding dynamic export problem.

[7] None of the results presented in this paper depend on the assumption that ν is normally distributed. The only restriction necessary for the moment inequalities introduced in Section 4.2 to be valid is that the distribution of ν is known to the researcher up to a scale parameter and log-concave.

2.3 Profits conditional on exporting

Conditional on entering a destination market j, every seller behaves as a monopolistically competitive firm. When setting the optimal price in each destination market in which a firm enters, firms know their demand function, transport costs and own marginal production costs. Therefore, the demand and supply assumptions above imply that the optimal price firm i sets in j is

$$p_{ijt} = \frac{\eta}{\eta - 1} \tau_{jt} a_{it} c_t. \tag{4}$$

As a result, the total revenue that i will obtain in country j is:

$$r_{ijt} = \left[\frac{\eta}{\eta - 1} \frac{\tau_{jt} a_{it} c_t}{P_{jt}} \right]^{1-\eta} Y_{jt} \tag{5}$$

and the export profit (gross of fixed costs) is $\eta^{-1} r_{ijt}$. Therefore, export profits conditional on entry are a function of (a) market size in the destination market, Y_{jt}; (b) competition by other suppliers, as captured by the price index, P_{ijt}; (c) production costs, c_{it}; (d) exporters' productivity, a_{it}; and, (e) transport costs, τ_{jt}. These variables are rarely observed in standard datasets. However, Appendix A.1 shows that, given the assumptions in Sections 2.1 and 2.2, we can rewrite the potential export revenue of i in j, r_{ijt}, as a function of variables that are typically observed in standard trade datasets: (a) the domestic revenues of every active firm i, r_{iht}; (b) the aggregate export flows from the home country h to any destination country j, R_{jt}; (c) an indicator for whether each of the active firms exports to j at t, d_{ijt}. Specifically, the revenue that firm i would obtain in country j at period t conditional on entering is

$$r_{ijt} = \frac{R_{jt}}{\sum_{s=1}^{N_t} d_{sjt}(r_{sht}/r_{iht}) ds}, \tag{6}$$

where N_t denotes the set of active firms in country h at period t. This expression allows us to obtain a measure of the revenues that each firm i would obtain in each destination j in period t conditional on entry. We can compute this measure both for firms that we observe exporting to j in t and for those who choose not to export.

2.4 Decision to export

Once we account for the fixed costs of exporting, the export profits that i will obtain in j are

$$\pi_{ijt} = \eta^{-1} r_{ijt} - f_{ijt}. \tag{7}$$

Firm i will decide to export to j if and only if $\mathbb{E}[\pi_{ijt}|\mathcal{J}_{ijt}] \geq 0$, where the vector $\mathcal{J}_{ijt}$ contains firm i's information about any variable affecting its potential profits from exporting to j in t, π_{ijt}, at the time it decides whether to export to j in year t.

Let $d_{ijt} = \mathbb{1}\{\mathbb{E}[\pi_{ijt}|\mathcal{J}_{ijt}] \geq 0\}$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function. Assuming that all determinants of fixed export costs are known to the firms at the time of deciding on export entry—i.e. $(dist_j, \nu_{ijt}) \in \mathcal{J}_{ijt}$—we can rewrite d_{ijt} as

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - f_{ijt} \geq 0\}, \tag{8}$$

where r_{ijt} is export revenue conditional on entry, f_{ijt} is the fixed export cost, and $\mathbb{E}[\cdot]$ denotes the expectation with respect to the data generating process (i.e. firms' expectations are rational).[8] Therefore, defining an agent's expectational error as ε_{ijt}, $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, it holds that

$$\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}] = 0. \tag{9}$$

Assuming that firms' expectations are rational implies that their expectational error in predicting export revenues is mean independent of any variable used to form this prediction.

Among all the variables contained in firms' information sets, only a subset of them will generally be used to predict export revenues conditional on entry. We denote this subset as $\mathcal{W}_{ijt}$. Therefore,

$$\mathcal{W}_{ijt} \subset \mathcal{J}_{ijt} \qquad \text{and} \qquad \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] = \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]. \tag{10}$$

For example, $\mathcal{W}_{ijt}$ will include any variable that firms might use to forecast either the demand level in j at t, Y_{jt}, or their own productivity a_{it}. Specifically, if, for example, firm i knows the exact demand level it will face in a country j, then $Y_{jt} \in \mathcal{W}_{ijt}$.[9]

A key assumption needed for the estimation approach introduced in Section 4.2 is that ν_{ijt} is independent of all other determinants of the export choice d_{ijt}:

$$\nu_{ijt} \perp (\mathcal{W}_{ijt}, dist_j). \tag{11}$$

From now on, for simplicity of notation and without loss of generality, we will assume that

[8] Equation (8) assumes that all firms know the demand elasticity η when deciding whether to export to j at t. This assumption is not crucial. The moment inequalities introduced in Section 4.2 are also valid in the case in which firms have imperfect information about the demand parameter η. Firms may also have different expectations about this common parameter. The key restriction is that all firms must face the same elasticity of demand in every country and time period.

[9] For now, we impose no restriction on the content of $\mathcal{W}_{ijt}$. Different estimation approaches will require imposing different restrictions on $\mathcal{W}_{ijt}$. In Section 4, we discuss the assumptions that different estimation approaches impose on the content $\mathcal{W}_{ijt}$.

$dist_j \in \mathcal{W}_{ijt}$. Given equations (3), (10), and (11), we can write the probability that i exports to j conditional on $(\mathcal{J}_{ijt}, dist_j)$ as

$$\mathcal{P}_{ijt} = \mathcal{P}(d_{ijt} = 1 | \mathcal{W}_{ijt}) = \Phi\big(\sigma^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big), \tag{12}$$

where $\mathcal{P}_{ijt} = \int_\nu \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu \geq 0\}\phi(\nu)d\nu$, and $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the standard normal probability density function and cumulative distribution function.

As it is clear from equation (12), even if we were to observe firms' actual expectations $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, data on export choices alone does not allow us to identify the scale of the parameter vector $(\sigma, \eta, \beta_0, \beta_1)$—that is, if we multiply these four parameters by the same positive number, the probability $\mathcal{P}_{ijt}$ remains constant. To normalize by scale the parameter vector in export entry models, researchers typically use additional data to estimate or calibrate the demand elasticity η. In our estimation, we set $\eta = 5$.[10] For simplicity of notation, from now on, we use θ to denote the remaining parameter vector $(\sigma, \beta_0, \beta_1)$. The choice of η will affect the magnitude of the parameter vector θ and, therefore, the value estimated for the fixed export costs in each destination country. However, in this paper we emphasize how sensitive these estimates are to different assumptions on the content of firms' information sets $\mathcal{W}_{ijt}$; the choice of η does not affect the ratio of the fixed export costs estimated under these different assumptions nor the counterfactuals.

2.5 Effect of change in export fixed costs

We study the effect of a policy that, for the firms located in country h, reduces the systematic part of export fixed costs by 40%. We denote the counterfactual value of β_0 as $\beta_0^1 = 0.6\beta_0$ and the counterfactual value of β_1 as $\beta_1^1 = 0.6\beta_1$. We assume that h is a small country and, therefore, for all possible destination countries j, the price index P_{jt} and the potential export revenues of every firm r_{ijt} are invariant to the change in (β_0, β_1). Therefore, the only variables affected by the policy are the set of export participation dummies, $\{d_{ijt}, i = 1, \ldots, N_t\}$ and, through them, the total exports from h to j, R_{jt}. We show in Section 6 how different assumptions on the content of firms' information sets $\mathcal{W}_{ijt}$ affect the model predictions for the effect that a reduction in export fixed costs has on the growth rate in the number of exporters and total exports. These growth rates are invariant to the choice of the normalizing constant η.

[10] 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.

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 1996 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.[11]

The firms in our dataset operate in one of two sectors: the manufacture of chemicals and food products.[12] These are the two largest Chilean export manufacturing sectors by volume. In Table 1, we report summary statistics, by year and sector, on the share of domestic firms exporting, the mean and median exports per exporting firm, the mean domestic revenues per firm and per exporting firm, and the mean and median number of markets the exporting firms enter. We focus our analysis on countries which saw at least five firms exporting to that destination 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. These are the set of destinations that we will use for estimation purposes in Section 5.

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 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.[13] 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 across years of \$7.7 million per exporter. The median exporter across all years exports approximately \$2.24 million. Relative to the chemicals sector, firms also typically export to a greater number of destination markets. The median exporting firm exports to 5-6 markets in the food sector. In the chemicals sector, the median exporting firm chooses to export to 3-4 countries.

We focus on export revenues for three countries—Argentina, Japan, and the United

[11] 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 around 80 percent of total export flows.

[12] 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.

[13] Every dollar value included in this paper is evaluated in year 2000 USD.

Table 1: Summary Statistics

Year	Share of exporters	Exports per exporter (avg)	Exports per exporter (med)	Home sales per firm (avg)	Home sales per exporter (avg)	Destinations per exporter (avg)
			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:

States—across all years of the data. We later illustrate our findings for these destinations as well. For the three countries, the total volume of exports across all years of the data in the chemicals sector equals \$205 million, \$112 million, and \$475 million and the total number of firms that export at least once during the sample period is 105, 13 and 61, respectively. The mean annual volume per exporter equals \$412,000, \$1.86 million, and \$2.48 million, respectively, for Argentina, Japan, and the United States. In the food sector, the total volume of exports and number of exporting firms is much larger to the United States (\$1,931 million and 122 unique exporters), and Japan (\$2,656 million and 126 unique exporters), but lower to Argentina (\$184 million and 85 unique exporters). The average per firm export volume to these three countries in the food sector equals \$484,000, \$4.09 million, and \$3.25 million, respectively.

Our data set includes both exporters and non-exporters. Furthermore, to minimize the possibility of selection bias in our estimates, 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.[14] Finally, we obtain information on the

[14]From our sample, we exclude only firms that appear in ENIA for less than three years or that appear during two or more discontinuous periods between 1995 and 2005 (i.e. firms that first disappear and later reappear in

distance from Chile to each destination market from CEPII.[15]

4 Empirical Approach

In the model we describe in Section 2, a firm i's export revenue to destination market j at time t, r_{ijt}, is a function of market size Y_{jt}; the degree of competition, as captured by the price index P_{jt}; firm i's productivity a_{it}; production costs at home c_t; transport costs τ_{jt}; and the elasticity of demand η. Firms might not know some or all of these variables when deciding whether to export to j at t. They instead form expectations of potential export revenues using their information set, $\mathcal{W}_{ijt}$. In the theoretical model, we did not impose assumptions on the content of the information set. To estimate the parameter vector θ and perform counterfactuals, researchers need to impose some restrictions on the content of $\mathcal{W}_{ijt}$. As Manski (1993) illustrates, without any restriction on $\mathcal{W}_{ijt}$, , we cannot identify the parameter vector θ.[16]

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 specify a model with perfect foresight. Under perfect foresight, the researcher assumes that all potential exporters know, at the time of their entry decision, the exact revenues they will obtain in each market if they choose to enter; i.e. r_{ijt} is a measurable function of $\mathcal{W}_{ijt}$.[17] Assuming perfect foresight implies assuming that exporters have all the information they need to perfectly predict export revenues upon entry and, therefore, face no uncertainty. We denote the information set imposed under this approach as $\mathcal{W}_{ijt} = Z^1_{ijt} = r_{ijt}$.

For most firms and in most destination countries, the set Z^1_{ijt} is likely to be strictly larger than firms' true information sets. At the time at which potential exporters start to make investments oriented to penetrate foreign markets, they are likely to face some uncertainty concerning the profits they will be able to obtain in such destination markets. In order to account for this uncertainty, second, we specify an empirical model in which all potential exporters forecast their potential export revenues in every foreign market using only information on their own lagged domestic sales, lagged aggregate exports to the destination country j, and distance from the home country to j. We denote this information set with a superscript 2,

the sample).

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

[16] Specifically, Manski (1993) shows that different assumptions on $\mathcal{W}_{ijt}$ might still generate likelihood functions that are identical functions of a given set of covariates, but in which each reduced form parameter has a different structural interpretation. This shows that we cannot always use measures of goodness of fit to discriminate across models that impose different assumptions on exporters' information sets.

[17] 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. (2014b), Antràs et al. (2014). The model described in Section 2 is partial equilibrium. Whether we can extend our flexible treatment of firms' information sets to general equilibrium models is left for future research.

$\mathcal{W}_{ijt} = Z^2_{ijt} = (r_{iht-1}, R_{it-1}, dist_j)$. This information set is likely to be strictly smaller than the actual information set firms possess when deciding to export. Besides, this model assumes no heterogeneity across firms and markets in their information sets. That is, all potential exporters base their entry decision on the same set of covariates.

These two first approaches are similar in that the researcher assumes that exporters' true information sets $\mathcal{W}_{ijt}$ correspond exactly to a vector of covariates Z_{ijt} observed by the researcher. Section 4.1 shows how to estimate the parameter vector θ under these assumptions.

Ideally, we would like to estimate the parameter vector, θ, and perform counterfactuals without imposing the strong assumptions above on firms' information sets. Thus, in our third approach, described in Section 4.2, we propose a moment inequality estimator to handle settings in which the econometrician observes only a *subset* of potential exporters' true information sets. Specifically, we let the observables in the information set equal Z^2_{ijt}, as defined above, but allow other unobserved variables to be in the actual information set of exporters, such that Z^2_{ijt} is a subset of their true information set. Instead of assuming that $\mathcal{W}_{ijt} = Z^2_{ijt}$, as in the second model above, we only assume that $Z^2_{ijt} \in \mathcal{W}_{ijt}$. The variables included in Z^2_{ijt} are either easily found in public documents (e.g. aggregate exports and distance between countries), or are variables that every firm keeps records of (e.g. domestic sales). Furthermore, unlike the two previous approaches that must specify the complete information set, this procedure permits the unobservable elements of firms' information sets to vary by firm and by export market, such that information sets need not be common to all exporters and destinations.

Independently of the assumptions that researchers impose on exporters' information sets, an additional complication is that researchers only observe a direct measure of the potential revenue from exporting r_{ijt} for those firms i, countries j and year t such that $d_{ijt} = 1$. Conversely, no direct measure of r_{ijt} is available for those firms, countries and time periods in which no actual exports happened. However, as equation (6) shows, under the assumptions described in Section 2, we can define a perfect proxy for r_{ijt} using information on variables that are observable independently of whether firm i actually exported to j in year t: (a) the domestic revenues of every active firm i, r_{iht}; (b) the aggregate export flows from the home country h to any destination country j, R_{jt}; (c) an indicator for whether each of the active firms exports to j at t, d_{ijt}. In Sections 4.1 and 4.2, we rely on the expression in equation (6) to generate an observed measure of r_{ijt} for every firm i, country j and year t.

An alternative approach to deal with the fact that r_{ijt} is unobserved for those observations with $d_{ijt} = 0$ is to assume that, for some vector of observable characteristics X_{ijt}, one can write $r_{ijt} = \delta X_{ijt} + e_{ijt} + u_{ijt}$, where δ is a vector of unknown parameters, u_{ijt} is unobserved and measurable in $\mathcal{W}_{ijt}$, and e_{ijt} is also unobserved and verifies $\mathbb{E}[e_{ijt}|\mathcal{W}_{ijt}] = 0$. In this case, in order to consistently estimate γ and avoid relying on equation (6) to construct a perfect proxy for r_{ijt}, one needs to deal with a sample selection problem. As long as we assume

that the information set of exporters $\mathcal{W}_{ijt}$ is equal to a vector Z_{ijt} of observed covariates, one may easily extend the estimation procedure described in Section 4.1 to consistently estimate both γ and θ (see Heckman (1979)). If the researcher only assumes that a vector of observed covariates Z_{ijt} is a subset of exporters true information sets $\mathcal{W}_{ijt}$, then Appendix A.7 shows how to extend the estimation procedure in Section 4.2 to also obtain bounds for γ. This estimation procedure is more general than that described in Section 4.2. However, given that the focus of this paper is on exporters' information sets at the time of deciding on the set of possible export destinations, we have opted to leave for the Appendix the discussion of sample selection issues and focus on the simpler case in which we observe a perfect proxy for r_{ijt}.

4.1 Perfect Knowledge of Exporters' Information Sets

Under the assumption that the econometrician's observed vector of covariates Z_{ijt} equals the firm's information set, $\mathbb{E}[r_{ijt}|Z_{ijt}]$ is a perfect proxy for $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and one can identify θ as the value of the parameter γ that maximizes the log-likelihood function

$$\mathcal{L}(\gamma|d, Z, dist) =$$
$$\mathbb{E}\Big[\sum_{j,t} d_{ijt} \log\big(\mathcal{P}(d_{jt}=1|r_{ijt}, Z_{ijt}, dist_j)\big) + (1-d_{ijt})\log\big(\mathcal{P}(d_{jt}=0|r_{ijt}, Z_{ijt}, dist_j)\big)\Big], \quad (13)$$

where the expectation is over individuals in the population, Z_{ijt} is the assumed information set of firm i at the time it decides whether to export to j at t, and

$$\mathcal{P}(d_{jt}=1|Z_{ijt}, dist_j) = \Phi\big(\gamma_2^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|Z_{ijt}] - \gamma_0 - \gamma_1 dist_j\big)\big) \quad (14)$$

The vector $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ denotes an unknown parameter vector whose true value is $\theta = (\beta_0, \beta_1, \sigma)$.[18]

Given that the researcher rarely observes firms' information sets and that the sets themselves are likely heterogeneous across agents, specifying the correct information set of each agent is notoriously complicated. If the information set specified by the researcher, Z_{ijt}, is such that $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] \neq \mathbb{E}[r_{ijt}|Z_{ijt}]$, then the estimator of θ that relies on the assumption $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] = \mathbb{E}[r_{ijt}|Z_{ijt}]$ will be biased. We denote the difference between the two revenue projections as ξ_{ij}: $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] = \mathbb{E}[r_{ijt}|Z_{ijt}] - \xi_{ij}$. In this case, one can identify θ as the

[18] In order to use the expressions in equations (13) and (14) to estimate the parameter vector γ, one first needs to estimate the function $\mathbb{E}[r_{ijt}|Z_{ijt}]$. See Manski (1991) and Ahn and Manski (1993) for additional details on this two-step estimation approach.

parameter that maximizes the likelihood function in equation (13) but with

$$\mathcal{P}(d_{ijt} = 1 | Z_{ijt}, dist_j) = \int_{\tilde{\nu}} \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|Z_{ijt}] - \gamma_0 - \gamma_1 dist_j - \tilde{\nu}_{ijt} \geq 0\} f(\tilde{\nu}|Z_{ijt}, dist_j) d\tilde{\nu}, \tag{15}$$

where $\tilde{\nu}_{ijt} = \eta^{-1}\xi_{ijt} + \nu_{ijt}$ and $f(\tilde{\nu}|Z, dist)$ denotes the density of $\eta^{-1}\xi + \nu$ conditional on Z and $dist$. When comparing equation (14) to the corresponding equation (15), it is clear that wrongly assuming that $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] = \mathbb{E}[r_{ijt}|Z_{ijt}]$ will generate biased estimates of θ unless $f(\tilde{\nu}|Z, dist)$ is normal with mean zero and variance σ_ν^2. This will only be true when $\xi_{ijt} = 0$ for every firm i, destination j and period t. Therefore, as long as the information set imposed by the researcher Z_{ijt} is such that the expectation of revenue conditional on it differs from the expectation conditional on the true information set, $\mathcal{W}_{ijt}$, the estimates of the parameters governing the fixed export costs, $(\beta_0, \beta_1, \sigma)$, that rely on equation (14) will be biased.

The direction of the bias for each element of θ depends on the shape of the distribution of $\eta^{-1}\xi + \nu$ conditional on Z and $dist$. Specifically, in the specific case in which ξ is normally distributed and independent of both ν, Z and $dist$, the presence of this additional error term will simply bias upwards the estimate of the variance of the composite error term. However, generally, ξ might not be independent of the term $\mathbb{E}[r_{ijt}|Z_{ijt}]$ and, in this case, the bias on the different parameters θ can take many different forms.

In the specific case in which researchers assume perfect foresight (i.e. assume that $r_{ijt} = \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$), we can analytically sign the bias on the estimates of β_0 and β_1. Applying the results in Yatchew and Griliches (1985) to this context, we can conclude that: if firms' true expectations are normally distributed, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] \sim \mathrm{N}(0, \sigma_e^2)$, and the expectational error is also normally distributed, $\xi_{ijt}|(\mathcal{W}_{ijt}, \nu_{ij}) \sim \mathrm{N}(0, \sigma_\xi^2)$; then, there is an upward bias in the estimates of the fixed costs parameters β_0, β_1 and σ_ν. Furthermore, the upward bias increases in the variance of the expectational error σ_ξ^2 relative to the variance of the true unobserved expectations σ_e^2. When either firms' true expectations or the expectational error are not normally distributed, there is no analytic expression for the bias term in the maximum likelihood estimate of θ. However, as the simulations presented in Appendix A.2 show, assuming perfect foresight when firms' expectations are actually imperfect generates an upward bias in the estimates of β_0 and β_1 under many different distributions of firms' true expectations and expectational error. This upward bias in β_0 and β_1 translates into an upward bias in the estimates of fixed export costs.

The tendency for upward bias in the maximum likelihood estimates of β_0 and β_1 caused by wrongly assuming perfect foresight shares the same basis as the well-known attenuation bias seen in linear models, where the bias affects the estimated coefficients on covariates affected by classical measurement error (see page 73 in Wooldridge (2002)). In our setting,

rational expectations implies that firms' expectational errors are mean independent of their true expectation and, therefore, correlated with the ex-post realization of export revenues. 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, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, would generate a downward bias on the coefficient on r_{ijt}. The probit model in equation (14) differs from this linear setting in two dimensions. First, we normalize the scale by setting the coefficient on the covariate measured with error, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, 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_ν, which compose the fixed cost estimate. 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. However, as we show in a simulation exercise in Appendix A.2, for a wide range of possible distributions, the positive bias in the estimates of the fixed costs parameters β_0 and β_1 persists.

Biased estimates of the structural parameter of interest θ will translate into incorrect predictions of the effect of the counterfactual changes in the environment described in Section 2.5. In Section 5, we illustrate the distinct estimates and counterfactual predictions found when assuming alternately that Z^1_{ijt} or Z^2_{ijt} (as defined in Section 4) perfectly describe the information set of potential exporters.

4.2 Partial Knowledge of Exporters' Information Sets

Finding a set of observed covariates that exactly correspond to agents' unknown information sets is, in most empirical applications, difficult. Conversely, it is usually quite simple to define a smaller vector of observed covariates that is contained in such information sets. For example, in each year, exporters will likely know past values of both their domestic sales, r_{iht-1}, and the aggregate exports from their home country to each destination market, R_{jt-1}. The former is a variable reported in all firms' accounting statements, and the latter is always included 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 both fixed and transport costs. Therefore, while it may be unrealistic to assume the agent's information set is exactly identical to the vector of observed covariates Z^2_{ijt} defined above, i.e. $Z^2_{ijt} = (r_{iht-1}, R_{it-1}, dist_j)$, the assumption that Z^2_{ijt} is contained in every potential exporters' information set may well be accurate. In this section, we show that, given a vector of observed covariates Z_{ijt} that is contained in the information set that every firm uses to forecast its gross export profits, i.e. $Z_{ijt} \subset \mathcal{W}_{ijt}$, we can form moment inequalities that partially identify the parameters of the firm's entry decision. In the model described in Section 2, these parameters compose the firm's fixed costs of exporting.

As we show 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, i.e. $Z_{ijt} \subset \mathcal{W}_{ijt}$, is not strong enough to point-identify the parameter vector θ. However, the assumption that researchers partially observe firms' information sets has enough power to identify a set that contains the true value of the parameter, θ. We describe below two new types of moment inequalities that define such a set.[19]

In Section 5, we will show further how one can use specification tests for partially identified models (e.g. Andrews and Soares (2010)) to test the null hypothesis that the model defined in Section 2, combined with different assumptions on the content of exporters' information sets, $\mathcal{W}_{ijt}$, is consistent with the data available to us. We will conclude that we cannot reject the null that $Z^2_{ijt} \in \mathcal{W}_{ijt}$ but we can reject both the null that exporters have perfect foresight, $r_{ijt} \in \mathcal{W}_{ijt}$, and the null that, besides the variables in the vector Z^2_{ijt}, they also have information about the productivity of other exporters.

4.2.1 Odds-based moment inequalities

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

$$\mathcal{M}(Z_{ijt};\gamma) = \mathbb{E}\left[\begin{array}{c} m_l(d_{ijt}, r_{ijt}, dist_j;\gamma) \\ m_u(d_{ijt}, r_{ijt}, dist_j;\gamma) \end{array} \middle| Z_{ijt} \right] \geq 0, \tag{16}$$

where the two moment functions are defined as

$$m_l(\cdot) = d_{ijt}\frac{1-\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt}-\gamma_0-\gamma_1 dist_j\big)\big)}{\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt}-\gamma_0-\gamma_1 dist_j\big)\big)} - (1-d_{ijt}), \tag{17a}$$

$$m_u(\cdot) = (1-d_{ijt})\frac{\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt}-\gamma_0-\gamma_1 dist_j\big)\big)}{1-\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt}-\gamma_0-\gamma_1 dist_j\big)\big)} - d_{ijt}. \tag{17b}$$

We denote set of all possible values of the parameter vector γ as Γ. 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 equations (16), (17a) and (17b):

Theorem 1 *For all $\theta \in \Gamma$, $\mathcal{M}(Z_{ijt};\theta) \geq 0$.*

Theorem 1 indicates that the odds-based inequalities are consistent with the true value of the parameter vector. A formal proof of Theorem 1 is in Appendix A.4. Heuristically, if we had introduced the true expectation $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ in the place of the ex post realization

[19] Whether the bounds defined by the moment inequalities in Sections 4.2.1 and 4.2.2 are sharp is an open question. However, as the results in Section 5 show, in our empirical application, they generate bounds that are small enough to be informative on two dimensions. We can learn both about biases in the parameter estimates that arise when misspecifying the agent's information set and about the effect of counterfactual changes in the economic environment on export participation and trade flows.

r_{ijt} in equations (17a) and (17b), then the moment conditions in equation (16) would hold as equalities at the true value of the parameter vector; i.e. for $\gamma = \theta$. This is clear from equations (17a) and (17b): if one replaces r_{ijt} with $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, the expressions represent the first order condition of the likelihood function conditional on Z_{ijt}. However, this score function is not useful to identify θ because the researcher does not observe the true expectations, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$. If we plug in the observed ex post realized revenue r_{ijt} instead of the unobserved expectations $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ in these rearranged score functions, we would obtain the expression:

$$\mathbb{E}\left[\left.\begin{array}{l} d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j-\eta^{-1}\varepsilon_{ijt}))}{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j-\eta^{-1}\varepsilon_{ijt}))}-(1-d_{ijt}) \\ (1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j-\eta^{-1}\varepsilon_{ijt}))}{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j-\eta^{-1}\varepsilon_{ijt}))}-d_{ijt} \end{array}\right| Z_{ijt}\right], \tag{18}$$

where ε_{ijt} is the expectational error firm i makes when forecasting the potential revenue from exporting to j at t when deciding whether to export. What we show in the proof is that if one were to drop the term involving ε_{ijt} in equation (18), the new expression, equal to equations (17a) and (17a), would be weakly larger than equation (18) and thus would be weakly larger than zero. This result is a direct application of Jensen's inequality. We need only two properties of the errors to generate the odds-based inequalities. First, firms have rational expectations and $Z_{ijt} \subset \mathcal{W}_{ijt}$, so that the expectational error ε_{ijt} has a mean equal to zero conditional on the vector Z_{ijt}. Second, we need $1 - \Phi(\cdot)/\Phi(\cdot)$ and $\Phi(\cdot)/(1 - \Phi(\cdot))$ to be globally convex.[20] This inequality holds at the true value of the parameter vector.

Even though both moment functions in equations (17a) and (17b) are derived from the score function, they are not redundant. In order to gain intuition about the identifying power of each of these moments, we can focus on identification of the parameter γ_0. Given observed values of d_{ijt}, r_{ijt}, and $dist_j$, and given any arbitrary value of the parameters γ_1 and γ_2, the moment function $m_l(\cdot)$ in equation (17a) is increasing in γ_0 and, therefore, will identify a lower bound on γ_0. With the same observed values, $m_u(\cdot)$ in equation (17b) is decreasing in γ_0 and will identify an upper bound on γ_0. Therefore, both moments are necessary to bound γ_0. The same intuition applies for identifying parameters γ_1 and γ_2.

In the particular case in which agents' expectations are perfect (i.e. $\varepsilon_{ijt} = 0$) and the vector of instruments Z_{ijt} includes all variables that agents use to predict either the ex post profits or the fixed export costs, i.e. $Z_{ijt} = \mathcal{W}_{ijt}$, the set Θ is a singleton and identical to the true value of the parameter vector, θ. The size of the set Θ increases monotonically in the variance of the expectational error—that is, in the variance of the difference between firms

[20] The assumption of normality of the structural error term is sufficient but not necessary for the existence of odds-based inequalities that correctly bound the true parameter vector. As long as the distribution of the structural error ν is log-concave, inequalities analogous to those in equation (17), with the correct cumulative distribution function $F_\nu(\cdot)$ instead of the normal cumulative distribution function $\Phi(\cdot)$, will also satisfy Theorem 1. The explanation of this result is that, for any log-concave distribution, both $F_\nu(\cdot)/(1 - F_\nu(\cdot))$ and $(1 - F_\nu(\cdot))/F_\nu(\cdot)$ are globally convex.

expected revenues $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and the ex post realization of such revenues r_{ijt}.

4.2.2 Generalized revealed-preference moment inequalities

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

$$\mathcal{M}^r(Z_{ijt};\gamma) = \mathbb{E}\left[\begin{array}{c} m_l^r(d_{ijt}, r_{ijt}, dist_j;\gamma) \\ m_u^r(d_{ijt}, r_{ijt}, dist_j;\gamma) \end{array} \middle| Z_{ijt}\right] \geq 0, \tag{19}$$

where the two moment functions are defined as

$$m_l^r(\cdot) = -(1-d_{ijt})\big(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j\big) + d_{ijt}\gamma_2 \frac{\phi\big(\gamma_2^{-1}(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}{\Phi\big(\gamma_2^{-1}(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}, \tag{20a}$$

$$m_u^r(\cdot) = d_{ijt}\big(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j\big) + (1-d_{ijt})\gamma_2 \frac{\phi\big(\gamma_2^{-1}(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}{1-\Phi\big(\gamma_2^{-1}(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}. \tag{20b}$$

We again denote set of all possible values of the parameter vector γ as Γ, and 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 equations (19), (20a) and (20b):

Theorem 2 *For all $\theta \in \Gamma$, $\mathcal{M}^r(Z_{ijt};\theta) \geq 0$.*

We provide a formal proof of Theorem 2 in Appendix A.5. Theorem 2 indicates that the generalized revealed-preference inequalities are consistent with the true value of the parameter vector, θ.[21] In general, the set of parameter values that satisfy both the generalized revealed-preference inequalities and the odds-based inequalities will contain values of the parameter vector γ other than the true parameter, θ. However, as we show in Section 5, in our empirical application, the set of parameter values that are consistent both with the odds-based and generalized revealed preference inequalities is small enough to allow us to draw economically meaningful conclusions.

Heuristically, the two moment functions in equations (20a) and (20b) are derived using standard revealed preference arguments. We focus our discussion on moment function (20b); the intuition behind the derivation of moment (20a) 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. $d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt}\big) \geq 0$. Taking the expectation of this

[21] The assumption of normality of the structural error term is sufficient but not necessary for the existence of generalized revealed-preference inequalities that correctly bound the true parameter vector. 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 analogous to those in equation (20) that also satisfy Theorem 2. For example, beyond the normal distribution, the type I extreme value distribution also satisfies this property.

inequality conditional on $(d_{ijt}, \mathcal{W}_{ijt})$, we obtain

$$d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + S_{ijt} \geq 0, \tag{21}$$

where $S_{ijt} = \mathbb{E}[-d_{ijt}\nu_{ijt}|d_{ijt}, \mathcal{W}_{ijt}]$. 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 not observed to the researcher; i.e. the term ν_{ijt} in the model described in Section 2.[22] We cannot directly use the inequality in equation (21) because it depends on the unobserved agents' expectations, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, both directly and through the term S_{ijt}. However, the inequality in equation (21) becomes weaker if we introduce the observed ex-post profits, r_{ijt}, in the place of the unobserved expectations and take the expectation of it conditional on W_{ijt}. Given the assumption that firms have rational expectations, the difference between the unobserved true firms' expectations and the realized revenues has mean equal to zero conditional on the vector W_{ijt}. We use this property of the expectational error combined with the fact that both $\phi(\cdot)/\Phi(\cdot)$ and $\phi(\cdot)/(1-\Phi(\cdot))$ are globally convex to apply Jensen's inequality and conclude that the inequality in equations (19) and (20b) holds at the true value of the parameter vector.

We label the moment functions in equations (20a) and (20b) as *generalized* revealed preference inequalities. These moments start with the baseline revealed preference inequalities introduced in Pakes (2010) and Pakes et al. (2015). The selection correction term we introduce in our setting allows us to account for an individual and choice-specific structural error ν_{ijt} with a non-zero variance–that is, we allow S_{ijt} to be different from zero. Given that $S_{ijt} \geq 0$ whenever $\sigma \neq 0$, if we generated revealed-preference inequalities without this individual and choice-specific structural error we would obtain weakly smaller identified sets than those found using the generalized revealed-preference inequalities in equations (19), (20a) and (20b).

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. In this very specific case, the generalized revealed preference inequalities do not have any additional identification power beyond that of the odds-based inequalities. However, in all other settings, these additional moments can provide additional identifying power.[23]

[22] Appendix A.5 shows that, under the assumptions in Section 2,

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

[23] How the identified sets found from each type of inequality compare in size is difficult to characterize generally. We show in simulations—available upon request—that there are cases in which the generalized revealed preference inequalities have additional identification power beyond that of the odds-based inequalities.

The odds-based and generalized revealed preference moment inequalities described in equations (16) and (19) condition on particular values of the instrument vector, Z. In empirical applications in which at least one of the variables in the vector Z is continuous, the sample analogue of these moment inequalities will likely involve an average over very few observations (if any). Therefore, for estimation, it is necessary to work with unconditional moment inequalities. Andrews and Shi (2013) and Armstrong (2015) define unconditional moments that imply no loss of information with respect to their conditional counterpart. We describe in Appendix A.6 the exact unconditional moments that we use to compute the estimates presented in Section 5.

As indicated above, both the odds-based and generalized revealed preference moment inequalities described in equations (16) and (19) rely on observing a measure of the potential revenue from exporting r_{ijt} for all firms, countries and time periods. In our specific setting, this is possible because the model described in Section 2 allows to rewrite r_{ijt} as a function of variables that are observed even for those firms i, countries j and periods t with zero exports. This might not be possible in other models. In order to deal with the selection problem that would arise in this case from the fact that r_{ijt} is only directly observable for those observations with $d_{ijt} = 1$, Appendix A.7 generalizes the moment inequality approach described in Sections 4.2.1 and 4.2.2 to deal with those cases in which the ex post revenues r_{ijt} are only observed for an endogenously selected subsample.

4.3 Deriving bounds on choice probabilities

As Sections 4.2.1 and 4.2.2 show, we can set identify and estimate the structural parameter vector, θ, without the need to fully specify and observe agents' information sets. However, beyond obtaining estimates of export fixed costs, another principal motivation for estimating export entry models, like that in Section 2, is to predict how changes to the economic environment will affect export participation and, through it, export volumes. In this section, we show that one can perform these counterfactual exercises without the need to impose any additional assumptions beyond those needed to define the odds-based and generalized revealed preference inequalities. Specifically, using the set estimates of the parameter vector θ and maintaining the assumption that only a subset of firms' true information sets are observed by the econometrician, we define bounds on the probability that a firm exports to each possible destination country.

Choice probabilities are not point identified in our setting for two reasons. First, even if we were to know the true value of the parameter vector, θ, the fact that we only observe a subset Z_{ijt} of the true information set, $\mathcal{J}_{ijt}$, implies that we cannot exactly compute firms' unobserved expectations $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and, therefore, we cannot compute the export probabilities in equation (12). Second, we do not recover the true value of the parameter vector θ in

our estimation, but only a set that includes it. As the following theorem shows, under these circumstances we may still derive bounds on the expected probability that firm i exports to country j at period t, conditional on Z_{ijt}. Here, Θ_{all} represents the set of parameter vectors that satisfy all of the inequalities—both the generalized revealed-preference and odds-based inequalities.

Theorem 3 *Suppose $Z_{ijt} \in \mathcal{J}_{ijt}$ and define $\mathcal{P}(Z_{ijt}) = \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]$, with $\mathcal{P}_{ijt}$ defined in equation* (12). *Then,*

$$\mathcal{P}_l(Z_{ijt}) \leq \mathcal{P}(Z_{ijt}) \leq \mathcal{P}_u(Z_{ijt}), \tag{22}$$

where

$$\mathcal{P}_l(Z_{ijt}) = \min_{\gamma \in \Theta_{all}} \frac{1}{1 + B_l(Z_{ijt};\gamma)}, \tag{23}$$

$$\mathcal{P}_u(Z_{ijt}) = \max_{\gamma \in \Theta_{all}} \frac{B_u(Z_{ijt};\gamma)}{1 + B_u(Z_{ijt};\gamma)}. \tag{24}$$

and

$$B_l(Z_{ijt};\gamma) = \mathbb{E}\left[\frac{1 - \Phi\big(\gamma_2^{-1}\big(\eta^{-1} r_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)}{\Phi\big(\gamma_2^{-1}\big(\eta^{-1} r_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)}\middle| Z_{ijt}\right], \tag{25}$$

$$B_u(Z_{ijt};\gamma) = \mathbb{E}\left[\frac{\Phi\big(\gamma_2^{-1}\big(\eta^{-1} r_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)}{1 - \Phi\big(\gamma_2^{-1}\big(\eta^{-1} r_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)}\middle| Z_{ijt}\right]. \tag{26}$$

The proof of Theorem 3 is in the Appendix A.8. By using the extreme points in the confidence set Θ_{all}^{α} instead of those in the identified set Θ_{all} in equations (23) and (24), we may build a confidence set for $\mathcal{P}(Z_{ijt})$. If we were to know the true value of the parameter vector, θ, then we could build bounds on $\mathcal{P}(Z_{ijt})$ as

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

The width of this interval is related to the fact that the researcher does not observe firms' true information sets but only the subset Z_{ijt}. By taking the minimum of the lower bound and the maximum of the upper bound in this interval over the values of the parameter vector in the set Θ_{all} we construct a larger interval that accounts for the fact that the parameter vector is only partially identified. Equation (22) defines bounds on export probabilities conditional on a particular value of the instrument vector Z_{ijt}. However, using equation (22) we may define bounds on the expected export probability for any subset of firms defined by a particular set

$\mathcal{Z}$ of values of the instrument vector Z_{ijt} as

$$\sum_{ijt} \mathcal{P}_l(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\} \leq \sum_{ijt} \mathcal{P}(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\} \leq \sum_{ijt} \mathcal{P}_u(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\}. \tag{27}$$

For example, if we define the $\mathcal{Z}$ to be a dummy variable selecting a particular country j^* and year t^*, $\mathcal{Z} = \mathbb{1}\{j = j^*, t = t^*\}$ equation (27) will yield bounds on the average export probability to country j^* in year t^*. In Section 5.1, we use the bounds in equation (27) to test the fit of the model for different countries and years. We show in Appendix A.9 how to use equation (27) to compute bounds for the counterfactual scenario described in Sections 2.5.

5 Results

We estimate the parameters of exporters' participation decisions using the three different empirical approaches discussed above. First, we use maximum likelihood to estimate the components of the exporter's fixed costs of serving a foreign market under perfect foresight: we assume the firm perfectly predicts the level of revenue it will earn upon entry. Second, we again use maximum likelihood methods, but under the two-step procedure described in Willis and Rosen (1979), Manski (1991) and Ahn and Manski (1993) in which we project a proxy for export revenue—here realized revenues—on the set of observable covariates that we assume compose a firm's information set. In practice, we include three variables in the information set the agent uses to predict revenues in a destination country j: the total aggregate exports in the prior year, R_{jt-1}, the distance to destination j, $dist_j$, and the firm's own domestic sales from the previous year, r_{jt-1}. Finally, third, we carry out our moment inequality approach. For comparison purposes, we assume, as in the two-step approach, that the firm knows lagged aggregate exports, its own lagged domestic revenue, and the distance to the export destination. However, unlike the two-step approach, the inequalities allow additional unobserved variables to enter the firm's information set, and these variables may vary idiosyncratically by firm, market, and time period.

We first discuss the parameter estimates and illustrate the baseline predictions of the models in comparison to the data. We then compare the estimates of the fixed costs of exporting under each of the three alternative methods.

5.1 Estimates and predicted exports

In Table 2, we report the estimates and the confidence regions for the parameters of our entry cost specification. The first coefficient, σ, represents the variance of the probit structural error affecting the fixed export costs. The remaining coefficients represent a constant component and the contribution of distance to the level of the fixed costs. We normalize the demand elasticity

of substitution, η, to equal a constant. From the raw coefficients, it is clear that the estimates from models that require full knowledge of the exporter's information sets produce much larger export participation 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 \$428,000 to \$479,000 when the export destination is 10,000 kilometers farther in distance. Under the two maximum likelihood procedures, the estimates of the added cost equals \$1,180,000 and \$812,000 for the same added distance.

We translate these coefficients into an estimate of the fixed costs of exporting by country and, for clarity of exposition, report the results in Table 3 for three countries out of the 22 destinations in the chemicals sector and 34 countries in the food sector used in our estimation. We focus on Argentina, Japan, and the United States. Total exports to these countries account for 29% of total exports of the Chilean chemicals sector and 56% of the food sector in the sample period and, as they span a wide range of possible distances to Chile, they are a good test for the implications of our estimates for countries that have very different fixed export costs. Under perfect foresight, we estimate the fixed costs in these three countries in the chemicals sector to equal \$894,000, \$2.80 million, and \$1.74 million respectively. Recall from Table 1, the mean volume of exports per exporting firm in this sector in Argentina, Japan, and the United States are only \$412,000, \$1.86 million, and \$2.48 million. Comparing the estimates under perfect foresight to the estimates from the two-step procedure, the latter produces entry cost estimates that are about 1/3 smaller in the chemicals sector. In the food sector, the fixed cost estimates under perfect foresight equal \$2.71 million, \$3.13 million, and \$2.90 million when exporting to Argentina, Japan, and the United States, respectively. The two-step procedure finds entry costs in the three countries that are about 20% smaller than the estimates under perfect foresight.

Under our moment inequality estimator, we find estimates of the fixed costs of exporting in the chemicals sector between \$270,000 and \$298,000 for Argentina, \$978,000 and \$1.06 million for Japan, and \$592,000 and \$632,000 for the United States. Across Argentina, Japan, and the US, the estimated bounds we find from the inequalities equal only a fraction of the perfect foresight estimates, with a level between 60% and 70% smaller than the perfect foresight values. The results are similar in the food sector: the fixed cost estimates from the moment inequality models are 80-85% smaller than the fixed cost estimates from perfect foresight. Comparing the bounds of the fixed costs from the inequalities to the estimates from the two-step approach, reported in Table 3, again the bounds are much smaller; the estimates of the fixed costs from the inequality approach are 50% smaller than those estimated under the two-step approach in the chemicals sector and about 75% smaller in the food sector. The results are in line with the discussion in Section 4.1 of the bias that arises if the researcher incorrectly assumes firms have perfect foresight. Here, we observe that specifying a specific

and limited information set also appears to drive an upward bias in the estimates of the fixed costs.

Finally, in Table 4, we report the observed level of export participation in our three comparison countries in the year 2005. Along with these observed values, we report the predictions from the export model under perfect foresight, the two-step approach, and from our inequalities. In part due to their high estimated levels of fixed costs and their high coefficient on distance, both the perfect foresight model and the two-step approach underestimate the number of entrants per country in 2005 on both the food and chemicals sectors. Interestingly, the predictions from these two approaches differ by country in the chemicals sector. For the United States, the two-step approach predicts a larger number of exporters than does the model that assumes perfect foresight. For Japan and Argentina, the perfect foresight model predicts greater entry than does the two-step approach.

For our inequality approach, the 95% confidence sets for the predicted number of exporters in both the chemicals and food sectors for Argentina, Japan, and the United States generally contain the observed number of exporters. The identified sets themselves also appear very close to the observed number of exporters. In Table 4, the one exception is the predicted number of exporters to the United States in the food sector. 48 firms chose to export, whereas the model predicts at least 64 would do so.

5.2 Robustness of inequality method

As discussed in Appendix A.6, translating the conditional moment restrictions in the inequality approach to unconditional moments offers a range of valid specifications of the set of instrument functions $g_a(\cdot)$. We illustrate the robustness of our inequality approach to various functional form assumptions on this set of functions in Figure 1.

To generate the confidence sets plotted in Figure 1, we re-run our moment inequality model under several alternative specifications of the instrument functions. Here, we use a consistent set of instrumental variables, identical to that in the main specification. Specifically, at the time the firms decides whether to enter a particular destination country in a given year, we assume the firm knows at least the aggregate exports to that country in the prior year, the distance to the country, and the firm's own domestic revenue in the prior year. With this set of variables—the same as Z^2_{ijt}, defined in Section 4—we compare alternative forms for the instrument functions $g_a(\cdot)$. Specifically, we use the functional form in Section **??**, but add moments in successive specifications that are weighted by the value of the instruments raised to different powers (i.e. using the notation in Section **??**, we explore how the set changes when we use different values for the constant a). The size of both the confidence set and the identified set varies across specifications, generally growing smaller with additional weighted moments, as illustrated in the figure. For each alternative instrument function, we carry out

the specification test suggested by Andrews and Soares (2010). The p-values are far from conventional significance levels. Thus, we fail to reject the null hypothesis that the confidence sets produced under our choice of moment inequalities contain the true value of the parameter.

In addition, we run additional moment inequality specifications in which we use the baseline model's set of instrument functions but vary the set of variables assumed to be in the information set of the firm when deciding whether to enter. We then run the model specification test of Andrews and Soares (2010) to test the null hypothesis that firms possess the information assumed in each specification at the time they choose whether to export. We run this test for two alternative information sets. In the first, we assume, as in perfect foresight, that the firm knows precisely the revenue it will earn upon exporting to a particular destination. We can reject, at conventional significance levels, that firms have this information set. We find a p-value of .05 and .01 in the chemicals and food sectors, respectively. In the second test, we estimate an inequality model in which we assume the firm knows an additional variable when it makes its entry decision. In addition to the total aggregate exports in the prior year, the distance to destination each destination, and the firm's own domestic sales from the previous year, we assume the firm also knows the average productivity of other exporters to a country in the prior year. With these four variables, and conditional on the model specification, we reject the model and the choice of information set in both the food and chemicals sector. We estimate a p-value of 0.04 and 0.01 in the food and chemicals sectors, respectively. Restating, the specification test rejects the model that includes the average productivity of other exporters to a country as an element of the firms' information set.

6 Counterfactuals

Beyond estimating the level of the fixed costs of exporting and testing the information set firms use in deciding whether to export, we can use our estimates to conduct counterfactual analyses. As introduced in Section 2.5, we simulate the effect of lowering the export fixed costs by 40%. We conduct the counterfactuals using only data from the year 2005, and compare the predictions from both our moment inequality approach and from a the models that require the researcher to specify the firm's information set. We report our counterfactual predictions from lowering export costs by 40% in Table 5.

We focus first on the predicted export participation and export volume under the perfect foresight model and under the model that requires the researcher to specify the firm's complete information set. Relative to the predictions from perfect foresight, the predicted export participation and volume under the two-step approach are lower for Argentina and Japan, but higher for the United States. That is, even when comparing the two approaches that assume that researchers observe firms' information sets, the predictions for how a policy will impact exports differs depending on the particular assumptions imposed on the content of these infor-

mation sets. Specifically, in the chemicals sector, we find the predicted export participation to be 1.1% and 11.0% lower under the two-step approach in Argentina and Japan, but is 12.6% higher under the two-step approach in the United States.

The moment inequality approach, which imposes weaker assumptions on the content of firms' information sets, produces larger predictions of the effect of the policy relative to either maximum likelihood approach. Specifically, the moment inequality estimator predicts growth in export participation in Japan that is between 22 and 24% higher than the perfect foresight prediction in the chemicals sector and between 33 and 37% higher in the food sector. Relative to the estimate from the two-step approach, the predicted number of exporters to Japan is between 37 and 40% higher in the chemicals sector and 14-18% higher in the food sector.

To illustrate the substantive findings on how export participation and trade flows vary with changes in the fixed costs of exporting, we use our preferred specification of the moment inequality model. Under this model, we assume a firm knows at least total aggregate exports in the prior year, the distance to destination each destination, and the firm's own domestic sales from the previous year. In Table 5. we report the predictions of export participation and export flows under the counterfactual policy for Argentina, Japan, and the United States. In Table 6, we report the percentage change in exports under the counterfactual policy relative to the predictions under the baseline level of fixed costs.

The estimates reveal substantive economic effects from the policy intervention. Decreasing export costs by 40% leads to a large increase in export participation in all three countries, particular in markets far from Chile. As a percentage of the baseline level, the policy which causes fixed costs to fall 40% leads to a 22 to 23% increase in export volume to Argentina in the chemicals sector. The 95% confidence set for this prediction suggests the increase may lie between 16 and 27%. 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 37 and 44%, with a confidence set ranging from 31 to 57%. We report the effects of the counterfactual policy on trade flows from Chile to Japan and the United States in both the food and chemicals sectors in Table 5.

Of course, the counterfactual predictions from our model do not account for the effect that a reduction in the fixed export costs could have on factor prices in Chile, in the demand level abroad, or in the degree of competition in destination markets. In this respect, they represent partial equilibrium effects, and might not capture the total effect of such changes in the economic environment on the number of exporters or aggregate exports. Our predicted changes, however, do illustrate the importance of a firm's fixed costs of exporting on the extensive margin of trade, which in turn affects the volume of trade. The relative precision in the estimates also illustrates that researchers can rely on only weak assumptions on firms' information sets and nonetheless provide policymakers meaningful counterfactual policy predictions.

7 Conclusion

We study the decision of individual firms in Chile to enter foreign export markets. This decision to participate in export markets drives much of the variation in the volume of trade. To be able to predict how firms will react to changes in the economic environment that affect exports, policymakers need a measure of the fixed costs of exporting. The traditional approach to recover these costs applies revealed preference analysis. Such analysis is difficult in this setting because researchers rarely observe firms' expectations about the revenue they will earn from exporting, and these expectations drive the decision to enter a foreign market.

We develop a new inequality estimator to recover the parameters of the firm's export decision that requires only weak assumptions on the information set firms use to predict revenue from exporting. We show that placing stronger assumptions on the information set, such as assuming firms have perfect foresight or specifying the exact set of variables the firm used to form its expectation, has consequences for the estimated parameters. Models requiring these assumptions produce large estimates of the fixed costs involved with exporting relative to the mean revenues of observed exporters. In contrast, our inequality approach allows the firm's expectations to be based on variables the econometrician does not observe. The estimated fixed costs from the inequalities 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 export fixed costs fall 40% also differ substantially across alternative methods.

On methodology, we provide applied economists a robust tool for estimating structural models of discrete choice when the decision maker's information set is unknown. We show how to apply these inequalities to bound the parameters of interest while imposing only weak assumptions on the information sets firm use to form expectations over future variables. Firms may rely on information the econometrician does not observe, and this set of information can differ across firms, markets, and time periods. In addition, we show how to use our inequalities to carry out counterfactual analyses. In our setting, we measure the change in export volume from a reduction in firms' fixed costs of exporting. The bounds we estimate for this effect are informative and provide substantive insight for policymakers.

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, 13, 22]

Ameriks, John, Joseph Briggs, Andrew Caplin, Matthew D. Shapiro, and Christopher Tonetti, "Resolving the Annuity Puzzle: Estimating Life-Cycle Models without (and with) Behavioral Data," *mimeo*, April 2014. [4]

Andrews, Donald W.K. and Gustavo Soares, "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 2010, *78*, 119–157. [3, 16, 25]

_ **and Xiaoxia Shi**, "Inference Based on Conditional Moment Inequalities," *Econometrica*, 2013, *81* (2), 609–666. [20, 49]

Antràs, Pol, Teresa Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S.," *mimeo*, December 2014. [1, 11]

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151–1199. [1, 11]

_ **, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple**, "Innovation and Production in the Global Economy," *mimeo*, December 2014. [1]

_ **, Sharat Ganapati, and Marc-Andreas Muendler**, "The Extensive Margin of Exporting Goods: A Firm-level Analysis," *mimeo*, September 2014. [3, 11]

Armstrong, Tim B., "Asymptotically Exact Inference in Conditional Moment Inequality Models," *Journal of Econometrics*, 2015, *186* (1), 51–65. [20, 49]

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

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott, "The Margins of US Trade," *American Economic Review*, 2010, *99*, 487–493. [1, 3]

Bilbiie, Florian, Fabio Ghironi, and Marc Melitz, "Endogenous Entry, Product Variety and Business Cycles," *Journal of Political Economy*, 2012, *120* (2), 304–345. [3]

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006, *121* (2), 541–585. [39]

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]

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]

Eaton, Jonathan, Samuel Kortum, and Francis Kramraz, "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 2011, *79* (5), 1453–1498. [1, 11]

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

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 1994, *84* (1), 157–177. [39]

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

Greenstone, Michael, Stephen Ryan, and Michael Yankovich, "The Value of a Statistical Life: Evidence from Military Retention Incentives and Occupation-Specific Mortality Hazards," *mimeo*, April 2014. [3]

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. [13]

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

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, 13, 22]

Manski, Charles F., "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, 11]

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

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

Mayer, Thierry and Soledad Zignago, "Notes on CEPII's Distances Measures," *mimeo*, May 2006. [11]

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]

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. [1, 2, 19]

— **, Jack R. Porter, Katherine Ho, and Joy Ishii**, "Moment Inequalities and their Application," *Econometrica*, 2015, *83* (1), 315–334. [2, 19]

Ruhl, Kim J. and Jonathan L. Willis, "New Exporter Dynamics," *mimeo*, July 2014. [1]

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

Willis, Robert J. and Sherwin Rosen, "Education and Self-Selection," *Journal of Political Economy*, 1979, *87* (5), 7–36. [2, 4, 22]

Wollman, Thomas G., "Trucks Without Bailouts: Equilibrium Product Characteristics for Commercial Vehicles," *mimeo*, November 2014. [2]

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, Cambridge MA: MIT Press, 2002. [14]

Yatchew, Adonis and Zvi Griliches, "Specification Error in Probit Models," *The Review of Economics and Statistics*, 1985, *67* (1), 134–139. [14, 38]

Table 2: Rescaled Parameter estimates from export model, alternative specifications

		Sector 24: chemical products			Sector 151: food products		
Model		sigma	constant	distance	sigma	constant	distance
Perfect Foresight	est	1074.0	760.9	1180.1	2039.2	2675.4	266.4
	se(est)	46.7	36.7	53.2	71.6	96.1	23.6
2-Step ML	est	701.9	502.2	812.3	1567.8	2087.2	198.1
	se(est)	24.3	20.2	30.0	48.4	65.7	18.1
MI	min(id set)	311.7	218.3	428.3	247.7	332.9	143.7
	max(id set)	341.0	245.8	479.0	264.9	380.1	173.4
	min(confidence set)	178.6	121.1	237.1	219.2	283.8	108.6
	max(confidence set)	465.1	316.3	651.2	307.7	473.8	191.9

Table 3: Estimates of export fixed costs, in $000s

	Sector 24: chemcial products			Sector 151: food products		
	Argentina	Japan	United States	Argentina	Japan	United States
Distance (in km)	1128	17247	8271	1128	17247	8271
Maximum Likelihood						
Est, Perfect Foresight ($000s)	894.0	2796.2	1736.9	2705.4	3134.9	2895.7
Std Error, Perfect Foresight ($000s)	242.3	708.6	438.7	543.5	612.1	566.6
Est, 2 Step ($000s)	593.8	1903.1	1174.0	2109.5	2428.8	2251.0
Std error, 2 Step ($000s)	109.2	315.9	190.9	328.6	363.7	336.9
Moment Inequalities						
Est, lower bound identified set	270.0	977.6	592.6	352.0	606.9	472.7
Est, upper bound identified set	298.2	1062.0	632.0	397.5	645.1	507.2
Std error, lower bound confidence set	163.0	627.5	408.8	304.2	552.2	430.0
Std error, upper bound confidence set	385.8	1420.8	836.3	489.7	730.7	590.3
Comparisons						
% Change, 2 step vs. perfect foresight	-0.34	-0.32	-0.32	-0.22	-0.23	-0.22
% Change, LB id set vs. PF	-0.67	-0.62	-0.64	-0.85	-0.79	-0.82
% Change, UB id set vs. PF	-0.70	-0.65	-0.66	-0.87	-0.81	-0.84
% Change, LB conf set vs. PF	-0.57	-0.49	-0.52	-0.82	-0.77	-0.80
% Change, UB conf set vs. PF	-0.82	-0.78	-0.76	-0.89	-0.82	-0.85

Table 4: Predicted number of exporters under alternative specifications, for select countries in the year 2005

Sector 24: chemical products							
	Observed no.			Based on identified set		Based on confidence set	
Destination Country	of exporters	Perfect Foresight	Two-step approach	Lower bound	Upper bound	Lower bound	Upper bound
Argentina	46	41.00	40.13	43.47	45.53	31.44	52.27
Japan	5	2.41	1.84	5.89	5.99	3.84	11.99
United States	24	15.69	19.04	22.16	22.64	15.00	30.86
Sector 151: food products							
	Observed no.			Based on identified set		Based on confidence set	
Destination Country	of exporters	Perfect Foresight	Two-step approach	Lower bound	Upper bound	Lower bound	Upper bound
Argentina	22	21.96	21.65	24.30	29.41	19.43	32.70
Japan	52	25.99	34.00	47.15	49.35	42.79	52.94
United States	48	39.13	40.53	78.13	81.30	72.74	84.69

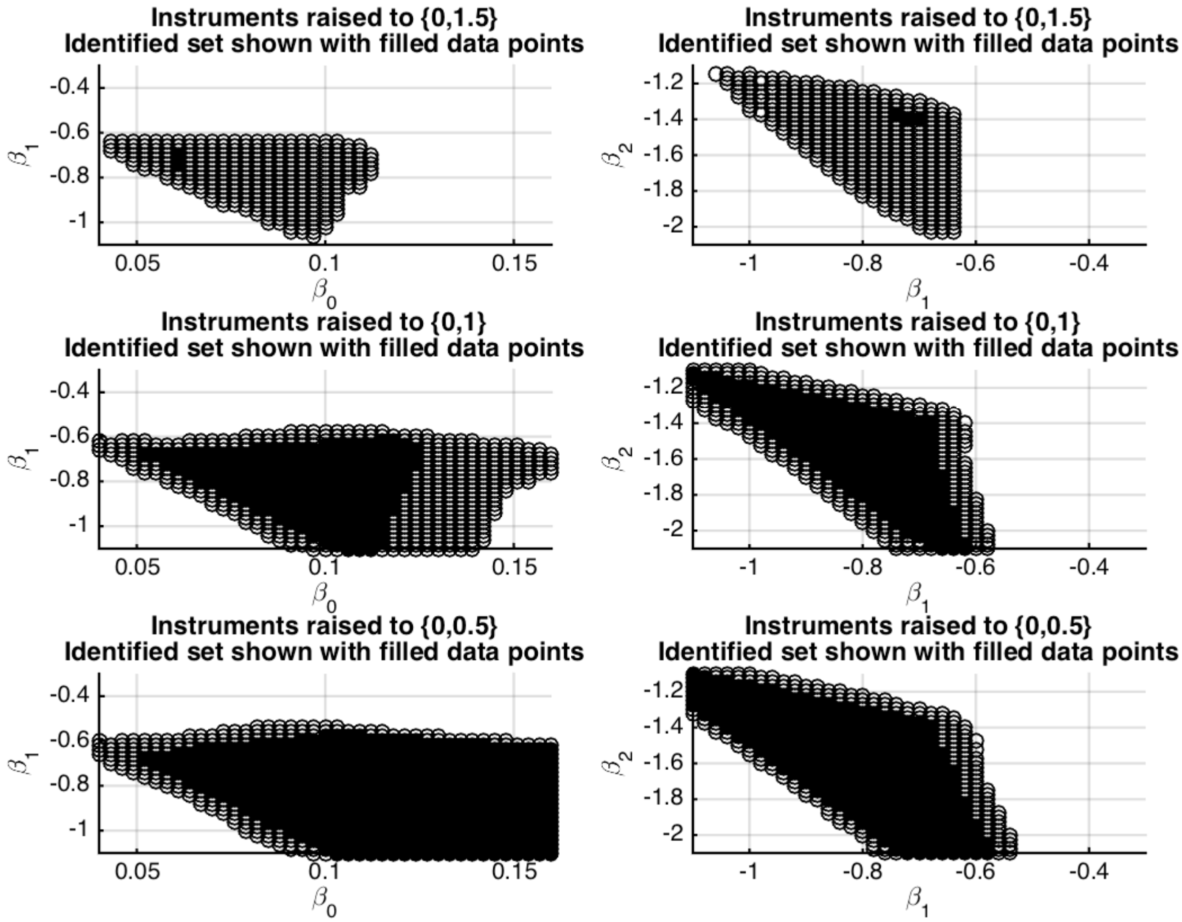
Table 5: Counterfactual predictions: Export participation and export volume after 40% decrease in fixed costs, in select countries in the year 2005

					Based on identified set		Based on confidence set	
		Observed data (N or Rev)	Perfect Foresight	Two-step approach	Lower bound	Upper bound	Lower bound	Upper bound
Sector 24: chemical products								
Export participation (no. of firms)	Argentina	46.00	61.64	60.96	65.45	66.96	55.97	73.31
	Japan	5.00	16.36	14.56	20.00	20.31	12.15	30.48
	United States	24.00	40.41	45.50	48.72	49.30	38.45	58.35
Export Volume ($ millions)	Argentina	27.60	17.86	17.26	24.92	25.22	22.42	29.20
	Japan	9.91	40.31	35.74	69.78	70.14	57.83	89.51
	United States	51.49	74.38	90.13	108.19	108.57	96.37	123.00
Sector 151: food products								
Export participation (no. of firms)	Argentina	22.00	50.05	49.97	54.80	60.10	49.20	63.44
	Japan	52.00	56.75	66.06	75.22	77.79	69.98	81.98
	United States	48.00	70.69	72.54	108.59	111.70	103.30	114.98
Export Volume ($ millions)	Argentina	10.95	14.69	15.35	30.19	31.90	27.25	33.77
	Japan	283.18	321.16	378.97	477.90	481.76	469.10	487.20
	United States	300.17	691.32	699.32	905.33	909.53	895.87	914.98

Table 6: Counterfactual predictions: Comparison of alternative specifications relative to the perfect foresight model, after 40% decrease in fixed costs in the year 2005

	Destination	% Increase from baseline pred, PF	% Increase from baseline pred, 2 Step ML	Based on identified set: % Increase from baseline pred, lb	Based on identified set: % Increase from baseline pred, ub	Based on confidence set: % Increase from baseline pred, lb	Based on confidence set: % Increase from baseline pred, ub
Sector 24: chemical products							
Export participation (no. of firms)	Argentina	50.3%	51.9%	47.1%	50.5%	42.5%	76.3%
	Japan	578.7%	692.3%	239.2%	239.5%	146.4%	298.4%
	United States	157.5%	139.0%	117.8%	119.8%	93.3%	139.3%
Export Volume ($ millions)	Argentina	40.5%	42.9%	22.0%	23.4%	16.3%	27.0%
	Japan	163.4%	192.3%	50.9%	51.1%	36.8%	56.2%
	United States	53.1%	40.5%	26.9%	27.3%	18.4%	34.0%
Sector 151: food products							
Export participation (no. of firms)	Argentina	127.9%	130.8%	104.4%	124.1%	94.6%	150.6%
	Japan	118.4%	94.3%	57.5%	60.1%	54.5%	66.3%
	United States	80.7%	79.0%	37.3%	39.0%	35.9%	41.9%
Export Volume ($ millions)	Argentina	114.3%	112.4%	36.6%	43.7%	31.1%	57.0%
	Japan	35.2%	22.9%	8.5%	9.2%	7.6%	10.8%
	United States	18.8%	18.9%	3.8%	4.1%	3.5%	4.6%

Figure 1: Confidence sets for export participation model, under alternative functional forms for the instrument set



Appendix

A.1 Proxy for export revenue: Details

We describe here how we can combine the structure introduced in Sections 2.1 and 2.2 with data on (i) aggregate exports from h to j in t, R_{jt}; (ii) domestic sales for every active firm, $\{r_{iht}; i = 1, \ldots, N_t\}$; and, (iii) the set of exporting firms, $\{d_{ijt}; i = 1, \ldots, N_t\}$, to define a perfect proxy for the export revenue that firm i would obtain in country j if it were to export to it in year t.

Given the expression for firm i's potential export revenue in j in equation (5), aggregating r_{ijt} across all firms located in country h that export to country j, we can write the aggregate exports from h to j in t as

$$R_{jt} = \sum_{i=1}^{N_t} d_{ijt} r_{ijt} di = \left[\frac{\eta}{\eta - 1} \frac{\tau_{jt} c_t}{P_{jt}} \right]^{1-\eta} Y_{jt} V_{jt}, \tag{28}$$

where V_{jt} is defined as

$$V_{jt} = \sum_{i=1}^{N_t} d_{ijt} a_{it}^{(1-\eta)} di. \tag{29}$$

Note that V_{jt} is the sum of the inverse physical productivity terms a_{it} (to the power of an exponent that depends on the demand elasticity η) across all firms that export to the destination country j in year t. We can therefore proxy for all the country-specific covariates in equation (5) by (R_{jt}/V_{jt}) and rewrite r_{ijt} as

$$r_{ijt} = \frac{a_{it}^{(1-\eta)}}{V_{jt}} R_{jt}. \tag{30}$$

The term $a_{it}^{(1-\eta)}/V_{jt}$ is the the unobserved firm-specific inverse physical productivity of firm i, a_{it}, relative to the sum of these physical productivities for all firms exporting to country j, V_{jt}. In order to proxy for this term, we use information on the domestic revenue of every firm $i = 1, \ldots, N_t$.

From equation (5), in the case in which $j = h$ and under the assumption that there are no domestic transport costs, $\tau_{iht} = 1$ for every firm i, it holds

$$r_{iht} = \left[\frac{\eta}{\eta - 1} \frac{a_{it} c_t}{P_{ht}} \right]^{1-\eta} Y_{ht}, \tag{31}$$

and, therefore, for any two firms i and i', we can write

$$\frac{a_{it}^{1-\eta}}{a_{i't}^{1-\eta}} = \frac{r_{iht}}{r_{i'ht}}. \tag{32}$$

Using this expression, we can rewrite the first term in equation (30) as

$$\frac{a_{it}^{(1-\eta)}}{V_{jt}} = \frac{1}{\frac{V_{jt}}{a_{it}^{(1-\eta)}}} = \frac{1}{\sum_{s=1}^{N_t} d_{sjt} \left(\frac{a_{st}}{a_{it}} \right)^{(1-\eta)} ds} = \frac{1}{\sum_{s=1}^{N_t} d_{sjt} (r_{sht}/r_{iht}) ds}. \tag{33}$$

Plugging back this expression into equation (30), we obtain the expression for r_{ijt} in terms of observable covariates in equation (6).

A.2 Bias in ML Estimates Under Perfect Foresight Assumption

In this section, we generate various simulated datasets for a simplified version of the binary probit export entry model in Section 2 under different assumptions on the distribution of firms' unobserved expectations and on the distribution of their expectational errors. Specifically, we assume that firm i decides whether to export to country j according to the model

$$d_{ij} = \mathbb{1}\{\psi_1 \mathbb{E}[r_{ij} | \mathcal{W}_{ij}] - \psi_2 - \nu_{ij}\},$$

where $d_{ij} = 1$ if firm i exports to j, $\psi_1 = \psi_2 = 0.5$, and $\nu_{ij} \sim \mathbb{N}(0, \sqrt{2})$ and independent of any other covariate. This export participation equation is identical to that in equation (8), except for the fact that, for the sake of simplicity, fixed export costs are assumed not to depend on distance. Mimicking the estimation problem described in Section 4.1, we assume that the researcher does not observe $\mathbb{E}[r_{ij}|\mathcal{W}_{ij}]$ but only r_{ij} such that

$$r_{ij} = \mathbb{E}[r_{ij}|\mathcal{W}_{ij}] + \varepsilon_{ij}.$$

In Table A.1 below, for different distributions of the true unobserved expectations, $\mathbb{E}[r_{ij}|\mathcal{W}_{ij}]$, and expectational error, ε_{ij}, we show the point estimates and standard errors from estimating ψ_1 and ψ_2 under the assumption of perfect foresight. Under this assumption, we can estimate ψ_1 and ψ_2 maximazing a likelihood function that relies on the individual likelihood

$$\mathcal{P}(d_{ij} = 1|r_{ij}) = \Phi((\sqrt{2})^{-1}(\beta_1 r_{ij} - \beta_2)).$$

Table A.1: Bias under Perfect Foresight

Model	Distribution of $\mathbb{E}[r_{ij}\|\mathcal{W}_{ij}]$	Distribution of ε_{ij}	$\hat{\psi}_1$	$\hat{\psi}_2$
1	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.25)$	0.4706 (0.0014)	0.4994 (0.0014)
2	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.5)$	0.3960 (0.0013)	0.4951 (0.0014)
3	$\mathbb{N}(0,1)$	$\mathbb{N}(0,1)$	0.2426 (0.0010)	0.4865 (0.0013)
4	t_2	t_2	0.1573 (0.0006)	0.4584 (0.0014)
5	t_5	t_5	0.2274 (0.0008)	0.4773 (0.0014)
6	t_{20}	t_{20}	0.2394 (0.0009)	0.4865 (0.0013)
7	t_{50}	t_{50}	0.2436 (0.0010)	0.4872 (0.0013)
8	$log\text{-}normal(0,1)$	$log\text{-}normal(0,1)$	0.1705 (0.0006)	0.5436 (0.0014)
9	$-log\text{-}normal(0,1)$	$-log\text{-}normal(0,1)$	0.1435 (0.0006)	0.4767 (0.0013)

Notes: All estimates in this table are normalized by scale by setting $var(\nu_{ij}) = 2$. In order to estimate each of the models, we generate 1,000,000 observations from the distributions of ν_{ij}, $\mathbb{E}[r_{ij}|\mathcal{W}_{ij}]$, and ε_{ij} described in columns 2 and 3 and in the main text. Whenever draws are generated from the log-normal distribution, we re-center them at zero. The true parameter values are $\psi_1 = \psi_2 = 0.5$.

The first three models in Table A.1 are specific examples of the general model studied in Yatchew and Griliches (1985). The results in columns 4 and 5 of Table A.1 show that there is downward bias in the estimate of ψ_1 and that the bias is larger as the variance of the expectational error, ε_{ij}, increases. This is consistent with the analytical formula for the bias term in Yatchew and Griliches (1985). In models 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 firms' expectations and the expectational errors are normally distributed. In models 4 to 7, we depart from the normal distribution by choosing a distribution both for the unobserved expectations and expectational errors that has fatter tails than the normal distribution. The downward bias in the estimate of ψ_1 persists and it is larger the higher the dispersion in the distribution of unobserved expectations and expectational errors. In models 8 and 9, we depart from the normal distribution by choosing distributions both for unobserved expectations and expectational errors that are asymmetric. Specifically, model 8 assumes distributions that are positively skewed, and model 9 distributions that are negatively skewed. In all cases, the estimate of the coefficient on the unobserved expectation is biased downwards.

The estimates shown in Table A.1 condition on the normalization $var(\nu_{ij}) = 2$. In practice, we never know what the variance of the structural error is. However, standard models of international trade as that described

in Section 2 imply that the coefficient on the expected export revenues is equal to the inverse of the price elasticity of demand, $1/\eta$. Furthermore, the literature in international trade provides multiple estimates of this price elasticity of demand (Feenstra, 1994; Broda and Weinstein, 2006), Accordingly, we choose the coefficient on expected revenue as the normalizing constant. Given the choice of a particular constant k as the value of ψ_1, we obtain rescaled estimates of the entry cost coefficient by multiplying our estimates of the fixed cost parameter, ψ_2, by $k/\hat{\psi}_1$. Given that the true value of k in our simulations is 0.5, the upward bias in the fixed costs parameters is given by the ratio

$$\frac{(\psi_1/\hat{\psi}_1)\hat{\psi}_2 - \psi_2}{\psi_2} = \frac{(0.5/\hat{\psi}_1)\hat{\psi}_2 - 0.5}{0.5}.$$

Table B.1 reports this number for the nine models described in Table A.1. The results show that assuming perfect foresight implies that we over estimate export fixed costs in a magnitude that varies between 6% (for the model in which the variance of the expectational error is minimal) and 219% (for a model in which the distribution of the expectational error is not symmetric).

Table B.1: Bias in fixed costs Estimates

Model	1	2	3	4	5	6	7	8	9
Bias	6%	25%	100%	191%	110%	103%	100%	219%	167%

A.3 Partial Identification: Example

The data are informative about the joint distribution of $(d_{ijt}, Z_{ijt}, r_{ijt})$ across i, j, and t. Consistently with the possible vectors of instruments discussed in Section 4, we assume that we always 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, let's use r^e_{ijt} to denote $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$. Note that 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 $(x_1, \ldots, x_K)$. In this note, we use $f(\cdot)$ to denote distributions that involve some variable that is not directly observable in the data; i.e. r^e_{ijt}. Using rules of conditional distributions, we can further write

$$\mathbb{P}(d_{ijt}, Z_{ijt}, r_{ijt}) = \int f^y(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}) f^y(r_{ijt}|r^e_{ijt}, Z_{ijt}) f^y(r^e_{ijt}|Z_{ijt}) \mathbb{P}(Z_{ijt}) dr^e_{ijt}, \tag{34}$$

where we use $\mathbb{P}(Z_{ijt})$ to denote that the marginal distribution of Z_{ijt} is 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 (34). The model in Section 2 imposes the following restriction on the elements of equation (34):

$$f^y(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}) = f(d_{ijt}|r^e_{ijt}, Z_{ijt}; \gamma^y) =$$
$$\Big(\Phi\big((\gamma^y_2)^{-1}\big(\eta^{-1} r^e_{ijt} - \gamma^y_0 - \gamma^y_1 dist_j\big)\big)\Big)^{d_{ijt}} \Big(1 - \Phi\big((\gamma^y_2)^{-1}\big(\eta^{-1} r^e_{ijt} - \gamma^y_0 - \gamma^y_1 dist_j\big)\big)\Big)^{1-d_{ijt}}. \tag{35}$$

Here, we show that γ is partially identified in a model that imposes restrictions that are stronger than those in Section 2. This means that there exists at least two structures S^y that imply different values of γ and that verify equation (34) even after we impose additional restrictions to those implied by the model in Section

2. Specifically, we impose the following additional restrictions on the elements of equation (34)

$$\gamma_1 \text{ is known and equal to } 0, \tag{36a}$$

$$r_{ijt} = r^e_{ijt} + \varepsilon_{ijt}, \qquad \varepsilon_{ijt}|(r^e_{ijt}, \xi_{ijt}) \sim \mathbb{N}(0, \sigma^2_\varepsilon) \tag{36b}$$

$$Z_{ijt} = r^e_{ijt} + \xi_{ijt} \qquad \xi_{ijt}|r^e_{ijt} \sim \mathbb{N}((\sigma_\xi/\sigma_{r^e})\rho_{\xi r^e}(r^e_{ijt} - \mu_{r^e}), (1-\rho^2_{\xi r^e})\sigma^2_\xi) \tag{36c}$$

$$r^e_{ijt} \sim \mathbb{N}(\mu_{r^e}, \sigma^2_{r^e}) \tag{36d}$$

where ξ_{ijt} denotes the different between the true unobserved expectation r^e_{ijt} and the instrument Z_{ijt}, $\mu_{r^e} = \mathbb{E}[r^e_{ijt}]$, $\sigma^2_{r^e} = var(R^e_{ijt})$, $\sigma^2_w = var(W_{ijt})$, and $\rho_{r^e w} = corr(r^e_{ijt}, W_{ijt})$. Equation (36a) restricts the model in Section 2 by assuming that distance does not affect fixed entry costs. Equation (36b) assumes that firms' expectational error is normally distributed and independent of both firms' unobserved expectations and the different 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}. Equation (36c) imposes a particular distributional assumption between 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. Finally, equation (36d) 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 (36) defines a model that is more restrictive than that on which we apply the moment inequalities defined in Section 4.2. However, as we show below, even after adding the assumptions in equation (36), we can still find at least two structures

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

that: (1) verify the restrictions in equations (35) and (36); (2) verify equation (34); and (3) $\gamma^{a_1} \neq \gamma^{a_2}$. If γ is partially identified in this stricter model, it will also be partially identified in the more general model described in Section 2.

Equation (36a) simplifies the identification exercise discussed here because the only parameters that are left to identify are (γ_0, γ_2); i.e. we can set $\gamma_1 = 0$ in equation (35). Equation (36b) assumes that the expectational error not only has mean zero and finite variance but is also normally distributed. It 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 (36c) and (36d) jointly determine the conditional density $f(r^e_{ijt}|Z_{ijt})$ entering equation (34).

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^{a_1} *and* S^{a_2} *for which*

1. *both* S^{y_1} *and* S^{y_2} *verify equations* (34), (35), *and* (36);
2. $\gamma^{y_1} \neq \gamma^{y_2}$.

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

Lemma A.3.1 *The parameter vector* (γ_0, γ_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 (35) as

$$\Big(\Phi\big(\eta^{-1}\frac{\sigma_{r^e}}{\gamma_2}r^e_{ijt} - \frac{\gamma_0}{\gamma_2}\big)\Big)^{d_{ijt}}\Big(1 - \Phi\big(\eta^{-1}\frac{\sigma_{r^e}}{\gamma_2}r^e_{ijt} - \frac{\gamma_0}{\gamma_2}\big)\Big)^{1-d_{ijt}}.$$

The parameter γ_2 only enters likelihood function in equation (34) either dividing σ_{r^e} or dividing γ_0. Therefore, we can only separately identify γ_0 and γ_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 (36b), (36c) and (36d), 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:

$$\begin{aligned} \sigma_r^2 &= \sigma_{r^e}^2 + \sigma_\varepsilon^2, \\ \sigma_z^2 &= \sigma_{r^e}^2 + \sigma_\xi^2 + 2\rho_{\xi r^e}\sigma_{r^e}\sigma_\xi, \\ \sigma_{rz} &= \sigma_{r^e}^2 + \rho_{\xi r^e}\sigma_{r^e}\sigma_\xi \end{aligned} \tag{37}$$

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 γ is point identified. If we allow $\rho_{\xi r^e}$ to be different from zero, this system of equations in equation (38) only allows to define bounds on $\sigma_{r^e}^2$. Note that we can rewrite the system of equations in equation (38) as

$$\begin{aligned} \sigma_r^2 &= \sigma_{r^e}^2 + \sigma_\varepsilon^2, \\ \sigma_z^2 &= \sigma_{r^e}^2 + \sigma_\xi^2 + 2\sigma_{\xi r^e} \\ \sigma_{rz} &= \sigma_{r^e}^2 + \sigma_{\xi r^e}. \end{aligned} \tag{38}$$

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

A.4 Proof of Theorem 1

For the sake of simplicity in the notation and consistent with the definition of potential exporters' information sets used earlier, in this section we assume that $dist_j \in \mathcal{W}_{ijt}$.

Lemma 1 *Let $L(d_{ijt}|\mathcal{W}_{ijt};\theta)$ denote the log-likelihood conditional on $\mathcal{W}_{ijt}$. Suppose equation* (12) *holds. Then:*

$$\frac{\partial L(d_{ijt}|\mathcal{W}_{ijt};\theta)}{\partial\theta} = \mathbb{E}\Big[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))} - (1-d_{ijt})\Big|\mathcal{W}_{ijt}\Big] = 0. \tag{39}$$

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

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

The score function is given by

$$\begin{aligned} &\frac{\partial L(d_{ijt}|\mathcal{W}_{ijt};\theta)}{\partial\theta} = \\ &\mathbb{E}\Big[d_{ijt}\frac{1}{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)}\frac{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)))}{\partial\theta} \\ &+(1-d)\frac{1}{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\frac{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{\partial\theta}\Big|\mathcal{W}_{ijt}\Big] = 0. \end{aligned} \tag{40}$$

Reordering terms

$$\begin{aligned} \frac{\partial L(d_{ijt}|\mathcal{W}_{ijt};\theta)}{\partial\theta} &= \mathbb{E}\Big[\frac{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta}{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\times \\ &\Big[d_{ijt}\frac{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\times \\ &\times\frac{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}{\partial\Phi(-\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta} + (1-d_{ijt})\Big|\mathcal{W}_{ijt}\Big] = 0. \end{aligned} \tag{41}$$

Given that

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

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

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

we can simplify:

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

Equation (39) follows by symmetry of the function $\Phi(\cdot)$. ■

Lemma 2 *Suppose the assumptions in equations* (9), (10), *and* (12) *hold. Then*

$$\mathbb{E}\left[d_{ijt}\frac{1 - \Phi(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j))}\middle|\mathcal{W}_{ijt}\right]$$
$$\geq$$
$$\mathbb{E}\left[d_{ijt}\frac{1 - \Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))}\middle|\mathcal{W}_{ijt}\right]. \tag{42}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and the assumptions in equations (9) and (10). From equations (2), (8) and the assumption that $dist_j \in \mathcal{W}_{ijt}$ it follows that d_{ijt} may be written as a function of the vector $(\mathcal{W}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{W}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt}, d_{ijt}] = 0$. Since

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

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

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

Equation (42) follows from the equality $\eta^{-1}r_{ijt} = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$. ■

Corollary 1 *Suppose $Z_{ijt} \in \mathcal{W}_{ijt}$. Then:*

$$\mathbb{E}\left[d_{ijt}\frac{1 - \Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))} - (1 - d_{ijt})\middle|Z_{ijt}\right] = 0. \tag{43}$$

and

$$\mathbb{E}\left[d_{ijt}\frac{1 - \Phi(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j))}\middle|Z_{ijt}\right] \geq \mathbb{E}\left[d_{ijt}\frac{1 - \Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))}\middle|Z_{ijt}\right]. \tag{44}$$

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

Lemma 3 *Let $L(d_{ijt}|\mathcal{W}_{ijt};\theta)$ denote the log-likelihood conditional on $\mathcal{W}_{ijt}$. Suppose equation* (12) *holds. Then:*

$$\frac{\partial L(d_{ijt}|\mathcal{W}_{ijt};\theta)}{\partial\theta} = \mathbb{E}\left[(1 - d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))}{1 - \Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j))} - d_{ijt}\middle|\mathcal{W}_{ijt}\right] = 0. \tag{45}$$

Proof: From equation (40), reordering terms

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

Given that

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

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

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

we can simplify:

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

Equation (45) follows by symmetry of the function $\Phi(\cdot)$. ■

Lemma 4 *Suppose the assumptions in equations* (9), (10), *and* (12) *hold. Then*

$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}\Bigg|\mathcal{W}_{ijt}\Bigg]$$
$$\geq$$
$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\Bigg|\mathcal{W}_{ijt}\Bigg]. \tag{46}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and the assumptions in equations (9) and (10). From equations (2), (8) and the assumption that $dist_j \in \mathcal{W}_{ijt}$ it follows that d_{ijt} may be written as a function of the vector $(\mathcal{W}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{W}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt}, d_{ijt}] = 0$. Since

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

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

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

Equation (46) follows from the equality $\eta^{-1}r_{ijt} = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$. ■

Corollary 2 *Suppose $Z_{ijt} \in \mathcal{W}_{ijt}$. Then:*

$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))} - d_{ijt}\Bigg|Z_{ijt}\Bigg] = 0. \tag{47}$$

and

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}\bigg|Z_{ijt}\right]$$
$$\geq$$
$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\bigg|Z_{ijt}\right]. \tag{48}$$

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

Proof of Theorem 1 Combining equations (43) and (44), we obtain the inequality defined by equations (16) and (17a). Combining equations (47) and (48), we obtain the inequality defined by equations (16) and (17b). ■

A.5 Proof of Theorem 2

For the sake of simplicity in the notation and consistent with the definition of potential exporters' information sets used earlier, in this section we assume that $dist_j \in \mathcal{W}_{ijt}$.

Lemma 5 *Suppose equations* (2) *and* (8) *hold. Then,*

$$\mathbb{E}[d_{ijt}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt})|\mathcal{W}_{ijt}]\geq 0. \tag{49}$$

Proof: From equations (2) and (8),

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

This implies

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

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

Lemma 6 *Suppose equations* (2), (3)*, and* (8) *hold. Then*

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j\big)+(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|\mathcal{W}_{ijt}\Big]\geq 0. \tag{50}$$

Proof: From equation (49),

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big]-\mathbb{E}\big[d_{ijt}\nu_{ijt}|\mathcal{W}_{ijt}]\geq 0. \tag{51}$$

Since the assumption in equation (3) implies that $\mathbb{E}[\nu_{ijt}|\mathcal{W}_{ijt}]=0$, it follows that

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

and we can rewrite equation (51) as

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j\big)|\mathcal{W}_{ijt}]+\mathbb{E}\big[(1-d_{ijt})\nu_{ijt}|\mathcal{W}_{ijt}]\geq 0. \tag{52}$$

Applying the Law of Iterated Expectations, it follows that

$$\mathbb{E}\big[(1-d_{ijt})\nu_{ijt}|\mathcal{W}_{ijt}\big] = \mathbb{E}\big[\mathbb{E}[(1-d_{ijt})\nu_{ijt}|d_{ijt},\mathcal{W}_{ijt}]|\mathcal{W}_{ijt}\big] = \mathbb{E}\big[(1-d_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt},\mathcal{W}_{ijt}]|\mathcal{W}_{ijt}\big] =$$
$$P(d_{ijt}=1|\mathcal{W}_{ijt}) \times 0 \times \mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + P(d_{ijt}=0|\mathcal{W}_{ijt}) \times 1 \times \mathbb{E}[\nu_{ijt}|d_{ijt}=0,\mathcal{W}_{ijt}] =$$
$$P(d_{ijt}=0|\mathcal{W}_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt}=0,\mathcal{W}_{ijt}] = \mathbb{E}[(1-d_{ijt})|\mathcal{W}_{ijt}]\mathbb{E}[\nu_{ijt}|d_{ijt}=0,\mathcal{W}_{ijt}] =$$
$$\mathbb{E}\big[(1-d_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt}=0,\mathcal{W}_{ijt}]|\mathcal{W}_{ijt}\big],$$

and we can rewrite equation (52) as

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + (1-d_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt}=0,\mathcal{W}_{ijt}]|\mathcal{W}_{ijt}\big] \geq 0. \tag{53}$$

Using the definition of d_{ijt} in equation (8), it follows

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

and, following equation (3), we can rewrite

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

Equation (50) follows by applying this equality to equation (53). ■

Lemma 7 *Suppose the assumptions in equations* (3) *and* (9) *hold. Then*

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] = \mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] \tag{54}$$

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

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] =$$
$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] + \mathbb{E}\Big[\eta^{-1}d_{ijt}\varepsilon_{ijt}\Big|\mathcal{W}_{ijt}\Big]. \tag{55}$$

From equations (3) and (9), $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt},\nu_{ijt}] = 0$. From equations (2), (8) and the assumption that $dist_j \in \mathcal{J}_{ijt}$ it follows that d_{ijt} is a function of the vector $(\mathcal{W}_{ijt},\nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{W}_{ijt},\nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt},d_{ijt}] = 0$ and, applying the Law of Iterated Expectations,

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

Applying this result to equation (55) yields equation (54).

Lemma 8 *Suppose the assumptions in equation* (9) *and* (3) *hold. Then*

$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{W}_{ijt}\Big]$$
$$\geq$$
$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{W}_{ijt}\Big] \tag{56}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and the assumptions in equations (9) and (3) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt},\nu_{ijt}] = 0$. From equations (2), (8) and the assumption that $dist_j \in \mathcal{W}_{ijt}$ it follows that d_{ijt} is a function of the vector $(\mathcal{W}_{ijt},\nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{W}_{ijt},\nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt},d_{ijt}] = 0$. Since

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

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

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

Equation (56) follows from the equality $\eta^{-1}r_{ijt} = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$. ■

Corollary 3 *Suppose $Z_{ijt} \in \mathcal{W}_{ijt}$ then*

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + (1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq 0, \tag{57}$$

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] = \mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|Z_{ijt}\Big], \tag{58}$$

and

$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big]$$
$$\geq$$
$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big]. \tag{59}$$

Proof: The results follow from Lemmas 6, 7 and 8 and the application of the Law of Iterated Expectations. ■

Lemma 9 *Suppose equations* (2) *and* (8) *hold. Then,*

$$\mathbb{E}[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt})|\mathcal{W}_{ijt}] \geq 0. \tag{60}$$

Proof: From equations (2) and (8),

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

This implies

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

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

Lemma 10 *Suppose equations* (2), (3), *and* (8). *Then*

$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{W}_{ijt}\Big] \geq 0. \tag{61}$$

Proof: From equation (60),

$$\mathbb{E}\big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] + \mathbb{E}\big[(1-d_{ijt})\nu_{ijt}\big|\mathcal{W}_{ijt}\big] \geq 0. \tag{62}$$

Since the assumption in equation (3) implies that $\mathbb{E}[\nu_{ijt}|\mathcal{W}_{ijt}] = 0$, it follows that

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

and we can rewrite equation (62) as

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] - \mathbb{E}\big[d_{ijt}\nu_{ijt}\big|\mathcal{W}_{ijt}\big] \geq 0. \tag{63}$$

Applying the Law of Iterated Expectations, it follows that

$$\begin{gathered}\mathbb{E}\big[d_{ijt}\nu_{ijt}\Big|\mathcal{W}_{ijt}\Big] = \mathbb{E}\big[\mathbb{E}[d_{ijt}\nu_{ijt}|d_{ijt}, \mathcal{J}_{ijt}]\big|\mathcal{W}_{ijt}\big] = \mathbb{E}\big[d_{ijt}\mathbb{E}[\nu_{ijt}|d_{ijt}, \mathcal{W}_{ijt}]\big|\mathcal{W}_{ijt}\big] = \\ P(d_{ijt} = 1|\mathcal{W}_{ijt}) \times 1 \times \mathbb{E}[\nu_{ijt}|d_{ijt} = 1, \mathcal{W}_{ijt}] + P(d_{ijt} = 0|\mathcal{W}_{ijt}) \times 0 \times \mathbb{E}[\nu_{ijt}|d_{ijt} = 0, \mathcal{W}_{ijt}] = \\ P(d_{ijt} = 1|\mathcal{W}_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt} = 1, \mathcal{W}_{ijt}] = \mathbb{E}[d_{ijt}|\mathcal{W}_{ijt}]\mathbb{E}[\nu_{ijt}|d_{ijt} = 1, \mathcal{W}_{ijt}] = \mathbb{E}\big[d_{ijt}\mathbb{E}[\nu_{ijt}|d_{ijt} = 1, \mathcal{W}_{ijt}]\big|\mathcal{W}_{ijt}\big],\end{gathered}$$

and we can rewrite equation (63) as

$$\mathbb{E}\big[-(1 - d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) - d_{ijt}\mathbb{E}[\nu_{ijt}|d_{ijt} = 1, \mathcal{W}_{ijt}]\big|\mathcal{W}_{ijt}\big] \geq 0. \tag{64}$$

Using the definition of d_{ijt} in equation (8), it follows

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

and, following equation (3), we can rewrite

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

Equation (61) follows by applying this equality to equation (64). ■

Lemma 11 *Suppose the assumptions in equation* (9) *and* (3) *hold. Then*

$$\mathbb{E}\Big[-(1 - d_{ijt})\big(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] = \mathbb{E}\Big[-(1 - d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] \tag{65}$$

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

$$\begin{gathered}\mathbb{E}\Big[-(1 - d_{ijt})\big(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] = \\ \mathbb{E}\Big[-(1 - d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] - \mathbb{E}\Big[k(1 - d_{ijt})\varepsilon_{ijt}\Big|\mathcal{W}_{ijt}\Big].\end{gathered} \tag{66}$$

From equations (9) and (3), $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt}, \nu_{ijt}] = 0$. From equations (2), (8) and the assumption that $dist_j \in \mathcal{W}_{ijt}$ it follows that d_{ijt} is a function of the vector $(\mathcal{W}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{W}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt}, d_{ijt}] = 0$ and, applying the Law of Iterated Expectations,

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

Applying this result to equation (66) yields equation (65).

Lemma 12 *Suppose the assumptions in equation* (9) *and* (3) *hold. Then*

$$\mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{W}_{ijt}\Big] \geq \mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{W}_{ijt}\Big] \tag{67}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and the assumptions in equations (9) and (3) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt}, \nu_{ijt}] = 0$. From equations (2), (8) and the assumption that $dist_j \in \mathcal{W}_{ijt}$ it follows

that d_{ijt} is a function of the vector $(\mathcal{W}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{W}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt}, d_{ijt}] = 0$. Since

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

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

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

Equation (67) follows from the equality $\eta^{-1}r_{ijt} = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] + \eta^{-1}\varepsilon_{ijt}$. ■

Corollary 4 *Suppose $Z_{ijt} \in \mathcal{W}_{ijt}$ then*

$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq 0. \tag{68}$$

$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|Z_{ijt}\Big] = \mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|Z_{ijt}\Big] \tag{69}$$

and

$$\mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq \mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \tag{70}$$

Proof of Theorem 2 Combining equations (57), (58), and (59) we obtain the inequality defined by equations (19) and (20a). Combining equations (68), (69), and (70) we obtain the inequality defined by equations (19) and (20b). ■

A.6 Deriving unconditional moments

The moment inequalities described in equations (16) and (19) 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 vector $g(\cdot)$, we derive unconditional moments that are consistent with our conditional moments:

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

where $m_l(\cdot)$, $m_u(\cdot)$, $m_l^r(\cdot)$, and $m_u^r(\cdot)$ are defined in equations (17) and (20), and Z_{ijt} is the same vector of observed covariates employed in defining the conditional moments in equations (16) and (19).

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

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

In words, for each of the scalar random variables 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 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 distance between the value of the instrument Z_{kijt} and the median value of this instrument. Specifically, in Section 5, we assume that

$Z_{ijt} = Z^2_{ijt} = (r_{iht-1}, R_{jt-1}, dist_j)$ 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 generalized revealed preference inequalities (in equations (17) and (20)), the total number of moments used in the estimation is equal to twenty-four for a given value of a, in addition to a constant vector. In the benchmark case we simultaneously use two different instrument functions, $g_a(Z_{ijt})$, for $a = 0, 1.5$, to define both an estimated set Θ_{all} and a confidence set Θ^{α}_{all} at significance level α. In Section 5.2, we show results for different vectors of instruments functions $g(Z_{ijt})$ that combine the functions $g_a(Z_{ijt})$ for different sets of values of a.

The unconditional moment inequalities proposed here generate a larger identified set than that defined by the conditional moments described in Sections 4.2.1 and 4.2.2. The main advantage of the moments proposed here is computational simplicity. Papers that define unconditional moments that imply no loss of information with respect to their conditional counterpart are Andrews and Shi (2013) and Armstrong (2015). The instrument functions suggested in these papers are computationally expensive in our setting.

A.7 Accounting for sample selection in r_{ijt}

Suppose

$$\begin{aligned} r_{it} &= \gamma X_{1it} + \nu_{1it} + \xi_{it}, \\ d_{it} &= \mathbb{1}\{\beta_1 \mathbb{E}[r_{it}|\mathcal{W}_{it}] + \beta_2 X_{2it} + \nu_{2it} \geq 0\}, \end{aligned}$$

the information set of firm i at period t at the time at which it decides whether to export is $\mathcal{J}_{it}$ with

$$(X_{2it}, \nu_{2it}, \nu_{1it}, \mathcal{W}_{it}) \in \mathcal{J}_{it}, \qquad \text{and} \qquad \mathbb{E}[\xi_{it}|\mathcal{J}_{it}] = 0$$

and

$$\begin{pmatrix} \nu_{1it} \\ \nu_{2it} \end{pmatrix} \sim \mathbb{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & 1 \end{pmatrix} \right),$$

Setting σ_2^2 to equal 1 is purely a normalization by scale. Crucially, note that we do not assume anything on whether X_{1it} is included in the information set of the exporter $\mathcal{J}_{it}$. While $\mathcal{J}_{it}$ encompasses everything firm i knows at t, $\mathcal{W}_{it}$ includes all those covariates that are used by i to predict export revenues at t.

We only observe $\{X_{1it}, X_{2it}, d_{it}, d_{it} r_{it}\}$. In words, we observe $\{X_{1it}, X_{2it}, d_{it}\}$ for all i and t, and r_{it} only for those observations i and t with $d_{it} = 1$. Let's denote

$$\rho_{12} = \frac{\sigma_{12}}{\sqrt{\sigma_1^2}\sqrt{\sigma_2^2}} = \frac{\sigma_{12}}{\sqrt{\sigma_1^2}} = \frac{\sigma_{12}}{\sigma_1},$$

where σ_1 denotes the standard deviation of ν_{1it}.

Given the definition of $\mathcal{J}_{it}$ as the true information set of agent i at period t and the assumption that

$(\mathcal{W}_{it}, X_{2it}, \nu_{it}) \in \mathcal{J}_{it}$, we can form the following conditional moment

$$\begin{aligned}
\mathbb{E}[r_{it}|d_{it}=1,\mathcal{J}_{it}] &= \gamma\mathbb{E}[X_{1it}|\mathcal{J}_{it}] + \mathbb{E}[\mathbb{E}[\nu_{1it}|d_{it}=1,\mathcal{W}_{it},X_{2it}]|\mathcal{J}_{it}] \\
&= \gamma\mathbb{E}[X_{1it}|\mathcal{J}_{it}] + \mathbb{E}\Big[\mathbb{E}\Big[\nu_{1it}\Big|\beta_1\mathbb{E}[r_{it}|\mathcal{W}_{it}] + \beta_2 X_{2it} + \nu_{2it} \geq 0, \mathcal{W}_{it}, X_{2it}\Big]\Big|\mathcal{J}_{it}\Big] \\
&= \gamma\mathbb{E}[X_{1it}|\mathcal{J}_{it}] + \mathbb{E}\Big[\mathbb{E}\Big[\nu_{1it}\Big|\nu_{2it} \geq -\Big(\beta_1\mathbb{E}[r_{it}|\mathcal{W}_{it}] + \beta_2 X_{2it}\Big), \mathcal{W}_{it}, X_{2it}\Big]\Big|\mathcal{J}_{it}\Big] \\
&= \gamma\mathbb{E}[X_{1it}|\mathcal{J}_{it}] + \rho_{12}\sigma_1\mathbb{E}\Big[\frac{\phi(\beta_1\mathbb{E}[r_{it}|\mathcal{W}_{it}] + \beta_2 X_{2it})}{\Phi(\beta_1\mathbb{E}[r_{it}|\mathcal{W}_{it}] + \beta_2 X_{2it})}\Big|\mathcal{J}_{it}\Big] \\
&= \gamma\mathbb{E}[X_{1it}|\mathcal{J}_{it}] + \sigma_{12}\mathbb{E}\Big[\frac{\phi(\beta_1\mathbb{E}[r_{it}|\mathcal{W}_{it}] + \beta_2 X_{2it})}{\Phi(\beta_1\mathbb{E}[r_{it}|\mathcal{W}_{it}] + \beta_2 X_{2it})}\Big|\mathcal{J}_{it}\Big].
\end{aligned}$$

We cannot directly use this moment for estimation purposes because we do not observe $\mathcal{W}_{it}$ and, therefore, we cannot construct a consistent estimator for $\mathbb{E}[r_{it}|\mathcal{W}_{it}]$ for every i and t. Similarly, we do not observe $\mathcal{J}_{it}$ and, therefore, we cannot construct $\mathbb{E}[X_{1it}|\mathcal{J}_{it}]$.

Under the assumption of rational expectations, we can write

$$r_{it} = \mathbb{E}[r_{it}|\mathcal{W}_{it}] + \varepsilon_{it}, \qquad \text{with} \qquad \mathbb{E}[\varepsilon_{it}|\mathcal{J}_{it}] = 0.$$

Therefore, we can rewrite

$$\mathbb{E}[r_{it}|d_{it}=1,\mathcal{J}_{it}] = \gamma\mathbb{E}[X_{1it}|\mathcal{J}_{it}] + \sigma_{12}\mathbb{E}\Big[\frac{\phi(\beta_1 r_{it} + \beta_2 X_{2it} - \beta_1\varepsilon_{it})}{\Phi(\beta_1 r_{it} + \beta_2 X_{2it} - \beta_1\varepsilon_{it})}\Big|\mathcal{J}_{it}\Big].$$

Given that $\phi(\cdot)/\Phi(\cdot)$ is globally convex and that $\mathbb{E}[\varepsilon_{it}|\mathcal{J}_{it}] = 0$, we can conclude that

$$\begin{gathered}
\mathbb{E}[r_{it}|d_{it}=1,\mathcal{J}_{it}] \\
\geq \\
\gamma\mathbb{E}[X_{1it}|\mathcal{J}_{it}] + \sigma_{12}\mathbb{E}\Big[\frac{\phi(\beta_1 r_{it} + \beta_2 X_{2it})}{\Phi(\beta_1 r_{it} + \beta_2 X_{2it})}\Big|\mathcal{J}_{it}\Big],
\end{gathered}$$

or, equivalently,

$$\mathbb{E}[r_{it}|d_{it}=1,\mathcal{J}_{it}] \geq \mathbb{E}\Big[\gamma X_{1it} + \sigma_{12}\frac{\phi(\beta_1 r_{it} + \beta_2 X_{2it})}{\Phi(\beta_1 r_{it} + \beta_2 X_{2it})}\Big|\mathcal{J}_{it}\Big],$$

and

$$\mathbb{E}\left[r_{it} - \gamma X_{1it} - \sigma_{12}\frac{\phi(\beta_1 r_{it} + \beta_2 X_{2it})}{\Phi(\beta_1 r_{it} + \beta_2 X_{2it})}\middle| d_{it}=1,\mathcal{J}_{it}\right] \geq 0.$$

Given a vector $Z_{it} \in \mathcal{J}_{it}$, we can further write

$$\mathbb{E}\left[r_{it} - \gamma X_{1it} - \sigma_{12}\frac{\phi(\beta_1 r_{it} + \beta_2 X_{2it})}{\Phi(\beta_1 r_{it} + \beta_2 X_{2it})}\middle| d_{it}=1,Z_{it}\right] \geq 0.$$

Besides this moment inequality, we can also derive revealed-preference and odds-based inequalities. When deriving these inequalities, we must take into account that now r_{it} is not observed for every firm i and period t. However, we can rely on the fact that X_{1it} is observed for every firm i and time period t, independently of their export status.

Specifically, we know that

$$\mathbb{E}[r_{it}|\mathcal{W}_{it}] = \mathbb{E}[\gamma X_{1it} + \nu_{1it} + \xi_{it}|\mathcal{W}_{it}] = \gamma\mathbb{E}[X_{1it}|\mathcal{W}_{it}] + \nu_{1it}.$$

Therefore, we can write

$$\begin{aligned}
d_{it} &= \mathbb{1}\{\beta_1\gamma\mathbb{E}[X_{1it}|\mathcal{W}_{it}] + \beta_1\nu_{1it} + \beta_2 X_{2it} + \nu_{2it} \geq 0\}, \\
&= \mathbb{1}\{\beta_1\gamma\mathbb{E}[X_{1it}|\mathcal{W}_{it}] + \beta_2 X_{2it} + \nu_{it} \geq 0\},
\end{aligned}$$

where $\nu_{it} = \beta_1 \nu_{1it} + \nu_{2it}$ and

$$\nu_{it} \sim \mathbb{N}(0, \sigma_\nu^2).$$

where $\sigma_\nu^2 = (\beta_1)^2 \sigma_1^2 + 1 + 2\sigma_{12}$. Normalizing the parameters driving d_{it} by scale, we can redefine

$$d_{it} = \mathbb{1}\{(\beta_1/\sigma_\nu)\gamma \mathbb{E}[X_{1it}|\mathcal{W}_{it}] + (\beta_2/\sigma_\nu)X_{2it} + \tilde{\nu}_{it} \geq 0\},$$

where $\tilde{\nu}_{it} \sim \mathbb{N}(0,1)$. Using this expression, we can derive the following odds-based inequalities

$$\mathbb{E}\left[d_{it} \frac{1 - \Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)}{\Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)} - (1-d_{it}) \middle| Z_{it} \right] \geq 0,$$
$$\mathbb{E}\left[(1-d_{it}) \frac{\Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)}{1 - \Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)} - d_{it} \middle| Z_{it} \right] \geq 0,$$

and the following revealed-preference inequalities

$$\mathbb{E}\left[-(1-d_{it})((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}) + d_{it}\sigma_\nu \frac{\phi((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it})}{\Phi((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it})} \middle| Z_{it} \right] \geq 0,$$
$$\mathbb{E}\left[d_{it}((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}) + (1-d_{it})\sigma_\nu \frac{\phi((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it})}{1 - \Phi((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it})} \middle| Z_{it} \right] \geq 0.$$

The parameters to identify are $(\gamma, \beta_1, \beta_2, \sigma_{12}, \sigma_1)$ and, for each observed covariate Z_{it} such that $Z_{it} \in \mathcal{J}_{it}$, the set of moment inequalities identifying these parameters is

$$\mathbb{E}\left[r_{it} - \gamma X_{1it} - \sigma_{12} \frac{\phi(\beta_1 r_{it} + \beta_2 X_{2it})}{\Phi(\beta_1 r_{it} + \beta_2 X_{2it})} \middle| d_{it} = 1, Z_{it} \right] \geq 0,$$
$$\mathbb{E}\left[d_{it} \frac{1 - \Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)}{\Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)} - (1-d_{it}) \middle| Z_{it} \right] \geq 0,$$
$$\mathbb{E}\left[(1-d_{it}) \frac{\Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)}{1 - \Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)} - d_{it} \middle| Z_{it} \right] \geq 0,$$
$$\mathbb{E}\left[-(1-d_{it})\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big) + d_{it}\sigma_\nu \frac{\phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)}{\Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)} \middle| Z_{it} \right] \geq 0,$$
$$\mathbb{E}\left[d_{it}\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big) + (1-d_{it})\sigma_\nu \frac{\phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)}{1 - \Phi\big((\beta_1/\sigma_\nu)\gamma X_{1it} + (\beta_2/\sigma_\nu)X_{2it}\big)} \middle| Z_{it} \right] \geq 0,$$

with $\sigma_\nu^2 = (\beta_1)^2\sigma_1^2 + 1 + 2\sigma_{12}$. It is immediate to see why we need to normalize σ_2^2. From the inequalities, the only parameter that depends on σ_2^2 that we can identify is σ_ν^2. However, we need to use the information on σ_ν^2 to identify σ_1^2.

In summary, our procedure does not need to take a stand on whether any of the observed or unobserved (to the econometrician) determinants of the export revenues conditional on exporting are known to the firm whenever it is taking the export decision. Specifically, it does not take a stand on whether the variables included in the vector X_{1it} are known to the exporter whenever it is taking the decision of whether to export, and it also allows both the variance of ν_{1it} and ε_{it} to be different from zero.

A.8 Proof of Theorem 3

Lemma 13 *Suppose the assumptions in equations* (9)*,* (3)*, and* (12) *hold. Then*

$$\mathbb{E}\left[\frac{1 - \Phi(\sigma^{-1}(\eta^{-1} r_{ijt} - \beta_0 - \beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1} r_{ijt} - \beta_0 - \beta_1 dist_j))} \middle| \mathcal{W}_{ijt} \right] \geq \mathbb{E}\left[\frac{1 - \mathcal{P}_{ijt}}{\mathcal{P}_{ijt}} \middle| \mathcal{W}_{ijt} \right]. \quad (71)$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and the assumptions in equations (9) and (3) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt}, \nu_{ijt}] = 0$. Since

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

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

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

Equation (71) follows from the equality $\eta^{-1}r_{ijt} = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]+\eta^{-1}\varepsilon_{ijt}$ and the definition of $\mathcal{P}_{ijt}$ in equation (12). ■

Lemma 14 *Suppose the assumptions in equations* (9), (3), *and* (12) *hold. Then*

$$\mathbb{E}\Big[\frac{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}\Big|\mathcal{W}_{ijt}\Big] \geq \mathbb{E}\Big[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\Big|\mathcal{W}_{ijt}\Big]. \tag{72}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and the assumptions in equations (9) and (3) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{W}_{ijt}, \nu_{ijt}] = 0$. 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

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

Equation (71) follows from the equality $\eta^{-1}r_{ijt} = \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]+\eta^{-1}\varepsilon_{ijt}$ and the definition of $\mathcal{P}_{ijt}$ in equation (12). ■

Lemma 15 *Suppose $Z_{ijt} \in \mathcal{W}_{ijt}$, then*

$$\mathbb{E}\Big[\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j+\eta^{-1}\varepsilon_{ijt}))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)+\eta^{-1}\varepsilon_{ijt})}\Big|\mathcal{W}_{ijt}\Big] \geq \mathbb{E}\Big[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big], \tag{73}$$

and

$$\mathbb{E}\Big[\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j+\eta^{-1}\varepsilon_{ijt}))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)+\eta^{-1}\varepsilon_{ijt})}\Big|\mathcal{W}_{ijt}\Big] \geq \mathbb{E}\Big[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big]. \tag{74}$$

Proof: It follows from lemmas 13 and 14 and the Law of Iterated Expectations. ■

Lemma 16 *Suppose Y is a variable with support in* $(0,1)$, *then*

$$\mathbb{E}\Big[\frac{1-Y}{Y}\Big] \geq \frac{1-\mathbb{E}[Y]}{\mathbb{E}[Y]}, \tag{75}$$

and

$$\mathbb{E}\Big[\frac{Y}{1-Y}\Big] \geq \frac{\mathbb{E}[Y]}{1-\mathbb{E}[Y]}. \tag{76}$$

Proof: We can rewrite the left hand side of equation (75) as

$$\mathbb{E}\Big[\frac{1-Y}{Y}\Big] = \mathbb{E}\Big[\frac{1}{Y}-1\Big] = \mathbb{E}\Big[\frac{1}{Y}\Big]-1, \tag{77}$$

and the right hand side of equation (75) as

$$\frac{1-\mathbb{E}[Y]}{\mathbb{E}[Y]} = \frac{1}{\mathbb{E}[Y]} - 1. \tag{78}$$

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

$$\mathbb{E}\Big[\frac{1}{Y}\Big] \geq \frac{1}{\mathbb{E}[Y]}. \tag{79}$$

Equations (77), (78), and (79) imply that equation (75) holds.

Define a random variable $X = 1 - Y$ and rewrite the left hand side of equation (76) 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 (77), (78), and (79) 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 (76) holds. ■

Corollary 5 *Suppose $\mathcal{P}_{ijt}$ is defined as in equation* (12)*, then*

$$\mathbb{E}\Big[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big] \geq \frac{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}, \tag{80}$$

and

$$\mathbb{E}\Big[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big] \geq \frac{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}. \tag{81}$$

Proof: Equation (12) implies that the support of $\mathcal{P}_{ijt}$ is the interval $(0,1)$. Therefore, Lemma 16 implies that equations (80) and (81) hold. ■

Lemma 17 *Suppose $Z_{ijt} \in \mathcal{J}_{ijt}$ and define $\mathcal{P}(Z_{ijt}) = \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]$, with $\mathcal{P}_{ijt}$ defined in equation* (12)*. Then,*

$$\frac{1}{1+B_l(Z_{ijt};\theta)} \leq \mathcal{P}(Z_{ijt}) \leq \frac{B_u(Z_{ijt};\theta)}{1+B_u(Z_{ijt};\theta)}, \tag{82}$$

where

$$B_l(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}\middle|Z_{ijt}\right]. \tag{83}$$

$$B_u(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}\middle|Z_{ijt}\right], \tag{84}$$

Proof: Combining equations (73) and (80),

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

and, reordering terms, we obtain the inequality

$$\frac{1}{1+B_l(Z_{ijt};\theta)} \leq \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]. \tag{85}$$

Combining equations (74) and (81),

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

and, reordering terms, we obtain the inequality

$$\frac{B_u(Z_{ijt};\theta)}{1+B_u(Z_{ijt};\theta)} \geq \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]. \tag{86}$$

Combining the inequalities in equations (85) and (86) we obtain equation (82). ■

A.9 Bounds on counterfactual choice probabilities

We may use equations (22), (23) and (24) to define bounds on expected export probabilities in the counterfactual scenarios described in Section 2.5.

Sections 2.5 describes a counterfactual scenario in which export fixed costs become

$$f_{ijt} = \beta_0^1 + \beta_1^1 dist_j + \nu_{ijt} = 0.6\beta_0 + 0.6\beta_1 dist_j + \nu_{ijt}.$$

In this case, the export probability is defined in equation (**??**) as $\mathcal{P}_{ijt}^1$. Using expressions analogous to equations (22), (23) and (24), we may define bounds the expectation of $\mathcal{P}_{ijt}^1$ conditional on any particular value or set of values of Z_{ijt} as follows

$$\underline{\mathcal{P}}^1(Z_{ijt}) \leq \mathcal{P}^1(Z_{ijt}) \leq \overline{\mathcal{P}}^1(Z_{ijt}), \tag{87}$$

where

$$\underline{\mathcal{P}}^1(Z_{ijt}) = \min_{\gamma\in\Theta_{all}} \frac{1}{1+B_2^1(Z_{ijt};\gamma)}, \tag{88}$$

$$\overline{\mathcal{P}}^1(Z_{ijt}) = \max_{\gamma\in\Theta_{all}} \frac{B_1^1(Z_{ijt};\gamma)}{1+B_1^1(Z_{ijt};\gamma)}, \tag{89}$$

with

$$B_1^1(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}\middle|Z_{ijt}\right], \tag{90}$$

$$B_2^1(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}\middle|Z_{ijt}\right]. \tag{91}$$

Besides computing the expected probability of exporting in actual and counterfactual scenarios, we may also define bounds on the ratio of expected export probabilities in these different scenarios. Specifically, for the counterfactual scenario described in Sections 2.5, we can compute bounds for the percentage growth of the expected export probability for the subset of observations with a given value of Z_{ijt} due to a 40% reduction in the fixed costs β_0 and β_1:

$$\min_{\gamma\in\Theta_{all}} \frac{1+B_2^1(Z_{ijt};\gamma)}{\frac{B_u(Z_{ijt};\gamma)}{1+B_u(Z_{ijt};\gamma)}} \leq \frac{\mathcal{P}_{ijt}^1(Z_{ijt})}{\mathcal{P}_{ijt}(Z_{ijt})} \leq \max_{\gamma\in\Theta_{all}} \frac{\frac{B_1^1(Z_{ijt};\gamma)}{1+B_1^1(Z_{ijt};\gamma)}}{1+B_l(Z_{ijt};\gamma)}, \tag{92}$$

where $B_u(Z_{ijt};\gamma)$ and $B_l(Z_{ijt};\gamma)$ are defined in equations (84) and (83), respectively; and $B_1^1(Z_{ijt};\gamma)$ and $B_2^1(Z_{ijt};\gamma)$ are defined in equations (90) and (91), respectively.