What do Exporters Know?*

Michael J. Dickstein

Stanford University and NBER

Eduardo Morales

Princeton University and NBER

June 23, 2015

Abstract

Much of the variation in international trade volume is driven by firms' extensive margin decision to participate in export markets. To understand this decision and predict the sensitivity of export flows to changes in trade costs, we estimate a standard model of firms' export participation. In choosing whether to export, firms weigh the fixed costs of exporting against the forecasted profits from serving a foreign market. We show that the estimated parameters and counterfactual predictions from the model depend heavily on how the researcher specifies firms' expectations over these profits. We therefore develop a novel moment inequality approach with weaker assumptions on firms' expectations. Our approach introduces a new set of moment inequalities—odds-based inequalities—and applies 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 require specifying firms' information sets, our approach generates estimates of fixed export costs that are 65-85% smaller. Counterfactual reductions in fixed costs generate gains in export participation that are 30% smaller, on average, than those predicted by existing approaches.

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

*We thank Tim Bresnahan, Lorenzo Caliendo, Jan De Loecker, Dave Donaldson, Liran Einav, Alon Eizenberg, Guido Imbens, Ariel Pakes, Esteban Rossi-Hansberg, James Tybout and seminar participants at the CEPR-JIE conference on Applied Industrial Organization, Dartmouth College, LMU, the NBER ITI meeting, Pennsylvania State University, Princeton University, the Stanford/Berkeley IO Fest, Stanford University, 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 2013, approximately 300,000 US firms chose to export to foreign markets.[1] The decision of these firms to sell abroad drives much of the variation in trade volume from the US.[2] Thus, to predict how aggregate exports may change with lower trade costs, exchange rate movements, or other policy or market fluctuations, researchers need to understand firms' extensive margin decisions to participate in export markets.

A large literature in international trade focuses on modeling firms' export decisions.[3] Empirical analyses of these decisions, however, face a serious data obstacle: the decision to export depends on a firm's expectations of the profits it will earn when serving a foreign market, which the researcher rarely observes. Absent direct data on firms' expectations, researchers must impose assumptions on how firms form these expectations. For example, researchers commonly assume firms' expectations are rational and depend on a set of variables observed in the data. The precise specification of agents' information, however, can importantly influence the overall measurement, as Manski (1993, 2004), and Cunha and Heckman (2007) show in the context of evaluating the returns to schooling. In the export setting, the assumptions on expectations may affect both the estimates of the costs firms incur when exporting and predictions of how firms will respond to counterfactual changes in these trade costs.

In this paper, we first document that estimates of the parameters underlying firms' export decisions depend heavily on how researchers specify the firm's expectations. We compare the predictions of a standard model in the international trade literature (Melitz, 2003; Helpman et al., 2008) under two different sets 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 to predict their own export profits. 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 these fixed costs by 40%. Finding important differences in the predictions from the two models, we then develop a new empirical model of export participation that places fewer restrictions on firms' expectations.

Under our new approach, firms may gather different signals about their productivity relative to competitors, or about the evolution of exchange rates, trade policy, political stability abroad, and 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 about their profits conditional on exporting. The

[1] Department of Commerce (2015)

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

[3] See for example Das et al. (2007), 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).

researcher must observe this subset, but need not observe any remaining variables that affect the firm's expectations. The set of unobserved variables may vary flexibly across firms, markets, and years. In contrast, standard estimation approaches require the researcher to fully specify and observe all variables in exporters' information sets. The trade-off from specifying only a subset of the firm's information is that we can only partially identify the true parameters of interest. To do so, we develop a new type of moment inequality, which we label the odds-based inequality, and combine it with inequalities based on revealed preference.[4] Using these inequalities, our empirical burden is twofold. First, we must show that placing fewer assumptions on expectations matters both for the estimates of the parameters of the exporter's problem and for the predictions of export flows under counterfactual trade policy. Second, our robust approach must generate bounds on the model's parameters and on predicted exports that are small enough to be informative.

We perform our empirical analysis in the context of a standard partial equilibrium, two period model of export participation.[5] 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 firms' profit forecasts. Specifically, 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 approach, developed in Willis and Rosen (1979), Manski (1991) and Ahn and Manski (1993), in which we assume that firms' expectations are rational and specify that firms form their expectations using only three variables: distance to the export market, aggregate exports from Chile to that market in the prior year, and the firm's own productivity from the prior year. The estimated fixed costs of exporting under this limited information 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 limited information approach reflects a possible bias in the estimation. Both the limited information 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

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

[5] 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 model à la Das et al. (2007).

more information or less— to predict their potential export profits, the estimates of the model parameters will generally be biased. Thus, our second key step is to employ our new types of moments inequalities to partially identify the exporter's fixed costs under weaker assumptions. Here, we again assume that firms know the distance to the export market, the aggregate exports to that market in the prior year, and their own productivity from the prior year. However, unlike the limited information approach described earlier, the inequalities we define do not restrict firms to use only these three variables when forecasting their potential export profits. We require only that firms know at least these variables. We chose this set of three variables in our specification because they are contained either in firms' own balance sheets or in official government statistics. It seems reasonable, therefore, to assume all firms might know at least these variables. We can, however, go further and test the null hypothesis that the potential exporters' information sets satisfy this minimal requirement. Specifically, conditional on the model, we use the specification test suggested in Andrews and Soares (2010) to test our assumption that these three variables are in the firm's information set.[6]

Under the traditional maximum likelihood methods, we estimate the fixed costs for exports from Chile to Argentina, for example, to equal \$594,000 or \$894,000 in the chemicals sector, depending on the specification of the information set. Using our inequalities approach, we find much lower 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 *at least* distance, lagged productivity, 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 and data, but placing one additional variable in the firm's information set. In this alternative, we assume the firm also knows the productivity of other firms that export 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. Similarly, we can also reject the assumption of perfect foresight at any generally used significance level.

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 approaches that require explicit assumptions on the exact content of each firm's information set, we find that the results differ substantially with these assumptions. For example, compared to predictions under perfect foresight, the predicted export participation under the alternative procedure that assumes firms know only distance, lagged aggregate

[6]Alternative specification tests for partially identified models defined by moment inequalities have been provided in Romano and Shaikh (2008), Andrews and Guggenberger (2009), and Bugni et al. (2015).

exports and lagged productivity is 3% and 20% higher for Argentina and Japan and 12% lower for the United States, in the chemicals sector. Comparing the predictions from these two models to those computed using our 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 manufacturing 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., 2015) 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 (Manski, 2004), our approach can recover bounds on the economic primitives of the agent's problem without imposing strong assumptions on agents' expectations.

We proceed in this paper by first describing our model of firm exports in Section 2, building up to an expression for firms' export participation decisions. 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 the researcher to have full knowledge 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 and with only partial knowledge of agents' information sets. 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 a reduction in fixed export costs. Section 8 concludes.

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 $i = 1, \ldots, N_t$.[7] We index the potential destination countries by $j = 1, \ldots, J$.

We model firms' export decisions using a two-period model. In the first period, firms choose

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

the set of countries to which they wish to export. To participate in 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:

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

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. Thus, 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.[8] We assume that

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

where σ_ν^2 measures the unobserved heterogeneity in fixed export costs across firms, countries and time periods.[9]

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_{jt}; (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}; and, (c) an indicator for whether each of the active firms exports to j at t, d_{ijt}. Specifically,

[8] 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.

[9] None of the results presented in this paper depends 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 is log-concave.

the revenue that firm i would obtain in country j at period t conditional on entering is

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

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 that choose not to export.

2.4 Decision to export

Once we account for the fixed costs of exporting, the remaining 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 all of the information firm i knows about the determinants of π_{ijt} and f_{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. We assume firms' expectations are rational, and thus $\mathbb{E}[\cdot]$ denotes the expectation with respect to the data generating process. Assuming further that all determinants of fixed export costs are known to firms when they decide whether to export—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 again r_{ijt} is export revenue conditional on entry and f_{ijt} is the fixed export cost.[10] We define an agent's expectational error as ε_{ijt}, where $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$. Under our assumptions, it holds that

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

That is, 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

[10]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. The key restriction is that all firms must face the same elasticity of demand in every country and time period.

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 $Y_{jt} \in \mathcal{W}_{ijt}$.[11]

We introduce $\mathcal{W}_{ijt}$ to highlight a key assumption needed for the moment inequality approach we introduce in Section 4.2. Recall, ν_{ijt} is the component of fixed export costs that firms know but researchers do not observe. In the inequality approach, ν_{ijt} must be independent of all other determinants of the export choice, d_{ijt}:

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

Thus, although $\nu_{ijt} \in \mathcal{J}_{ijt}$, it is independent of the elements of the information set used to form expectations. As Section 4.1 shows, this independence assumption is also imposed in standard discrete choice models (e.g. probit or logit).

For simplicity of notation going forward and without loss of generality, we will assume that $dist_j \in \mathcal{W}_{ijt}$. Given equations (3), (10), and (11), we can write the probability that i exports to j conditional on $\mathcal{W}_{ijt}$ as

$$\mathcal{P}(d_{ijt} = 1|\mathcal{W}_{ijt}) = \int_{\nu} \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu \geq 0\}\phi(\nu)d\nu \tag{12}$$

$$= \Phi\big(\sigma^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big), \tag{13}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the standard normal probability density function and cumulative distribution function. Equation (13) indicates that, after integrating over the unobserved heterogeneity ν_{ijt}, we can write the probability that a firm i exports to a country j at period t as a probit model whose index depends on firm i's expectations of the gross profits it will earn in j at t upon entry. Equations (6) and (13) capture all implications of the export model that will be important both for estimation and computing counterfactuals.

As is clear from equation (13), even if we were to observe firms' actual expectations, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, data on export choices alone do 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

[11] For now, we impose no restriction on the content of $\mathcal{W}_{ijt}$. In Section 4, we describe the requirements alternative empirical models impose on $\mathcal{W}_{ijt}$.

elasticity η. In our estimation, we set $\eta = 5$.[12] For simplicity of notation going forward, we use θ to denote the remaining parameter vector $(\sigma, \beta_0, \beta_1)$.[13]

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 the policy affects are the set of export participation dummies, $\{d_{ijt}, i = 1, \dots, N_t\}$ and, through them, the total exports from h to j, R_{jt}. We show in Section 6 how different assumptions on the information firms use to form expectations, $\mathcal{W}_{ijt}$, lead to different predictions of the number of exporters and total exports under the counterfactual policy.

With our counterfactual policy, we capture in a stylized way the effect of export promotion programs on the fixed costs of exporting and ultimately on export participation. Such programs are common. Van Biesebroeck et al. (2015) discuss Canadian Trade Commissioner Service measures that lower entry barriers to increase export participation. Volpe Martincus and Carballo (2008) and Volpe Martincus et al. (2010) document similar measures in Peru and Uruguay targeting the extensive margin decision of firms to export. According to Lederman et al. (2009), typical programs include country image building (advertising, promotional events, advocacy), export support services (exporter training, technical assistance on logistics, customs, and packaging), and marketing (trade fairs, follow-up services offered by representatives abroad). It is hard to quantify the precise savings in fixed export costs that these services imply; our choice of a 40% reduction in fixed costs in the counterfactual is meant to illustrate one possible level. We then emphasize how the assumptions that researchers impose on the content of the information set $\mathcal{W}_{ijt}$ will affect the policy predictions.

3 Data

Our data come from two separate sources. The first is an extract of the Chilean customs database, which covers the universe of exports of Chilean firms from 1995 to 2005. The second is the Chilean Annual Industrial Survey (*Encuesta Nacional Industrial Anual*, or ENIA), which

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

[13] 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 the sensitivity of our estimates 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 does it affect the comparison across counterfactuals.

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.[14]

The firms in our dataset operate in one of two sectors: the manufacture of chemicals and food products.[15] 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. We use this set of destinations in our estimation 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.[16] 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.

In our empirical exercises, we illustrate our findings using three destination countries—Argentina, Japan, and the United States. 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

[14] 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.

[15] 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.

[16] The dollar values we report in this paper are US dollars in year 2000 terms.

Table 1: Summary Statistics

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

Notes: All variables (except "share of exporters") are reported in millions of USD in year 2000 terms.

food sector, the total volume of exports and number of exporting firms is lower for Argentina ($184 million and 85 unique exporters) but much larger for the United States ($1,931 million and 122 unique exporters) and Japan ($2,656 million and 126 unique exporters). The average per firm export volume in the food sector to these three countries 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. Finally, we obtain information on the distance from Chile to each destination market from CEPII.[17]

4 Empirical Approach

In the model we describe in Section 2, 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

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

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 may know only some of these variables when deciding whether to export to j at t. They therefore form expectations of potential export revenues using only their information set, $\mathcal{W}_{ijt}$. In the theoretical model, we did not impose assumptions on the content of the information set. In estimation, however, we need to place restrictions on $\mathcal{W}_{ijt}$ to identify the parameter vector θ and perform counterfactuals, as Manski (1993) demonstrates.[18]

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, exporters face no uncertainty; when deciding whether to export, firms have all the information they need to predict perfectly the gross profits they'll earn upon entry.[19] We denote the information set imposed under this approach as Z^1_{ijt}. Here, $\mathcal{W}_{ijt} = Z^1_{ijt} = r_{ijt}$, where r_{ijt} is the export revenue the firm would earn upon entry.

For most firms and in most destination countries, the set Z^1_{ijt} is likely to be strictly larger than firms' true information sets. That is, at the time firms decide whether to serve a foreign market, they often lack perfect knowledge of the revenue they'll earn upon entry. Thus, we specify a second empirical model in which potential exporters forecast their export revenues in every foreign market using only information on three variables: (1) their own lagged domestic sales, which serves as a proxy for productivity; (2) lagged aggregate exports to the destination country j; and (3) distance from the home country to j. We denote this information set with a superscript 2, $\mathcal{W}_{ijt} = Z^2_{ijt} = (r_{iht-1}, R_{it-1}, dist_j)$. This information set is likely to be strictly smaller than the actual information set firms possess when deciding to export. In addition, this model assumes all potential exporters base their entry decision on the same set of covariates. It does not permit firms to vary in the types of information they use.

Under both of these approaches, the researcher assumes that exporters' true information sets $\mathcal{W}_{ijt}$ correspond exactly to a vector of covariates Z_{ijt} that the researcher observes. Section 4.1 shows how to estimate the parameter vector θ under these assumptions. Ideally, one would like to estimate θ and perform counterfactuals without imposing such strong assumptions on firms' information sets. Thus, in our third approach, described in Section 4.2, we propose a moment inequality estimator that can handle settings in which the econometrician observes only a *subset* of the elements contained in firms' true information sets. That is, instead of

[18] Specifically, Manski (1993) shows that different assumptions on $\mathcal{W}_{ijt}$ might still generate identical likelihood functions for a given set of covariates. Under these different assumptions, each reduced form parameter has a different structural interpretation. Thus, one cannot use common goodness-of-fit measures to discriminate among models that impose different assumptions on agents' information sets.

[19] 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.

assuming $\mathcal{W}_{ijt} = Z^2_{ijt}$, as in the second model above, we require only that $Z^2_{ijt} \subset \mathcal{W}_{ijt}$. The remaining elements in $\mathcal{W}_{ijt}$ need not be observed by the researcher. Those unobservable elements of firms' information sets can vary flexibly by firm and by export market.

Finally, before proceeding to estimation, we comment on the availability of data on export revenue. To carry out any of the three procedures above, the researcher needs a direct measure of r_{ijt}, the potential export revenue for firm i in country j and year t. We observe revenue for periods and markets in which i chooses to export. We have no direct measure of r_{ijt} when the firm chooses not to export.

We rely on the assumptions of the model described in Section 2 to define a perfect proxy for r_{ijt} using the following information: (a) the domestic revenues of firm i in year t, r_{iht}; (b) the aggregate export flows from the home country h to the destination country j in year t, R_{jt}; (c) a vector of indicators for whether each of the active firms in year t exports to j at t, $\{d_{ijt}; i = 1, \ldots, N_t\}$. We observe these variables both for firms that choose to export and for those that do not. In Sections 4.1 and 4.2, we use the expression in equation (6) to generate an observed measure of r_{ijt} for every firm i, country j and year t.

As an alternative to the theoretical model, one can instead use the observed data on export revenue, r_{ijt}, observed for those firms that choose to export in market j and time t. This procedure would involve estimating a new equation for revenue as a function of observable covariates, such as the firm's domestic sales. One could then predict the potential revenue for firms that choose not to export as a function of the estimated parameters and the observables of the non-exporting firms. However, this approach will generate a selection problem: those firms that choose to export likely have a larger unobserved determinant of export revenue. How to fix this problem depends on the researcher's assumptions about the firm's information. When the researcher assumes she observes all variables that enter the firm's information set—i.e. $\mathcal{W}_{ijt}$ is equal to a vector Z_{ijt}—Heckman (1979) provides a procedure to recover the parameters of both the revenue equation and the equation for the entry decision. If the researcher instead assumes only that the vector of observed covariates Z_{ijt} is a subset of exporters' true information sets $\mathcal{W}_{ijt}$, we show in Appendix A.7 how to extend the estimation procedure in Section 4.2 to obtain bounds for the parameters in both equations.

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 unknown parameter γ that maximizes a standard log-likelihood function

$$\mathcal{L}(\gamma|d, Z, dist) =$$
$$\mathbb{E}\Big[\sum_{j,t} d_{ijt} \log\big(\mathcal{P}(d_{jt} = 1|Z_{ijt}, dist_j)\big) + (1 - d_{ijt}) \log\big(\mathcal{P}(d_{jt} = 0|Z_{ijt}, dist_j)\big)\Big], \tag{14}$$

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

$$\mathcal{P}(d_{jt} = 1|Z_{ijt}, dist_j) = \Phi\big(\gamma_2^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|Z_{ijt}] - \gamma_0 - \gamma_1 dist_j\big)\big) \tag{15}$$

The vector $\gamma = (\gamma_0, \gamma_1, \gamma_2)$ denotes an unknown parameter vector whose true value is $\theta = (\beta_0, \beta_1, \sigma)$.[20]

The key assumption underlying this procedure is that the researcher correctly specifies the agent's information set. When the researcher's choice of information set, Z_{ijt}, is incorrect—that is, when $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] \neq \mathbb{E}[r_{ijt}|Z_{ijt}]$—then the estimator of θ under this procedure will be biased. We denote the difference between the two revenue projections as ξ_{ijt}: $\mathbb{E}[r_{ijt}|Z_{ijt}] = \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] + \xi_{ijt}$. In this case, one can identify θ as the parameter that maximizes the likelihood function in equation (14) but with

$$\mathcal{P}(d_{ijt} = 1|Z_{ijt}, dist_j) =$$
$$\int_{\tilde{\nu}} \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|Z_{ijt}] - \gamma_0 - \gamma_1 dist_j - \tilde{\nu} \geq 0\} f(\tilde{\nu}|Z_{ijt}, dist_j) d\tilde{\nu}, \tag{16}$$

where $\tilde{\nu} = \eta^{-1}\xi + \nu$ and $f(\tilde{\nu}|Z, dist)$ denotes the density of $\eta^{-1}\xi + \nu$ conditional on Z and $dist$. When comparing equation (15) to equation (16), 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 is true only when $\xi_{ijt} = 0$ for every firm i, destination j and period t. 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 $dist$.

Biased estimates of the structural parameter of interest θ will translate into biased estimates of fixed export costs and into incorrect predictions of the effect of the counterfactual changes in the environment. We show in Section 5 that wrongly assuming a specific information sets generates an upward bias in the estimated fixed export costs. We provide some intuition for the direction of the bias here.

Take, for example, the case in which researchers assume perfect foresight. Under perfect foresight, $r_{ijt} = \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$. Yatchew and Griliches (1985) provide an analytical form

[20] In order to use the expressions in equations (14) and (15) to estimate the θ, 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.

for the bias under a particular distribution for the error between the true expectations and the researcher's assumption. Specifically, if firms' true expectations are normally distributed, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] \sim \mathbb{N}(0, \sigma_e^2)$, and the expectational error is also normally distributed, $\xi_{ijt}|(\mathcal{W}_{ijt}, \nu_{ij}) \sim \mathbb{N}(0, \sigma_\xi^2)$, there is an upward bias in the estimates of the fixed costs parameters β_0, β_1 and σ. 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. That is, the worse the researcher's proxy for the true expectations, the greater the bias.

When either firms' true expectations or the expectational error are not normally distributed, there is no analytic expression for the bias of the maximum likelihood estimator of θ. However, we present simulations in Appendix A.2 to show that the upward bias appears fairly general. 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.[21]

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 firms' accounting statements, and the latter is 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 that the agent's information set is exactly identical to the vector of observed covariates $Z_{ijt}^2 = (r_{iht-1}, R_{it-1}, dist_j)$, for example, the assumption that Z_{ijt}^2 is contained in every potential exporters' information set may well be

[21] The intuition for upward bias in the maximum likelihood estimates of β_0, β_1, and σ caused by wrongly assuming perfect foresight shares the same basis as the well-known attenuation bias affecting OLS estimates in linear models when a covariate is affected by classical measurement error (see page 73 in Wooldridge (2002)). Rational expectations implies that firms' expectational errors are mean independent of their unobserved true expectation and, therefore, correlated with the ex-post realization of the variable whose expectation affects firms' decisions. In our setting, this implies that the expectational error is correlated with the realized 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 (15) 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. 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.

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.[22]

In Section 5, we further show 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 test our main specification that presumes $Z^2_{ijt} \in \mathcal{W}_{ijt}$. We also test two models that impose different informational assumptions. In the first, we add the productivity of other exporters to Z_{ijt}, and then test whether firms in fact know their competitors' productivities when deciding whether to export. In the second, we test the perfect foresight assumption, under which $r_{ijt} \in \mathcal{W}_{ijt}$.

4.2.1 Odds-based moment inequalities

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

$$\mathcal{M}(Z_{ijt};\gamma) = \mathbb{E}\left[\begin{array}{c} m_l(d_{ijt}, r_{ijt}, dist_j;\gamma) \\ m_u(d_{ijt}, r_{ijt}, dist_j;\gamma) \end{array} \middle| Z_{ijt}\right] \geq 0, \tag{17}$$

where the two moment functions are defined as

$$m_l(\cdot) = d_{ijt}\frac{1-\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt}-\gamma_0-\gamma_1 dist_j\big)\big)}{\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt}-\gamma_0-\gamma_1 dist_j\big)\big)} - (1-d_{ijt}), \tag{18a}$$

$$m_u(\cdot) = (1-d_{ijt})\frac{\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt}-\gamma_0-\gamma_1 dist_j\big)\big)}{1-\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt}-\gamma_0-\gamma_1 dist_j\big)\big)} - d_{ijt}. \tag{18b}$$

We denote the 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

[22] Whether the bounds defined by the moment inequalities in Sections 4.2.1 and 4.2.2 are sharp is left for future research. However, as the results in Section 5 show, in our empirical application, they generate bounds that are small enough to be informative 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 on export participation and trade volume from counterfactual changes in the economic environment.

main property of the inequalities defined in equations (17), (18a) and (18b):

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. Briefly, the inequalities arise because we do not observe the true expectation, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$. If we did, we could simply take the first order condition of the likelihood function conditional on $\mathcal{W}_{ijt}$. Its expectation should be equal to zero at the true value of the parameter vector:

$$\mathbb{E}\left[\begin{array}{c|c} d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j-\eta^{-1}\varepsilon_{ijt}))}{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j-\eta^{-1}\varepsilon_{ijt}))} - (1-d_{ijt}) & \\ (1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j-\eta^{-1}\varepsilon_{ijt}))}{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j-\eta^{-1}\varepsilon_{ijt}))} - d_{ijt} & \end{array} Z_{ijt}\right] = 0, \tag{19}$$

where here we replace $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ in the score function with the equivalent expression, $r_{ijt} - \varepsilon_{ijt}$. The score function looks identical to the inequalities in (18a) and (18b), but with the ε_{ijt} term included, where ε_{ijt} is the expectational error firm i makes when forecasting the potential revenue from exporting to j at t. In the proof in Appendix A.4, we show that equation (17) is weakly larger than equation (19) and thus would be weakly larger than zero when evaluated at the true value of the parameter vector. This result is a direct application of Jensen's inequality. We rely on two properties of the distributions of the errors to apply Jensen's inequality. First, the expectational error ε_{ijt} has a mean equal to zero conditional on the vector Z_{ijt}. This follows from the assumptions that firms have rational expectations and $Z_{ijt} \subset \mathcal{W}_{ijt}$. Second, both $1 - \Phi(\cdot)/\Phi(\cdot)$ and $\Phi(\cdot)/(1 - \Phi(\cdot))$ are globally convex.[23]

Even though both moment functions in equations (18a) and (18b) 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 (18a) 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 (18b) 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

[23]The assumption of normality of the structural error term is not necessary for the existence of odds-based inequalities. As long as the distribution of the structural error ν is log-concave, inequalities analogous to those in equation (18), 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.

variance of the expectational error. That is, as the variance of the difference between firms' expected revenues $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and the ex post realization of such revenues r_{ijt} grows, so does the size of Θ .

4.2.2 Revealed preference moment inequalities

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

$$\mathcal{M}^r(Z_{ijt};\gamma) = \mathbb{E}\left[\begin{array}{c} m_l^r(d_{ijt}, r_{ijt}, dist_j;\gamma) \\ m_u^r(d_{ijt}, r_{ijt}, dist_j;\gamma) \end{array} \middle| Z_{ijt}\right] \geq 0, \tag{20}$$

where the two moment functions are defined as

$$m_l^r(\cdot) = -(1-d_{ijt})\big(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j\big) + d_{ijt}\gamma_2 \frac{\phi\big(\gamma_2^{-1}(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}{\Phi\big(\gamma_2^{-1}(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}, \tag{21a}$$

$$m_u^r(\cdot) = d_{ijt}\big(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j\big) + (1-d_{ijt})\gamma_2 \frac{\phi\big(\gamma_2^{-1}(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}{1-\Phi\big(\gamma_2^{-1}(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}. \tag{21b}$$

We again denote the set of all possible values of the unknown parameter vector γ as Γ, and 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 (20), (21a) and (21b):

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 revealed preference inequalities are consistent with the true value of the parameter vector, θ.[24] In general, the set of parameter values that satisfies both the 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 our empirical application in Section 5, the set of parameter values that are consistent both with the odds-based and revealed preference inequalities is small enough to allow us to draw economically meaningful conclusions.

Heuristically, the two moment functions in equations (21a) and (21b) are derived using standard revealed preference arguments. We focus our discussion on moment function (21b); the intuition behind the derivation of moment (21a) 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

[24] The assumption of normality of the structural error term is sufficient but not necessary for the existence of revealed preference inequalities analogous to those in equations (20), (21a) and (21b) 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 that also satisfy Theorem 2. For example, in addition to the normal distribution, the type I extreme value distribution also satisfies this property.

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

$$d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + S_{ijt} \geq 0, \tag{22}$$

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 by the researcher, including the term ν_{ijt} in the model described in Section 2.[25] We cannot directly use the inequality in equation (22) 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 (22) becomes weaker if we introduce the observed ex-post profits, r_{ijt}, in the place of the unobserved expectations $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and take the expectation of the resulting expression. Similar to the odds-based inequalities, we then apply Jensen's inequality and conclude that the inequalities in equations (20) and (21b) hold at the true value of the parameter vector. Again, to apply Jensen's inequality, we rely on two conditions. First, the difference between the unobserved true firms' expectations and the realized revenues has mean equal to zero conditional on the vector $\mathcal{W}_{ijt}$. Second, $\phi(\cdot)/\Phi(\cdot)$ and $\phi(\cdot)/(1-\Phi(\cdot))$ are globally convex .

The moment functions in equations (21a) and (21b) follow the revealed preference inequalities introduced in Pakes (2010) and Pakes et al. (2015), and previously applied in Eizenberg (2014) and Morales et al. (2015). We apply the inequalities to a setting with a specific error structure not present in previous empirical examples. The novelty of the revealed preference moment inequalities introduced in equations (20) and (21) is that they allow for structural errors ν_{ijt} that may vary across i, j and t and that might have unbounded support. The cost of allowing for this flexibility is that we need to assume the distribution of ν_{ijt}, up to a scale parameter. In addition, the inequalities we define only apply to binary choice problems.[26]

The distribution of ν_{ijt} affects the functional form of the selection correction term, S_{ijt}. In our empirical application, we find $\sigma > 0$. Therefore, accounting for the selection correction term S_{ijt} in our empirical application is important. Given that $S_{ijt} \geq 0$ whenever $\sigma > 0$, if we had generated revealed preference inequalities without the term S_{ijt}, we would have obtained weakly smaller identified sets than those found using the revealed preference inequalities in equations (20) and (21).

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

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

[26] It is an open question whether the odds-based and revealed preference inequalities introduced in Sections 4.2.1 and 4.2.2 could be extended to discrete choice problems in which the choice set contains three or more elements. We leave this for future research.

4.2.3 Combining inequalities for estimation

For our estimation approach, we combine the odds-based and revealed preference moment inequalities described in equations (17) and (20). 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 revealed preference inequalities do not have any additional identification power beyond that of the odds-based inequalities. However, in all other settings, the revealed preference moments can provide additional identifying power beyond that provided by the odds-based inequalities.[27]

The set of inequalities we define in equations (17) and (20) 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.

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, in addition to measuring export fixed costs, another key motivation for estimating export entry models is to predict export participation and trade volume in counterfactual environments. In this section, we show that one can perform these counterfactual exercises without imposing additional assumptions beyond those needed to define the odds-based and revealed preference inequalities.

We derive bounds on the probability that a firm exports to each market. 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, we only observe a subset Z_{ijt} of the variables in the true information set firms use to predict export revenues. Thus, we cannot compute firms' unobserved expectations, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, exactly and therefore cannot compute the export probabilities in equation (13). 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

[27] How the identified sets defined by 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 revealed preference inequalities have additional identification power beyond that of the odds-based inequalities.

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 revealed preference and odds-based inequalities.

Theorem 3 *Suppose $Z_{ijt} \in \mathcal{W}_{ijt}$ and define $\mathcal{P}(Z_{ijt}) = \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]$, with $\mathcal{P}_{ijt}$ defined in equation* (13)*. Then,*

$$\mathcal{P}_l(Z_{ijt}) \leq \mathcal{P}(Z_{ijt}) \leq \mathcal{P}_u(Z_{ijt}), \tag{23}$$

where

$$\mathcal{P}_l(Z_{ijt}) = \min_{\gamma \in \Theta_{all}} \frac{1}{1 + B_l(Z_{ijt};\gamma)}, \tag{24}$$

$$\mathcal{P}_u(Z_{ijt}) = \max_{\gamma \in \Theta_{all}} \frac{B_u(Z_{ijt};\gamma)}{1 + B_u(Z_{ijt};\gamma)}. \tag{25}$$

and

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

$$B_u(Z_{ijt};\gamma) = \mathbb{E}\left[\frac{\Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)}{1 - \Phi\big(\gamma_2^{-1}\big(\eta^{-1}r_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)}\bigg| Z_{ijt}\right]. \tag{27}$$

The proof of Theorem 3 is in the Appendix A.8. We highlight two features of these bounds. First, as described above, even if we were to know the true value of the parameter vector, θ, then would still find bounds on the probability of exporting, $\mathcal{P}(Z_{ijt})$:

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

Second, equation (23) defines bounds on export probabilities conditional on a particular value of the instrument vector Z_{ijt}. However, using equation (23) we may also 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

$$\sum_{ijt} \mathcal{P}_l(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\} \leq \sum_{ijt} \mathcal{P}(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\} \leq \sum_{ijt} \mathcal{P}_u(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\}. \tag{28}$$

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 (28) will yield bounds on the expected number of exporters to country j^* in year t^*. In Section 5.1, we use the bounds in equation (28) to test the fit of the model for different countries and years. We show in Appendix A.9 how to use equation (28) 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 in Section 4.1 and 4.2. 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 realized revenues on the set of observable covariates that we assume compose a firm's information set. In practice, we include the three variables described as the vector Z^2_{ijt} in Section 4.1: (a) the total aggregate exports from Chile to the destination country in the prior year, (b) the distance from Chile to the destination country, j, and (c) the firm's own domestic sales from the previous year. Finally, third, we carry out our moment inequality approach. For comparison purposes, we assume the firm knows the same three variables as in the two-step approach. However, unlike the two-step approach, the inequalities allow additional unobserved variables to enter the firm's true information set, 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 explore the robustness of our moment inequality estimator.

5.1 Estimates and predicted exports

In Table 2, we report the estimates and the confidence regions for the parameters of our entry model. 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 five. 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 fixed export costs than does our moment inequality approach. For example, consider the coefficient on the distance variable in models estimated using data from the chemicals sector. Under the moment inequality approach, the set of parameter values that satisfy the moments imply an added cost of \$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 equal \$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

Table 2: Parameter estimates

Estimator	Chemicals			Food		
	σ	β_0	β_1	σ	β_0	β_1
Perfect Foresight (MLE)	1074.0 (46.7)	760.9 (36.7)	1180.1 (53.2)	2039.2 (71.6)	2675.4 (96.1)	266.4 (23.6)
Limited Info. (MLE)	701.9 (24.3)	502.2 (20.2)	812.3 (30.0)	1567.8 (48.4)	2087.2 (65.7)	198.1 (18.1)
Moment Inequality	[311.7, 341.0] (178.6, 465.1)	[218.3, 245.8] (121.1, 316.3)	[428.3, 479.0] (237.1, 651.2)	[247.7, 264.9] (219.2, 307.7)	[332.9, 380.1] (283.8, 473.8)	[143.7, 173.4] (108.6, 191.9)

Notes: All variables are reported in thousands of year 2000 USD. For the two ML estimators, standard errors are reported in parentheses. For the moment inequality estimates, extreme points of the identified set are reported in square brackets and extreme points of the confidence set are reported in parentheses. Confidence sets are computed using Andrews and Soares (2010).

for 29% of total exports of the Chilean chemicals sector and 56% of the food sector in the sample period. In addition, these three nations span a wide range of possible distances to Chile and thus provide a good illustration of how export participation and volume relate to 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. 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 generate an upward bias in the estimates of the fixed costs.

Table 3: Export fixed costs

Estimator	Chemicals Argentina	 Japan	 United States	Food Argentina	 Japan	 United States
Perfect Foresight (MLE)	894.0 (242.3)	2796.2 (708.6)	1736.9 (438.7)	2705.4 (543.5)	3134.9 (612.1)	2895.7 (566.6)
Limited Info. (MLE)	593.8 (109.2)	1903.1 (315.9)	1174.0 (190.9)	2109.5 (328.6)	2428.8 (363.7)	2251.0 (336.9)
Moment Inequality	[270.0, 298.2] (163.0, 385.8)	[977.6, 1062.0] (627.5, 1420.8)	[592.6, 632.0] (408.8, 836.3)	[352.0, 397.5] (304.2, 489.7)	[606.9, 645.1] (552.2, 730.7)	[472.7, 507.2] (430.0, 590.3)

Notes: All variables are reported in thousands of year 2000 USD. For the two ML estimators, standard errors are reported in parentheses. For the moment inequality estimates, extreme points of the identified set are reported in square brackets and extreme points of the confidence set are reported in parentheses. Confidence sets are computed using the procedure described in Andrews and Soares (2010).

Finally, in Table 4, we report the observed level of export participation in our three comparison countries in the final year of our data, 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 in 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 that at least 78 would do so.

6 Counterfactuals

Beyond estimating the level of the fixed costs of exporting, we can use our estimates to conduct counterfactual analyses. As introduced in Section 2.5, we simulate the effect of lowering the fixed costs of exporting 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 the models that require the researcher to specify the exact set of covariates included in firms' information sets. We report our counterfactual predictions in Table 5.

Table 4: Goodness of fit

Estimator	Chemicals Argentina	Chemicals Japan	Chemicals United States	Food Argentina	Food Japan	Food United States
Observed	46	5	24	22	52	48
Perfect Foresight (MLE)	41.00 (0.44)	2.41 (0.60)	15.69 (1.52)	21.96 (0.20)	25.99 (2.29)	39.13 (2.73)
Limited Info. (MLE)	40.13 (0.40)	1.84 (0.52)	19.04 (1.88)	21.65 (0.23)	34.00 (2.92)	40.53 (2.92)
Moment Inequality	[43.47, 45.53] (31.44, 52.27)	[5.89, 5.99] (3.84, 11.99)	[22.16, 22.64] (15.00, 30.86)	[24.30, 29.41] (19.43, 32.70)	[47.15, 49.35] (42.79, 52.94)	[78.13, 81.30] (72.74, 84.69)

Notes: All variables are reported in thousands of year 2000 USD. For the two ML estimators, standard errors are reported in parentheses. For the moment inequality estimates, all points in the identified set and confidence sets are used to compute the counterfactual changes. The corresponding minimum and maximum predicted values obtained using all parameter values contained in the identified set are reported in square brackets; the minimum and maximum values obtained using all parameter values contained in the confidence set are reported in parentheses.

We focus first on the predicted change in 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 researchers observe firms' information sets, the predictions for how a policy will affect exports differs depending on the assumptions imposed on the content of these information sets. Specifically, in the chemicals sector, we find the predicted change in export participation to be 3% and 19.6% higher under the two-step approach in Argentina and Japan, but is 11.8% lower under the two-step approach in the United States.[28]

The moment inequality approach, which imposes weaker assumptions on the content of firms' information sets, produces smaller predictions of the effect of the policy relative to either maximum likelihood approach. Specifically, the moment inequality estimator predicts growth in export participation in the United States that is between 23.9 and 25.2% lower than the perfect foresight prediction in the chemicals sector and between 51.7 and 53.7% lower in the food sector. Relative to the estimate from the two-step approach, the predicted number of exporters to the United States is between 13.8 and 15.3% lower in the chemicals sector and between 50.6 and 52.7% lower in the food sector. These differences are large and likely to be important for the evaluation of any export promotion policy.

The differences across the predictions generated by the three alternative models are much larger for countries that are far away from Chile (e.g. Japan and the United States) than for countries that are close to Chile (e.g. Argentina). This distinction across countries in the

[28]Even though these differences are sizable, the counterfactual predictions in Table 5 appear more robust to alternative specifications of the exporters' information sets than do the parameter estimates in Table 3. The parameter estimates generated under the two different ML estimation approaches differed by approximately 30%.

Table 5: Impact of 40% Reduction in Fixed Costs

Estimator	Chemicals Argentina	Chemicals Japan	Chemicals United States	Food Argentina	Food Japan	Food United States
		% Change in Number of Exporters				
Perfect Foresight (MLE)	50.3 (0.4)	578.7 (129.9)	157.5 (14.0)	127.9 (0.6)	118.4 (9.8)	80.7 (6.0)
Limited Info. (MLE)	51.9 (0.4)	692.3 (184.9)	139.0 (13.4)	130.8 (0.8)	94.3 (8.5)	79.0 (6.2)
Moment Inequality	[47.1, 50.5] (42.5, 76.3)	[239.2, 239.5] (146.4, 298.4)	[117.8, 119.8] (93.3, 139.3)	[104.4, 124.1] (94.6, 150.6)	[57.5, 60.1] (54.5, 66.3)	[37.3, 39.0] (35.9, 41.9)
		% Change in Volume of Exports				
Perfect Foresight (MLE)	40.5 (2.8)	163.4 (90.8)	53.1 (26.3)	114.3 (4.0)	35.2 (9.6)	18.8 (5.0)
Limited Info. (MLE)	42.9 (2.9)	192.3 (172.3)	40.5 (17.3)	112.4 (5.5)	22.9 (6.3)	18.9 (5.3)
Moment Inequality	[22.0, 23.4] (16.3, 27.0)	[50.9, 51.1] (36.8, 56.2)	[26.9, 27.3] (18.4, 34.0)	[36.6, 43.7] (31.1, 57.0)	[8.5, 9.2] (7.6, 10.8)	[3.8, 4.1] (3.5, 4.6)

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

sensitivity of the predictions reflects two factors: (a) biases in the estimated contribution of distance to total fixed costs matter less for predictions in countries close to Chile, and (b) both the level of uncertainty in export revenues and the heterogeneity in this uncertainty across exporters are likely to be smaller in countries that are very similar to the exporter's domestic markets.

The estimates reveal substantive economic effects from a hypothetical export promotion measure that reduces fixed export costs.[29] 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 that 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

[29] See the references in Section 2.5 for examples of this kind of policy.

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. Our results also illustrate how researchers' assumptions on exporters' expectations translate into distinct measurement in policy counterfactuals. The relative precision in the moment inequality estimates also illustrates that researchers can rely on weaker assumptions on firms' information sets and nonetheless provide policymakers meaningful counterfactual predictions.

7 What do Exporters Know?

In addition to estimating the fixed costs of exporting and conducting counterfactuals, we use the moment inequalities introduced in Sections 4.2.1 and 4.2.2 to test the content of potential exporters' information sets. Specifically, conditional on the structure imposed by equations (6) and (13), we can use the specification test suggested in Andrews and Soares (2010) to test the assumption that a set of observed covariates is contained in the firm's information set.[30]

The intuition behind these specification tests is the following: given the structure imposed by equations (6) and (13), if the vector of observed covariates Z_{ijt} we use to define the odds-based and revealed preference inequalities are in the exporter's information set, $\mathcal{W}_{ijt}$, Theorems 1 and 2 imply that the set of parameter values consistent with these inequalities must be non-empty. If, in our finite sample, we find an empty set during estimation, there may be two reasons: (a) the population moment inequalities also yield an empty identified set and, consequently, at least one of the set of assumptions under which they are derived does not hold in the population, or, (b) the set defined by the population moment inequalities is non-empty but, due to sampling noise, its sample analogue is empty. The test suggested in Andrews and Soares (2010) allows us to compute a p-value for the null hypothesis that the identified set is nonempty.

We run three tests of the information set. In the first, we test our main specification of the moment inequalities that includes three covariates in the vector of instruments: the aggregate exports from Chile to each destination market in the previous year, the distance to each market, and the firm's own domestic sales in the previous year. We fail to reject the null that the model is correctly specified at conventional significance levels. Thus, we fail to reject the null hypothesis that potential exporters know at least the three covariates when forming expectations over export profits.

In a second test, we re-run the same inequality model as in our main specification, but we add an additional variable to the vector of instruments. In this specification we assume

[30] Alternative specification tests for partially identified models defined by moment inequalities have been provided in Romano and Shaikh (2008), Andrews and Guggenberger (2009), and Bugni et al. (2015).

firms also know the average productivity of other exporters to a market in the prior year. When we include the four variables in the specification, we reject the model and the choice of information set in both the food and chemicals sector. We find 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.[31]

Finally, we run our moment inequality procedure under the assumption of perfect foresight. We presume the firm knows precisely the revenue it will earn upon exporting to a particular destination. We can reject, at conventional significance levels, that firms know their exact future revenue when deciding whether to export. We find a p-value of 0.05 and 0.01 in the chemicals and food sectors, respectively.

8 Conclusion

We study the extensive margin decision of firms to enter foreign export markets. This decision to participate in export markets drives much of the variation in trade volume. Thus, to predict how trade volume will adjust to changes in the economic environment, policymakers first need a measure of the determinants of firms' decisions to engage in exporting.

In developing an empirical model of firms' export decisions, however, researchers face a data obstacle. The decision to export depends on firms' expectations about the profits they will earn from exporting, which researchers rarely observe. In the standard approach, researchers specify a set of observable covariates that enter the exporters' information sets. We show that the precise specification of the information set matters both for the estimates of structural parameters as well as for the model's predictions of how firms will respond to changes in trade costs. Specifically, we prove formally that the assumption that firms have perfect foresight will bias fixed export costs upwards, and the size of this bias increases with the degree of uncertainty that firms face at the time they decide whether to export.

To handle firm's unobserved expectations, we develop a new moment inequality approach to estimate structural models of binary choice when the decision maker's information set is unknown. We recover the parameters of the firm's export decision requiring only that researchers specify a subset of variables included in exporters' information sets. Our inequality

[31]Many recent papers have put forward different micro-foundations for exporters' information sets. As an example, Dasgupta and Mondria (2014) introduce rationally inattentive importers that endogenously process different amounts of information across different countries; Allen (2014) models the search process that exporters follow to acquire information about prices in different markets; Albornoz et al. (2010) show evidence consistent with exporters learning from their previous export experience; and Fernandes and Tang (2014) develop a model in which firms update their beliefs about demand in foreign markets based on the number of neighbors exporting to these destinations. These papers have implications for the relationship between exporters' information sets and variables beyond those studied in this paper. In ongoing research, we conduct a broader examination of the set of variables different types of firms use in their decision to enter foreign markets.

estimator is consistent with exporters knowing more than what the researcher specifies. The estimated fixed costs from our 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. The bounds we estimate for the effect of this counterfactual on export participation and export volume are sufficiently tight to inform policy.

Finally, we show how to use a specification test of our inequality model to test alternative assumptions on the content of the information sets firms use in their export decision—that is, to test what exporters know. We reject a model that presumes firms possess perfect foresight. We can also reject a model that presumes firms know their competitors' past productivity when deciding whether to export today.

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, 14, 22]

Albornoz, Facundo, Héctor F. Calvo Pardo, Gregory Corcos, and Emanuel Ornelas, "Sequential Exporting," *Journal of International Economics*, 2010, *88* (1), 1–24. [28]

Allen, Treb, "Information Frictions in Trade," *Econometrica*, 2014, *82* (6), 2041–2083. [28]

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

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, 23, 24, 27]

— **and Patrik Guggenberger**, "Validity of Subsampling and Plug-in Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 2009, *25*, 669–709. [3, 27]

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

Antràs, Pol, Teresa Fort, and Felix Tintelnot, "The Margins of Global Sourcing: Theory and Evidence from U.S.," *mimeo*, December 2014. [1, 12]

Arkolakis, Costas, "Market Penetration Costs and the New Consumers Margin in International Trade," *Journal of Political Economy*, 2010, *118* (6), 1151–1199. [1, 12]

— **, 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. [4, 12]

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

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

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, 4]

Biesebroeck, Johannes Van, Emily Yu, and Shenjie Chen, "The Impact of Trade Promotion Services on Canadian Exporter Performance," *Canadian Journal of Economics*, 2015, *forthcoming.* [9]

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

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006, *121* (2), 541–585. [36]

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi, "Alternative Specification Tests for Partially Identified Models Defined by Moment Inequalities," *Journal of Econometrics*, 2015, *185* (1), 259–282. [3, 27]

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]

Dasgupta, Kunal and Jordi Mondria, "Inattentive Importers," *mimeo*, August 2014. [28]

Department of Commerce, "A Profile of U.S. Importing and Exporting Companies, 2012-2013," *Department of Commerce*, 2015. [1]

Eaton, Jonathan, Samuel Kortum, and Francis Kramraz, "An Anatomy of International Trade: Evidence from French Firms," *Econometrica*, 2011, *79* (5), 1453–1498. [1, 12]

Eizenberg, Alon, "Upstream Innovation and Product Variety in the U.S. Home PC Market," *Review of Economic Studies*, 2014, *81*, 1003–1045. [2, 19]

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 1994, *84* (1), 157–177. [36]

Fernandes, Ana and Heiwei Tang, "Learning to Export from Neighbors," *Journal of International Economics*, 2014, *94* (1), 87–94. [28]

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. [4]

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. [9]

Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161. [13, 46]

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]

Lederman, Daniel, Marcelo Olarreaga, and Lucy Payton, "Export Promotion Agencies Revisited," *World Bank Policy Research Working Paper*, 2009, *5125.* [9]

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, 14, 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, 12]

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

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

Martincus, Christian Volpe and Jerónimo Carballo, "Is Export Promotion Effective in Developing Countries? Firm-Level Evidence on the Intensive and the Extensive Margins of Exports," *Journal of International Economics*, 2008, *76* (1), 89–106. [9]

_ , _ , and Pablo Garcá, "Entering New Country and Product Markets: Does Export Promotion Help?," *Review of World Economics*, 2010, *146* (3), 437–467. [9]

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, 19]

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

Pakes, Ariel, "Alternative Models for Moment Inequalities," *Econometrica*, 2010, *78* (6), 1783–1822. [2, 19]

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

Romano, Joseph P. and Azeem M. Shaikh, "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 2008, *138* (9), 2786–2807. [3, 27]

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. [9]

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. [15]

Yatchew, Adonis and Zvi Griliches, "Specification Error in Probit Models," *The Review of Economics and Statistics*, 1985, *67* (1), 134–139. [14, 35]

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, $\{r_{iht}; i = 1, \dots, N_t\}$; and, (iii) the set of exporting firms, $\{d_{ijt}; i = 1, \dots, N_t\}$, 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

$$R_{jt} = \sum_{i=1}^{N_t} d_{ijt} r_{ijt} di = \left[\frac{\eta}{\eta - 1} \frac{\tau_{jt} c_t}{P_{jt}} \right]^{1-\eta} Y_{jt} V_{jt}, \tag{29}$$

where V_{jt} is defined as

$$V_{jt} = \sum_{i=1}^{N_t} d_{ijt} a_{it}^{(1-\eta)} di. \tag{30}$$

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

$$r_{ijt} = \frac{a_{it}^{(1-\eta)}}{V_{jt}} R_{jt}. \tag{31}$$

The term $a_{it}^{(1-\eta)}/V_{jt}$ is 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 $i = 1, \dots, N_t$.

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 that:

$$r_{iht} = \left[\frac{\eta}{\eta - 1} \frac{a_{it} c_t}{P_{ht}} \right]^{1-\eta} Y_{ht}, \tag{32}$$

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

$$\frac{a_{it}^{1-\eta}}{a_{i't}^{1-\eta}} = \frac{r_{iht}}{r_{i'ht}}. \tag{33}$$

Using this expression, we can rewrite the first term in equation (31) as

$$\frac{a_{it}^{(1-\eta)}}{V_{jt}} = \frac{1}{\frac{V_{jt}}{a_{it}^{(1-\eta)}}} = \frac{1}{\sum_{s=1}^{N_t} d_{sjt} \left(\frac{a_{st}}{a_{it}} \right)^{(1-\eta)} ds} = \frac{1}{\sum_{s=1}^{N_t} d_{sjt} (r_{sht}/r_{iht}) ds}. \tag{34}$$

Plugging back this expression into equation (31), 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 consider the bias generated by wrongly assuming perfect foresight in cases in which exporters are uncertain about export profits upon entry. In order to do so, we consider a simplified version of the export model described in Section 2. Specifically, assume that firm i decides whether to export to country j according to the model

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

where $\eta^{-1} = \beta_0 = 0.5$, and $\nu_{ijt} \sim \mathbb{N}(0, 2)$ and independent of any other covariate. This export participation condition is identical to that in equation (8) except that we do not include distance as a covariate (i.e. we assume β_1 equals 0). 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, $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, and expectational error, ε_{ijt}, we show the point estimates and standard errors that researchers would obtain if they were to estimate η^{-1} and β_0 under the assumption of perfect foresight. Under this assumption, we can estimate η^{-1} and β_0 as the values of the unknown parameter vector (γ_0, γ_1) that maximize a likelihood function that relies on the individual likelihood

$$\mathcal{P}(d_{ij} = 1|r_{ij}) = \Phi((\sqrt{2})^{-1}(\gamma_1 r_{ij} - \gamma_2)).$$

Note that this individual likelihood is normalized by scale using the correct variance of ν_{ijt}. In the main text, we normalize by scale by fixing η to a known value. The reason is that, when using actual data, the true value of σ is unknown and external estimates of η are more reliable than those of σ. In this simulation, we decide to normalize by σ purely for simplicity.

Table A.1: Bias under Perfect Foresight

Model	Distribution of $\mathbb{E}[r_{ij}\|\mathcal{W}_{ij}]$	Distribution of ε_{ij}	$\hat{\eta}^{-1}$	$\hat{\beta}_0$
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	$\textit{log-normal}(0,1)$	$\textit{log-normal}(0,1)$	0.1705 (0.0006)	0.5436 (0.0014)
9	$-\textit{log-normal}(0,1)$	$-\textit{log-normal}(0,1)$	0.1435 (0.0006)	0.4767 (0.0013)

Notes: All estimates in this table are normalized by scale by setting $var(\nu_{ijt}) = 2$. In order to estimate each of the models, we generate 1,000,000 observations from the distribution $\nu_{ijt} \sim \mathbb{N}(0, 2)$ and from the distributions of $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, and ε_{ijt} described in columns 2 and 3. Whenever draws are generated from the log-normal distribution, we re-center them at zero. The true parameter values are $\eta^{-1} = \beta_0 = 0.5$.

The first three rows 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 η^{-1} and that the bias is larger as the variance of the expectational error, ε_{ijt}, increases. This is consistent with the analytical formula for the bias term in Yatchew and Griliches (1985). In rows 4 to 10, we explore departures from the setting studied in Yatchew and Griliches (1985). Specifically, we depart from the assumption that both the unobserved firms' expectations and the expectational errors are normally distributed. In rows 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 η^{-1} persists and it is larger the higher the dispersion in the distribution of unobserved expectations and expectational errors. In rows 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 η^{-1} 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, in Section 5 we choose η as the normalizing constant. Given the choice of a particular constant k as the value of η^{-1}, we obtain rescaled estimates of the entry cost coefficient by multiplying our estimates of the fixed cost parameter, β_0, by $k/\hat{\eta}_1^{-1}$. 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

$$\frac{(0.5/\hat{\eta}_1^{-1})\hat{\beta}_0 - 0.5}{0.5}.$$

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. Consistent with the alternative vectors of instruments discussed in Section 4, we define Z_{ijt} such that $dist_j \in Z_{ijt}$. We denote the joint distribution of the vector $(d_{ijt}, Z_{ijt}, r_{ijt})$ as $\mathbb{P}(d_{ijt}, Z_{ijt}, r_{ijt})$. In this section, we use $\mathbb{P}(\cdot)$ to denote distributions that may be directly estimated given the available data on $(d_{ijt}, Z_{ijt}, r_{ijt})$. For the sake of simplicity in the notation, we use r^e_{ijt} to denote $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$. Without loss of generality, we can write

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

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

$$\mathbb{P}(d_{ijt}, Z_{ijt}, r_{ijt}) = \int f^y(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}) f^y(r_{ijt}|r^e_{ijt}, Z_{ijt}) f^y(r^e_{ijt}|Z_{ijt}) \mathbb{P}(Z_{ijt}) dr^e_{ijt}, \tag{35}$$

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 (35). The model in Section 2 imposes the following restriction on the elements of equation (35):

$$f^y(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}) = f(d_{ijt}|r^e_{ijt}, Z_{ijt}; \gamma^y) =$$
$$\Big(\Phi\big((\gamma^y_2)^{-1}\big(\eta^{-1} r^e_{ijt} - \gamma^y_0 - \gamma^y_1 dist_j\big)\big)\Big)^{d_{ijt}} \Big(1 - \Phi\big((\gamma^y_2)^{-1}\big(\eta^{-1} r^e_{ijt} - \gamma^y_0 - \gamma^y_1 dist_j\big)\big)\Big)^{1-d_{ijt}}. \tag{36}$$

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 (35) 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 (35)

$$\gamma_1 \text{ is known and equal to 0,} \tag{37a}$$

$$r_{ijt} = r^e_{ijt} + \varepsilon_{ijt}, \qquad \varepsilon_{ijt}|(r^e_{ijt}, \xi_{ijt}) \sim \mathbb{N}(0, \sigma^2_\varepsilon) \tag{37b}$$

$$Z_{ijt} = r^e_{ijt} + \xi_{ijt} \qquad \xi_{ijt}|r^e_{ijt} \sim \mathbb{N}((\sigma_\xi/\sigma_{r^e})\rho_{\xi r^e}(r^e_{ijt} - \mu_{r^e}), (1 - \rho^2_{\xi r^e})\sigma^2_\xi) \tag{37c}$$

$$r^e_{ijt} \sim \mathbb{N}(\mu_{r^e}, \sigma^2_{r^e}). \tag{37d}$$

Equation (37a) restricts the model in Section 2 by assuming that distance does not affect fixed export costs. Equation (37b) assumes that firms' expectational error is normally distributed and independent of both firms' unobserved expectations and the difference between the instrument and the unobserved expectations, ξ_{ijt}. By contrast, the model in Section 2 only imposes mean independence between ε_{ijt} and r^e_{ijt}. Equation (37c) imposes a particular assumption on the joint distribution of firms' unobserved true expectations r^e_{ijt} and the subset of the variables used by firms to form those expectations that are observed to the researcher, Z_{ijt}. The model in Section 2 does not impose any assumption on this relationship. Finally, equation (37d) 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 (37) defines a model that is more restrictive than that defined in Section 2. However, as we show below, even after imposing the assumptions in equation (37), 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 (36) and (37); (2) verify equation (35); and (3) $\gamma^{y_1} \neq \gamma^{y_2}$. 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 (37a) 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 (36). Equation (37b) assumes that the expectational error not only has mean zero and finite variance but is also normally distributed. This implies that the conditional density $f(r_{ijt}|r^e_{ijt}, Z_{ijt})$ is normal:

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

By applying Bayes' rule, both equations (37c) and (37d) jointly determine the conditional density $f(r^e_{ijt}|Z_{ijt})$ entering equation (35).

Result A.3.1 *There exists empirical distributions of the vector of observable variables* (d, Z, X), $\mathbb{P}(d, Z, X)$, *such that there are at least two structures* S^{y_1} *and* S^{y_2} *for which*

1. *both* S^{y_1} *and* S^{y_2} *verify equations* (35), (36), *and* (37);
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 (36) 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 (35) 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 (37b), (37c) and (37d), 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:

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

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, the system of equations in equation (39) only allows us to define bounds on $\sigma_{r^e}^2$. We can rewrite the system of equations in equation (39) as

$$\begin{aligned}\sigma_r^2 &= \sigma_{r^e}^2 + \sigma_\varepsilon^2,\\ \sigma_z^2 &= \sigma_{r^e}^2 + \sigma_\xi^2 + 2\sigma_{\xi r^e}\\ \sigma_{rz} &= \sigma_{r^e}^2 + \sigma_{\xi r^e}. \end{aligned} \tag{39}$$

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* (13) *holds. Then:*

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

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

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

Reordering terms

$$\begin{aligned}\frac{\partial L(d_{ijt}|\mathcal{W}_{ijt};\theta)}{\partial\theta} = \mathbb{E}\Big[&\frac{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta}{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\times\\ &\Big[d_{ijt}\frac{\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\times\\ &\times\frac{\partial(1-\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}{\partial\Phi(-\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta}+(1-d_{ijt})\Big]|\mathcal{W}_{ijt}\Big] = 0.\end{aligned} \tag{42}$$

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

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

we can simplify:

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

Equation (40) follows by symmetry of the function $\Phi(\cdot)$. ■

Lemma 2 *Suppose the assumptions in equations* (9), (10), *and* (13) *hold. Then*

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

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 (43) 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:*

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}-(1-d_{ijt})\middle|Z_{ijt}\right]=0. \tag{44}$$

and

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

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* (13) *holds. Then:*

$$\frac{\partial L(d_{ijt}|\mathcal{W}_{ijt};\theta)}{\partial\theta}=\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}-d_{ijt}\middle|\mathcal{W}_{ijt}\right]=0. \tag{46}$$

Proof: From equation (41), 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 (46) follows by symmetry of the function $\Phi(\cdot)$. ■

Lemma 4 *Suppose the assumptions in equations* (9), (10), *and* (13) *hold. Then*

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

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 (47) 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:*

$$\mathbb{E}\Bigg[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))} - d_{ijt}\Bigg|Z_{ijt}\Bigg] = 0. \quad (48)$$

and

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}\middle|Z_{ijt}\right]$$
$$\geq$$
$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j))}\middle|Z_{ijt}\right]. \tag{49}$$

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

Proof of Theorem 1 Combining equations (44) and (45), we obtain the inequality defined by equations (17) and (18a). Combining equations (48) and (49), we obtain the inequality defined by equations (17) and (18b). ■

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,*

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

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

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

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

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

Proof: From equation (50),

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

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 (52) as

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

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 (53) as

$$\mathbb{E}\big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j\big)+(1-d_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt}=0,\mathcal{W}_{ijt}]|\mathcal{W}_{ijt}\big]\geq 0. \quad (54)$$

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 (51) follows by applying this equality to equation (54). ■

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

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

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

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] =$$
$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big]+\mathbb{E}\Big[\eta^{-1}d_{ijt}\varepsilon_{ijt}\Big|\mathcal{W}_{ijt}\Big]. \quad (56)$$

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 (56) yields equation (55).

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

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

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 (57) 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*

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j\big)+(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|Z_{ijt}\Big]\geq 0, \tag{58}$$

$$\mathbb{E}\Big[d_{ijt}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] = \mathbb{E}\Big[d_{ijt}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|Z_{ijt}\Big], \tag{59}$$

and

$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j)\big)}\Big|Z_{ijt}\Big]$$
$$\geq$$
$$\mathbb{E}\Big[(1-d_{ijt})\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}{1-\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)\big)}\Big|Z_{ijt}\Big]. \tag{60}$$

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,*

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

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*

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

Proof: From equation (61),

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

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 (63) as

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

Applying the Law of Iterated Expectations, it follows that

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

and we can rewrite equation (64) as

$$\mathbb{E}\big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) - d_{ijt}\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}]\big|\mathcal{W}_{ijt}\big] \geq 0. \tag{65}$$

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 (62) follows by applying this equality to equation (65). ■

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

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

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

$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] =$$
$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{W}_{ijt}\Big] - \mathbb{E}\Big[k(1-d_{ijt})\varepsilon_{ijt}\Big|\mathcal{W}_{ijt}\Big]. \tag{67}$$

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 (67) yields equation (66).

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

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

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 (68) 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*

$$\mathbb{E}\Big[-(1-d_{ijt})\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq 0. \tag{69}$$

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

and

$$\mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq \mathbb{E}\Big[d_{ijt}\sigma\frac{\phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \tag{71}$$

Proof of Theorem 2 Combining equations (58), (59), and (60) we obtain the inequality defined by equations (20) and (21a). Combining equations (69), (70), and (71) we obtain the inequality defined by equations (20) and (21b). ■

A.6 Deriving unconditional moments

The moment inequalities described in equations (17) and (20) condition on particular values of the instrument vector, Z. From these conditional moments, we can derive unconditional moment inequalities. Each of these unconditional moments is defined by an *instrument function*. Specifically, given an instrument function $g(\cdot)$, we derive unconditional moments that are consistent with our conditional moments:

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

where $m_l(\cdot)$, $m_u(\cdot)$, $m_l^r(\cdot)$, and $m_u^r(\cdot)$ are defined in equations (18) and (21), and Z_{ijt} is the same vector of observed covariates employed in defining the conditional moments in equations (17) and (20).

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

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

In words, for each of scalar random variable Z_{kijt} included in the instrument vector $Z_{ijt} = (Z_{1ijt}, \ldots, Z_{kijt}, \ldots, Z_{Kijt})$, the function $g_a(\cdot)$ builds two moments by splitting the observations into two groups depending on whether the value of the instrument variable for that observation is above or below its median. Within each moment, each observation is weighted differently depending on the value of a and on the 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 revealed preference inequalities (in equations (18) and (21)), 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.[32]

A.7 Sample selection in the estimation of determinants of export revenues

Assume a setting characterized by the the following three equations

$$r_{ijt} = \delta X_{ijt} + u_{ijt} + e_{ijt}, \qquad \mathbb{E}[e_{ijt}|\mathcal{J}_{ijt}] = 0,\ u_{ijt} \in \mathcal{W}_{ijt},\ \mathcal{W}_{ijt} \in \mathcal{J}_{ijt}, \tag{72}$$

$$\mathbb{E}[X_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[X_{ijt}|\mathcal{W}^x_{ijt}], \qquad \mathcal{W}^x_{ijt} \in \mathcal{J}_{ijt} \tag{73}$$

$$\left(\begin{array}{c} u_{ijt} \\ \nu_{ijt} \end{array} \right) \Bigg| \mathcal{W}^x_{ijt} \sim \mathbb{N}\left(\left(\begin{array}{c} 0 \\ 0 \end{array} \right), \left(\begin{array}{cc} \sigma^2_u & \sigma_{\nu u} \\ \sigma_{u\nu} & \sigma^2_\nu \end{array} \right) \right), \tag{74}$$

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0\}, \qquad (dist_j, \nu_{ijt}) \in \mathcal{J}_{ijt} \tag{75}$$

where $(d_{ijt}, d_{ijt}r_{ijt}, X_{ijt}, dist_j)$ are observed by the researcher, $(u_{ijt}, e_{ijt}, \mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}])$ are unobserved, and we normalize by scale by setting $\sigma^2_\nu = 1$. Therefore,

$$\rho_{u\nu} = \frac{\sigma_{u\nu}}{\sqrt{\sigma^2_u}\sqrt{\sigma^2_\nu}} = \frac{\sigma_{u\nu}}{\sqrt{\sigma^2_u}} = \frac{\sigma_{u\nu}}{\sigma_u}.$$

In this model, we need to keep track of three different information sets: (a) $\mathcal{J}_{ijt}$ is the true information set of potential exporter i at period t about potential determinants of profits in country j; (b) $\mathcal{W}_{ijt}$ is the subset of $\mathcal{J}_{ijt}$ that is useful to predict r_{ijt}; (c) $\mathcal{W}^x_{ijt}$ is the subset of $\mathcal{W}_{ijt}$ that is useful to predict the vector X_{ijt}.

As in Section 2, we do not fully specify the content of the information set $\mathcal{W}^x_{ijt}$. Specifically, we do not assume anything on whether X_{ijt} is included in the information set of the exporter, $\mathcal{W}^x_{ijt}$. Therefore, this model allows for a very flexible relationship between the information set that exporters have at the time they decide whether to export and the set of determinants of export revenue observed by the researcher, X_{ijt}.

The model described by equations (72), (73), (74), and (75) corresponds to that in Section 2 if we assume that $u_{ijt} = e_{ijt} = 0$ and δ is known. In this case, equation (6) is simply a specific case of equation (72), and equation (74) collapses to equation (3).

The model described by equations (72), (73), (74), and (75) also nests the censored regression model in Heckman (1979). Specifically, the two-step estimator for $(\delta, \beta_0, \beta_1, \sigma^2_u, \sigma_{u\nu})$ in Heckman (1979) would apply to the model in equations (72), (73), (74), and (75) if we were to assume that $e_{ijt} = 0$ and X_{ijt} is measurable in $\mathcal{W}_{ijt}$.

[32] We have recomputed the tables presented in Section 5 using alternative definitions of the instrument function $g_a(Z_{ijt})$. Even though the boundaries of both identified and confidence sets depend on the instrument functions, the main conclusions are robust. The exact results are available upon request.

Given that r_{ijt} is observed only in those cases in which d_{ijt},

$$\begin{aligned}
\mathbb{E}[r_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] &= \delta\mathbb{E}[X_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + \mathbb{E}[u_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + \mathbb{E}[e_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] \\
&= \delta\mathbb{E}[X_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + \mathbb{E}\Big[u_{ijt}\Big|\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt} \geq 0, \mathcal{W}_{ijt}\Big] \\
&= \delta\mathbb{E}[X_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + \mathbb{E}\Big[u_{ijt}\Big|\nu_{ijt} \leq \eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j, \mathcal{W}_{ijt}\Big] \\
&= \delta\mathbb{E}[X_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + \mathbb{E}\Big[u_{ijt}\Big|\nu_{ijt} \geq -\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j\big), \mathcal{W}_{ijt}\Big] \\
&= \delta\mathbb{E}[X_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + \rho_{12}\sigma_u \mathbb{E}\Big[\frac{\phi(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)}{\Phi(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - \beta_0 - \beta_1 dist_j)}\Big|\mathcal{W}_{ijt}\Big],
\end{aligned}$$

where the first equality comes from equation (72); the second equality comes from equation (75) and $\mathbb{E}[e_{ijt}|\mathcal{J}_{ijt}] = 0$; the third equality simply rearranges terms; the fourth equality uses symmetry in the distribution of ν_{ijt}; and the fifth equality computes the inverse Mills ratio. We cannot directly use this moment for estimation purposes because we do not observe $\mathcal{W}_{ijt}$ and, therefore, we cannot construct a consistent estimator for $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$ and $\mathbb{E}[X_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}]$ for every i and t.

In order to deal with the unobserved term $\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]$, we follow an approach identical to that in Section 4.2.2. Under the assumption of rational expectations, we can write

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

Therefore, we can rewrite

$$\mathbb{E}[r_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] = \delta\mathbb{E}[X_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + \rho_{12}\sigma_u \mathbb{E}\Big[\frac{\phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j - \eta^{-1}\varepsilon_{ijt})}{\Phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j - \eta^{-1}\varepsilon_{ijt})}\Big|\mathcal{W}_{ijt}\Big].$$

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

$$\mathbb{E}[r_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] \leq \delta\mathbb{E}[X_{ijt}|d_{ijt}=1,\mathcal{W}_{ijt}] + \rho_{12}\sigma_u \mathbb{E}\Big[\frac{\phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)}{\Phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)}\Big|\mathcal{W}_{ijt}\Big],$$

or, equivalently,

$$\mathbb{E}[r_{it}|d_{it}=1,\mathcal{W}_{it}] \leq \mathbb{E}\Big[\delta X_{ijt} + \rho_{12}\sigma_u \frac{\phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)}{\Phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)}\Big|d_{ijt}=1,\mathcal{W}_{ijt}\Big],$$

and

$$\mathbb{E}\Big[r_{it} - \delta X_{ijt} - \rho_{12}\sigma_u \frac{\phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)}{\Phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)}\Big|d_{ijt}=1,\mathcal{W}_{ijt}\Big] \leq 0.$$

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

$$\mathbb{E}\Big[r_{it} - \delta X_{ijt} - \rho_{12}\sigma_u \frac{\phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)}{\Phi(\eta^{-1}r_{ijt} - \beta_0 - \beta_1 dist_j)}\Big|d_{ijt}=1,Z_{ijt}\Big] \leq 0. \tag{76}$$

Besides the moment inequality in equation (76), we can also derive revealed preference and odds-based inequalities analogous to those introduced in Sections 4.2.1 and 4.2.2. When deriving these inequalities, we must take into account that, contrary to the case described in the main text, in the model considered here, r_{ijt} is not observed for every firm i and period t. However, we can rely on the fact that X_{ijt} is observed for every firm i, country j and time period t, independently of the value of d_{ijt}.

Specifically, we know that

$$\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] = \mathbb{E}[\delta X_{ijt} + u_{ijt} + e_{ijt}|\mathcal{W}_{ijt}] = \delta\mathbb{E}[X_{ijt}|\mathcal{W}_{ijt}] + u_{ijt},$$

and, using equation (73), we can write

$$\mathbb{E}[X_{ijt}|\mathcal{W}_{ijt}] = \mathbb{E}[X_{ijt}|\mathcal{W}^x_{ijt}].$$

Imposing this equality into equation (75), we can write

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\delta\mathbb{E}[X_{ijt}|\mathcal{W}^x_{ijt}] - \beta_0 - \beta_1 dist_j + \eta^{-1}u_{ijt} - \nu_{ijt} \geq 0\},$$

and, analogously,

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\delta\mathbb{E}[X_{ijt}|\mathcal{W}^x_{ijt}] - \beta_0 - \beta_1 dist_j + v_{ijt} \geq 0\},$$

where $v_{it} = \eta^{-1}u_{ijt} - \nu_{ijt}$ and

$$v_{it} \sim \mathbb{N}(0, \sigma_v^2),$$

where

$$\sigma_v^2 = (\eta^{-1})^2\sigma_u^2 + 1 + 2\eta^{-1}\sigma_{u\nu} = (\eta^{-1})^2\sigma_u^2 + 1 + 2\eta^{-1}\rho_{u\nu}\sigma_u. \tag{77}$$

Analogously, we can write

$$d_{ijt} = \mathbb{1}\{(\eta^{-1}/\sigma_v)\delta\mathbb{E}[X_{ijt}|\mathcal{W}^x_{ijt}] - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j + (v_{ijt}/\sigma_v) \geq 0\}, \tag{78}$$

with $(v_{ijt}/\sigma_v) \sim \mathbb{N}(0, 1)$. Using equation (78) and following the steps in Sections 4.2.1 we can derive the odds-based inequalities

$$\mathbb{E}\left[d_{ijt}\frac{1 - \Phi\big((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j\big)}{\Phi\big((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j\big)} - (1 - d_{ijt})\middle| Z_{ijt}\right] \geq 0, \tag{79}$$

and

$$\mathbb{E}\left[(1 - d_{ijt})\frac{\Phi\big((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j\big)}{1 - \Phi\big((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j\big)} - d_{ijt}\middle| Z_{ijt}\right] \geq 0. \tag{80}$$

Similarly, using equation (78) and following the steps in Sections 4.2.2 we can derive the revealed preference inequalities

$$\begin{aligned}
&\mathbb{E}\Bigg[- (1 - d_{ijt})((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j) \\
&\qquad + d_{ijt}\frac{\phi((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j)}{\Phi((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j)}\Bigg| Z_{ijt}\Bigg] \geq 0, \\
&\mathbb{E}\Bigg[d_{ijt}((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j) \\
&\qquad + (1 - d_{ijt})\frac{\phi((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j)}{1 - \Phi((\eta^{-1}/\sigma_v)\delta X_{ijt} - (\beta_0/\sigma_v) - (\beta_1/\sigma_v)dist_j)}\Bigg| Z_{ijt}\Bigg] \geq 0,
\end{aligned}$$

or, equivalently,

$$\mathbb{E}\left[- (1 - d_{ijt})(\eta^{-1}\delta X_{ijt} - \beta_0 - \beta_1 dist_j) + d_{ijt}\sigma_v\frac{\phi(\sigma_v^{-1}(\eta^{-1}\delta X_{ijt} - \beta_0 - \beta_1 dist_j))}{\Phi(\sigma_v^{-1}(\eta^{-1}\delta X_{ijt} - \beta_0 - \beta_1 dist_j))}\middle| Z_{ijt}\right] \geq 0, \tag{81}$$

$$\mathbb{E}\left[d_{ijt}(\eta^{-1}\delta X_{ijt} - \beta_0 - \beta_1 dist_j) + (1 - d_{ijt})\sigma_v\frac{\phi(\sigma_v^{-1}(\eta^{-1}\delta X_{ijt} - \beta_0 - \beta_1 dist_j))}{1 - \Phi(\sigma_v^{-1}(\eta^{-1}\delta X_{ijt} - \beta_0 - \beta_1 dist_j))}\middle| Z_{ijt}\right] \geq 0. \tag{82}$$

The parameters to identify are $(\delta, \eta, \beta_0, \beta_1, \rho_{12}, \sigma_u)$ and, for each observed covariate Z_{it} such that $Z_{it} \in \mathcal{J}_{it}$, the set of moment inequalities identifying these parameters are those given in equations (76), (79), (80), (81), (82), plus the bounds

$$-1 \leq \rho_{u\nu} \leq 1,$$
$$0 \leq \sigma_u,$$

with σ_v defined in equation (77).

A.8 Proof of Theorem 3

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

$$\mathbb{E}\left[\frac{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}\middle|\mathcal{W}_{ijt}\right] \geq \mathbb{E}\left[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\middle|\mathcal{W}_{ijt}\right]. \tag{83}$$

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

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

Equation (83) 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 (13). ■

Lemma 14 *Suppose the assumptions in equations* (9), (3), *and* (13) *hold. Then*

$$\mathbb{E}\left[\frac{\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma^{-1}(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j))}\middle|\mathcal{W}_{ijt}\right] \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\middle|\mathcal{W}_{ijt}\right]. \tag{84}$$

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

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

Equation (83) 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 (13). ■

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

$$\mathbb{E}\left[\frac{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j+\eta^{-1}\varepsilon_{ijt}))}{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)+\eta^{-1}\varepsilon_{ijt})}\middle|\mathcal{W}_{ijt}\right] \geq \mathbb{E}\left[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\middle|Z_{ijt}\right], \tag{85}$$

and

$$\mathbb{E}\left[\frac{\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j+\eta^{-1}\varepsilon_{ijt}))}{1-\Phi(\sigma^{-1}(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}]-\beta_0-\beta_1 dist_j)+\eta^{-1}\varepsilon_{ijt})}\middle|\mathcal{W}_{ijt}\right] \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\middle|Z_{ijt}\right]. \tag{86}$$

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*

$$\mathbb{E}\Big[\frac{1-Y}{Y}\Big] \geq \frac{1-\mathbb{E}[Y]}{\mathbb{E}[Y]}, \tag{87}$$

and

$$\mathbb{E}\Big[\frac{Y}{1-Y}\Big] \geq \frac{\mathbb{E}[Y]}{1-\mathbb{E}[Y]}. \tag{88}$$

Proof: We can rewrite the left hand side of equation (87) as

$$\mathbb{E}\Big[\frac{1-Y}{Y}\Big] = \mathbb{E}\Big[\frac{1}{Y} - 1\Big] = \mathbb{E}\Big[\frac{1}{Y}\Big] - 1, \tag{89}$$

and the right hand side of equation (87) as

$$\frac{1-\mathbb{E}[Y]}{\mathbb{E}[Y]} = \frac{1}{\mathbb{E}[Y]} - 1. \tag{90}$$

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

$$\mathbb{E}\Big[\frac{1}{Y}\Big] \geq \frac{1}{\mathbb{E}[Y]}. \tag{91}$$

Equations (89), (90), and (91) imply that equation (87) holds.

Define a random variable $X = 1 - Y$ and rewrite the left hand side of equation (88) 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 (89), (90), and (91) 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 (88) holds. ■

Corollary 5 *Suppose $\mathcal{P}_{ijt}$ is defined as in equation (13), then*

$$\mathbb{E}\Big[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big] \geq \frac{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}, \tag{92}$$

and

$$\mathbb{E}\Big[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big] \geq \frac{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}. \tag{93}$$

Proof: Equation (13) implies that the support of $\mathcal{P}_{ijt}$ is the interval $(0,1)$. Therefore, Lemma 16 implies that equations (92) and (93) 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 (13). Then,*

$$\frac{1}{1+B_l(Z_{ijt};\theta)} \leq \mathcal{P}(Z_{ijt}) \leq \frac{B_u(Z_{ijt};\theta)}{1+B_u(Z_{ijt};\theta)}, \tag{94}$$

where

$$B_l(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}\bigg|Z_{ijt}\right]. \tag{95}$$

$$B_u(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt}-\beta_0-\beta_1 dist_j\big)\big)}\bigg|Z_{ijt}\right], \tag{96}$$

Proof: Combining equations (85) and (92),

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

and, reordering terms, we obtain the inequality

$$\frac{1}{1+B_l(Z_{ijt};\theta)} \leq \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]. \tag{97}$$

Combining equations (86) and (93),

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

and, reordering terms, we obtain the inequality

$$\frac{B_u(Z_{ijt};\theta)}{1+B_u(Z_{ijt};\theta)} \geq \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]. \tag{98}$$

Combining the inequalities in equations (97) and (98) we obtain equation (94). ■

A.9 Bounds on counterfactual choice probabilities

We may use equations (23), (24) and (25) 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

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

In this counterfactual, the export probability is defined as

$$\begin{aligned}\mathcal{P}^c_{ijt} = \mathcal{P}^c(d_{ijt}=1|\mathcal{W}_{ijt}) &= \int_\nu \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - 0.6\beta_0 - 0.6\beta_1 dist_j - \nu \geq 0\}\phi(\nu)d\nu \\ &= \Phi\big(\sigma^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{W}_{ijt}] - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big),\end{aligned}$$

Using expressions analogous to equations (23), (24) and (25), we may define bounds the expectation of $\mathcal{P}^c_{ijt}$ conditional on any particular value or set of values of Z_{ijt} as follows

$$\underline{\mathcal{P}}^c(Z_{ijt}) \leq \mathcal{P}^c(Z_{ijt}) \leq \overline{\mathcal{P}}^c(Z_{ijt}), \tag{99}$$

where

$$\underline{\mathcal{P}}^c(Z_{ijt}) = \min_{\gamma\in\Theta_{all}} \frac{1}{1+B^c_l(Z_{ijt};\gamma)}, \tag{100}$$

$$\overline{\mathcal{P}}^c(Z_{ijt}) = \max_{\gamma\in\Theta_{all}} \frac{B^c_u(Z_{ijt};\gamma)}{1+B^c_u(Z_{ijt};\gamma)}, \tag{101}$$

with

$$B_u^c(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}\middle| Z_{ijt}\right], \tag{102}$$

$$B_l^c(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1-\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}{\Phi\big(\sigma^{-1}\big(\eta^{-1}r_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}\middle| Z_{ijt}\right]. \tag{103}$$

In addition to 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:

$$\min_{\gamma\in\Theta_{all}} \frac{\frac{1}{1+B_l^c(Z_{ijt};\gamma)}}{\frac{B_u(Z_{ijt};\gamma)}{1+B_u(Z_{ijt};\gamma)}} \leq \frac{\mathcal{P}_{ijt}^c(Z_{ijt})}{\mathcal{P}_{ijt}(Z_{ijt})} \leq \max_{\gamma\in\Theta_{all}} \frac{\frac{B_u^c(Z_{ijt};\gamma)}{1+B_u^c(Z_{ijt};\gamma)}}{\frac{1}{1+B_l(Z_{ijt};\gamma)}}, \tag{104}$$

where $B_u(Z_{ijt};\gamma)$ and $B_l(Z_{ijt};\gamma)$ are defined in equations (96) and (95), respectively; and $B_u^c(Z_{ijt};\gamma)$ and $B_l^c(Z_{ijt};\gamma)$ are defined in equations (102) and (103), respectively.