What do Exporters Know?*

Michael J. Dickstein

Stanford University and NBER

Eduardo Morales

Princeton University and NBER

February 13, 2015

[PRELIMINARY AND INCOMPLETE]

Abstract

The decision of firms to participate in export markets drives much of the variation in the volume of trade. To understand this decision, and in particular to predict how firms will react to currency devaluations or changes in trade policies, policymakers need a measure of the costs firms incur when entering foreign markets. Prior estimates of these costs are often large relative to most exporters' observed revenues. We show that these estimates depend heavily on how the researcher specifies firms' expectations over the potential revenue they would earn upon entry. In response, we develop a novel moment inequality approach that allows us to (1) recover entry costs placing weaker assumptions on firms' expectations and (2) quantify the effects of counterfactual policies. Our approach both introduces a new set of moment inequalities, odds-based inequalities, and generalizes the revealed-preference inequalities introduced in Pakes (2010). We use data from Chilean exporters to show that, relative to methods that must specify firms' information sets, our approach generates estimates of entry costs that are approximately 70% smaller than previous measures. We predict gains in export volume from reductions in entry costs and currency devaluations that are between 30% and 60% larger, respectively, than those predicted by existing approaches.

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

*We thank Tim Bresnahan, Jan De Loecker, Dave Donaldson, Guido Imbens, Ariel Pakes, Esteban Rossi-Hansberg and seminar participants at Pennsylvania State University, Princeton University and the Stanford/Berkeley IO Fest for helpful suggestions. All errors are our own. Email: mjd@stanford.edu, ecmorale@princeton.edu.

1 Introduction

In 2012, approximately 300,000 US firms chose to export to foreign markets. The decision of these firms to sell abroad drives the volume of trade from the US—according to Bernard et al. (2010), 70% of the cross-sectional variation in exports comes from firms entering or exiting a market rather than changing the intensity of their export volume. Thus, to predict export flows and economic output, policymakers must focus on firm-level entry decisions, which depend on the costs firms incur when entering export markets.

A long literature in international trade seeks to measure the entry costs involved with exporting to a new market and to quantify the likely effects from different export promotion programs and from fluctuations in a nation's currency.[1] All empirical work faces one complication, however: the decision to export depends on a firm's expectations of the revenue it will earn upon entry into a foreign market. Without knowing precisely the contents of the firm's information set—how it views its own productivity and the future evolution of exchange rates, trade policy, and political stability abroad—it is difficult to measure export entry costs and predict how the firm will respond to changes in the economic environment.

The exporter's problem is not unique. Many other discrete decisions require the firm or the consumer to forecast a key independent variable. 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)). Similarly, 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 expectations about the riskiness of the task assigned. Similarly, a retiree's decision to purchase a private annuity (Ameriks et al. (2014)) depends on her expectations about life expectancy, and a consumer's decision to buy a durable good depends on the timing of future product updates (Gowrisankaran and Rysman (2012)). In these settings, the econometrician rarely observes the agent's forecast and often cannot even collect the set of information the agent used in his forecast. Thus, when estimating a structural model of discrete choice, the researcher typically substitutes a proxy. The exact form of that proxy, and the assumptions the researcher imposes on agents' expectations, affects the parameter estimates recovered from the discrete choice model. In the context of the exporter's decision, we illustrate that the estimates of the structural parameters prove quite sensitive to these assumptions.

We start our analysis of the exporter's problem with a standard partial equilibrium, two period model of export participation. Following Melitz (2003), we specify a model that features a demand function with constant elasticity of substitution, constant marginal cost, and monopolistic competition between firms. The model provides us a proxy for the firm's ex

[1] See for example Burstein et al. (2005).

post revenue upon exporting to a particular market as a function of the firm's domestic sales, aggregate exports to the market, and the average domestic sales of those firms that chose to export to the market. The decision to export depends on firms' expectations about these ex post revenues and on entry costs, which may depend on the distance to the export market. We estimate these entry costs for 22 destination countries using the model and firm-level data from the Chilean chemical sector. To do so, we depart from the existing empirical international trade literature in that we do not require the researcher to have full or perfect knowledge of the content of firms' information sets at the time of their entry decisions.[2] Instead, we develop new types of moment inequalities that allow us to identify the value of export entry costs and conduct counterfactuals with only partial knowledge of exporters' information sets.

We have three main contributions. First, we show that estimates of export entry costs and the model-based predictions of firms' export decisions and export volumes are sensitive to assumptions the researcher places on firms' information sets. Specifically, maintaining the assumption that firms' expectations are rational and using maximum likelihood methods, we compare two alternative assumptions on firms' knowledge. First, we impose perfect foresight. That is, we assume firms can predict perfectly the revenues they will earn ex post. In this case, we find export costs from Chile to Argentina, the United States, and Japan to equal \$894,000, \$1.7 million, and \$2.8 million, respectively. These cost estimates appear large, given that the mean firm-level export revenues to these countries across all years of the data equal only \$430,000, \$2.31 million, and \$1.95 million. Second, we follow the two-step procedure of Willis and Rosen (1979), Manski (1991) and Ahn and Manski (1993) in which, in a first stage, we specify the set of variables firms use to form their unobserved expectations. We assume a firm's unobserved information set includes distance to the export market, a measure of the firm's productivity in the prior year, and aggregate exports to the market in the prior year. The export entry costs in this two-step approach are lower, but still generally above or close to the mean revenues per firm: the costs of exporting to Argentina, the United States, and Japan are \$594,000, \$1.2 million, and \$1.9 million, respectively. That is, the average exporter to Argentina and Japan is expected to earn negative or close to zero profits upon entry.

The two-step procedure, a common method for handling unobserved expectations, requires the researcher to specify precisely the content of the agent's information set. In the above example, we choose three variables contained in the Chilean customs data. If firms actually employ a different set of variables—either more information or less—these entry cost estimates may differ from the true parameters. The direction of the bias that arises when wrongly assuming that firms have perfect foresight or a specific but incorrect information set is difficult to characterize generally. The tendency for attenuation in the coefficient that is measured with error is similar to that seen in the linear model case, but in a probit model we can only provide

[2]See Roberts and Tybout (1997), Das et al. (2007), Arkolakis (2010), Cherkashin et al. (forthcoming), Moxnes (2010), Eaton et al. (2011), Ruhl and Willis (2014), Arkolakis et al. (2014).

an analytical form for the bias in special cases (see Yatchew and Griliches (1985)). We show empirically in the export case a downward bias in the coefficient on revenue, which translates into larger estimates of the entry cost parameters.

As a second main contribution of this paper, we overcome this potential for bias with a new empirical approach. We employ two new types of moments inequalities, which we label odds-based inequalities and generalized revealed preference inequalities.[3] Intuitively, we use the observed decisions of firms to export to a particular market as evidence that the expected returns to exporting for those firms—in terms of gross expected profits less any fixed entry costs—must exceed the expected returns from not serving that destination market. However, Manski (2002) shows that preference parameters and unobserved expectations are not separately identified from the distribution of choices alone. We must therefore make further assumptions.

We compare the assumptions under our methodology with those common to methods used in the empirical trade literature on export participation. In the literature related to unobserved expectations, researchers typically employ ex-post realizations of the variables for which the firm forms expectations, here export revenue, as a proxy for unobserved expectations. In the solution Willis and Rosen (1979), Manski (1991) and Ahn and Manski (1993) propose, for example, the researcher assumes agents are rational and forms expectations by projecting the ex post realization on a set of variables the researcher observes. Using the inequalities we develop, we also impose rationality on the firms. However, we allow the firm's expectations to depend on variables the researcher does not observe. The form of our inequalities, however, restrict us to binary decision problems, a limitation relative to Willis and Rosen (1979). In addition, in comparison to the existing empirical literature that defines inequalities from revealed preference, notably Pakes (2010) and Pakes et al. (forthcoming), we allow idiosyncratic structural errors, such as a probit error, to affect the firm's binary choices.

In our data and using our inequalities, we find set-identified estimates of exporters' entry costs in the Chilean chemical sector. We compare our estimates of the costs of exporting from Chile to Argentina, for example, to the maximum likelihood estimates reported above. Depending on the researcher's assumptions on the firms' information sets, the entry costs ranged from \$594,000 to \$894,000 under traditional methods. Using our approach, we find a much lower range of entry costs, between approximately \$270,000 and \$298,000. Our estimates of the fixed costs for Argentina lie far below the mean level of exports per firm; at the upper bound, the fixed costs represent 69% of the mean level of exports. For the United States and Japan our fixed costs estimates represent an even smaller fraction of the average exports per firm, at most 54% for Japan and 27% for the United States.

[3]The generalized revealed preference inequalities extend the revealed preference inequalities introduced in Pakes (2010). Pakes (2010) allows for structural errors in specific cases: (a) when such errors are common across individual and/or choices (i.e. fixed effects); and (b) ordered choice models. The generalized revealed preference inequalities allow for an individual and choice-specific structural error in binary choice models.

In arriving at these fixed cost estimates, we assume that firms know the distance to the export destination, the aggregate exports to that market in the prior year, and their own domestic sales from the prior year, which we use in the model as a measure of the firm's productivity. Importantly, our model does not restrict firms to use only these variables, but requires that the firm know at least these variables. We can use the specification test of Andrews and Soares (2010) to test the null hypothesis that the moments we specify produce a confidence set that contains the true value of the parameter vector. That is, we can test our specification that presumes these variables are in the firm's information set. In our estimation, we cannot reject the null hypothesis that exporters know distance, lagged domestic sales, and lagged aggregate exports when making their export decisions.

Finally, as a third main contribution, we provide measurement of firms' response to two counterfactual policies. To do so, we first show how to bound predictions of the exporter's behavior in counterfactual environments. We provide a novel—and very simple—procedure for conducting counterfactual analyses using our inequalities. With this procedure to predict firms' export participation decisions, we examine how firms would respond to (1) a policy that reduces entry costs by 40% and (2) a currency depreciation of 20%.

We first compare our counterfactual predictions under different assumptions on firms' information sets. Relative to estimates from a model that assumes perfect foresight, the estimates from the two-step procedure differ substantially. After a 40% reduction in entry costs and assuming firms have perfect foresight, the predicted export revenue to Argentina, Japan, and the United States equals \$17.9 million, \$40.3 million, and \$74.4 million, respectively. Under the two-step procedure, the predicted export volume is 3% and 11% lower for Argentina and Japan and 21% higher for the United States. That is, we would predict the relative effect of the policy on exports to Argentina and the United States to be different depending on our specification of the information set. Comparing the estimates under perfect foresight to the set identified by the moment inequalities, in the latter the predicted range of export volumes to Argentina, Japan, and the United States are higher by 29-45%, 56-77% and 35-45%, respectively.

We use our inequalities to examine the substantive effect of the reduction in entry costs, possibly via a launch-aid subsidy, and the effect of a currency devaluation. Comparing trade flows from Argentina to Chile in the baseline environment to the counterfactual change in entry costs, the number of exporters increases between 14% and 21% and export volume increases anywhere from 3% to 16%. Under a 20% currency depreciation, the predicted number of exporters from Chile to Argentina increases between 2% and 12% and the volume of exports increases between 37% and 49%.

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 empirical setting in more detail and outline the maximum likelihood procedures. In Section

5, we introduce our moment inequality estimator, built to provide inference in settings in which the researcher may not observe the same variables as the firm. We discuss how to build these inequalities as well as conduct counterfactuals with possibly set-identified parameters. Section 6 contains the results from estimating models in which we must specify the firm's information set and results from our moment inequality approach. We compare the estimates from the alternative approaches. In Section 7, we use our inequality model to predict the effect on export participation and export volume from changes to the economic environment. Section 8 concludes.

2 Export Model

We begin with a model of a firm's export behavior. All firms in our dataset are located in a single country h but may sell in every country. We index the firms located in h and active at period t by $i = 1, \ldots, N_t$.[4] We index the potential destination countries by $j = 1, \ldots, J$.

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 set by a firm in the destination country, Y is the total expenditure in country j 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. We define p_{ijt}, P_{jt} and Y_{jt} in country j's currency. This specification implies that every firm faces a constant demand elasticity in country j equal to η. We assume that the parameter η is constant across countries and time periods.

2.2 Supply

Firm i produces one unit of output with a cost-minimizing combination of inputs that cost $a_{it}c_t$, where c represents the cost of this bundle in country h's currency and a_{it} measures the number of bundles of inputs that firm i uses to produce one unit of output. The cost c will be affected by factor prices in h. The inverse of a_{it} denotes firms i's physical productivity level

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

in t. We assume that 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 destination market $j \neq h$, it must pay the production cost $a_{it}c_t$ 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. We assume that fixed export costs defined in terms of country h currency are

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

where $dist_{jt}$ denotes the distance in kilometers from country h to country j (constant over time), and ν_{ijt} is an aggregate of all remaining determinants of f_{ijt}.[5]

2.3 Profits conditional on exporting

We assume that every seller in market j behaves as a monopolistically competitive firm. The demand and supply assumptions above imply that the optimal price firm i sets in j is

$$p_{ijt} = \frac{\eta}{\eta - 1} \frac{\tau_{jt} a_{it} c_t}{e_{jt}}, \tag{3}$$

where e_{jt} denotes the price in units of home currency of one unit of country j's currency. As a result, the total revenue that i will obtain in any country j in country h's currency is:

$$r_{ijt} = \left[\frac{\eta}{\eta - 1} \frac{\tau_{jt} a_{it} c_t}{P_{jt}} \right]^{1-\eta} Y_{jt} e_{jt}^{\eta} \tag{4}$$

and the export profit (gross of fixed costs) is $\eta^{-1} r_{ijt}$.

2.4 Decision to export

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

$$\pi_{ijt} = \eta^{-1} r_{ijt} - f_{ijt}. \tag{5}$$

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

[5] We assume that the fixed export costs f_{ijt} are independent of previous export experience of i in country j. However, given the very flexible specification of the time process of the term a_{it}, our model will be able to match any observed persistence in export status.

$\mathbb{1}\{\mathbb{E}[\pi_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] \geq 0\}$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function. Assuming that $dist_j \in \mathcal{J}_{ijt}$, we can rewrite d_{ijt} as

$$d_{ijt} = \mathbb{1}\{\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] - f_{ijt} \geq 0\}, \tag{6}$$

where r_{ijt} is defined in equation (4), f_{ij} is defined in equation (2), and $\mathbb{E}[\cdot]$ denotes the expectation with respect to the data generating process. Therefore, defining agents' expectational error as ε_{ijt}, $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}]$, it holds that

$$\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0. \tag{7}$$

Given equations (2) and (6) and the assumption that

$$\nu_{ijt}|(r_{ijt}, \mathcal{J}_{ijt}) \sim \mathbb{N}(0, \sigma_\nu^2), \tag{8}$$

we can write the probability that i exports to j conditional on $\mathcal{J}_{ijt}$ as

$$\mathcal{P}_{ijt} = \mathcal{P}(d_{ijt} = 1|\mathcal{J}_{ijt}) = \Phi\big(\sigma_\nu^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big), \tag{9}$$

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

2.5 Effect of change in export entry costs

We study the effect of a policy that, for the firms located in country h, reduces export entry 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 $\beta_1^1 = .6\beta_1$. We assume that, for all firms and countries, τ, c, e, P, Y, and a remain invariant to the change in (β_0, β_1). From equation (4), this implies that r_{ijt} is invariant to the change in (β_0, β_1). Therefore, the only variables affected by the policy are the set of export participation dummies, $\{d_{ijt}, i = 1, \ldots, N\}$ and, through them, the total exports from h to j, R_{jt}. We denote with superscript 1 the value of these variables for the case in which $(\beta_0, \beta_1) = (\beta_0^1, \beta_1^1)$. Using g_{jt}^1 to denote the gross export growth due to the change in export costs, we can write

$$g_{jt}^1 = \frac{R_{jt}^1}{R_{jt}} = \frac{\int_{i\in N_t} d_{ijt}^1 r_{ijt} di}{\int_{i\in N_t} d_{ijt} r_{ijt} di} = \frac{\int_{i\in N_t} \mathcal{P}_{ijt}^1 r_{ijt} di}{\int_{i\in N_t} \mathcal{P}_{ijt} r_{ijt} di}, \tag{10}$$

where

$$\mathcal{P}_{ijt}^1 = \mathcal{P}^1(d_{ijt} = 1|\mathcal{J}_{ijt}) = \Phi\big(\sigma_\nu^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big). \tag{11}$$

2.6 Effect of currency depreciation

We also study the effect of a 20% currency devaluation in country h. Denoting with a superscript 2 the value of the different variables after the counterfactual change in e_j, $e_j^2 = 1.2e_j$. We assume that, for all countries and firms, τ, c, P, Y, and a remain invariant to the change in e. From equation (4), this implies that the currency devaluation changes export revenues conditional on exporting such that $r_{ijt}^2 = r_{ijt}(1.2)^\eta$. Using g_{jt}^2 to denote the gross export growth due to the currency devaluation we can write

$$g_{jt}^2 = \frac{R_{jt}^2}{R_{jt}} = \frac{\int_{i\in N_t} d_{ijt}^2 r_{ijt}^2 di}{\int_{i\in N_t} d_{ijt} r_{ijt} di} = \frac{\int_{i\in N_t} \mathcal{P}_{ijt}^2 r_{ijt}(1.2)^\eta di}{\int_{i\in N_t} \mathcal{P}_{ijt} r_{ijt} di}, \tag{12}$$

where the third equality uses equation (4) and

$$\mathcal{P}_{ijt}^2 = \mathcal{P}^1(d_{ijt} = 1|\mathcal{J}_{ijt}) = \Phi\big(\sigma_\nu^{-1}\big(\eta^{-1}\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}](1.2)^\eta - \beta_0 - \beta_1 dist_j\big)\big). \tag{13}$$

2.7 Normalization

Given equations (9), (10), and (11), the effect of the counterfactual changes described in Sections 2.5 and 2.6 is independent of the scale of the parameter vector $(\sigma_\nu, \eta, \beta_0, \beta_1)$—that is, if we multiply these four parameters by the same positive constant, the probabilities $\mathcal{P}_{ijt}$, $\mathcal{P}_{ijt}^1$, and $\mathcal{P}_{ijt}^2$ and the changes in export revenues g_{jt}^1 and g_{jt}^2 remain the same. This implies that we cannot use observed data on export participation to identify the scale parameter. However, this scale parameter is irrelevant for the outcomes of our counterfactual exercises; if we maintain a constant scale parameter, we can still compare the relative magnitudes of the estimates of $(\sigma_\nu, \eta, \beta_0, \beta_1)$ that we obtain through different estimation methods.

In order to normalize by scale the parameter vector in export entry models, researchers in international trade typically calibrate η to a given constant. We follow that approach and fix η^{-1} to a positive constant k and rewrite $\mathcal{P}_{ij}$, $\mathcal{P}_{ij}^1$, and $\mathcal{P}_{ij}^2$ as a function of $(\sigma_\nu, \beta_0, \beta_1)$

$$\mathcal{P}_{ijt} = \Phi\big(\sigma_\nu^{-1}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big), \tag{14}$$

$$\mathcal{P}_{ijt}^1 = \Phi\big(\sigma_\nu^{-1}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big), \tag{15}$$

$$\mathcal{P}_{ijt}^2 = \Phi\big(\sigma_\nu^{-1}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}](1.2)^\eta - \beta_0 - \beta_1 dist_j\big)\big). \tag{16}$$

For simplicity of notation, from now on, we use θ to denote the parameter vector $(\sigma_\nu, \beta_0, \beta_1)$. Following standard estimates in the literature, we will set $k = 0.2$, which implies an elasticity of substitution equal to $\eta = 5$.

3 Empirical Setting

3.1 Data

Our data come from two separate sources. The first is an extract of the Chilean customs database, which covers the universe of exports of Chilean firms from 1996 to 2005. The second is the Chilean Annual Industrial Survey (*Encuesta Nacional Industrial Anual*, or ENIA), which includes all manufacturing plants with at least 10 workers for the same years. We merge these two data sets using firm identifiers, allowing us to exploit information on the export destinations of each firm and on their domestic activity.[6]

These firms operate in the 19 different 2-digit ISIC sectors that deal with manufacturing.[7] We restrict our analysis to one sector: the manufacture of chemicals and chemical products. This is the second largest export manufacturing sector in Chile.[8] In Table 1, we report summary statistics on the number of exporters, the volume of exports, as well as the intensity of exports per firm. 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. We observe 266 unique firms manufacturing products in the chemicals sector across all years; on average, 102 of these firms participate in at least one export market in a given year. In Panel 1 of Table 1, we report the total annual exports in this sector, which are on average \$1.184 billion but fluctuate, with a low in 2001 of \$673 million. The mean level of exports per firm, across all export destinations, varies across years as well, but is roughly \$580,000 per firm and destination. We focus in Panel 2 of Table 1 on export revenues for three countries—Argentina, Japan, and the United States—across all years of the data. We focus on these destinations in later counterfactual exercises. For these three countries, the total volume of exports across all years of the data equals \$368 million, \$1.489 billion, and \$2.093 billion, respectively.[9] The mean annual volume per exporter equals \$430,000, \$1.95 million, and \$2.31 million, respectively, for Argentina, Japan, and the United States

Our data set includes both exporters and non-exporters. Furthermore, in order to minimize the risk of selection bias in our estimates, we use an unbalanced panel that includes not only

[6]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 label 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 value.

[7]ENIA encompasses class D (sectors 15 to 36) of the ISIC rev.3.1 industrial classification. The Chilean chemicals sector is sector 24.

[8]The largest export manufacturing sector is food and food products. The estimation of fixed costs for this sector are in progress.

[9]In 2005, the export volume for Argentina, Japan, and the United States equals \$42 million, \$176 million, and \$150 million, respectively.

those firms that appear in ENIA in every year between 1995 and 2005 but also those that were created or disappeared during this period.[10] Finally, we obtain information on the distance from Chile to each destination market from CEPII.[11]

3.2 Proxy for export revenue

The first step common in the empirical literature on export participation is to define a proxy for the revenue that every active firm i would obtain in any country j if it were to export in t, r_{ijt}. In some datasets, these revenues are directly observed by the researcher but only in the case of those firm-country-year combinations with positive exports. 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 any home country h to any destination country j, R_{jt}; (c) the set of firms located in h that actively export to j in t, $\{d_{ijt}, i = 1, \ldots, N_t\}$. Specifically:

$$r_{ijt} = \frac{R_{jt}}{\int_{s \in N} d_{sjt}(r_{sht}/r_{iht})ds}. \tag{17}$$

From equations (10), (12), (14), (15), and (16), we can predict the effect on aggregate exports from country h to a given country j in t from the policy counterfactuals described in Sections 2.5 and 2.6. To do so, we need data on the measure of potential export revenues in equation (17), the value of the parameter vector θ and the information sets $\mathcal{J}_{ijt}$ that a potential exporter i has about the export revenue it might earn if it decides to sell its product in country j in year t.

Typical datasets in international trade contain data needed to construct the proxy for export revenues in equation (17) for multiple home and destination countries. The information sets exporters use to predict revenue, however, are rarely available to researchers. Therefore, a key hurdle to overcome in order to perform the policy exercises described in Sections 2.5 and 2.6 is to estimate θ without observing all elements of $\mathcal{J}_{ijt}$.

3.3 Exporters' Information Sets

In the model we describe in Section 2, potential exporters may be uncertain about the ex-post revenues they would earn upon entering a market. We did not, however, impose assumptions on the content of the information set, $\mathcal{J}_{ijt}$ that firm i uses to predict its potential export

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

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

revenues. Here, we discuss several alternatives.

In the existing literature, to recover the unknown parameter vector, θ, researchers must specify a vector of observed covariates, Z_{ijt}, that equal the information set, $\mathcal{J}_{ijt}$. Researchers choose this set Z_{ijt} depending on the covariates in their theoretical model. In the model we develop in Section 2, the variable that a potential exporter i needs to predict, r_{ijt}, is a function of (1) the firm's domestic sales, (2) aggregate exports to the destination country j and (3) the domestic sales of all the firms that will end up exporting to j. Therefore, with this model, the researcher would define an information set containing observed covariates capable of predicting any of these three sets of variables.

In Section 4, we discuss two specific definitions of exporters' information sets, Z_{ijt}. First, we describe a model with perfect foresight. We denote this set with a superscript 1, $Z^1_{ijt} = (r_{ijt}, dist_j)$. Under perfect foresight, potential exporters know, at the time of their entry decision, the exact revenues they will obtain in each market if they choose to enter. In most empirical settings, the set Z^1_{ijt} is likely to be strictly larger than firms' true information sets. Second, we describe a setting in which exporters forecast their potential export revenues using only information on their own lagged domestic sales, lagged aggregate exports to the destination country j, and distance from the home country to j. We denote this information set with a superscript 2, $Z^2_{ijt} = (r_{iht-1}, R_{it-1}, dist_j)$. This information set is likely to be strictly smaller than the actual information set firms possess when deciding whether to export. Under either definition of the information set, researchers typically assume these sets are common across all firms. That is, all potential exporters base their entry decision on the same set of covariates.

Ideally, we would like to estimate the parameter vector, θ, and perform counterfactuals without imposing the strong assumptions above on firms' information sets. In Section 5, we propose a moment inequality estimator that allows both estimation and counterfactual simulations in scenarios in which the econometrician observes only a *subset* of potential exporters' true information sets. Specifically, we let the observables in the information set equal Z^2_{ijt}, as defined above, but allow other unobserved variables to be in the information set, such that Z^2_{ijt} is a subset of the true information set. Furthermore, unlike the approaches that must specify the complete information set, this procedure permits the unobservable elements to vary by firm, such that information sets need not be common to all exporters.

4 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 $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and one can identify θ as the value of the parameter γ that maximizes the log-likelihood function

$$\mathcal{L}(\gamma|d,r,Z) =$$
$$\mathbb{E}\Big[\sum_t \sum_j d_{ijt} \log\big(\mathcal{P}(d_{jt}=1|r_{ijt}, Z_{ijt})\big) + (1-d_{ijt}) \log\big(\mathcal{P}(d_{jt}=0|r_{ijt}, Z_{ijt})\big)\Big], \tag{18}$$

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

$$\mathcal{P}(d_{jt}=1|r_{ijt}, Z_{ijt}) = \Phi\big(\gamma_2^{-1}\big(k\mathbb{E}[r_{ijt}|Z_{ijt}] - \gamma_0 - \gamma_1 dist_j\big)\big) \tag{19}$$

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

Given that the researcher rarely observes firms' information sets and that the sets themselves are likely heterogeneous across agents, specifying the correct information set of each agent is notoriously complicated. If the information set specified by the researcher, Z_{ijt}, is such that $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] \neq \mathbb{E}[r_{ijt}|Z_{ijt}]$, then the identified value of θ under the assumption that $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[r_{ijt}|Z_{ijt}]$ will be biased. We denote the difference between the two revenue projections as ξ_{ij}: $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[r_{ijt}|Z_{ijt}] - \xi_{ij}$. In this case, one can identify θ as the parameter that maximizes the likelihood function in equation (18) but with

$$\mathcal{P}(d_{ijt}=1|r_{ijt}, Z_{ijt}) =$$
$$\int_{k\xi+\nu} \mathbb{1}\{k\mathbb{E}[r_{ijt}|Z_{ijt}] - \gamma_0 - \gamma_1 dist_j - (k\xi_{ijt} + \nu_{ijt}) \geq 0\} f((k\xi+\nu)|Z_{ijt}) d(k\xi+\nu), \tag{20}$$

where $f(k\xi+\nu|Z)$ denotes the density of $k\xi+\nu$ conditional on Z. When comparing equation (19) to the corresponding equation (20), it is clear that wrongly assuming that $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = \mathbb{E}[r_{ijt}|Z_{ijt}]$ will generate biased estimates of θ unless $f(k\xi+\nu|Z_{ijt})$ is normal with mean zero and variance σ_ν^2. The direction of the bias for each element of θ depends on the shape of the distribution of $k\xi+\nu$ conditional on Z_{ijt}. Specifically, in the specific case in which ξ is normally distributed and independent of both ν and Z, the presence of this additional error term will simply biased upwards the estimate of the variance of the composite error term. However, generally, ξ might be correlated with the term $\mathbb{E}[r_{ijt}|Z_{ijt}]$ and, in this case, the bias on the different parameters θ might take many different forms.

In the specific case in which we assume perfect foresight (i.e. we assume that $r_{ijt} = \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$), we can analytically sign the bias on the estimates of β_0 and β_1. Applying the results in Yatchew and Griliches (1985) to this context, we can conclude that: if firms' true expectations are normally distributed, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] \sim \mathbb{N}(0, \sigma_e^2)$, and the expectational error is also normally distributed, $\xi_{ij}|(\mathcal{J}_{ijt}, \nu_{ij}) \sim \mathbb{N}(0, \sigma_\xi^2)$; then, there is an upward bias in the estimates of β_0 and β_1. Specifically, the ML estimates of these entry costs parameters converge to $\beta_0(\sigma_e^2 + \sigma_\xi^2)/\sigma_e^2$ and $\beta_1(\sigma_e^2 + \sigma_\xi^2)/\sigma_e^2$. Therefore, the upward bias increases in the variance of the expectational error relative to the variance of the true unobserved expectations. When either firms' true expectations or the expectational error are not normally distributed, there is no analytic expression for the bias term in β_0 and β_1. However, as the simulations presented in Appendix A.2 show, assuming perfect foresight when firms' expectations are actually imperfect generates an upward bias in β_0 and β_1 under many different distributions of firms' true expectations and expectational error.

One may gain intuition for the upward bias in the ML estimates of β_0 and β_1 caused by wrongly assuming perfect foresight from the well-known result on the downward bias of estimates of covariates affected by classical measurement error in linear models (see page 73 in Wooldridge (2002)). Rational expectations implies that firms' expectational errors are mean independent of their true expectation and, therefore, correlated with the ex-post realization of export revenues. Consequently, in linear regression models, wrongly assuming perfect foresight and using the ex-post realized revenue r_{ijt} as a regressor instead of the unobserved expectation $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ will generate a downward bias on the coefficient on r_{ijt}. The probit model in equation (19) differs from this linear setting in two dimensions. First, we perform the due normalization by scale by setting the coefficient on the covariate measured with error, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, to k. This implies that the bias generated by the correlation between the expectational error, ε_{ijt}, and r_{ijt} will be reflected in an upward bias in the estimates of the entry costs parameter β_0 and β_1. Second, the direction of the bias depends not only on the correlation between ε_{ijt} and r_{ijt} but also on the functional form of the distribution of unobserved expectations and expectational error. However, as Appendix A.2 shows, for a wide range of possible distributions, the positive bias in the estimates of the entry costs parameters β_0 and β_1 persists.

Biased estimates of the structural parameter of interest θ will translate into incorrect predictions of the effect of the counterfactual changes in the environment described in Sections 2.5 and 2.6. In Section 6, we illustrate the distinct estimates and counterfactual predictions found when assuming alternately that Z_{ijt}^1 or Z_{ijt}^2 (as defined in Section 3.3) perfectly describe the information set of potential exporters.

5 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 vector of observed covariates that is contained in such information sets. For example, in each year, exporters will generally know their domestic sales as well as aggregate exports from their home country to each destination market in the previous year. These variables are also usually observed in standard datasets and, therefore, they form a vector Z_{ijt} such that $Z_{ijt} \subset \mathcal{J}_{ijt}$.

As Appendix A.3 shows, given the model described in Section 2 and available data on Z_{ijt} and d_{ijt}, the assumption that $Z_{ijt} \subset \mathcal{J}_{ijt}$ is not strong enough to point-identify the parameter vector θ. However, this assumption has enough power to identify a set that contains the true value of the parameter θ. Specifically, we present two new types of moment inequalities that, under the assumption that $Z_{ijt} \subset \mathcal{J}_{ijt}$, will define such a set.

5.1 Odds-based moment inequalities

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

$$\mathcal{M}(Z_{ijt};\gamma) = \mathbb{E}\left[\begin{array}{c} m_l(d_{ijt}, r_{ijt}, dist_j;\gamma) \\ m_u(d_{ijt}, r_{ijt}, dist_j;\gamma) \end{array} \middle| Z_{ijt} \right] \geq 0, \tag{21}$$

where the two moment functions are defined as

$$m_l(\cdot) = d_{ijt} \frac{1 - \Phi\big(\gamma_2^{-1}\big(kr_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)}{\Phi\big(\gamma_2^{-1}\big(kr_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)} - (1 - d_{ijt}), \tag{22a}$$

$$m_u(\cdot) = (1 - d_{ijt}) \frac{\Phi\big(\gamma_2^{-1}\big(kr_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)}{1 - \Phi\big(\gamma_2^{-1}\big(kr_{ijt} - \gamma_0 - \gamma_1 dist_j\big)\big)} - d_{ijt}. \tag{22b}$$

We denote as Θ the set of all values of the parameter vector γ in the parameter space Γ_β that verify the inequalities defined in equations (21), (22a) and (22b). The following theorem contains the main property of Θ.

Theorem 1 *$\beta \in \Theta$ for any $\beta \in \Gamma_\beta$.*

The proof of the Theorem 1 is in Appendix A.4. Theorem 1 indicates that the odds-based inequalities are consistent with the true value of the parameter vector.

Intuitively, the two moment functions in equations (22a) and (22b) are derived from the score function of a likelihood function in which we replace the unknown expectation $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ with the observed ex post revenue, r_{ijt}. The key difference between our moment inequality approach and the maximum likelihood approach that assumes perfect foresight, described in Section (4), is that here we include an error term that arises when the perfect foresight assumption is inaccurate. This error term reflects the difference between the unobserved expectation that enters in the firms' entry rule, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, and the observed ex post revenue, r_{ijt}. Given the assumption that firms have rational expectations and that $Z_{ijt} \subset \mathcal{J}_{ijt}$, this expectational error has a mean equal to zero conditional on the vector Z_{ijt}. We use this property of the expectational error combined with the fact that both $1 - \Phi(\cdot)/\Phi(\cdot)$ and $\Phi(\cdot)/(1 - \Phi(\cdot))$ are globally convex to apply Jensen's inequality and conclude that the inequality in equation (21) should hold at the true value of the parameter vector.[12] [13]

Even though both moment functions in equations (22a) and (22b) are derived from the score function, they are not redundant. In order to gain intuition into 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}, Z_{ijt}, and of the parameters γ_1 and γ_2, the moment $m_l(\cdot)$ in equation (22a) is increasing in γ_0 and, therefore, will identify a lower bound on γ_0. The opposite is true for the moment $m_u(\cdot)$. Therefore, both moments are necessary to identify both upper and lower bounds on γ_0. The same intuition goes through for the parameters γ_1 and γ_2.

In the particular case in which agents' expectations are perfect and the vector of instruments Z_{ijt} happens to be identical to the agents' information set, $\mathcal{J}_{ijt}$, the set Θ is a singleton and identical to the true value of the parameter vector, θ. The size of the set Θ increases monotonically in the variance of the expectational error—that is, in the difference between firms expected revenues $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the ex post realization of such revenues r_{ijt}.

5.2 Generalized revealed-preference moment inequalities

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

$$\mathcal{M}^r(Z_{ijt};\gamma) = \mathbb{E}\left[\begin{array}{c} m_l^r(d_{ijt}, r_{ijt}, dist_j;\gamma) \\ m_u^r(d_{ijt}, r_{ijt}, dist_j;\gamma) \end{array}\middle| Z_{ijt}\right] \geq 0, \tag{23}$$

where the two moment functions are defined as

$$m_l^r(\cdot) = -(1-d_{ijt})\big(kr_{ijt} - \gamma_0 - \gamma_1 dist_j\big) + d_{ijt}\gamma_2 \frac{\phi\big(\gamma_2^{-1}(kr_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}{\Phi\big(\gamma_2^{-1}(kr_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}, \tag{24a}$$

$$m_u^r(\cdot) = d_{ijt}\big(kr_{ijt} - \gamma_0 - \gamma_1 dist_j\big) + (1-d_{ijt})\gamma_2 \frac{\phi\big(\gamma_2^{-1}(kr_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}{1-\Phi\big(\gamma_2^{-1}(kr_{ijt} - \gamma_0 - \gamma_1 dist_j)\big)}. \tag{24b}$$

We denote as Θ^r the set of all values of the parameter vector γ in the parameter space Γ_β that verify the inequalities defined in equations (23), (24a) and (24b). The following theorem contains the main property of Θ^r.

Theorem 2 *$\beta \in \Theta^r$ for any $\beta \in \Gamma_\beta$.*

[12] Put differently, under the perfect foresight assumption, one identifies the true parameter vector using moment functions (22a) and (22b), where the corresponding moments hold as equalities instead of the inequality introduced in equation (21).

[13] The assumption of normality of the structural error term is sufficient but not necessary for the existence of odds-based inequalities that correctly bound the true parameter vector. As long as the distribution of the structural error ν is log-concave, inequalities analogous to those in equation (22), with the correct cumulative distribution function $F_\nu(\cdot)$ instead of the normal cumulative distribution function $\Phi(\cdot)$, will also satisfy Theorem (1). The intuition for 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.

The proof of the Theorem 2 is in the Appendix A.5. Theorem 2 indicates that the generalized revealed-preference inequalities are consistent with the true value of the parameter vector.[14] In general, both sets Θ and Θ^r will also contain values of γ other than its true value, β. However, as we show in Section 6, in our empirical application, the interaction of both sets Θ and Θ^r is small enough so that economically meaningful conclusions may be drawn from combining the odds-based and generalized revealed-preference inequalities.

Intuitively, the two moment functions in equations (24a) and (24b) are derived from standard revealed preference arguments. We focus our discussion on moment function (24b); the intuition behind the formation of moment (24a) is analogous. If firm i decides to export to j in period t, so that $d_{ijt} = 1$, then by revealed preference, the firm must expect to earn positive returns: $d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \gamma_0 - \gamma_1 dist_j - \nu_{ijt}\big) \geq 0$. Substituting $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] = r_{ijt} - \varepsilon_{ijt}$, and taking an expectation of this inequality conditional on $(d_{ijt}, \mathcal{J}_{ijt})$, we obtain the inequality,

$$d_{ijt}\big(kr_{ijt} - \gamma_0 - \gamma_1 dist_j\big) + S(\mathcal{J}_{ijt}) \geq 0, \tag{25}$$

where $S(\mathcal{J}_{ijt}) = \mathbb{E}[-d_{ijt}\nu_{ijt}|d_{ijt}, \mathcal{J}_{ijt}]$. The term $kr_{ijt} - \gamma_0 - \gamma_1 dist_j$ accounts for factors that are observed to the econometrician and that determine the export choice of firm i in country j at t. The term $S(\cdot)$ is a selection correction term and accounts for the fact that firms might decide whether to export based on determinants of profits that are not observed to the researcher; i.e. the term ν_{ijt} in the model described in Section 2.[15]

We label the moment functions in equations (24a) and (24b) as *generalized* revealed preference inequalities. These moments start with the baseline revealed preference inequalities introduced in Pakes (2010) and Pakes et al. (forthcoming), and add an allowance for a structural error ν_{ijt} with a non-zero variance–that is, we allow $S(\mathcal{J}_{ijt})$ to be different from zero. Given that $S(\mathcal{J}_{ijt}) \geq 0$ whenever there is an individual-specific structural error, revealed-preference inequalities that ignore this term will always define a weakly smaller set than the generalized revealed-preference inequalities in equations (23), (24a) and (24b).

[14] The assumption of normality of the structural error term is sufficient but not necessary for the existence of generalized revealed-preference inequalities that correctly bound the true parameter vector. As long as the distribution of the structural error ν is such that both $f_\nu(\cdot)/F_\nu(\cdot)$ and $f_\nu(\cdot)/(1 - F_\nu(\cdot))$ are globally convex, we may write inequalities analogous to those in equation (24) that also satisfy Theorem 2. Besides the normal distribution, the type I extreme value distribution also satisfies this property.

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

$$S(\mathcal{J}_{ijt}) = (1 - d_{ijt})\gamma_2 \frac{\phi\big(\gamma_2^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \gamma_0 - \gamma_1 dist_j)\big)}{1 - \Phi\big(\gamma_2^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \gamma_0 - \gamma_1 dist_j)\big)}$$

We cannot directly use the term $S(\cdot)$ in our inequalities because they depend on the unobserved agents' expectations, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$. However, the inequality in equation 2 becomes weaker if we substitute the proxy for ex-post profits, r_{ijt}, in $S(\mathcal{J}_{ijt})$ in place of the unobserved term, $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$. The reason is that the expectational error has a mean equal to zero conditional on the vector Z_{ijt}. We use this property of the expectational error combined with the fact that both $\phi(\cdot)/\Phi(\cdot)$ and $\phi(\cdot)/(1-\Phi(\cdot))$ are globally convex to apply Jensen's inequality.

As indicated in Section 5.1, the set defined by the odds-based inequalities contains only the true value of the parameter vector whenever firms' expectations are perfect and the vector of instruments Z_{ijt} is identical to firms' information sets. Therefore, in this very specific case, the generalized revealed preference inequalities do not have any additional identification power beyond that of the odds-based inequalities. In all other settings, these additional moments can provide identifying power.[16]

5.3 General applicability of our moment inequalities

Both the odds-based and the generalized revealed-preference moment inequalities defined in equations (21) and (23) identify the true value of the parameter vector in any binary choice model that has the following properties:

1. The dummy variable d capturing the choice is determined following the equation $d = \mathbb{1}\{\beta X^* + \nu \geq 0\}$, where the econometrician observes d but does not observe either X^* or ν;

2. The econometrician observes a variable X such that $X = X^* + \varepsilon$, and $\mathbb{E}[\varepsilon|X^*, Z] = 0$;

3. The term ν is independent of both the unobserved term X^* as well an instrument vector Z that is observed by the econometrician;

4. The marginal distribution of ν is log-concave.

The economic model described in Section 2 is analogous to this statistical model. In the notation of this model, X^* represents firms' unobserved expectations, $X^* = \mathbb{E}[r|\mathcal{J}]$; X captures the ex-post observed realization of revenue, $X = r$; ε equals firms' expectational error; ν captures the unobserved component of the export entry costs; and Z being a subset of firms' information sets. Even though the model in Section 2 assumes that ν is normally distributed, the inequalities in Sections 5.1 and 5.2 apply more broadly; the only distributional requirement on ν is that it should be log-concave.[17]

As far as we know, this statistical model has not been studied in the literature. Specifically, the previous literature has imposed alternative restrictions on the relationship between the unobserved covariate X^* and its instrument vector Z. For example, Willis and Rosen (1979),

[16] How Θ compares to Θ^r 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.

[17] A random variable y has a log concave distribution if its density function f_y satisfies that $f_y(\lambda y_1 + (1-\lambda)y_2) \geq [f_y(y_1)]^{\lambda}\ [f_y(y_2)]^{1-\lambda}$, $0 \leq \lambda \leq 1$, for any given values y_1 and y_2 in the support of y. Some general references on log concave density functions are Pratt (1981), Heckman and Honoré (1990), and Bagnoli and Bergstrom (2005). Heckman and Honoré (1990) clarify that the class of log concave densities includes the normal, logistic, uniform, exponential, extreme value and laplace (or double exponential) densities. Under some parameter restrictions, it also includes the power function, Weibull, gamma, chi-squared and beta distributions.

Manski (1991) and Ahn and Manski (1993) assume that the unobserved variable X^* may be written as a deterministic function of the observed instrument vector Z; i.e. $X^* = m(Z)$, where $m(\cdot)$ may be unknown. In the setting of the model described in Section 2, this implies that the researcher fully observes the information set of firms, $\mathcal{J}$, which is equal to the vector Z. The estimation procedure in Schennach (2007) imposes a slightly weaker restriction on the relationship between X^* and Z. Specifically, she imposes that $X^* = m(Z) + W$, where W is a covariate the researcher does not observe that is independent of the observed covariate Z and whose expectation is zero— i.e. $W \perp Z$, $\mathbb{E}(W) = 0$. In the context of our economic model, this assumption implies that all the variables in agents' informations sets that the researcher does not observe must verify two conditions: (a) enter additively in agents' expectations; and, (b) be independent of any observed variables in agents' information sets. We discuss three other related methods to identify the parameters of binary choice models with endogenous regressors in Appendix A.8.

5.4 Deriving unconditional moments

The moment inequalities described in equations (21) and (23) 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 will be more useful to work with unconditional moment inequalities. Each of the unconditional moment inequalities is defined by an *instrument function*. Specifically, given an instrument vector $g(\cdot)$, we derive unconditional moments that are consistent with our conditional moments:

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

where $m_l(\cdot)$, $m_u(\cdot)$, $m_l^r(\cdot)$, and $m_u^r(\cdot)$ are defined in equations (22) and (24). The unconditional moment inequalities proposed here generate a larger identified set than that defined by the conditional moments described in Section 5.1 and 5.2. The main advantage of the moments proposed here is computational simplicity. Papers that define unconditional moments that imply no loss of information with respect to their conditional counterpart are Armstrong (2014) and Andrews and Shi (2013). The instrument functions suggested in these papers are computationally expensive in our setting. Instead, in Section 6, we present results based on a set of instrument functions $g(Z_{ijt}) = (g_0(Z_{ijt}), g_{0.5}(Z_{ijt}), g_1(Z_{ijt}), g_{1.5}(Z_{ijt}))$ with

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

and $Z_{ijt} = (r_{iht-1}, R_{jt-1}, dist_j)$. Therefore, for any given value of a,

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

Given that each particular instrument function $g_a(Z_{ijt})$ contains six instruments and there are four basic odds-based and generalized revealed preference inequalities (in equations (22) and (24)), the total number of instruments used in the estimation is equal to twenty-four times the number of different values of a that are combined to form the instrument vector $g(Z_{ijt})$, in addition to a constant vector. In the benchmark case we simultaneously use two different instrument functions, $g_a(Z_{ijt})$, for $a = 0, 1.5$, to define both an estimated set Θ_{all} and a confidence set Θ^{α}_{all} at significance level α. In Section 6.2, we show results for different vectors of instruments functions $g(Z_{ijt})$ that combine the functions $g_a(Z_{ijt})$ for different sets of values of a.

5.5 Deriving bounds on choice probabilities

As Sections 5.1, 5.2 and 5.4 show, we can set identify and estimate the structural parameter vector, θ, without the need to fully specify and observe agents' information sets. However, beyond obtaining estimates of export entry costs, a main motivation of estimating the export entry model of Section 2 is to make predictions about how changes to the economic environment will affect export participation and, through it, export volumes. In this section, we show that one can use the same data and assumptions imposed in the estimation routine to define bounds on a firm's probability of exporting.

Choice probabilities are not point identified in our setting for two reasons. First, even if we were to know the true value of the parameter vector, θ, the fact that we only observe a subset Z_{ijt} of the true information set, $\mathcal{J}_{ijt}$, implies that we cannot compute the export probabilities in equation (14). Second, we do not recover the true value of the parameter vector in our estimation, but only a set that includes it. As detailed in Appendix A.6, under these circumstances we may still derive bounds on the expected probability conditional on Z_{ijt} that a firm i exports to country j at period t.

Suppose that the true value of the parameter vector, θ, is known. Then Theorem 3 defines bounds on the expectation of the export probability $\mathcal{P}_{ijt}$ (see equation (14) for its definition) conditional on a vector Z_{ijt} such that $Z_{ijt} \in \mathcal{J}_{ijt}$.

Theorem 3 *Suppose $Z_{ijt} \in \mathcal{J}_{ijt}$ and define $\mathcal{P}(Z_{ijt}) = \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]$, with $\mathcal{P}_{ijt}$ defined in equation* (14)*. Then,*

$$\frac{1}{1 + B_2(Z_{ijt};\theta)} \leq \mathcal{P}(Z_{ijt}) \leq \frac{B_1(Z_{ijt};\theta)}{1 + B_1(Z_{ijt};\theta)}, \tag{26}$$

where

$$B_1(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}\bigg| Z_{ijt}\right], \tag{27}$$

$$B_2(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}{\Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}\bigg| Z_{ijt}\right]. \tag{28}$$

The proof of the Theorem 3 is in the Appendix A.6. Note that we can use information on the realized export revenues r_{ijt}, distance $dist_j$ and the instrument vector Z_{ijt} to compute consistent estimates of $B_1(Z_{ijt};\theta)$ and $B_2(Z_{ijt};\theta)$ at any particular value of $Z_{ijt} = z$,

$$\hat{B}_1(z;\theta) = \sum_{ijt}\left[\frac{\Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}\right]\frac{\mathbb{1}\{Z_{ijt} = z\}}{\sum_{ijt}\mathbb{1}\{Z_{ijt} = z\}}, \tag{29}$$

$$\hat{B}_2(z;\theta) = \sum_{ijt}\left[\frac{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}{\Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\big)}\right]\frac{\mathbb{1}\{Z_{ijt} = z\}}{\sum_{ijt}\mathbb{1}\{Z_{ijt} = z\}}, \tag{30}$$

where, for simplicity in the notation, we use $\sum_{ijt}$ to denote $\sum_i \sum_j \sum_t$. Using the estimators (29) and (30) and plugging them into equation (26) we define bounds on the average export probability across those firms i countries j and time periods t such that their instrument vector Z_{ijt} is equal to z. Specifically, by appropriately

The bounds in equation (26) depend on the true value of the parameter vector, θ, and, therefore, cannot be computed. However, we may use the information on the points included in either the identified set Θ_{all} or the confidence set Θ_{all}^{α} to build an identified set or confidence set, respectively, for $\mathcal{P}(Z_{ijt})$ that does not depend on the true value of the parameter vector θ.

Corollary 1 *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* (9)*. Then,*

$$\underline{\mathcal{P}}(Z_{ijt}) \leq \mathcal{P}(Z_{ijt}) \leq \overline{\mathcal{P}}(Z_{ijt}), \tag{31}$$

where

$$\underline{\mathcal{P}}(Z_{ijt}) = \min_{\gamma \in \Theta_{all}} \frac{1}{1 + B_2(Z_{ijt}; \gamma)}, \tag{32}$$

$$\overline{\mathcal{P}}(Z_{ijt}) = \max_{\gamma \in \Theta_{all}} \frac{B_1(Z_{ijt}; \gamma)}{1 + B_1(Z_{ijt}; \gamma)}. \tag{33}$$

and $B_1(Z_{ijt}; \gamma)$ *and* $B_2(Z_{ijt}; \gamma)$ *are defined in equations* (27) *and* (28)*, respectively.*

The proof of Corollary 1 is immediate from Theorem 3. We can define consistent estimates of $\underline{\mathcal{P}}(Z_{ijt})$ and $\overline{\mathcal{P}}(Z_{ijt})$ simply by substituting $B_1(Z_{ijt}; \gamma)$ and $B_2(Z_{ijt}; \gamma)$ in these equations by their consistent estimators $\hat{B}_1(Z_{ijt}; \gamma)$ and $\hat{B}_2(Z_{ijt}; \gamma)$, as defined in equations (29) and (30), respectively.

Equation (31) defines bounds on export probabilities conditional on a particular value of the instrument vector Z_{ijt}. However, using equation (31) we may define bounds on the expected export probability for any subset of firms defined by a particular set $\mathcal{Z}$ of values of the instrument vector Z_{ijt} as

$$\sum_{ijt} \underline{\mathcal{P}}(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\} \leq \sum_{ijt} \mathcal{P}(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\} \leq \sum_{ijt} \overline{\mathcal{P}}(Z_{ijt}) \mathbb{1}\{Z_{ijt} \in \mathcal{Z}\}. \tag{34}$$

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 (34) will yield bounds on the average export probability to country j^* in year t^*. In Section 6.1, we use the bounds in equation (34) to test the fit of the model for different countries and years. We show in Appendix A.7 how to use equation (34) to compute bounds for the counterfactual scenarios described in Sections 2.5 and 2.6.

6 Results

We estimate the parameters of the exporters' participation decisions using three different approaches discussed above. First, we use maximum likelihood to estimate the components of the exporter's costs of entering a new market under perfect foresight—we assume the firm perfectly predicts the level of revenue it will earn upon entry. Second, the again use maximum likelihood methods, but under the two-step procedure in which we use the realized revenue as a proxy in a first stage. In this approach, we specify the agent's information set in the first stage to include the total aggregate exports in the prior year, the distance to the destination country, and the firm's own domestic sales from the previous year. Finally, third, we carry out our inequality estimation. For comparison purposes, we assume, as in the two-step approach, that the firm knows lagged aggregate exports, its own lagged domestic revenue, and the distance to the export destination. However, unlike the two-step approach, the inequalities allow additional unobserved variables to influence the decision to export.

We first discuss the parameter estimates and illustrate the fit of the models in comparison to the data. We then compare the estimates of the costs of entry under each of the three alternative methods.

6.1 Predictions and fit

In Table 2, we report the estimates and the confidence regions for the parameters of our entry cost specification. The first coefficient, σ, represents the variance of the probit structural error in the model of the export participation under the normalization discussed in Section 2.7. The remaining coefficients represent a constant component and the contribution of distance to the level of the costs of entry under this normalization. From the raw coefficients, it is clear that both the model involving perfect foresight and the two-step approach of Willis and Rosen (1979), Manski (1991) and Ahn and Manski (1993) produce much larger estimates of the costs to participate in an export market than does our moment inequality approach. For example, looking at the set identified parameter on the distance variable from the moment inequalities, the identified set ranges from \$428,000 to \$479,000 added cost when the export destination is 10,000 kilometers farther in distance. Under the two maximum likelihood procedures, the estimates of the distance coefficient equals \$1,180,000 and \$812,000 for the same added distance.

We translate these coefficients into an estimate of the entry costs of exporting and report the results in Table 4. For clarity of exposition, we focus on three countries out of the 22 destinations to which Chilean firms export chemical products: Argentina, Japan, and the United States. These three countries offer prototypical examples of how export volume and export participation differs by location and market size. Under perfect foresight, we estimate the entry costs in these three countries to equal \$894,000, \$2.80 million, and \$1.74 million, respectively. Recall from Table 1, the mean volume of exports per firm in Argentina, Japan, and the United States are only \$430,000, \$1.95 million, and \$2.3 million. Comparing the estimates under perfect foresight to the estimates from the two-step procedure, the two-step procedure produces entry cost estimates that about 1/3 smaller.

Under our moment inequality estimator, we find estimates of the entry costs from exporting of between \$270,000 and \$298,000 for Argentina, \$977,000 and \$1.06 million for Japan, and between \$592,000 and \$632,000 for the United States. Across Argentina, Japan, and the United States, the estimated bounds we find from the inequalities equal only a fraction of the perfect foresight estimates, with a level between 30% and 38% of the perfect foresight values. Comparing the bounds of the entry costs from the inequalities to the estimates from the two-step approach, again the bounds are much smaller; the estimates of the entry costs from the inequality approach equal about half those estimated under the two-step approach.

The results appear in line with the discussion in Section 4 of the bias that may arise if the researcher incorrectly specifies the firm's information set. Here, specifying a specific and limited information set appears to drive an upward bias in the estimates of the entry costs, as it pushes downward the estimated coefficient on the mismeasured revenue variable. We see this effect most strongly when comparing the estimates from our moment inequalities to the estimates of entry costs from a model in which firms have perfect foresight or forecast revenue using a specific set of covariates. This result is not guaranteed analytically; the same tendency that produces attenuation bias in a linear model with measurement error still exists, but the nonlinear errors in the probit model can produce a force in the opposite direction.

Furthermore, comparing the estimates from the three methods to the summary statistics on export revenue per firm provides additional support for the direction of the bias under the assumption that the firm observes the full information set. For Argentina, Japan, and the United States, the estimated entry costs from perfect foresight lie well above the median level of per-firm exports. Specifically, at the estimated level of entry costs, roughly 89%, 75%, and 71% of exporters to Argentina, Japan, and the United States have annual export revenues below the entry costs, meaning these firms would lose money on entry. The two-step procedure finds lower levels of the entry costs. Nonetheless, anywhere from 63% to 81% of exporters would lose money from participating in export markets given this level of costs. Using our inequality model, we estimate entry costs closer to the median per-firm revenue in Argentina Japan, and the United States. In all countries, the entry costs from the inequality approach fall below the country and year-specific averages level of exports per firm.

Finally, in Table 6, we report the observed level of export participation in our three comparison countries in the year 2005. Along with these observed values, we report the predictions from the export model under perfect foresight, the two-step approach, and from our inequalities. In part due to their high estimated levels of entry 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. Interestingly, the predictions from these two approaches differ by country. For the United States, the two-step approach predicts a larger number of exporters than does the model that assume perfect foresight. For Japan and Argentina, the perfect foresight model predicts greater entry than does the two-step approach. Comparing these estimates to those from the inequality approach, it is clear that estimates of export participation based on perfect knowledge of the information set differ importantly from the observed level of participation. In contrast, the inequality model's predictions contain the observed numbers of exporters in the estimated identified set.

6.2 Robustness of inequality method

As discussed in Section 5.4, translating the conditional moment restrictions in the inequality approach to unconditional moments offers a range of valid specifications of the set of instruments. We illustrate the robustness of our inequality approach to various functional form assumptions on the set of instruments.

In Table 3 we report confidence sets for several alternative specifications of the functional form for the instruments in the moment inequality model. Here, we use a consistent set of instrumental variables, as in the main specification. Specifically, at the time the firms decides whether to enter a particular destination country in a given year, we assume the firm knows the aggregate exports to that country in the prior year, the distance to the country, and the firm's own domestic revenue in the prior year, which functions as proxy for its productivity relative to other exporters. With this set of variables, we compare alternative forms for the instrument functions. Specifically, we use the form in Section 5.4, but add moments in successive specifications that are weighted by the value of the instruments raised to different powers. We report the end points of the confidence set over the parameters from these specifications in Table 3 and illustrate the confidence sets in two dimensions in Figure 1. The size of the confidence set varies across specifications, generally growing smaller with additional weighted moments, as illustrated in the figure. For each alternative instrument function, we carry out the specification test suggested by Andrews and Soares (2010). The p-values reported in the final column of Table 3 illustrate that at conventional significance levels, we fail to reject the null hypothesis that the confidence sets produced under our sets of moment inequalities contain the true value of the parameter. We also report the fixed costs estimates under different specifications. In Table 5, again the confidence sets grow smaller with larger numbers of weighted moments included in the specification.[18]

7 Counterfactuals

As introduced in Section 2.5 and Section 2.6, we conduct two distinct counterfactual analyses using the estimates from our alternative methodologies. In the first counterfactual, we simulate the effect of lowering the export entry costs by 40%. There are multiple mechanisms—some legal, some illegal—that policymakers might use to lower export entry costs, including launch-aid subsidies. In the second counterfactual, we simulate the effect of a 20% depreciation of

[18] In addition, we run additional moment inequality specifications in which we vary both the set of instrument functions and the set of variables assumed to be in the information set of the firm when deciding whether to enter. We conduct a specification test of an inequality model in which we assume the firm also knows the average productivity of other exporters to a country in the prior period. This is a new variable added to the set of three instruments included in our main specification. With p-value of .97, 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.

the Chilean peso relative to foreign currency. We conduct the counterfactuals using only data from the year 2005.

The predictions of the counterfactual in which we lower export costs by 40% appear in Table 7. In Table 8, we transform the estimates to illustrate the comparison across the methodologies. 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 maximum likelihood approaches, the predictions for how a policy will impact exports differs between the two methods depending on the export destination. The moment inequality approach produces larger predictions of the effect of the policy relative to either maximum likelihood approach.

We report the predictions from the second counterfactual—in which we simulate a currency depreciation of 20%—in Table 9. Again, the moment inequality estimator generally predicts larger effects on export participation and export volume from the policy.

We compare the baseline level of exports to the counterfactual predictions in Table 10. To quantify the impact of the policy interventions, we use only the main inequality specification. The estimates reveal substantive economic effects from the two policy interventions. Decreasing export costs by 40% leads to a large increase in export participation in all three countries, particular in markets far from Chile. As a percentage of the baseline level, the policy which causes entry costs to fall 40% leads to between a 14% and 21% increase in exports to Argentina. The percentage increases in Japan and the United States are larger, between 38% and 69% and between 27% and 43%, respectively. Export volume also increases, but not by quite as much, as the new exporters are likely smaller firms on average.

The counterfactual in which the Chilean currency falls by 20% produces a significant increase in export participation relative to the baseline, though the effects are strongest in Japan. Compared to the predictions when entry costs fall, the currency change produces a relatively stronger effect on exports, as it affects the dollar returns of all exporters, even those not on the margin of participation. For all three countries, the predicted volume of exports in dollars is predicted to increase by between 37% and as much as 72% in the case of Japan.

8 Conclusion

We develop a new inequality estimator to recover the parameters of a firm's export decision when the firm must form expectations over the revenue it will earn upon entry. Many discrete choice settings in economics fit this mold, in which the researcher does not observe the firm's expectations but has a proxy, often the ex post realization of those expectations.

The prior empirical literature generally followed one of two approaches in the case in which researchers do not observe the firm's expectation: either assume the firm has perfect foresight or assume the econometrician observes the exact set of variables the firm used to form its expectation. We show that, in the context of export participation decisions, both methods lead to large estimates of the entry costs involved with exporting. The entry costs found exceed the mean revenues of the observed exporters. In contrast, our inequality approach allows the firm's expectations to be based on variables the econometrician does not observe. The estimates of entry costs from the inequalities are between one third and one half the size of the costs found using the approaches common in the international trade literature. The predictions of two counterfactual economic environments—in which export entry costs fall 40% and the local currency depreciates 20%—differ substantially across alternative methods. Even when comparing the two prior approaches, the predictions of how exports to each destination country change after a fall in entry costs differs in direction depending on the assumptions the researcher places on the firm's unobserved expectations.

On methodology, our novel inequalities approach improves on the existing empirical literature in three ways. First, we provide bounds on the parameters of interest that are robust to alternative assumptions on the information set the firm used to form its expectations over future variables. Second, at least for binary decisions, we relax a restriction common to past empirical papers that use inequality estimators—namely, that there is no individual level structural error that differs across firms, markets, and time periods. Finally, we show how to use our inequalities to carry out counterfactual analyses. Jointly, these contributions provide applied economists a robust tool for estimating structural models of discrete choice when the decision maker's information set is unknown.

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.

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

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

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

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

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

Armstrong, Tim B., "Asymptotically Exact Inference in Conditional Moment Inequality Models," *mimeo*, December 2014.

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.

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

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.

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

Blundell, Richard W. and James L. Powell, "Endogeneity in Nonparametric and Semiparametric Regression Models," in Mathias F. Dewatripont, Lars P. Hansen, and Stephen J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress, Vol. II*, Cambridge: Cambridge University Press, 2003.

_ **and** _ , "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 2004, *71*, 655–679.

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

Burstein, Ariel, Martin S. Eichenbaum, and Sergio Rebelo, "Large Devaluations and the Real Exchange Rate.," *Journal of Political Economy*, 2005, *113* (3), 742–784.

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 Economy*, forthcoming.

Chesher, Andrew, "Identification in Nonseparable Models," *Econometrica*, 2003, *71* (2), 1401–1444.

_ , "Nonparametric Identification under Discrete Variation," *Econometrica*, 2005, *73* (2), 1525–1550.

_ , "Instrumental Values," *Journal of Econometrics*, 2007, *139* (1), 15–34.

_ , "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 2010, *75* (2), 575–601.

_ , "Semiparametric Structural Models of Binary Response: Shape Restrictions and Partial Identification," *mimeo*, 2011.

_ , Adam Rosen, and Konrad Smolinski, "An Instrumental Variable Model of Multiple Discrete Choice," *mimeo*, 2011.

_ and _ , "An Instrumental Variable Random Coefficients Model for Binary Outcomes," *mimeo*, 2012.

_ and Konrad Smolinski, "Sharp identified sets for discrete variable IV models," *mimeo*, 2010.

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

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

Evdokimov, Kirill and Halbert White, "Some Extensions of a Lemma of Kotlarski," *Econometric Theory*, 2012, *28* (4), 925–932.

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

Florens, Jean-Pierre, James J. Heckman, Costas Meghir, and Edward Vytlacil, "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 2008, *76* (5), 1191–1206.

Gowrisankaran, Gautam and Marc Rysman, "Dynamics of Consumer Demand for New Durable Goods," *Journal of Political Economy*, 2012, *120* (2), 1173–1219.

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.

Heckman, James J. and Bo E. Honoré, "The Empirical Content of the Roy Model," *Econometrica*, 1990, *58* (5), 1121–1149.

Imbens, Guido W. and Whitney K. Newey, "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 2009, *77* (5), 1481–1512.

Lewbel, Arthur, "Semiparametric Qualitative Response Model Estimation With Unknown Heteroskedasticity or Instrumental Variables," *Journal of Econometrics*, 2000, *97*, 145–177.

Magnac, Thierry and Eric Maurin, "Identification and Information in Monotone Binary Models," *Journal of Econometrics*, 2007, *139*, 76–104.

_ and _ , "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *The Review of Economic Studies*, 2008, *75* (3), 835–864.

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.

Manski, Charles F., "Identification of Decision Rules in Experiments on Simple Games of Proposal and Response," *European Economic Review*, 2002, *46* (4-5), 880–891.

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

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

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

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

_ , Jack R. Porter, Katherine Ho, and Joy Ishii, "Moment Inequalities and their Application," *Econometrica*, forthcoming.

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

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

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

Schennach, Susanne M., "Instrumental Variable Estimation of Nonlinear Errors-in-Variables Models," *Econometrica*, 2007, *75* (1), 201–239.

Shaikh, Azeem M. and Edward J. Vytlacil, "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 2011, *79* (3), 949–955.

Vytlacil, Edward J. and Nese Yildiz, "Dummy Endogenous Variables in Weakly Separable Models," *Econometrica*, 2007, *75* (3), 757–779.

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

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

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

Table 1: Summary statistics on export participation and volume, years 1996-2005

Panel 1: Across all firms and countries							
		Total	Summary statistics on the annual volume of exports per exporting firm, across all countries (in $million/firm)				
Year	No. of unique exporters	volume of exports, in $ millions	Mean	Std Deviation	Median	95th percentile	Max
1996	92	1,190.14	0.52	1.41	0.11	2.15	13.96
1997	107	1,359.71	0.57	1.86	0.11	2.72	31.02
1998	105	1,556.54	0.54	1.36	0.13	2.48	14.60
1999	101	1,218.27	0.51	1.40	0.12	2.21	20.95
2000	109	919.85	0.49	1.10	0.14	2.39	10.30
2001	105	673.89	0.47	1.11	0.12	1.83	8.94
2002	102	1,488.95	0.61	1.58	0.16	2.62	16.21
2003	106	1,071.21	0.61	1.51	0.17	2.48	21.34
2004	106	1,144.68	0.70	1.52	0.23	2.73	17.77
2005	94	1,220.10	0.78	1.73	0.20	3.61	16.00

Panel 2: For select countries, across all years							
		Total	Summary statistics on the annual volume of exports per export firm, by country (in $million/firm)				
Country	No. of unique exporters in all years	volume of exports in all years, in $ millions	Mean	Std Deviation	Median	95th percentile	Max
Argentina	106	368.35	0.43	0.74	0.18	1.60	5.52
Japan	12	1,489.27	1.95	2.96	0.36	8.70	10.64
United States	63	2,092.53	2.31	4.24	0.70	12.05	31.02

Notes:
Table includes exporters and export volume in the Chilean chemical sector. Across all years, a total of 266 unique firms manufacture products in the chemical sector. We focus on exports to 22 countries, which include all export destinations with at least five unique firms exporting to the country in all years between 1996 and 2005. Data on exporters and export volume collected from the Chilean customs database.

Table 2: Parameter estimates from export model, alternative specifications

Panel A: Parameter estimates for fixed costs specification (in $000s)				
Method		sigma	constant	distance
Maximum likelihood, perfect foresight	Estimate	1,074.03	760.88	1,180.08
	Std error	46.74	36.67	53.21
Maximum likelihood, two-step approach	Estimate	701.91	502.20	812.25
	Std error	24.33	20.15	30.01
Moment Inequalities	Lower bound of identified set	311.74	218.34	428.32
	Upper bound of identified set	340.96	245.82	478.96
Panel B: Confidence regions for fixed costs specification (in $000s)				
Method		sigma	constant	distance
Maximum likelihood, perfect foresight	Lower bound, 95% Conf Int	982.43	689.00	1,075.79
	Upper bound, 95% Conf Int	1,165.63	832.76	1,284.38
Maximum likelihood, two-step approach	Lower bound, 95% Conf Int	654.22	462.70	753.44
	Upper bound, 95% Conf Int	749.61	541.70	871.06
Moment Inequalities	Lower bound, 95% Conf Set	177.78	120.45	242.11
	Upper bound, 95% Conf Set	470.59	317.65	647.06

Table 3: Robustness: moment inequalities using alternative instrument functions

	sigma		constant		distance		
	lower	upper	lower	upper	lower	upper	
Specification	bound	bound	bound	bound	bound	bound	P-value
Includes moments with weights raised to {0,1}	166.67	571.43	100.00	371.43	183.33	800.00	0.00
Includes moments with weights raised to {0,1.5}	177.78	470.59	120.45	317.65	242.11	647.06	0.00
Includes moments with weights raised to {0,1,2}	190.48	470.59	130.95	317.65	300.00	647.06	0.28
Includes moments with weights raised to {0,2}	275.86	444.44	193.10	305.56	378.57	622.22	0.47

Table 4: Estimates of export entry costs, in $000s

	Via maximum likelihood, perfect foresight			Via maximum likelihood, two-step approach			Via moment inequalities			
Destination country	Estimate	Lower bound (95% CI)	Upper bound (95% CI)	Estimate	Lower bound (95% CI)	Upper bound (95% CI)	Lower bound, estimate	Upper bound, estimate	Lower bound (95% Conf Set)	Upper bound (95% Conf Set)
Argentina	894.03	799.05	989.01	593.85	551.04	636.65	269.99	298.15	162.00	388.00
Japan	2,796.22	2,518.45	3,073.99	1,903.12	1,779.29	2,026.95	977.63	1,061.96	637.37	1,421.89
United States	1,736.93	1,564.98	1,908.89	1,174.02	1,099.20	1,248.83	592.55	632.03	414.85	841.07

Table 5: Robustness: export entry costs from moment inequalities using alternative instrument functions, in $000s

	Includes moments with weights raised to {0,1}		Includes moments with weights raised to {0,1.5}		Includes moments with weights raised to {0,1,2}		Includes moments with weights raised to {0,2}	
Destination country	Lower bound (95% Conf Set)	Upper bound (95% Conf Set)	Lower bound (95% Conf Set)	Upper bound (95% Conf Set)	Lower bound (95% Conf Set)	Upper bound (95% Conf Set)	Lower bound (95% Conf Set)	Upper bound (95% Conf Set)
Argentina	139.02	456.86	162.00	388.00	172.32	389.33	239.79	373.25
Japan	493.29	1,729.79	637.37	1,421.89	699.23	1,427.77	868.00	1,362.06
United States	307.73	1,011.68	414.85	841.07	424.95	846.95	527.52	805.45

Table 6: Fit measure: predicted number of exporters under alternative specifications, for select countries in the year 2005

Destination Country	Observed Data	Predictions from Model					
		Via Maximum Likelihood, under Perfect Foresight	Via Maximum Likelihood, under two-step approach	Via moment inequalities, using identified set		Via moment inequalities, using confidence set	
				Lower bound	Upper Bound	Lower bound	Upper Bound
Argentina	46	41.00	40.13	42.00	46.90	31.08	52.24
Japan	5	2.41	1.84	4.51	7.20	2.31	12.66
United States	24	15.69	19.04	19.26	24.28	12.52	34.24

Table 7: Counterfactual predictions: Export participation and export volume after 40% decrease in entry costs, in select countries in the year 2005

	Predictions from Model					
Destination Country	Via Maximum Likelihood, under Perfect Foresight	Via Maximum Likelihood, under two-step approach	Via moment inequalities, using identified set		Via moment inequalities, using confidence set	
			Lower bound	Upper Bound	Lower bound	Upper Bound
Panel 1: Export participation						
Argentina	61.64	60.96	64.35	68.18	55.52	73.48
Japan	16.36	14.56	17.45	21.90	10.28	32.69
United States	40.41	45.50	44.91	52.26	35.01	61.66
Panel 2: Export volume (in $ millions)						
Argentina	17.86	17.26	23.08	25.98	20.51	28.95
Japan	40.31	35.74	62.90	71.47	49.94	89.96
United States	74.38	90.13	100.01	107.93	88.36	120.40

Table 8: Counterfactual predictions: Comparison of alternative specifications relative to the perfect foresight model, after 40% decrease in entry costs in the year 2005

Destination Country	Prediction Via Maximum Likelihood, under Perfect Foresight	% change in prediction when comparing perfect foresight: Maximum Likelihood, under two-step approach	% change in prediction when comparing perfect foresight: Moment inequalities, using identified set: Lower bound	% change in prediction when comparing perfect foresight: Moment inequalities, using identified set: Upper Bound
Panel 1: Export participation				
Argentina	61.64	-1.10%	4.39%	10.61%
Japan	16.36	-11.02%	6.67%	33.87%
United States	40.41	12.58%	11.13%	29.31%
Panel 2: Export volume (in $ millions)				
Argentina	17.86	-3.36%	29.24%	45.51%
Japan	40.31	-11.31%	56.07%	77.32%
United States	74.38	21.18%	34.46%	45.11%

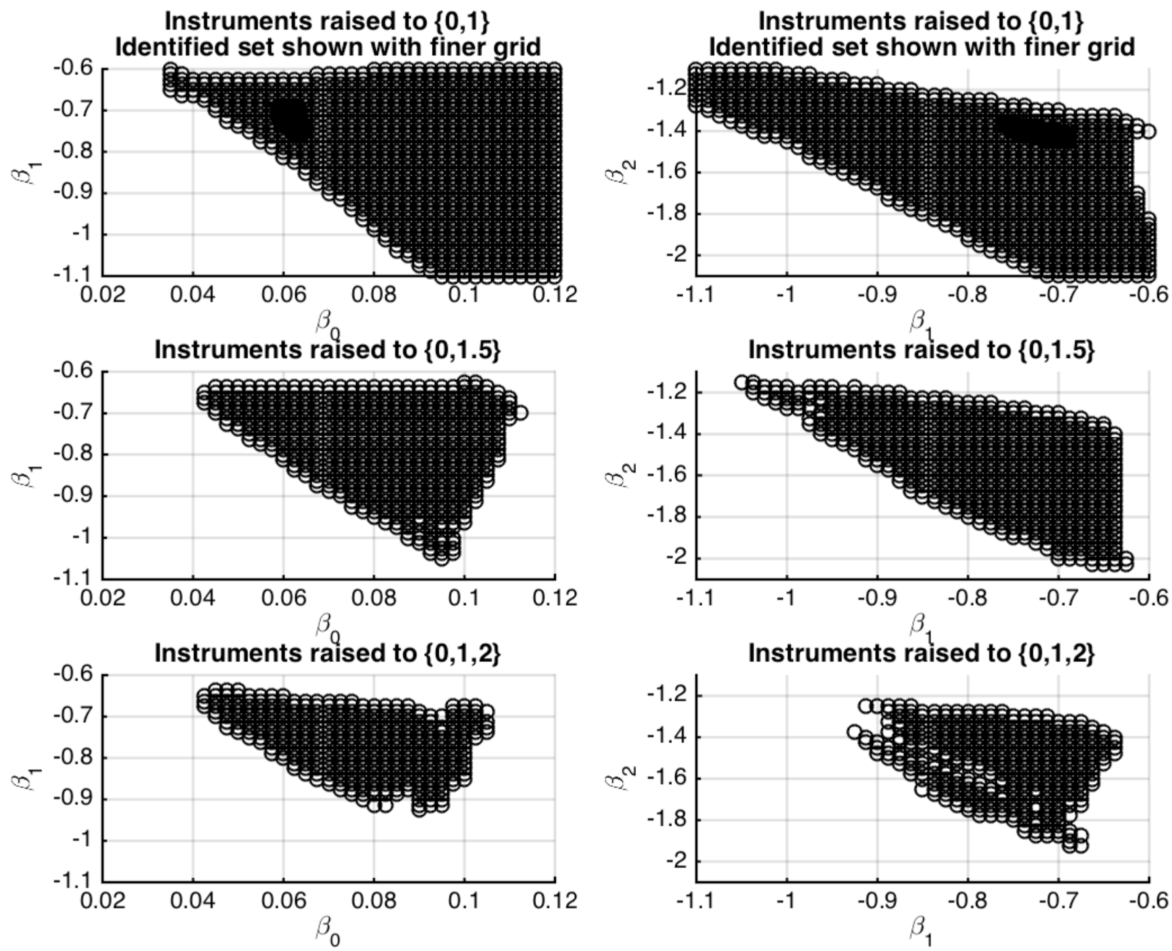
Table 9: Counterfactual predictions: Export participation and export volume after 20% Chilean peso depreciation, in select countries in the year 2005

	Predictions from Model					
Destination Country	Via Maximum Likelihood, under Perfect Foresight	Via Maximum Likelihood, under two-step approach	Via moment inequalities, using identified set		Via moment inequalities, using confidence set	
			Lower bound	Upper Bound	Lower bound	Upper Bound
Panel 1: Export participation						
Argentina	43.91	42.89	49.05	54.96	37.43	65.14
Japan	5.64	4.59	12.84	20.86	8.78	29.93
United States	22.16	31.08	30.06	44.00	22.11	68.94
Panel 2: Export volume (in $ millions)						
Argentina	35.69	33.41	50.70	55.33	43.99	64.51
Japan	83.75	71.69	140.87	175.25	115.91	202.80
United States	153.36	201.03	217.14	238.89	195.01	255.04

Table 10: Counterfactual predictions: Percentage changes in export participation and volume under moment inequalities, in select countries in the year 2005

Destination Country	Observed data	Predictions at baseline using identified set		% Change relative to baseline prediction from a 40% decrease in export entry costs		% Change relative to baseline prediction from a 20% decrease in the real exchange rate	
		Lower bound	Upper Bound	Lower bound	Upper Bound	Lower bound	Upper Bound
Panel 1: Export participation							
Argentina	46.00	45.53	46.90	14%	21%	2%	12%
Japan	5.00	5.30	7.20	38%	69%	25%	67%
United States	24.00	20.79	24.28	27%	43%	9%	36%
Panel 2: Export volume (in $ millions)							
Argentina	27.60	18.91	21.75	3%	16%	37%	49%
Japan	9.91	37.97	51.35	9%	33%	44%	72%
United States	51.49	80.41	86.32	6%	14%	40%	49%

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



Appendix

A.1 Proxy for export revenue: Details

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

Given the expression for firm i's potential export revenue in j in equation (4), 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 at t as

$$R_{jt} = \int_{i \in N_t} d_{ijt} r_{ijt} di = \left[\frac{\eta}{\eta - 1} \frac{\tau_{jt} c_t}{P_{jt}} \right]^{1-\eta} Y_{jt} e_{jt}^{\eta} V_{jt}, \tag{35}$$

where V_{jt} is defined as

$$V_{jt} = \int_{i \in N_t} d_{ijt} a_{it}^{(1-\eta)} di. \tag{36}$$

Note that V_{jt} is simply the sum of the physical productivity terms, a_{it}, (to the power of an exponent that depends on η) across all firms exporting to the destination country j in year t. We can therefore proxy for all the country specific covariates in equation (4) by (R_{jt}/V_{jt}) and rewrite r_{ijt} as

$$r_{ijt} = \frac{a_{it}^{(1-\eta)}}{V_{jt}} R_{jt} \tag{37}$$

In order to proxy for the unobserved firm specific physical productivity of firm i relative to the sum of these physical productivities for all firms exporting to country j, we will use information on the domestic revenue of every firm $i = 1, \ldots, N_t$. Note that, from equation (4), in the case in which $j = h$ and under the standard assumption in trade models that there are no domestic transport costs, $\tau_{iht} = 1$ for every firm i, it holds

$$r_{iht} = \left[\frac{\eta}{\eta - 1} \frac{a_{it} c_t}{P_{ht}} \right]^{1-\eta} Y_{ht}, \tag{38}$$

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

$$\frac{a_{it}^{1-\eta}}{a_{i't}^{1-\eta}} = \frac{r_{iht}}{r_{i'ht}}. \tag{39}$$

Using this expression, we can rewrite the first term in equation (37) as

$$\frac{a_{it}^{(1-\eta)}}{V_{jt}} = \frac{1}{\frac{V_{jt}}{a_{it}^{(1-\eta)}}} = \frac{1}{\int_{s \in N_t} d_{sjt} \left(\frac{a_{st}}{a_{it}} \right)^{(1-\eta)} ds} = \frac{1}{\int_{s \in N_t} d_{sjt} (r_{sht}/r_{iht}) ds}. \tag{40}$$

Plugging back this expression into equation (37), we obtain the expression for r_{ijt} in terms of observable covariates in equation (17).

A.2 Bias in ML Estimates Under Perfect Foresight Assumption

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

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

where $d_{ij} = 1$ if firm i exports to j, $\psi_1 = \psi_2 = 0.5$, and $\nu_{ij} \sim \mathbb{N}(0, \sqrt{2})$ and independent of any other covariate. Mimicking the estimation problem described in Section 4, we assume that the researcher does not observe

$\mathbb{E}[r_{ij}|\mathcal{J}_{ij}]$ but only r_{ij},

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

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

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

Table A.1: Bias in ML Estimates

Model	Distribution of $\mathbb{E}[r_{ij}\|\mathcal{J}_{ij}]$	Distribution of ε_{ijt}	$\hat{\psi}_1$	$\hat{\psi}_2$
1	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.25)$	0.4706 (0.0014)	0.4994 (0.0014)
2	$\mathbb{N}(0,1)$	$\mathbb{N}(0,0.5)$	0.3960 (0.0013)	0.4951 (0.0014)
3	$\mathbb{N}(0,1)$	$\mathbb{N}(0,1)$	0.2426 (0.0010)	0.4865 (0.0013)
4	t_2	t_2	0.1573 (0.0006)	0.4584 (0.0014)
5	t_5	t_5	0.2274 (0.0008)	0.4773 (0.0014)
6	t_{20}	t_{20}	0.2394 (0.0009)	0.4865 (0.0013)
7	t_{50}	t_{50}	0.2436 (0.0010)	0.4872 (0.0013)
8	$\textit{log-normal}(0,1)$	$\textit{log-normal}(0,1)$	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 $var(\nu_{ij}) = 2$. In order to estimate each of the models, we generate 1,000,000 observations from the distributions of ν_{ij}, $\mathbb{E}[r_{ij}|\mathcal{J}_{ij}]$, and $\mathbb{N}(0, 0.25)$ described in columns 2 and 3 and in the main text. Whenever draws are generated from the log-normal distribution, we re-center them at zero. The true parameter values are $\psi_1 = \psi_2 = 0.5$.

The first three models in Table A.1 are specific examples of the general model studied in Yatchew and Griliches (1985). The results in columns 4 and 5 of Table A.1 show that there is downward bias in the estimate of ψ_1 and that the bias is larger as the variance of the expectational error, ε_{ijt} increases. This is consistent with the analytical formula for the bias term in Yatchew and Griliches (1985). In models 4 to 10, we explore departures from the setting studied in Yatchew and Griliches (1985). Specifically, we depart from the assumption that both the unobserved firms' expectations and the expectational errors are normally distributed. In models 4 to 7, we depart from the normal distribution by choosing a distribution both for the unobserved expectations and expectational errors that has fatter tails than the normal distribution. The downward bias in the estimate of ψ_1 persists and it is larger the higher the dispersion in the distribution of unobserved expectations and expectational errors. In models 8 and 9, we depart from the normal distribution by choosing distributions both for unobserved expectations and expectational errors that are asymmetric. Specifically, model 8 assumes distributions that are positively skewed, and model 9 distributions that are negatively skewed. In all cases, the estimate of the coefficient on the covariate that we are measuring with error (i.e. affected by expectational error) is biased downwards.

The estimates shown in Table A.1 condition on the normalization $var(\nu_{ij}) = 2$. In practice, we never know what the variance of the structural error is. However, standard models of international trade as that described in Section 2 imply that the coefficient on the expected export revenues is equal to the inverse of the price elasticity of demand, $1/\eta$. Furthermore, the literature in international trade provides multiple estimates of this price elasticity of demand (Feenstra, 1994; Broda and Weinstein, 2006), Accordingly, we choose the coefficient on expected revenue as the normalizing constant. Given the choice of a particular constant k as the value of ψ_1, we obtain rescaled estimates of the entry cost coefficient by multiplying our estimates $\hat{\psi}_2$ by $k/\hat{\psi}_1$. Given that the true value of k in our simulations is 0.5, the upward bias in the entry costs parameters is given by the ratio

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

Table B.1 reports this number for the nine models described in Table A.1. The results show that, in the different models, assuming perfect foresight implies that we over estimate export entry 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 Entry Costs Estimates

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

A.3 Partial Identification: Example

The data are informative about the joint distribution of $(d_{ijt}, Z_{ijt}, r_{ijt})$ across i, j, and t. Consistently with the possible vectors of instruments discussed in the main text, we assume that we always define Z_{ijt} such that $dist_j \in Z_{ijt}$. We denote the joint distribution of the vector $(d_{ijt}, Z_{ijt}, r_{ijt})$ as $\mathbb{P}(d_{ijt}, Z_{ijt}, r_{ijt})$. For the sake of simplicity in the notation, let's use r^e_{ijt} to denote $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$. Note that we can write

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

where, for any vector $(x_1, \ldots, x_K)$, we use $f(x_1, \ldots, x_K)$ to denote the joint distribution of $(x_1, \ldots, x_K)$. We can further write

$$\mathbb{P}(d_{ijt}, Z_{ijt}, r_{ijt}) = \int f^a(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}) f^a(r_{ijt}|r^e_{ijt}, Z_{ijt}) f^a(r^e_{ijt}|Z_{ijt}) \mathbb{P}(Z_{ijt}) dr^e_{ijt}, \tag{41}$$

where we use $\mathbb{P}(Z_{ijt})$ to denote that the marginal distribution of Z_{ijt} is directly observable in the data. Any structure $S^a \equiv \{f^a(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}),\ f^a(r_{ijt}|r^e_{ijt}, Z_{ijt}),\ f^a(r^e_{ijt}|Z_{ijt})\}$ is admissible as long as it verifies the restrictions imposed in Section 2 and equation (41). The model in Section 2 imposes the following two restrictions on the elements of equation (41). First,

$$\begin{gathered} f^a(d_{ijt}|r^e_{ijt}, r_{ijt}, Z_{ijt}) = f(d_{ijt}|r^e_{ijt}, Z_{ijt}; \gamma^a) = \\ \Big(\Phi\big((\gamma^a_2)^{-1}\big(kr^e_{ijt} - \gamma^a_0 - \gamma^a_1 dist_j\big)\big)\Big)^{d_{ijt}} \Big(1 - \Phi\big((\gamma^a_2)^{-1}\big(kr^e_{ijt} - \gamma^a_0 - \gamma^a_1 dist_j\big)\big)\Big)^{1-d_{ijt}}. \end{gathered} \tag{42}$$

Second, $Z_{ijt} \subset \mathcal{J}_{ijt}$ and, therefore, given the definition of r^e_{ijt} as $\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$, the expectation of the distribution $f(r_{ijt}|r^e_{ijt}, Z_{ijt})$ is equal to r^e_{ijt}. The model presented in Section 2 does not imply any additional restrictions on the elements of equation (41).

Here, we show that γ is partially identified in a model that imposes restrictions that are stronger than those in Section 2. Specifically, we impose the following additional restrictions on the elements of equation (41)

$$\gamma_1 \text{ is known and equal to } 0, \tag{43a}$$

$$r_{ijt} = r^e_{ijt} + \varepsilon_{ijt}, \qquad \varepsilon_{ijt}|(r^e_{ijt}, W_{ijt}) \sim \mathbb{N}(0, \sigma^2_\varepsilon) \tag{43b}$$

$$Z_{ijt} = r^e_{ijt} + W_{ijt} \qquad W_{ijt}|r^e_{ijt} \sim \mathbb{N}((\sigma_w/\sigma_{r^e})\rho_{r^e w}(r^e_{ijt} - \mu_{r^e}), (1 - \rho^2_{r^e w})\sigma^2_w) \tag{43c}$$

$$r^e_{ijt} \sim \mathbb{N}(\mu_{r^e}, \sigma^2_{r^e}) \tag{43d}$$

where $\mu_{r^e} = \mathbb{E}[r^e_{ijt}]$, $\sigma^2_{r^e} = var(R^e_{ijt})$, $\sigma^2_w = var(W_{ijt})$. Below, we show that, even after adding the assumptions in equation (43), we can still find at least two structures

$$S^{a_1} \equiv \{(\gamma_0^{a_1}, \gamma_2^{a_1}), f^{a_1}(r_{ijt}|r^e_{ijt}, Z_{ijt}), f^{a_1}(r^e_{ijt}|Z_{ijt})\},$$
$$S^{a_2} \equiv \{(\gamma_0^{a_2}, \gamma_2^{a_2}), f^{a_2}(r_{ijt}|r^e_{ijt}, Z_{ijt}), f^{a_2}(r^e_{ijt}|Z_{ijt})\},$$

that: (1) verify the restrictions in equations (42) and (43); (2) verify equation (41); and (3) $\gamma^{a_1} \neq \gamma^{a_2}$. If γ is partially identified in this stricter model, it will also be partially identified in the more general model described in Section 2.

Equation (43a) simplifies the identification exercise discussed here because the only parameters that are left to identify are (γ_0, γ_2) –we can set $\gamma_1 = 0$ in equation (42)–. Equation (43b) assumes that the expectational error not only has mean zero and finite variance but is also normally distributed. It implies that the conditional density $f(r_{ijt}|r^e_{ijt}, Z_{ijt})$ is normal:

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

By applying Bayes' rule, both equations (43c) and (43d) jointly determine the conditional density $f(r^e_{ijt}|Z_{ijt})$ entering equation (41).

If we had imposed the assumption that $\rho_{r^e w} = 0$ and, therefore, W_{ijt} is independent of r^e_{ijt}, then the model described in Section 2 and the additional restrictions in equations (43b) and (43c) would have become a simple model of repeated measurements of the unobserved covariate r^e_{ijt}. In this context, as Evdokimov and White (2012) show, one may apply Kotlarski's identity and recover the distribution of r^e_{ijt}. The feature that makes the model described in Section 2 and equations (43b) and (43c) depart from a setting with repeated measurements is that we allow the distribution of W_{ijt} to freely depend on the value of r^e_{ijt}. As we show below, in this case, the parameter vector γ is only partially identified.

Result A.3.1 *There exists empirical distributions of the vector of observable variables (d, Z, X), $\mathcal{P}(d, Z, X)$, such that there are at least two structures S^{a_1} and S^{a_2} for which*

1. *both S^{a_1} and S^{a_2} verify equations (41), (42), and (43);*
2. *$\gamma^{a_1} \neq \gamma^{a_2}$.*

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 (42) as

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

The parameter γ_2 only enter likelihood function in equation (41) 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 $\rho_{r^e w}$ is assumed to be equal to zero.*

Proof: From equations (43b), (43c) and (43d), 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:

$$\begin{aligned}
\sigma_r^2 &= \sigma_{r^e}^2 + \sigma_\varepsilon^2,\\
\sigma_z^2 &= \sigma_{r^e}^2 + \sigma_w^2 + 2\rho_{r^e w}\sigma_{r^e}\sigma_w,\\
\sigma_{rz} &= \sigma_{r^e}^2 + \rho_{r^e w}\sigma_{r^e}\sigma_w
\end{aligned} \tag{44}$$

The left hand side of these three equations is directly observed in the data. If we impose the assumption that $\rho_{r^e w} = 0$, then $\sigma_{rz} = \sigma^2_{r^e}$ and, therefore, from Lemma A.3.1, the vector γ is point identified. If we allow $\rho_{r^e w}$

to be different from zero, this system of equations in equation (45) only allows to define bounds on $\sigma^2_{r^e}$. Note that we can rewrite the system of equations in equation (45) as

$$\begin{aligned}\sigma_r^2 &= \sigma^2_{r^e} + \sigma^2_\varepsilon, \\ \sigma_z^2 &= \sigma^2_{r^e} + \sigma^2_w + 2\sigma_{r^e w} \\ \sigma_{rz} &= \sigma^2_{r^e} + \sigma_{r^e w}.\end{aligned} \tag{45}$$

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

A.4 Proof of Theorem 1

Lemma 1 *Let $L(d_{ijt}|\mathcal{J}_{ijt};\theta)$ denote the log-likelihood conditional on $\mathcal{J}_{ijt}$. Suppose the assumptions in equations (2), (6), and (8) hold and we impose the normalization $\eta^{-1} = k$. Then:*

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

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

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

The score function is given by

$$\begin{aligned}&\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt};\theta)}{\partial\theta} = \\ &\mathbb{E}\Bigg[d_{ijt}\frac{1}{1-\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)}\frac{\partial(1-\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)))}{\partial\theta} \\ &+(1-d)\frac{1}{\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}\frac{\partial\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}{\partial\theta}\Bigg|\mathcal{J}_{ijt}\Bigg] = 0.\end{aligned} \tag{47}$$

Reordering terms

$$\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt};\theta)}{\partial\theta} = \mathbb{E}\Bigg[\frac{\partial\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))/\partial\theta}{\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}\Bigg[d_{ijt}\frac{\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}{1-\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))}\times \tag{48}$$

$$\times\frac{\partial(1-\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)))/\partial\theta}{\partial\Phi(-k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j))/\partial\theta} + (1-d_{ijt})\Bigg|\mathcal{J}_{ijt}\Bigg] = 0. \tag{49}$$

Given that

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

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

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

we can simplify:

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

Equation (46) follows by symmetry of the function $\Phi(\cdot)$. ■

Lemma 2 *Suppose the assumptions in equations* (7) *and* (8) *hold. Then*

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma_\nu^{-1}(kr_{ijt}-\beta_0-\beta_1 dist_j))}{\Phi(\sigma_\nu^{-1}(kr_{ijt}-\beta_0-\beta_1 dist_j))}\bigg|\mathcal{J}_{ijt}\right] \geq \mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\bigg|\mathcal{J}_{ijt}\right]. \quad (50)$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the assumptions in equations (7) and (8) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. From equations (2), (6) and the assumption that $dist_j \in \mathcal{J}_{ijt}$ it follows that d_{ijt} may be written as a function of the vector $(\mathcal{J}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, d_{ijt}] = 0$. Since

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

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

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j+k\varepsilon_{ijt}))}{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)+k\varepsilon_{ijt})}\bigg|\mathcal{J}_{ijt}\right]$$
$$\geq$$
$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\bigg|\mathcal{J}_{ijt}\right].$$

Equation (50) follows from the equality $kr_{ijt} = k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] + k\varepsilon_{ijt}$. ■

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

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-(1-d_{ijt})\bigg|Z_{ijt}\right]=0. \quad (51)$$

and

$$\mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma_\nu^{-1}(kr_{ijt}-\beta_0-\beta_1 dist_j))}{\Phi(\sigma_\nu^{-1}(kr_{ijt}-\beta_0-\beta_1 dist_j))}\bigg|Z_{ijt}\right] \geq \mathbb{E}\left[d_{ijt}\frac{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|Z_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\bigg|Z_{ijt}\right]. \quad (52)$$

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

Lemma 3 *Let* $L(d_{ijt}|\mathcal{J}_{ijt};\theta)$ *denote the log-likelihood conditional on* $\mathcal{J}_{ijt}$*. Suppose the assumptions in equations* (2), (6), *and* (8) *hold and we impose the normalization* $\eta^{-1} = k$*. Then:*

$$\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt};\theta)}{\partial\theta} = \mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}-d_{ijt}\bigg|\mathcal{J}_{ijt}\right]=0. \quad (53)$$

Proof: From equation (47), reordering terms

$$\frac{\partial L(d_{ijt}|\mathcal{J}_{ijt};\theta)}{\partial\theta} = \mathbb{E}\left[\frac{\partial(1-\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}{1-\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\left[d_{ijt}+\right.\right.$$
$$\left.\left.+(1-d_{ijt})\frac{1-\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\frac{\partial\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))/\partial\theta}{\partial(1-\Phi(-\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j)))/\partial\theta}\bigg|\mathcal{J}_{ijt}\right]=0.$$

Given that

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

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

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

we can simplify:

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

Equation (53) follows by symmetry of the function $\Phi(\cdot)$. ■

Lemma 4 *Suppose the assumptions in equations* (7) *and* (8) *hold. Then*

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma_\nu^{-1}(kr_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma_\nu^{-1}(kr_{ijt}-\beta_0-\beta_1 dist_j))}\middle|\mathcal{J}_{ijt}\right]$$
$$\geq$$
$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\middle|\mathcal{J}_{ijt}\right]. \tag{54}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the assumptions in equations (7) and (8) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt},\nu_{ijt}] = 0$. From equations (2), (6) and the assumption that $dist_j \in \mathcal{J}_{ijt}$ it follows that d_{ijt} may be written as a function of the vector $(\mathcal{J}_{ijt},\nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt},\nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt},d_{ijt}] = 0$. Since

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

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

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

Equation (54) follows from the equality $kr_{ijt} = k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] + k\varepsilon_{ijt}$. ■

Corollary 3 *Suppose $Z_{ijt} \in \mathcal{J}_{ijt}$. Then:*

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))} - d_{ijt}\middle|Z_{ijt}\right] = 0. \tag{55}$$

and

$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma_\nu^{-1}(kr_{ijt}-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma_\nu^{-1}(kr_{ijt}-\beta_0-\beta_1 dist_j))}\middle|Z_{ijt}\right]$$
$$\geq$$
$$\mathbb{E}\left[(1-d_{ijt})\frac{\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|Z_{ijt}]-\beta_0-\beta_1 dist_j))}{1-\Phi(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j))}\middle|Z_{ijt}\right]. \tag{56}$$

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

Proof of Theorem 1 Combining equations (51) and (52), we obtain the inequality defined by equations (21) and (22a). Combining equations (55) and (56), we obtain the inequality defined by equations (21) and (22b). ■

A.5 Proof of Theorem 2

Lemma 5 *Suppose equations* (2) *and* (6) *hold. Then,*

$$\mathbb{E}[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j - \nu_{ijt}\big)|\mathcal{J}_{ijt}] \geq 0. \tag{57}$$

Proof: From equations (2) and (6),

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

This implies

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

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

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

$$\mathbb{E}\Big[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + (1-d_{ijt})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}\Big] \geq 0. \tag{58}$$

Proof: From equation (57),

$$\mathbb{E}\big[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\big|\mathcal{J}_{ijt}\big] - \mathbb{E}\big[d_{ijt}\nu_{ijt}\big|\mathcal{J}_{ijt}\big] \geq 0. \tag{59}$$

Since the assumption in equation (8) implies that $\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}] = 0$, it follows that

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

and we can rewrite equation (59) as

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

Applying the Law of Iterated Expectations, it follows that

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

and we can rewrite equation (60) as

$$\mathbb{E}\big[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + (1-d_{ijt})\mathbb{E}[\nu_{ijt}|d_{ijt}=0,\mathcal{J}_{ijt}]\big|\mathcal{J}_{ijt}\big] \geq 0. \tag{61}$$

Using the definition of d_{ijt} in equation (6), it follows

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

and, following equation (8), we can rewrite

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

Equation (58) follows by applying this equality to equation (61). ■

Lemma 7 *Suppose the assumptions in equation* (7) *and* (8) *hold. Then*

$$\mathbb{E}\Big[d_{ijt}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] = \mathbb{E}\Big[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] \tag{62}$$

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

$$\mathbb{E}\Big[d_{ijt}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] = \mathbb{E}\Big[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] + \mathbb{E}\Big[kd_{ijt}\varepsilon_{ijt}\Big|\mathcal{J}_{ijt}\Big]. \quad (63)$$

From equations (7) and (8), $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. From equations (2), (6) and the assumption that $dist_j \in \mathcal{J}_{ijt}$ it follows that d_{ijt} is a function of the vector $(\mathcal{J}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, d_{ijt}] = 0$ and, applying the Law of Iterated Expectations,

$$\mathbb{E}\big[kd_{ijt}\varepsilon_{ijt}\big|\mathcal{J}_{ijt}\big] = \mathbb{E}\big[kd_{ijt}\mathbb{E}\big[\varepsilon_{ijt}\big|\mathcal{J}_{ijt}, d_{ijt}\big]\big|\mathcal{J}_{ijt}\big] = \mathbb{E}\big[kd_{ijt} \times 0\big|\mathcal{J}_{ijt}\big] = 0.$$

Applying this result to equation (63) yields equation (62).

Lemma 8 *Suppose the assumptions in equation* (7) *and* (8) *hold. Then*

$$\begin{gathered}\mathbb{E}\Big[(1-d_{ijt})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}\Big] \\ \geq \\ \mathbb{E}\Big[(1-d_{ijt})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}\Big]\end{gathered} \quad (64)$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the assumptions in equations (7) and (8) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. From equations (2), (6) and the assumption that $dist_j \in \mathcal{J}_{ijt}$ it follows that d_{ijt} is a function of the vector $(\mathcal{J}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, d_{ijt}] = 0$. Since

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

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

$$\begin{gathered}\mathbb{E}\Big[(1-d_{ijt})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + k\varepsilon_{ijt})\big)}{1-\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j + k\varepsilon_{ijt})\big)}\Big|\mathcal{J}_{ijt}\Big] \\ \geq \\ \mathbb{E}\Big[(1-d_{ijt})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}\Big]\end{gathered}$$

Equation (64) follows from the equality $kr_{ijt} = k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] + k\varepsilon_{ijt}$. ■

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

$$\mathbb{E}\Big[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + (1-d_{ijt})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq 0, \quad (65)$$

$$\mathbb{E}\Big[d_{ijt}\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] = \mathbb{E}\Big[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|Z_{ijt}\Big], \quad (66)$$

and

$$\begin{gathered}\mathbb{E}\Big[(1-d_{ijt})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \\ \geq \\ \mathbb{E}\Big[(1-d_{ijt})\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{1-\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big].\end{gathered} \quad (67)$$

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* (6) *hold. Then,*

$$\mathbb{E}[-(1-d_{ijt})\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j-\nu_{ijt}\big)|\mathcal{J}_{ijt}]\geq 0. \tag{68}$$

Proof: From equations (2) and (6),

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

This implies

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

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

Lemma 10 *Suppose equations* (2)*,* (6) *and* (8) *hold. Then*

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

Proof: From equation (68),

$$\mathbb{E}\big[-(1-d_{ijt})\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\big|\mathcal{J}_{ijt}\big]+\mathbb{E}\big[(1-d_{ijt})\nu_{ijt}\big|\mathcal{J}_{ijt}\big]\geq 0. \tag{70}$$

Since the assumption in equation (8) implies that $\mathbb{E}[\nu_{ijt}|\mathcal{J}_{ijt}]=0$, it follows that

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

and we can rewrite equation (70) as

$$\mathbb{E}\big[d_{ijt}\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\big|\mathcal{J}_{ijt}\big]-\mathbb{E}\big[d_{ijt}\nu_{ijt}\big|\mathcal{J}_{ijt}\big]\geq 0. \tag{71}$$

Applying the Law of Iterated Expectations, it follows that

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

and we can rewrite equation (71) as

$$\mathbb{E}\big[-(1-d_{ijt})\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)-d_{ijt}\mathbb{E}[\nu_{ijt}|d_{ijt}=1,\mathcal{J}_{ijt}]\big|\mathcal{J}_{ijt}\big]\geq 0. \tag{72}$$

Using the definition of d_{ijt} in equation (6), it follows

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

and, following equation (8), we can rewrite

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

Equation (69) follows by applying this equality to equation (72). ■

Lemma 11 *Suppose the assumptions in equation* (7) *and* (8) *hold. Then*

$$\mathbb{E}\Big[-(1-d_{ijt})\big(kr_{ijt}-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big]=\mathbb{E}\Big[-(1-d_{ijt})\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]-\beta_0-\beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] \tag{73}$$

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

$$\mathbb{E}\Big[-(1-d_{ijt})\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] =$$
$$\mathbb{E}\Big[-(1-d_{ijt})\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|\mathcal{J}_{ijt}\Big] - \mathbb{E}\Big[k(1-d_{ijt})\varepsilon_{ijt}\Big|\mathcal{J}_{ijt}\Big]. \tag{74}$$

From equations (7) and (8), $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. From equations (2), (6) and the assumption that $dist_j \in \mathcal{J}_{ijt}$ it follows that d_{ijt} is a function of the vector $(\mathcal{J}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, d_{ijt}] = 0$ and, applying the Law of Iterated Expectations,

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

Applying this result to equation (74) yields equation (73).

Lemma 12 *Suppose the assumptions in equation* (7) *and* (8) *hold. Then*

$$\mathbb{E}\Big[d_{ijt}\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}\Big] \geq \mathbb{E}\Big[d_{ijt}\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|\mathcal{J}_{ijt}\Big] \tag{75}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the assumptions in equations (7) and (8) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. From equations (2), (6) and the assumption that $dist_j \in \mathcal{J}_{ijt}$ it follows that d_{ijt} is a function of the vector $(\mathcal{J}_{ijt}, \nu_{ijt})$; i.e. $d_{ijt} = d(\mathcal{J}_{ijt}, \nu_{ijt})$. Therefore, $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, d_{ijt}] = 0$. Since

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

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

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

Equation (75) follows from the equality $kr_{ijt} = k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] + k\varepsilon_{ijt}$. ■

Corollary 5 *Suppose $Z_{ijt} \in \mathcal{J}_{ijt}$ then*

$$\mathbb{E}\Big[-(1-d_{ijt})\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big) + d_{ijt}\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq 0. \tag{76}$$

$$\mathbb{E}\Big[-(1-d_{ijt})\big(kr_{ijt} - \beta_0 - \beta_1 dist_j\big)\Big|Z_{ijt}\Big] = \mathbb{E}\Big[-(1-d_{ijt})\big(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j\big)\Big|Z_{ijt}\Big] \tag{77}$$

and

$$\mathbb{E}\Big[d_{ijt}\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \geq \mathbb{E}\Big[d_{ijt}\sigma_\nu \frac{\phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}{\Phi\big(\sigma_\nu^{-1}(k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] - \beta_0 - \beta_1 dist_j)\big)}\Big|Z_{ijt}\Big] \tag{78}$$

Proof of Theorem 2 Combining equations (65), (66), and (67) we obtain the inequality defined by equations (23) and (24a). Combining equations (76), (77), and (78) we obtain the inequality defined by equations (23) and (24b). ■

A.6 Proof of Theorem 3

Lemma 13 *Suppose the assumptions in equations* (7), (8), *and* (9) *hold. Then*

$$\mathbb{E}\Big[\frac{1 - \Phi(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j))}{\Phi(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j))}\Big|\mathcal{J}_{ijt}\Big] \geq \mathbb{E}\Big[\frac{1 - \mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\Big|\mathcal{J}_{ijt}\Big]. \tag{79}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the assumptions in equations (7) and (8) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. Since

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

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

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

Equation (79) follows from the equality $kr_{ijt} = k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] + k\varepsilon_{ijt}$ and the definition of $\mathcal{P}_{ijt}$ in equation (9). ■

Lemma 14 *Suppose the assumptions in equations* (7), (8), *and* (9) *hold. Then*

$$\mathbb{E}\left[\frac{\Phi(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j))}{1-\Phi(\sigma_\nu^{-1}(kr_{ijt} - \beta_0 - \beta_1 dist_j))}\middle|\mathcal{J}_{ijt}\right] \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\middle|\mathcal{J}_{ijt}\right]. \tag{80}$$

Proof: It follows from the definition of ε_{ijt} as $\varepsilon_{ijt} = r_{ijt} - \mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}]$ and the assumptions in equations (7) and (8) that $\mathbb{E}[\varepsilon_{ijt}|\mathcal{J}_{ijt}, \nu_{ijt}] = 0$. Since

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

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

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

Equation (79) follows from the equality $kr_{ijt} = k\mathbb{E}[r_{ijt}|\mathcal{J}_{ijt}] + k\varepsilon_{ijt}$ and the definition of $\mathcal{P}_{ijt}$ in equation (9). ■

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

$$B_2(Z_{ijt};\theta) \geq \mathbb{E}\left[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\middle|Z_{ijt}\right], \tag{81}$$

and

$$B_1(Z_{ijt};\theta) \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\middle|Z_{ijt}\right]. \tag{82}$$

Proof: It follows from lemmas 13 and 14, the definitions of $B_1(Z_{ijt};\theta)$ and $B_2(Z_{ijt};\theta)$ in equations (27) and (28), and the Law of Iterated Expectations. ■

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

$$\mathbb{E}\Big[\frac{1-Y}{Y}\Big] \geq \frac{1-\mathbb{E}[Y]}{\mathbb{E}[Y]}, \tag{83}$$

and

$$\mathbb{E}\Big[\frac{Y}{1-Y}\Big] \geq \frac{\mathbb{E}[Y]}{1-\mathbb{E}[Y]}. \tag{84}$$

Proof: We can rewrite the left hand side of equation (83) as

$$\mathbb{E}\Big[\frac{1-Y}{Y}\Big] = \mathbb{E}\Big[\frac{1}{Y} - 1\Big] = \mathbb{E}\Big[\frac{1}{Y}\Big] - 1, \tag{85}$$

and the right hand side of equation (83) as

$$\frac{1-\mathbb{E}[Y]}{\mathbb{E}[Y]} = \frac{1}{\mathbb{E}[Y]} - 1. \tag{86}$$

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

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

Equations (85), (86), and (87) imply that equation (83) holds.

Define a random variable $X = 1 - Y$ and rewrite the left hand side of equation (84) 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 (85), (86), and (87) 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 (84) holds. ■

Corollary 6 *Suppose $\mathcal{P}_{ijt}$ is defined as in equation* (9)*, then*

$$\mathbb{E}\Big[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big] \geq \frac{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}, \tag{88}$$

and

$$\mathbb{E}\Big[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\Big|Z_{ijt}\Big] \geq \frac{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}. \tag{89}$$

Proof: Equation (9) implies that the support of $\mathcal{P}_{ijt}$ is the interval $(0,1)$. Therefore, Lemma 16 implies that equations (88) and (89) hold. ■

Proof of Theorem 3 Combining equations (81) and (88),

$$B_2(Z_{ijt};\theta) \geq \mathbb{E}\left[\frac{1-\mathcal{P}_{ijt}}{\mathcal{P}_{ijt}}\middle|Z_{ijt}\right] \geq \frac{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]},$$

and, reordering terms, we obtain the inequality

$$\frac{1}{1+B_2(Z_{ijt};\theta)} \leq \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]. \tag{90}$$

Combining equations (82) and (89),

$$B_1(Z_{ijt};\theta) \geq \mathbb{E}\left[\frac{\mathcal{P}_{ijt}}{1-\mathcal{P}_{ijt}}\middle|Z_{ijt}\right] \geq \frac{\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}{1-\mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]}$$

and, reordering terms, we obtain the inequality

$$\frac{B_1(Z_{ijt};\theta)}{1+B_1(Z_{ijt};\theta)} \geq \mathbb{E}[\mathcal{P}_{ijt}|Z_{ijt}]. \tag{91}$$

Combining the inequalities in equations (90) and (91) we obtain equation (26). ■

A.7 Bounds on counterfactual choice probabilities

We may use equations (31), (32) and (33) to define bounds on expected export probabilities in the counterfactual scenarios described in Sections 2.5 and 2.6.

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

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

In this case, the export probability is defined in equation (15) as $\mathcal{P}^1_{ijt}$. Using expressions analogous to equations (31), (32) and (33), we may define bounds the expectation of $\mathcal{P}^1_{ijt}$ conditional on any particular value or set of values of Z_{ijt} as follows

$$\underline{\mathcal{P}}^1(Z_{ijt}) \leq \mathcal{P}^1(Z_{ijt}) \leq \overline{\mathcal{P}}^1(Z_{ijt}), \tag{92}$$

where

$$\underline{\mathcal{P}}^1(Z_{ijt}) = \min_{\gamma \in \Theta_{all}} \frac{1}{1 + B_2^1(Z_{ijt};\gamma)}, \tag{93}$$

$$\overline{\mathcal{P}}^1(Z_{ijt}) = \max_{\gamma \in \Theta_{all}} \frac{B_1^1(Z_{ijt};\gamma)}{1 + B_1^1(Z_{ijt};\gamma)}, \tag{94}$$

with

$$B_1^1(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}\bigg| Z_{ijt}\right], \tag{95}$$

$$B_2^1(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}{\Phi\big(\sigma_\nu^{-1}\big(kr_{ijt} - 0.6\beta_0 - 0.6\beta_1 dist_j\big)\big)}\bigg| Z_{ijt}\right]. \tag{96}$$

Section 2.6 describes a counterfactual scenario in which, due to a currency depreciation, the potential revenue from exporting becomes

$$r_{ijt}^2 = r_{ijt}(1.2)^\eta$$

In this case, the export probability is defined in equation (16) as $\mathcal{P}^2_{ijt}$. Using expressions analogous to equations (31), (32) and (33), we may define bounds the expectation of $\mathcal{P}^2_{ijt}$ conditional on any particular value or set of values of Z_{ijt} as follows

$$\underline{\mathcal{P}}^2(Z_{ijt}) \leq \mathcal{P}^2(Z_{ijt}) \leq \overline{\mathcal{P}}^2(Z_{ijt}), \tag{97}$$

where

$$\underline{\mathcal{P}}^2(Z_{ijt}) = \min_{\gamma \in \Theta_{all}} \frac{1}{1 + B_2^2(Z_{ijt};\gamma)}, \tag{98}$$

$$\overline{\mathcal{P}}^2(Z_{ijt}) = \max_{\gamma \in \Theta_{all}} \frac{B_1^2(Z_{ijt};\gamma)}{1 + B_1^1(Z_{ijt};\gamma)}, \tag{99}$$

with

$$B_1^2(Z_{ijt};\theta) = \mathbb{E}\left[\frac{\Phi\big(\sigma_\nu^{-1}\big(kr_{ijt}(1.2)^\eta - \beta_0 - \beta_1 dist_j\big)\big)}{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ijt(1.2)^\eta} - \beta_0 - \beta_1 dist_j\big)\big)}\bigg| Z_{ijt}\right], \tag{100}$$

$$B_2^2(Z_{ijt};\theta) = \mathbb{E}\left[\frac{1 - \Phi\big(\sigma_\nu^{-1}\big(kr_{ijt}(1.2)^\eta - \beta_0 - \beta_1 dist_j\big)\big)}{\Phi\big(\sigma_\nu^{-1}\big(kr_{ijt}(1.2)^\eta - \beta_0 - \beta_1 dist_j\big)\big)}\bigg| Z_{ijt}\right]. \tag{101}$$

Besides computing expected probabilities of export 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 entry costs β_0 and β_1 using the expressions in Theorem 3:

$$\min_{\gamma\in\Theta_{all}} \frac{1+B_2^1(Z_{ijt};\gamma)}{\frac{B_1(Z_{ijt};\gamma)}{1+B_1(Z_{ijt};\gamma)}} \leq \frac{\mathcal{P}_{ijt}^1(Z_{ijt})}{\mathcal{P}_{ijt}(Z_{ijt})} \leq \max_{\gamma\in\Theta_{all}} \frac{\frac{B_1^1(Z_{ijt};\gamma)}{1+B_1^1(Z_{ijt};\gamma)}}{1+B_2(Z_{ijt};\gamma)}, \tag{102}$$

where $B_1(Z_{ijt};\gamma)$ and $B_2(Z_{ijt};\gamma)$ are defined in equations (27) and (28), respectively; and $B_1^1(Z_{ijt};\gamma)$ and $B_2^1(Z_{ijt};\gamma)$ are defined in equations (95) and (96), respectively. For the counterfactual scenario described in Section 2.6, we can construct bounds analogous to that in equation (102) using $B_1^2(Z_{ijt};\gamma)$ and $B_2^2(Z_{ijt};\gamma)$ instead of $B_1^1(Z_{ijt};\gamma)$ and $B_2^1(Z_{ijt};\gamma)$.

A.8 Related econometric literature on discrete choice with endogenous regressors

There are three additional alternative models that build on conditional independence assumptions to identify the parameters of binary choice models with endogenous regressors: (1) the IV model of Chesher (2010) and Chesher (2011); (2) the triangular system model that motivates the use of control function methods; and, (3) the special regressor approach.

As Blundell and Powell (2003) show, even when the econometrician observes an excluded variable that is independent of the error term in the random utility function, semi-parametric and non-parametric binary response models are generally not point identified. Chesher (2010) shows that this result holds even if we impose parametric restrictions both on the random utility function and on the marginal distribution of the error term. Chesher (2010) provides the inequalities that sharply define the identified set under the assumption that the econometrician observes a excluded variable that is independent of the error term (i.e. fully independent instrument). While Chesher (2010) focuses on the case in which the endogenous variable is continuous, Chesher (2011) performs an analogous exercise for the case in which it is discrete.[19] Following our notation, Chesher (2010) and Chesher (2011) assume that $(\nu+\varepsilon)|Z \sim (\nu+\varepsilon)$.

Our model is stricter than the one proposed in Chesher (2010) in that we formally define the error term as the sum of two different unobserved components—a structural error, ν, and an expectational error, ε—and we only allow for endogeneity that is due to expectational error. However, in another sense, our model can be viewed as more flexible than that in Chesher (2010) because our identification strategy does not assume that the aggregate error term, $(\nu+\epsilon)$, is fully independent of the instrument vector, Z. We only need to impose mean independence between this instrument and ε. This weaker independence assumption of our statistical model matches the assumptions common to economic models of agents with rational expectations.[20]

The triangular system control function model is attractive because, under certain conditions, it point identifies the parameters of interest. In particular, Blundell and Powell (2004) obtain point identification by applying a control-function approach. This approach assumes that the endogenous variables are determined by an equation $X = m(Z, W)$ such that there is a one-to-one mapping from the latent variables, W, to the endogenous variables, X, at each value of the instrument vector, Z.[21] The latent variable W is assumed to verify: $(\nu+\varepsilon)|X, Z \sim (\nu+\varepsilon)|X, W \sim (\nu+\varepsilon)|W$. In contrast, our model is a single-equation model: there is no specification of any structural equation that would imply that the error term $(\nu+\varepsilon)$ is independent of the endogenous regressor X, conditional on some latent variable, W.[22]

The special regressor approach assumes that the aggregate unobservable component, $(\nu+\varepsilon)$, is distributed independently of a continuously distributed explanatory variable (i.e. special regressor) and impose a particular index restriction (Lewbel (2000)). This model is point identified only if the special regressor has large support.

[19] Other papers that explore this IV approach are Chesher and Smolinski (2010), Chesher et al. (2011), Chesher and Rosen (2012).

[20] The only type of independence that the rational expectations assumption imposes on the definition of the expectational error is *mean* independence between this error and any variable contained in the information set of the agent.

[21] This restriction rules out cases in which there are discrete endogenous variables. In our case, we allow for discrete endogenous variables as long as the measurement error verifies the restriction $\mathbb{E}[\varepsilon|X^*, Z] = 0$.

[22] Other papers that explore the use of control function methods for the identification of binary choice models in semi- and non-parametric settings are Blundell and Powell (2003), Chesher (2003), Chesher (2005), Chesher (2007), Vytlacil and Yildiz (2007), Florens et al. (2008), Imbens and Newey (2009), and Shaikh and Vytlacil (2011).

In an application in which the only source of endogeneity is expectational error, this approach implies that one covariate, the special regressor with large support, is measured without error.[23] Our statistical model allows all the regressors to contain expectational error.

[23] Other papers that explore the special regressor approach are Magnac and Maurin (2007) and Magnac and Maurin (2008).