Quantifying the Extensive Margin of Trade*

James E. Anderson
Boston College
NBER

Yoto V. Yotov
Drexel University

December 4, 2020

Abstract

We propose a short-run model of the extensive margin of trade and deploy it to distinguish and quantify domestic and cross-border margins. Our empirical focus is on the *domestic extensive margin* of trade (domestic distribution of a product) and its importance for quantifying globalization effects on the international extensive margin of trade. We build a dataset that combines data on the domestic extensive margin and the standard international extensive margin. It reveals significant and intuitive variation in the domestic extensive margin across countries and over time. We quantify the extensive margin effects of European Union (EU) integration, 2008-2018, and demonstrate that these effects cannot be identified without the domestic extensive margin. We find strong and highly heterogeneous effects, both across countries and directionally.

JEL Classification Codes: F13, F14, F16
Keywords: Extensive Margin, Domestic Extensive Margin, Globalization, Gravity

*We thank João Santos Silva for kindly sharing the original programming codes for the FLEX estimator. We also thank Rebecca Freeman and Stephan-Alfons Nolte for very helpful advice and guidance to the data. Contact information: James E. Anderson, Department of Economics, Boston College, Chestnut Hill, MA 02467, USA. Yoto V. Yotov, School of Economics, Drexel University, Philadelphia, PA 19104, USA.

"Trade diversification is a national imperative for the Government of Canada. Over the next six years, starting in 2018-19, Canada's export diversification strategy will invest $1.1 billion to help Canadian businesses access new markets."

(Government of Canada, March 3, 2020)

"Increased diversification is associated with lower output volatility and greater macroeconomic stability [in low-income countries]. There is both a growth payoff and a stability payoff to diversification, underscoring the case for paying close attention to policies that facilitate diversification and structural transformation."

(IMF, March, 2014)

1 Introduction

The opening quotes highlight export diversification as a policy imperative for developing and developed countries alike. Measuring and analyzing export diversification have thus been important objectives for most international organizations, e.g., the World Bank, the International Monetary Fund (IMF), and the Inter-American Development Bank (IDB).[1] The analytic image of export diversification is the extensive margin of trade, the subject of prominent academic studies: from a theory perspective, e.g., Helpman et al. (2008a); from an estimation perspective, e.g., Santos Silva et al. (2014); from a policy perspective, e.g., Cadot et al. (2011); and from a measurement/index perspective, e.g., Hummels and Klenow (2005). We contribute to the export diversification literature by proposing a simple model of adjustment on both domestic (range of products) and international (range of destinations) margins. We deploy the model to distinguish and quantify globalization effects on domestic and cross-border margins of trade.

Our focus is on the *domestic extensive margin* of trade (domestic distribution of a product) and its importance for quantifying the international extensive margin of trade. We build

[1]The traditional export diversification indexes and the Hummels-Klenow (product and country) extensive margin indexes are featured prominently in the World Bank's World Integrated Trade Solution interface (https://wits.worldbank.org/wits/wits/witshelp/Content/Utilities/e1.trade_indicators.htm). Similarly, the International Monetary Fund developed and maintains the "The Diversification Toolkit: Export Diversification and Quality Databases" (https://www.imf.org/external/np/res/dfidimf/diversification.htm), while the Inter-American Development Bank has devoted significant effort to study and promote export diversification in Latin America (https://www.iadb.org/en/news/idb-recommends-latin-america-accelerate-export-diversification).

a dataset that combines data on the standard international extensive margin and the novel domestic extensive margin. We document significant but intuitive variation in the domestic extensive margin across countries and over time. We quantify the impact of European Union (EU) integration, 2008-2018, and demonstrate that these effects cannot be identified without the domestic extensive margin. We find strong but widely heterogeneous effects of EU integration on the extensive margin of trade. Our methods have implications for policy evaluation and for the measurement of export diversification and the extensive margin of trade.

On the theory side, we introduce the domestic extensive margin of trade in a short run structural gravity model that features action on the extensive margin of international trade (new export destinations) and domestic trade (new products). Capital is sector- and destination-specific. Investment on the extensive margins is selected when the expected return exceeds the product of the opportunity cost of capital and an adjustment cost factor.[2] Since much of this capital is unobservable, its behavior is inferred by fixed effects modeled consistently with the theoretical implications of the model. A key implication of our model is that proper quantification of the international extensive margin (the set of partners any sector exports to) should also take into account the domestic margin of trade (the set of sectors with positive production).

The application is based on a novel dataset that covers the extensive margin of trade in mining and manufacturing goods for 32 European countries over the period 2008-2018. The most important and unique dimension of our dataset is the *domestic extensive margin*. The original source for data on the domestic extensive margin is Eurostat's *Production Communautaire* (PRODCOM) database. It is combined with Eurostat's COMEXT data to construct the database. The combination of PRODCOM and COMEXT allowed us to

[2]The baseline for our theory is the intensive margin short-run gravity model of Anderson and Yotov (2020). However, the general approach to model investment in bilateral trade links is also in the spirit of Arkolakis (2010), Head et al. (2010), Chaney (2014), Mion and Opromolla (2014), Sampson (2016), and Crucini and Davis (2016). Our innovations in relation to these papers are (i) developing the model on the extensive margin, and (ii) the focus on the domestic extensive margin.

build an estimating sample that covers consistently constructed data on the external and the domestic extensive margins for 32 European countries and about 3300-3400 products, 2008-2018. We also experiment with several alternative estimating samples, which demonstrate the robustness of our main findings.

To explore the variation in the domestic extensive margin we construct a novel index, the *Domestic Extensive Margin (DEM)*. It is defined as the ratio of the number of products actually produced by a given country in a given year and the total number of possible products that could have been produced by the same country and in the same year. We document several interesting patterns in the DEM variation across countries and over time. First, the domestic extensive margin varies widely (but intuitively) across countries. The countries with the lowest DEM indexes are smaller and/or poorer economies (e.g., Iceland and North Macedonia), while the countries with the largest DEMs are large and rich economies (e.g., Germany and France).[3] We also document significant DEM variation over time. A number of countries have experienced an increase/improvement on the domestic extensive margin (e.g., Hungary and the Netherlands). On the other end of the spectrum are Portugal, Croatia, Finland, Italy, and the United Kingdom. A third group of countries (e.g., Spain and Germany) have not experienced significant changes on the domestic extensive margin.

We demonstrate the use of our methods by quantifying the impact of EU integration on the extensive margin of trade. This application is attractive for three reasons. First, it highlights a key argument that one cannot identify the desired EU integration effects in a theory-consistent econometric specification without data on the domestic extensive margin. Second, from an econometric perspective, the focus on Europe (2008-2018) allows us to obtain estimates of the EU integration effects within a simple, flexible, and robust econometric specification with fixed effects only.[4] The fixed effects treatment is convenient because it

[3]This observation is consistent with and complements the policy argument for the importance of the international extensive margin from the development literature, according to which the (international) extensive margin of trade is a more important indicator for developing/poorer countries because their exports are less diverse. This makes them dependent on exports of a few products and, therefore, these countries are more vulnerable to terms of trade changes.

[4]Technically, we do have controls for membership in the World Trade Organization (WTO) and in Eco-

enables us to obtain a series of EU integration estimates (across time and for individual countries) while, at the same time, the rich fixed effects structure of our model diminishes omitted variable and endogeneity concerns. Finally, the proposed application is interesting and relevant for its potential implications for export diversification strategies.

We rely on three different strands of the literature to specify our econometric model. First, the theory developed in this paper extends the CES structural gravity model to a closed form that features both domestic and cross-border extensive margins of trade. The model motivates our reduced-form empirical specification that identifies these margins. Second, the reduced form specification achieves identification with a rich set of fixed effects following recent developments in the empirical gravity literature on the intensive margin of trade. Third, the fixed effects representation of the theoretical model is estimated with the Santos Silva et al. (2014) FLEX estimator. FLEX is designed to consistently deal with the boundedness above and below of the extensive margin dependent variable. We also demonstrate the robustness of our main findings to the use of alternatives estimators including Tobit, OLS, and the Poisson Pseudo Maximum Likelihood (PPML) of Santos Silva and Tenreyro (2006, 2011). We show below that identification of the EU integration effects with the theory-consistent specification requires the use of data on the domestic extensive margin, regardless of the estimator.

The empirical analysis starts with a benchmark specification that imposes common globalization effects across all countries in the sample. The main result from this analysis is that globalization has had a significant positive impact on the international extensive margin of trade relative to the domestic extensive margin for the European economies. Intensive integration processes within Europe are the natural explanation for this result. This result

nomic Integration Agreements (EIAs) in our main specifications. However, given the specifics of sample (i.e., covering only European economies) and the use of country-pair fixed effects, the estimates of the EIA and WTO covariates are identified of very few observations and the introduction of these variables does not affect our main results. For example, Montenegro is the only country from our sample that became a WTO member during the period of investigation (in 2012), while all the variation in the EIA covariate come from the trade agreements of very few countries including Bosnia and Herzegovina, Croatia, Montenegro, and Serbia.

should be important from a policy perspective because there is plenty of anecdotal evidence that the impact of globalization on the intensive margin stalled during the years after the great recession. In contrast, our results indicate that the impact of globalization on the extensive margin during the same period has been economically strong and statistically significant. Our preferred specification implies that, on average across the countries in our sample, the number of internationally traded products increased by about 271 relative to the number of domestically traded products during the period of investigation, or about 7.2 percent of the total number of possibly traded products in 2018.

Highly heterogeneous effects emerge when we allow for country-specific effects of EU integration. Three main findings stand out from the country-specific analysis. The first is the wide heterogeneity of EU integration effects. Second, the estimates suggest that the effects on the extensive margin have been the strongest for the recent and new EU members, while the large EU economies have experienced relatively small extensive margin gains. Third, a decomposition of the effects on the new EU members reveals that the impact of EU integration on the extensive margin on their trade has been positive but very asymmetric, with significantly stronger effects on the imports of new from old EU members, and still positive but significantly smaller effects on the exports of the new to the old EU members. The intuitive explanation is that the new EU members were not able to position their (possibly inferior) products well in the more developed West-European market. In contrast, the new EU members responded strongly to the opportunity to increase varieties from the Western European countries.

Our work complements and extends two strands of the literature. Naturally, most closely related is the literature on the extensive margin of trade. Helpman et al. (2008a) and Chaney (2008) are two prominent examples of theoretical contributions to this literature. From an empirical and application perspective, see Hummels and Klenow (2005) for an important study on the extensive margin at the sector/product level, and Helpman et al. (2008a) for an influential analysis of the extensive margin at the country level. Finally, from an estima-

tion point of view, Santos Silva et al. (2014) summarize and extend the latest econometric developments in the estimation of the extensive margin of trade. Their FLEX estimator is used to obtain our main results. Our main innovations in relation to this literature are the modeling of the extensive margin in the short run and the introduction of the *domestic extensive margin*. As we demonstrate below, our contribution has implications for quantifying the effects of various policies as well as for the measurement and the construction of indexes on the extensive margin of trade.

The other branch of related literature includes papers that emphasize the importance of proper account for domestic trade flows on the intensive margin of trade. For example, Yotov (2012) uses domestic trade flows to resolve 'the distance puzzle' in international trade. Ramondo et al. (2016) demonstrate that when domestic trade flows are taken into account, two other gravity literature puzzles are resolved: (i) that larger countries should be richer than smaller countries and (ii) that real income per capita increases too steeply with country size. Agnosteva et al. (2019) employ domestic trade flows to estimate heterogeneous domestic trade costs. Finally, Heid et al. (forthcoming) show that the use of domestic trade allows for identification of unilateral and non-discriminatory trade policies in intensive margin structural gravity models. Our contribution is that we offer a theoretical motivation and empirical evidence for the importance of the domestic extensive margin for quantifying the extensive margin of trade. Our methods open avenues for many extensions and applications, e.g., estimating the impact of country-specific policies and characteristics (e.g., export promotion, institutional quality, etc.). We elaborate on some of these ideas in the concluding section of the paper.

The rest of the paper is structured as follows. Section 2 develops our theoretical model and then translates it into an econometric specification. Section 3 describes the data sources and our methods to construct the data. Section 4 reports and discusses our estimates of the impact of globalization and the results from a series of robustness experiments. Finally, Section 5 summarizes our contributions and findings and points to a series of additional

implications and extensions.

2 Quantifying the Extensive Margin of Trade

Subsection 2.1 combines and extends three prominent strands of the trade literature to derive a short-run gravity theory on the extensive margin(s) of trade. Our key contributions in relation to the existing literature are the derivation of the *short-run extensive margin of international trade* and the introduction of the notion of *domestic extensive margin*. Subsection 2.2 capitalizes on a number of developments in the empirical literature on the extensive and the intensive margins of trade to translate our theory into an econometric specification.

2.1 Short Run Gravity and the Extensive Margin

Three influential strands of the trade literature are building blocks erected here in a closed form structural short-run gravity model of the extensive margin. First, our model nests the standard gravity equation, c.f., Anderson (1979), Eaton and Kortum (2002), Anderson and van Wincoop (2003), and Arkolakis et al. (2012). Second, we incorporate bilateral investment/dynamics in the spirit of Arkolakis (2010), Head et al. (2010), Chaney (2014), Mion and Opromolla (2014), Sampson (2016), Crucini and Davis (2016), and Anderson and Yotov (2020). Third, we account for action on the extensive margin of trade following Melitz (2003), Helpman et al. (2008a) and Chaney (2008). The novelties are (i) our treatment of the extensive margin in the short run and (ii) the explicit account for the domestic extensive margin. Since the key building blocks of our theory are relatively standard in the literature, we relegate all derivations to the Appendix. This section summarizes our assumptions, presents the resulting model, and provides intuition behind each of its components with emphasis on the novel elements.

The world consists of many countries that produce their own product varieties (Armington, 1969) and trade with each other. Production in each of them is split in sectors and

carried on by heterogeneous firms, which allocate capital to production and to distribution to a set of destinations.[5] The capital becomes specific once allocated. Subsequently, the firms draw productivities from a Pareto distribution, demand shocks are realized and labor is efficiently allocated. The firms face iceberg trade frictions in distribution as well as the cost of resources needed to serve the destinations. The iceberg frictions include a fixed cost for each destination. The firms that can make operating profits hire labor from a national market and deploy it efficiently to production and to distribution to the various destinations. The firms use capital and labor for production and distribution subject to sectoral Cobb-Douglas technology. This production structure is combined with a standard Armington-CES demand system for all buyers in the multi-country world. Imposing perfect spatial arbitrage subject to the trade costs (both iceberg and endogenous) yields the short run equilibrium structural gravity model for each sector k:[6]

$$X_{ij,t}^k = \underbrace{\frac{Y_{i,t}^k E_{j,t}^k}{Y_t^k}\left(\frac{t_{ij,t}^k}{\Pi_{i,t}^k P_{j,t}^k}\right)^{(1-\sigma^k)\rho^k}}_{Structural\ Gravity} \underbrace{\times (\lambda_{ij,t}^k)^{1-\rho^k}}_{Short\ Run} \underbrace{\times (\phi_{i,t}^k)^{(\rho^k-1)/\rho^k}}_{Extensive\ Margin} \quad \forall i,j \tag{1}$$

Equation (1) can be decomposed into four structural components. We label the first term 'Structural Gravity' because, as famously demonstrated by Arkolakis et al. (2012), it can be derived from a very wide class of theoretical economic micro-foundations. The intuition behind this term is standard and simple, i.e., bilateral trade flows between two countries ($X_{ij,t}^k$) are proportional to the product of their sizes ($Y_{i,t}^k$ and $E_{j,t}^k$, output and expenditure, respectively) as a share of world output (Y_t^k), and inversely proportional to the trade frictions between them, which consist of direct bilateral trade costs ($t_{ij,t}^k$) and general equilibrium trade frictions captured by the multilateral resistances ($\Pi_{i,t}^k$ and $P_{j,t}^k$, outward and inward, respectively), and where σ^k is the elasticity of substitution.

[5]The description of the model is in terms of integrated production and distribution firms but it applies equally to arms length relations between production and distribution.

[6]See the Appendix for detailed derivations.

Following Anderson and Yotov (2020), we label the second term in equation (1) 'Short Run'. This term consists of two structural elements. In our model, $\lambda_{ij,t}^k$ is a bilateral capacity variable – the fraction of country i's capital allocated *ex ante* to destination j. More broadly, the notion of bilateral capacity here is consistent with the network link dynamics modeled by Chaney (2014), the link between managers' experience in previous firms and the export performance of their current company described by Mion and Opromolla (2014), and with the 'marketing capital' of Head et al. (2010). The second element in the 'Short Run' term is the parameter $\rho^k \in (0,1)$, whose structural interpretation is as buyers' short run incidence elasticity, the fraction of a 1% rise in trade cost that is paid by buyers in short run equilibrium.[7] Alternatively, ρ^k can be interpreted as the proportion by which the short run trade elasticity is reduced from the long run trade elasticity.[8] In a hypothetical long-run equilibrium, represented by $\rho^k = 1$, the 'Short Run' term will disappear and equation (1) collapses to the standard cross-section structural gravity model, naturally interpreted as a long run model.

We label the third structural term in specification (1) the 'Extensive Margin'. This term combines two elements: one traditional and one novel. The traditional element is captured by $\phi_{i,t}^k$, which for simplicity in our model is interpreted as an iceberg fixed entry cost, becoming sunk after entry.[9] The novelty is that $\phi_{i,t}^k$ is raised to the power of $(\rho^k - 1)/\rho^k$, which arises because we have combined an extensive margin model, e.g., of Melitz (2003), with

[7] In the current setting with heterogeneous firms, ρ^k is micro-founded as a combination of the elasticity of substitution along with supply parameters including the capital share and the Pareto shape parameter, all allowed to vary at the sectoral level. See the Appendix for detailed derivations and more discussion.

[8] To see this intuition, note that ρ^k enters multiplicatively the power of the trade cost element in the 'Structural Gravity' term, thus driving a wedge between the short-run and the long-run trade elasticity. Anderson and Yotov (2020) obtain estimates of ρ around 0.25, a static structural rationale for the difference between the larger (long-tun) trade elasticity values from the trade literature vs. the smaller (short-run) estimates from the IRBC macro literature.

[9] The interpretation of $\phi_{i,t}^k$ is that for any new market k served by i the creation of 1 unit of destination specific capital requires $\phi_{i,t}^k$ units drawn from alternative uses. $\phi_{i,t}^k > 1$ could represent the cost of learning to manage production and distribution to destination i in sector k. Because each firm has a fixed per period cost for each destination market that it serves, there is generally an interior extensive margin of distribution, independently of the sunk cost $\phi_i^k - 1$. Another operational mechanism behind an active interior extensive margin is the diminishing marginal product of marketing capital. In general equilibrium this interacts with the force of rising opportunity cost of capital being drawn from other projects.

the short run gravity model of Anderson and Yotov (2020). Given the theoretical bounds for $\rho^k \in (0,1)$, the relationship between the sunk export entry cost ($\phi_{i,t}^k$) and trade flows is inverse, as expected. Furthermore, note that when ρ^k approaches its upper limit of 1, the 'Extensive Margin' term would disappear. The intuition for this is that we have reached the hypothetical long-run equilibrium, where serving all potential links on the extensive margin have been efficiently exhausted. As noted before, when $\rho^k = 1$ the 'Short Run' term will also vanish, thus our model will collapse to the standard, long-run, 'Structural Gravity' model. Identifying the short run extensive margin effects, as captured by $(\phi_{i,t}^k)^{(\rho^k-1)/\rho^k}$, will be the main focus of our empirical analysis.

Finally, we emphasize that (1) holds equally for international and domestic links, i.e., both for $i \neq j$ and for $\forall i = j$. Thus equation (1) captures two distinct forms of the extensive margin of trade. First is the standard external (cross-border) margin of trade whereby the production and distribution for export changes. Second is the *sectoral* margin whereby production and distribution of a product begins or ends. This is what we call the *'Domestic Extensive Margin' (DEM)*, a key focus of our empirical analysis.

There are two sources of time series variation in the two extensive margins. One is due to cyclic volatility of service in bilateral links. The other is secular change (growth or decline) in the number of markets served. Both sources of action on the extensive margins are potentially active and quantitatively important.[10] Both forms of the extensive margin are described by the simple selection mechanism of the heterogeneous firm model embedded in specification (1).

The empirical analysis below demonstrates that proper econometric accounting for the domestic extensive margin may have significant implications for identifying the impact of a number of determinants of the external extensive margin. The lesson is consistent with the closed form model (1) but more broadly suggests the importance of simultaneously

[10]Besedes and Prusa (2006) document the high volatility over time of 10 digit HS level bilateral US exports. Chaney (2014) describes the rich dynamics of French exporting firms focused on their entry into new bilateral markets.

accounting for both extensive margins.

2.2 From Theory to Empirics: Estimating the Extensive Margin

Motivated by our theoretical equation (1), the objective of this section is to specify a corresponding econometric model. To this end, we proceed in three steps and rely on three different strands of the literature. First, we translate our theory into an econometric model, which is broadly consistent with other structural models on the extensive margin of trade, e.g., Helpman et al. (2008b). Second, following the recommendations of Santos Silva et al. (2014) we select their *FLEX* estimator to obtain our main results. Finally, guided by the empirical literature on the intensive and on the extensive margins of trade and by our key contribution (i.e., the introduction of the domestic extensive margin), we select the covariates in our empirical model.

We start by translating our theory into an econometric model.[11] Let $N^k_{ij,t}$ be an indicator equal to one when at least one firm exports k from i to j at time t. In order for this to be the case, there should be at least one firm in this sector that finds it profitable to produce and export, i.e., $\pi^k_{ij,t}(\varrho) > 0$. This implies that the probability for a given sector to be exported from origin i to destination j at time t is:

$$\Pr(N^k_{ij,t} = 1|x_{ij,t}) = \Pr(\pi^k_{ij,t}(\varrho) > 0) = F^k(x'_{ij,t}\beta) \tag{2}$$

Letting $N_{ij,t} = \sum_k N^k_{ij,t}$ the total number of sectors exported from i to j at time t, the previous expression implies:

$$E(N_{ij,t}|x_{ij,t}) = \sum_k \Pr(N^k_{ij,t} = 1|x_{ij,t}) = \sum_k F^k(x'_{ij,t}\beta) = N_{i,t}F(x'_{ij,t}\beta), \tag{3}$$

where $N_{i,t}$ is the total number of sectors available in origin i, and $F(x'_{ij,t}\beta) = \left(F^k(x'_{ij,t}\beta)\right)/N_{i,t}$

[11] As demonstrated by Santos Silva et al. (2014), the same steps can be applied to translate Helpman et al. (2008b) into corresponding econometric models.

is interpreted as the probability that a randomly selected sector k will be exported from country i to country j at time t.

Next, to select the functional form for $F(x'_{ij,t}\beta)$, we follow Santos Silva et al. (2014):

$$F(x'_{ij,t}\beta) = 1 - \left(1 + \omega \exp\left(x'_{ij,t}\beta\right)\right)^{\frac{-1}{\omega}}.$$

As introduced by Santos Silva (2001), this functional form has two main advantages for our purposes. First, consistent with the fact that our dependent variable is bounded from above and from below, the proposed function is double-bounded too.[12] Second, this specification is flexible (hence, the FLEX estimator) as there are no prior constraints imposed on the shape parameter ω, apart from it being positive, i.e. $\omega > 0$. Thus, as noted by Santos Silva et al. (2014), the implied distribution can be symmetric ($\omega = 1$), left-skewed ($\omega < 1$), or right-skewed ($\omega > 1$), as dictated by the data. The flexible functional form is potentially very important to capture the distribution of the extensive margin of trade, where the larger number of observations is clustered in the lower tail of the distribution and they will determine the shape of the estimated function and lead to bad fit of the upper tail of the distribution due to its low weight in the objective function. Following Santos Silva et al. (2014), we will estimate the model by Bernoulli pseudo-maximum likelihood, which is easy to implement and it is consistent under very general conditions.[13]

To demonstrate the robustness of our main results, we also experiment with three alternative estimators. First, we employ a double-bounded Tobit estimator. In addition, following the best current practices in the intensive margin gravity literature, we also experiment with the Poisson Pseudo Maximum Likelihood (PPML), which has the attractive properties of being a count multiplicative model, which can take into account the information contained in the zero observations in our sample. PPML established itself as the leading gravity esti-

[12]The functional form is in fact bounded between 0 and 1, but this will be consistent when the dependent variable in our model is transformed from a level to a share.

[13]The refer the reader to Gourieroux et al. (1984) and Papke and Wooldridge (1996) for feasible alternatives and further details.

mator due to the seminal work of Santos Silva and Tenreyro (2006), and we refer the reader to Santos Silva and Tenreyro (2006) and Santos Silva and Tenreyro (2011) for excellent discussions of the attractive features of PPML for gravity estimations on the intensive margin of trade. Finally, despite its limitations in the current setting, we also obtain robustness estimates with the OLS estimator. As demonstrated in the sensitivity analysis, our main results and conclusions are robust to the use of alternative estimators.

The third and final step to complete our econometric setup is to explicitly define the covariates in our model. To this end, we rely on the numerous contributions to the empirical literature on the intensive and on the extensive margins of trade, as well as on our key contribution, i.e. the introduction of the domestic extensive margin. Taking into account the latest developments in the estimation of gravity equations (on the extensive and on the intensive margin), we start by defining:

$$\exp\left(x'_{ij,t}\beta\right) = \exp\left(\pi_{i,t} + \chi_{j,t} + \gamma_{ij} + BIPOL_{ij,t}\beta_1\right), \quad \forall i \neq j \tag{4}$$

Equation (4) includes three sets of fixed effects. $\pi_{i,t}$ and $\chi_{j,t}$ are exporter-time and importer-time fixed effects, which would control for and absorb the multilateral resistance terms from our theoretical model, as well as any other country-specific time-varying characteristics that may affect the bilateral extensive margin, on the exporter and on the importer side, respectively. γ_{ij} denotes a set of country-pair fixed effects, whose purpose is to account for all time-invariant bilateral determinants of the extensive margin of trade. Finally, $BIPOL_{ij,t}$ is a vector of time-varying bilateral determinants of trade, e.g., trade agreements, tariffs, etc.

An important feature of all existing empirical papers on the extensive margin of trade, as captured by equation (4), is that, without any exception, all of the existing extensive margin analyses are performed exclusively with international trade data only and without taking into account the domestic extensive margin. As we demonstrate next, proper/theory-consistent account for the domestic extensive margin may have significant implications for

estimating the impact of numerous determinants of the extensive margin of trade. To see this, note that once the domestic extensive margin is introduced, equation (4) becomes:

$$\begin{aligned} \exp\left(x'_{ij,t}\beta\right) &= \exp\left(\pi_{i,t} + \chi_{j,t} + \gamma_{ij} + BIPOL_{ij,t}\beta_1 + EXS_{i,t}{\times}BRDR_{ij}\beta_2\right) \times \\ &= \exp\left(IMP_{j,t}{\times}BRDR_{ij}\beta_3 + CNTRY_{j,t}{\times}BRDR_{ij}\beta_4\right) \qquad \forall i,j \qquad (5) \end{aligned}$$

Three new terms, which could not be identified from specification (4), appear in equation (5) due to the introduction of the domestic extensive margin. $EXS_{i,t}$ is a vector of non-discriminatory export support policies, e.g., export subsidies, trade fairs, etc. $BRDR_{ij}$ is an indicator variable for cross-border trade, equal to 0 for domestic trade. Note that when the extensive margin is defined only based on cross-border observations, the effects of any non-discriminatory export policies cannot be identified because they are absorbed by the exporter-time fixed effects. $EXS_{i,t}{\times}BRDR_{ij}$, is time-varying and bilateral and, therefore, it can be identified in the presence of all fixed effects from (5).

The second new term in (5) is $IMP_{j,t}{\times}BRDR_{ij}$, and it is constructed as an interaction between a vector of non-discriminatory import protection policies, $IMP_{j,t}$, and the international border dummy. Similar to the case of export support, the impact of any non-discriminatory import protection policies cannot be identified in the presence of importer-time fixed effects without the domestic extensive margin. The third new term in (5) is $CNTRY_{j,t}{\times}BRDR_{ij}$, and it is constructed as an interaction between a vector of country-specific characteristics and policies, e.g. institutional quality, technical barriers to trade etc., $CNTRY_{j,t}$, and the international border dummy. Once again, the impact of such policies cannot be identified without the domestic extensive margin. The difference between this term and the directional (export and import policies) is that we can only identify the differential impact of such policies on international relative to internal trade, however not depending on the direction of trade flows, e.g., not on the impact of exports vs. imports.

Finally, we note that the introduction of the domestic extensive margin has two poten-

tially important implications for the estimates of the effects of any bilateral trade policies, which are included in vector $BIPOL_{ij,t}$. Consider, for example, the impact of regional trade agreements. The introduction of the domestic extensive margin allows for an explicit account that, consistent with Melitz (2003), trade liberalization may lead to decrease in the number of products that are produced domestically.[14] The implication for the estimates of the impact of RTAs in that scenario is that they may be biased downward without accounting for the domestic extensive margin. Alternatively, if one believes that trade liberalization leads to production in more sectors, i.e., an increase in the domestic extensive margin, then the implication for RTA estimates that are obtained without account for that is that they may be biased upward. Next, consider the impact of WTO membership on the extensive margin of trade. In addition to allowing for the possibility to capture a possible decrease in the number of domestically produced varieties, the introduction of the domestic extensive margin allows for the identification of country-specific WTO effects for each member country. This is not possible without the domestic extensive margin because the country-specific WTO effects would be absorbed by the exporter-time and/or importer-time fixed effects in the econometric specification.

3 Data: Construction and Sources

To conduct the empirical analysis we construct a novel dataset that covers the extensive margin for mining and manufacturing goods for 32 exporters and 212 importers over the period 2008-2018. A very important and unique dimension of our dataset is the *domestic extensive margin*. As described in more detail next, availability of data on the domestic extensive margin is what predetermined the dimensions of our estimating sample(s). Guided by theory, and in an attempt to utilize as much of the available data as possible, we construct and experiment with several alternative samples by extending the data coverage across the

[14] Our econometric model does not take a stand on whether trade liberalization leads to an increased or a decreased number of domestic varieties. However, we believe that this is an interesting empirical question, which can be viewed as a direct test of one of the key implications of Melitz's landmark theory.

product and country dimensions. We construct the estimating sample(s) in three steps that are described in the three corresponding subsections of this section.

3.1 The Domestic Extensive Margin

The original data source that we use to construct the *domestic extensive margin* is PRODCOM. This database is developed, maintained, and hosted by Eurostat.[15] PRODCOM consists of yearly files that include statistics on the value of production (in thousands of Euro) for 35 European countries and about 3800 product categories in mining, quarrying, and manufacturing. A list of the countries covered in PRODCOM, along with their 2-letter ISO alpha codes, can be found in the first two columns of Table 1. Data for most countries are available and balanced throughout the period of investigation, 2008-2018. However, there are several exceptions. Specifically, inspection of the data reveals that (i) Bosnia and Herzegovina, Northern Macedonia; Serbia, and Montenegro appear in the database only in 2011, i.e., data for these countries are not available for the period 2008-2010. (ii) Turkey has no data after 2011, i.e., data for Turkey are available only for three years, including 2008, 2009, and 2010; and (iii) data for Croatia are not available in 2012. Most probably due to the focus on mining and manufacturing, three countries (Cyprus, Luxembourg, and Malta) do not produce/sell any of the products covered in PRODCOM. These countries are eliminated from our sample. As a result, we were able to construct an unbalanced domestic extensive margin panel for 32 countries.

Turning to the product dimension, PRODCOM offers production data for the broad categories of mining, quarrying, and manufacturing (with the exception of military products and some energy products). Due to the invention of new products and discontinued production of others, the number of categories covered in PRODCOM varies across years, but, on average, PRODCOM covers about 3800 categories of mining, quarrying, and manufacturing

[15]The title 'PRODCOM' comes from the French "PRODuction COMmunautaire" (Community Production). We accessed the original PRODCOM files at https://ec.europa.eu/eurostat/web/prodcom/data/excel-files-nace-rev.2.

products. The number of all possible PRODCOM products in a given year appears in the last row of Table 1. There are 25 steel products in the original PRODCOM dataset for which data were missing for all countries and in all years. In addition, we noticed that Finland had missing observations for 5 products in some years. After investigating the raw data, we concluded that (i) Finland must produce in four of these categories, e.g., because there was production of the same products in years that were neighboring to the missing observations, and (ii) Finland did not produce in one category, e.g., because the observations for all but the missing year were zeroes. This information was sufficient to construct the missing values for Finland's extensive margin.

In addition to actual reported values of production, the original PRODCOM database includes observations labeled as Confidential (:C), Estimated (:E), or Confidential/Estimated (:CE). These observations account for a total of approximately 20% of the original data. The observations labeled Confidential (:C) or Confidential/Estimated (:CE) account for more than 19%, while the Estimated (:E) observations were less than 1%. While the presence of confidential and/or estimated observations could have been potentially problematic for an analysis on the *intensive margin* of trade, they are not such a big concern in our case, where the focus is on the *extensive margin* and all we need to know is whether there is production or not in a given category. To take advantage of the information contained in the confidential and the estimated observations we proceed in three steps. First, we assign a value of one on the extensive margin for any estimated or confidential observations for which there were positive production values in the same category but in other years in the original data. Second, we assign a value of zero on the extensive margin for any estimated or confidential observations for which the non-missing production values in the same categories in all other years in the original data are zeroes. Finally, if the observations for all years for a given country and product category were classified as confidential and/or estimated, we assign a value of one on the extensive margin.

The last two steps in the construction of the domestic extensive margin are (i) to replace

the positive reported production values with ones, and (ii) to sum them for each country and year in the sample. For consistent comparisons (since the number of possible products varies across years), we define our novel index of the *Domestic Extensive Margin (DEM)* as the ratio between the number of products actually produced by a given country in a given year, $D_{i,t}$, and the total number of possible products that could have been produced by the same country and in the same year, $N_{i,t}$:

$$DEM_{i,t} = \frac{D_{i,t}}{N_{i,t}}.$$

The domestic extensive margin indexes for all countries and all years in our sample appear in Table 1. The total number of possible products are reported in the last row of the table. The last column of the table reports percentage changes for each country between the first and the last year for which data are available. The exception is Serbia, for which the initial year for the percentage change in the last column is 2012. As can be seen from Table 1, the domestic extensive margin index for Serbia in 2011 is very different from the relatively stable indexes in the subsequent years. In combination with the fact that 2011 is the first year for which Serbia was included in PRODCOM, we conclude that the 2011 data for Serbia are not reliable and, therefore, for the remainder of the analysis we treat the observations for Serbia in 2011 as missing.

Several interesting patterns regarding the heterogeneity of the *DEM* index across the countries in our sample as well as the evolution of the index over time stand out from Table 1. First, and most important for our identification purposes, we see that the domestic extensive margin varies widely across countries. For simplicity, focus on the column for the last year in the sample, 2018. The variation that we observe makes intuitive sense. For example, the countries with the lowest domestic extensive margin indexes are smaller and poorer economies (e.g., Montenegro, Iceland, North Macedonia, Bosnia and Herzegovina, and Latvia), while the countries with the largest indexes are large and rich economies (e.g.,

Germany, France, Spain, Italy, and the United Kingdom). This observation is consistent with and complements the policy argument for the importance of the international extensive margin from the development literature, according to which the (international) extensive margin of trade is a more important indicator for developing/poorer countries because their exports are less diverse. This makes them dependent on exports of a few products and, therefore, these countries are more vulnerable to terms of trade changes.

The second notable finding in Table 1 is the significant variation in the domestic extensive margin within countries and over time. Even though not crucial for our purposes, this variation will further aid identification. Three patterns stand out from Table 1 and, to analyze them, we focus on the percentage changes that are reported in the last column of the table. First, we see that a number of countries have experienced an increase/improvement on the domestic extensive margin. The countries with the largest increases are Hungary, Netherlands, Lithuania, Slovenia, and Greece. Apart from the Netherlands, a possible explanation for such favorable ranking is that these countries have benefitted from their integration in the European Union. On the other side of the spectrum we find Portugal, Croatia, Finland, Italy, and the United Kingdom. Finally, a third group of countries have not experienced significant changes on the domestic extensive margin. These countries include Norway, Spain, Germany, and Iceland. Interestingly, two of these countries (e.g., Germany and Spain) have very large indexes, while the other two countries (e.g., Norway and Iceland) are among the ones with the smallest indexes.

This section presented the *Domestic Extensive Margin* index. The accompanying analysis revealed wide heterogeneity in the *DEM* indexes across the countries in our sample as well as significant variation of *DEM* over time. This variation is useful for identification of heterogeneous EU integration effects on domestic and international margins below, controlling for size effects and multilateral resistance effects consistent with the structural gravity model.

3.2 Matching the Domestic & International Extensive Margins

We rely on the COMEXT database to construct the international extensive margin of trade. According to the official Eurostat web site *"COMEXT is Eurostat's reference database for detailed statistics on international trade in goods"*, and the dataset offers very detailed statistics according to the Combined Nomenclature (CN) classification system.[16] We follow the standard method to construct the extensive margin, i.e., first, we assign values of one to the positive product-level flows in COMEXT, and then we sum them for each pair-year combination. The result is a time-varying bilateral variable, which is defined as the number of products exported from i to j at year t. The structure of COMEXT, in combination with the design of PRODCOM, presented several opportunities to construct and experiment with alternative estimating samples. We describe those opportunities and our choices next.

For each reporting country COMEXT includes separate data on exports and on imports. Based on this information, we construct and experiment with three alternative measures of the extensive margin of trade. First, following most of the literature on the intensive margin of trade flows, our main extensive-margin variable is constructed based on the average between the import and the export flows in COMEXT. Alternatively, we also construct a sample, where we start with the reported export values and we replace the missing exports observations with the corresponding non-missing import values. We call this sample the *"Exporter-based Sample"*, and we experiment with it in the sensitivity analysis. Similarly, we also construct a measure where we start with the reported import values and we replace the missing import observations with the corresponding non-missing export values. We call this sample the *"Importer-based Sample"*. As we demonstrate later, estimates obtained with the three alternative samples are very similar to each other.

The key novelty of our analysis is the introduction of the *the domestic extensive margin*. Thus, it is very important for our purposes to construct a consistent correspondence between

[16] We accessed the historical version of the Comext data at https://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?sort=1&dir=comext.

the domestic and the international extensive margins. To this end, we benefited tremendously from the fact that the two main underlying databases (PRODCOM and COMEXT) that we used to build our estimating samples were designed to be consistent with each other by construction. Specifically, as noted in the PRODCOM user guide, *"[b]efore data collection could begin, it was necessary to draw up a common list of products to be covered ... As PRODCOM statistics have to be comparable with external trade statistics, which are based on the Combined Nomenclature (CN), there had to be a close relationship between the two nomenclatures."* We took advantage of the close matching and existing concordances between the PRODCOM and the CN classifications to construct consistent estimating samples that cover both the domestic and the international extensive margins.[17]

While the matching between PRODCOM and CN was intended to be very close by design, *"it was felt by the PRODCOM committee that there were instances where the CN classification gave too much detail in how it broke down products within a specific category, but equally instances when it did not give enough detail to meet the needs of the likely end users of PRODCOM data."* (p.6, PRODCOM Guide). As a result, the matching between the PRODCOM classification and the Combined Nomenclature includes one-to-one matches, many CN to one PRODCOM matches, one CN to many PRODCOM matches, and many CN to many PRODCOM matches. There was also a small fraction of products of the PRODCOM categories that did not have a match in the Combined Nomenclature.[18]

As expected, an investigation of the matching patterns between PRODCOM and CN reveals that most of the cases are one-to-one matches and the second largest share includes many CN to one PRODCOM matches. In combination, these two types of matches cover between 78.5% and 100% of the PRODCOM product categories for which there is a CN match.[19] Therefore, we constructed and experimented with two alternative estimating sam-

[17]The original PRODCOM to CN concordance files come from the Eurostat RAMON site at: https://ec.europa.eu/eurostat/ramon/relations/index.cfm?TargetUrl=LST_REL&StrLanguageCode=EN&IntCurrentPage=4.

[18]Specifically, the fraction of PRODCOM products that could not be matched to the CN classification varies between 6.57%, in 2016, and 9.58% in 2010.

[19]Specifically, they cover more than 97% of the PRODCOM products in 2008 and 2009, 100% of the

ples based on the underlying product matching and coverage between the PRODCOM and the CN classifications. The first sample includes only the products for which we have one-to-one matching. The number of products that we cover this way is around 2000 in each year of our sample. We label this sample the *"Conservative-product sample"*, and we use it in the robustness analysis. The second sample, which is the one used in the main analysis, is our *"Extended-product sample"* because it covers all products for which we have one-to-one or one-to-many matching between PRODCOM and CN, i.e., in the latter case there are multiple CN products corresponding to a single PRODCOM category. The number of products in this extended sample varies between 3276 and 3513, thus covering almost all (between 93% and 100%) possible products in the original PRODCOM classification for which there was a CN match.

Based on the *"Extended-product sample"* used to obtain our main results, in the last column of Table 1 we report the percentage change in the total number of exported products for each country in our sample during the period 2008-2018, i.e., on the international extensive margin (IEM) of trade. While we use bilateral international extensive margin data in our econometric analysis, the percentage changes in the total number of exported products that we report here are informative for at least two reasons. First, according to the indexes in the last column of Table 1, the countries in our sample can be classified in three distinct and sufficiently large groups: (i) We see some countries that experienced a very significant increase in the international extensive margin between 2008 and 2018, e.g., Montenegro and North Macedonia, followed by Croatia and Hungary; (ii) The second group of countries did not experience significant change in the number of exported products between 2008 and 2018. Some examples include, Lithuania, Bosnia and Herzegovina and Bulgaria; (iii) Finally, a number of countries saw a decrease in the number of exported products during the period of investigation. The decrease is moderate and occurs mostly in developed countries, e.g., France, Germany, UK, Austria. Based on these insights, we draw the intuitive conclu-

PRODCOM products in 2010 and 2011, more than 80% of the PRODCOM products between 2012 and 2016, and 78.5% of the PRODCOM products in 2017 and 2018.

sion that the countries that have benefitted the most on the international extensive margin of trade are smaller and poorer European economies, while the larger and more developed countries have actually contracted the number of products that they export.

Second, comparisons between the percentage changes in the last two columns of Table 1 reveal some interesting patterns of the relationship between the evolution of the international and the domestic extensive margins of trade. This is important for our purposes and for better understanding of our estimation results because what we would be identifying in our econometric analysis is the relative impact of globalization and European integration on the international relative to the domestic margin of trade. Based on the percentage changes in the last two columns of Table 1, we notice the following patterns: (i) Faster positive growth on the external margin and slower positive growth on the domestic margin, e.g., North Macedonia; (ii) No change on the external margin but an increase on the domestic margin, e.g., Lithuania; (iii) Decrease on the international extensive margin and no change on the domestic extensive margin, e.g., Germany; (iv) Growth on the international extensive margin and decrease on the domestic extensive margin, e.g., Montenegro; (v) Decrease on the domestic extensive margin and stronger decrease on the international extensive margin, e.g., the United Kingdom. (vi) Faster positive growth on the domestic extensive margin and slower positive growth on the international margin, e.g., Hungary. The case of Hungary is interesting with respect to our methodological contributions because, based on the international extensive margin indexes, one would conclude that Hungary has benefitted a lot from globalization. However, this is not the case relative to the domestic extensive margin for Hungary. We will return to some of these descriptive patterns when we interpret our econometric finings.

Even though COMEXT is based on data reported by European Union members only, the database allowed us to construct the international extensive margin for a very wide number of countries due to the fact that each EU reporter offered information both on its imports from and on its exports to all other countries in the world. This feature of COMEXT has two

implications for our analysis. First, it enabled us to construct the extensive margin of trade for the few non-EU countries (e.g., Turkey, Bosnia and Herzegovina, Northern Macedonia, etc.) from the PRODCOM database. This determined the 32×32 country dimension of our main estimating sample, where we have consistently constructed domestic and international margins of trade for all countries in the sample. In addition, we capitalized on the extensive country coverage of COMEXT to construct and experiment with an alternative (32×75) *"Extended-country sample"*, which includes domestic extensive margin for the 32 PRODCOM countries, as well as the (international) extensive margin of trade between the 32 PRODCOM countries and 75 additional importers from COMEXT.[20]

Finally, inspection of COMEXT revealed that there were export data for Cyprus, Luxembourg, and Malta, even though, as discussed earlier, these countries were not present in PRODCOM. We also noticed that for many countries the number of exported products in COMEXT exceeded the number of domestically produced products in PRODCOM. This motivated us to implement an alternative procedure to construct the domestic margin of trade, which further demonstrated the robustness of our main findings. Specifically, for each year-country combination we constructed the domestic margin of trade as the total number of products exported by this country to any other country in the world. The implicit assumption that we make when implementing this procedure is that any product that is produced in a give country is exported to at least one trading partner. We view this assumption as plausible for two reasons. First, because our sample covers mining and manufacturing (and not agriculture and especially services, where localized consumption is a more significant problem for trade). And second, because usually every country declares exports to its most

[20]The COMEXT dataset includes information for a total of 212 importers. However, we note that the wide importer coverage does not add to the domestic extensive margin and, therefore, it does not lead to improvements in the data for our identification purposes. At the same time, an extended country sample is much more computationally intensive with the non-linear estimators that we will employ. This is why we only use 75 importers and we employ this extended sample in the robustness analysis. We selected the 75 importers as follows. First, we identified the 60 countries with the largest GDPs in 2018. Together they account for more than 99.9% of world GDP in 2018. The 60 largest economies did not include 15 of the smallest countries from the PRODCOM data. Therefore, we added these small economies to end up with a total of 75 countries in our *"Extended-country sample"*.

closely related partner in almost every category.

The proposed procedure also has several important advantages. First, by construction, it ensures that the number of internationally-traded products will always be smaller and in rare cases equal (i.e, when the maximum number of products exported to a specific trading partner is the same as the total number of exported products). Second, it will enable us to construct the domestic margin of trade for all 35 countries and all years that are covered in the original PRODCOM database. Third, the procedure allows for the construction of the domestic extensive margin based only on international trade data. Thus, in principle, it can be used to construct the domestic extensive margin for a very large number of countries, as long as the underlying international trade data are available for all pairs. This is not the case in our sample, because COMEXT does not include trade between non-EU countries. Therefore, we only experiment with a sample that covers the original 35 countries from PRODCOM and the extended product list.

In sum, data availability enabled us to construct and experiment with several estimating samples. Our main estimating sample has the following dimensions and characteristics: (i) It covers 32 exporters and 32 importers for which there are consistently constructed data on the domestic extensive margin and on the international extensive margin of trade; (ii) The international extensive margin for the main sample is based on averaged data on export and import trade flows; and, finally, (iii) The number of underlying products used to construct our main sample is about 3300-3400 products, including the products for which there is one-to-one and one-to-many matching between PRODCOM and CN.

In addition, we also experiment with the following alternative estimating samples, each of which differs from the main sample in one dimension only: (i) A sample that is constructed without averaging but instead using the underlying export values as the base for the extensive margin (*the "Exporter-based Sample"*); (ii) A sample that uses the underlying imports as the base for the extensive margin (*the "Importer-based Sample"*); (iii) A sample that takes advantage of the extended importer coverage in COMEXT, i.e. with 32 exporters and 75 im-

porters (*the "Extended-country Sample"*); and (iv) A sample that uses about 2000 underlying products that match uniquely between the PRODCOM and the COMEXT databases (*the "Conservative-product Sample"*). Finally, (v) we construct a sample based on the alternative procedure to construct the domestic extensive margin as the total number of products that are exported. We label this sample *"Export-based DEM Sample"*. As demonstrated below, our results are robust to the use of any of these estimating samples.

3.3 Additional Data and Sources

To perform our main analysis, we rely on a demanding specification with a rich set of fixed effects. Specifically, we use exporter-time and importer-time fixed effects, which will control for all possible country-specific determinants of the extensive margin on the exporter and on the importer side, respectively. In addition, we employ country-pair fixed effects, which will absorb and control for all possible time-invariant bilateral determinants of the extensive margin. Finally, we also control for additional time-varying bilateral variables (e.g., economic integration agreements, EIAs, and membership in the world trade organization, WTO). These control variables come from the dynamic gravity database of the U.S. International Trade Commission, c.f., Gurevich and Herman (2018). We do note, however, that given the specifics of sample (i.e., covering only European economies) and the use of country-pair fixed effects, the estimates of the EIA and WTO covariates would be identified of very few observations. For example, Montenegro is the only country from our sample that became a WTO member during the period of investigation (in 2012).[21] Similarly, all the variation in the EIA covariate could come from the trade agreements of very few countries including Bosnia and Herzegovina, Croatia, Montenegro, and Serbia. Thus, the estimates of the WTO and the EIA effects might be subject to influence of outliers and we do not expect that these covariates would have significant impact for our main estimates. Nevertheless, we do include them as control variables in our analysis, but we will not pay close attention to their

[21]For this reason, the WTO estimate will be dropped due to perfect collinearity when we obtain our country-specific globalization effects.

estimates.[22]

4 Globalization and the Extensive Margin of Trade

The implications of structural model (7) and the specifics of the data sample(s) of Section 3 naturally suggest application to quantify the impact of globalization on the extensive margin of trade in Europe.[23] This application has several advantages for our purposes. First, it is interesting and relevant from a policy perspective, because it will enable us to answer an important question: What is the impact of European integration on the extensive margin of trade? Our estimates reveal that despite anecdotal evidence for stalled globalization on the intensive margin of trade, the impact of globalization on the extensive margin in Europe has been strong and widely (but intuitively) heterogeneous.

Second, from an econometric perspective, the focus on Europe during the period 2008-2018 will allow us to capture the desired globalization effects within a simple and robust reduced form econometric specification with fixed effects only. The fixed effects treatment enables us to obtain a series of globalization estimates while the rich fixed effects structure of our model will diminish omitted variable and, more broadly, endogeneity concerns.

Third, and most important for us, the application highlights our key argument/contribution that the effects of globalization on the extensive margin on trade should be measured relative to the domestic extensive margin. Specifically, model (7) requires fixed effect controls for

[22]Even though this is beyond the scope of this paper, we note that our methods can be applied to obtain country-specific effects of the impact of WTO, or of the effects of any other country-specific policy (e.g., export promotion or institutional quality) on the extensive margin of trade, even in the presence of the full set of exporter(-time) and importer(-time) fixed effects. Such analysis would not be possible without the use of the observations on the domestic extensive margin. Thus, our methods and key innovation open avenues for a series of potentially interesting and important applications that academic researchers and policy makers were not able to address in properly specified (i.e., with exporter-time and importer-time fixed effects) extensive margin gravity models.

[23]Given the composition of the sample, the analysis can also be interpreted as a quantification of the impact of European integration on the extensive margin of trade. Apart from data limitations, the focus on Europe is useful from a methodological perspective, because EU membership and globalization forces within the European Union would be the main driving sources for our results by definition. The standard policy variables such as WTO membership and the formation of free trade agreements play almost no role in our analysis due to the presence of the country-pair fixed effects.

the origin-time and destination-time multilateral resistances which will absorb globalization effects on the extensive margin in the absence of controls on the domestic extensive margin. Another virtue of the application and the corresponding econometric specification is that it enables a flexible estimation of the effects of globalization across time and across individual countries.

For the application to the impact of globalization in the EU, the econometric setup (5) in Section 2.2 is simplified to the following estimating equation:

$$\frac{N_{ij,t}}{N_{i,t}} = 1 - \left(1 + \omega \exp\left(\pi_{i,t} + \chi_{j,t} + \gamma_{ij} + \delta_1 WTO_{ij,t} + \delta_2 EIA_{ij,t} + \sum_{i,t} \beta_{i,t} GLOB_{i,t} \times BRDR_{ij}\right)\right)^{\frac{-1}{\omega}} + \epsilon_{ij,t}. \tag{6}$$

The estimator, the dependent variable, and all fixed effects in specification (6) were defined earlier. There are three new covariates. Specifically, $WTO_{ij,t}$ is an indicator for membership in the World Trade Organization and $EIA_{ij,t}$ is a dummy variable for economic integration agreements (EIAs).

The third new covariate, $\sum_{i,t} \beta_{i,t} GLOB_{i,t} \times BRDR_{ij}$, is the focus of the model. It comprises time-invariant cross-border dummies $BRDR_{ij}$ interacted with origin-time globalization dummies $GLOB_{i,t}$. This setup captures the impact of globalization on the external extensive margin of trade for each country-year combination in our sample. In the lens of theoretical model (7), suppressing the sector notation, the $\{GLOB_{i,t} BRDR_{ij}\}$ dummies control for $\lambda_{ij,t}^{1-\rho} \phi_{i,t}^{(\rho-1)/\rho}$.

The resulting set of globalization dummies $\{GLOB_{i,t} BRDR_{ij}\}$ absorb and fully control for any non-discriminatory policy or country-specific characteristic that my affect the external/cross-border extensive margin differentially relative to the domestic extensive margin. Even though the set of country-year-specific globalization dummies does not allow us to identify the effects of specific policies, we find their use appropriate to capture the integration processes in Europe. From a methodological perspective, note that none of these effects could be identified without the use of observations on the domestic extensive margin.

A finding that $\beta_{i,t} > 0$ implies that there is a *relative* increase in the international extensive margin relative to the domestic extensive margin. In principle, a positive estimate of $\beta_{i,t}$ may reflect several scenarios, e.g., (i) faster growth on the external margin and slower growth on the domestic margin; (ii) no change on the external margin but a decrease on the domestic margin, (iii) growth on the international extensive margin and no change on the domestic extensive margin; (iv) growth on the international extensive margin and decrease on the domestic extensive margin; (v) decrease on the international extensive margin and faster decrease on the domestic extensive margin, etc. Based on the DEM and the IEM indexes and their relationship that we discussed in the data section, we saw examples of each of those scenarios and we will return to them when we interpret our results. In sum, what we can identify is the effects of globalization/European integration on the international *relative* to the domestic extensive margin. Finally, we note that, due to perfect collinearity with the country-pair fixed effects, we have to omit the border estimate for one year for each country when we obtain the country-specific estimates. The year we select is the first year of the sample, 2008. Thus, the globalization estimates that we obtain should be interpreted as deviations from the corresponding border effects in 2008.

We start the analysis with benchmark results which impose a common globalization effect across all countries for each year in our sample, i.e., we constrain the country-specific globalization effects to be common across all countries in our sample, $GLOB_t = \sum_i GLOB_{i,t}$ and $\beta_{i,t} = \beta_t$. The results are reported in column (1) of Table 2. From the methodological persepective, note that we obtain estimates of the impact of globalization on the extensive margin even in the presence of the full set of exporter-time and importer-time fixed effects. These globalization effects cannot be identified if the model is estimated without the domestic extensive margin observations.[24] Note also that identification does not come from any variation over time – identification of border effects in the presence of exporter and importer fixed effects even in a cross section setting so long as the dataset includes domestic trade.

[24]In that case, the $GLOB_t$ variables would have been perfectly collinear with and, therefore, absorbed by the country-time fixed effects.

Second, from a policy perspective, we find that, with the exception of the estimate on $GLOB\_2009$, the globalization estimates in column (1) of Table 2 are (i) all positive and (ii) they are increasing monotonically over time. In combination, the implication of these results is that the impact of international borders on the international relative to the domestic extensive margin has fallen significantly between 2008 and 2018 for the countries in our sample. In other words, globalization has had a significant positive impact on the international extensive margin of trade relative to the domestic extensive margin for the European economies. Intensive integration processes within Europe are the natural explanation for this result. We find this result to be important from a policy perspective because there is plenty of anecdotal evidence that the impact of globalization on the intensive margin stalled during the years after the great recession, while our results indicate that the impact of globalization on the extensive margin during the same period has been economically strong and statistically significant.

For a meaningful interpretation with respect to the extensive margin, we obtain the marginal effect of the globalization estimate in 2018, which captures the total impact of globalization during the period of investigation. The marginal effect is 270.797 (std.err. 38.050), which means that, on average, the number of internationally traded products increased by about 271 relative to the number of domestically traded products during the period of investigation, or about 7.2 percent of the total number of possibly traded products in 2018.

The rest of the columns in Table 2 offer estimates from a series of sensitivity experiments designed to test the robustness of our main findings. Broadly, we split our robustness checks in two categories: (i) alternative estimators, which are reported in panel B of Table 2; and (ii) alternative samples, which are reported in panel C of Table 2. Specifically, the results in columns (2) to (4) of panel B are obtained with the Tobit, the PPML, and the OLS estimators, respectively. We also experiment with the following alternative estimating samples: (i) the "Exporter-based Sample", as defined in the Data section, in column (5);

(ii) the "Importer-based Sample", in column (6); (iii) the main sample but only with positive observations, in column (7); (iv) a three-year interval sample, in column (8); (v) the "Conservative-product Sample", in column (9); and (vi) the "Extended-country Sample", in column (10). Based on the results from these experiments, we conclude that our main findings of the impact of globalization on the extensive margin of trade are robust to the use of alternative estimators and alternative estimating samples.

Finally, the estimates in the last two columns of Table 2 are obtained with the *"Export-based DEM Sample"*, where the domestic extensive margin is constructed as the total number of products exported by this country to any other country in the world. Two main findings stand out. First, we note that, overall, the results in column (11) confirm our main conclusions. However, second, we notice that the monotonic increase in the globalization estimates is violated in 2017, where the estimate on $GLOB\_2017$ is still statistically significant but smaller as compared to the estimates on $GLOB\_2016$. Inspection of the underlying domestic margin data reveals some unusual patterns. Specifically, as illustrated in Figure 1, which graphs the yearly percentage changes in the domestic extensive margin for all countries in our sample, there are four unusual spikes in 2017, which are for Cyprus, Malta, Iceland, and Montenegro. Column (12) of Table 1 reproduces the results from column (11) but without the outliers. The monotonically increasing pattern of the globalization estimates is restored.

Consistent with the theory, the main specification allows for differential, country-specific effects of globalization, i.e., we employ $\sum_{i,t} \beta_{i,t} GLOB_{i,t} \times BRDR_{ij}$, where the globalization estimates, $\hat{\beta}_{i,t}$, now vary not only for each year but also for each country in our sample. We remind the reader that, due to perfect collinearity with the country-pair fixed effects, we need to drop one border estimate for each country and our choice are the country effects for 2008. Thus, all other country-specific globalization estimates should be interpreted as deviations from the corresponding border effect for the same country in 2008 and, by construction, the estimates for 2018 would capture the total (cumulated) effects during the period 2008-2018. Our results appear in Table 3.

The main implication of the estimates in Table 3 is wide heterogeneity of estimated globalization effects, mostly statistically significant. To facilitate discussion, we focus on the cumulative effects for 2018 from the last column of Table 3, and we plot them in Figure 2. The figure enables us to group the countries in our sample in four categories.

The first group includes five countries, Austria, Belgium, Sweden, France and Germany – the only countries for which we obtain negative globalization estimates for 2018. In fact, Austria is the only country with a negative and statistically significant estimate of the impact of globalization on the international extensive margin. (See Table 3 for statistical significance). The negative estimates for five countries only (with only one of them statistically significant) contrasts with the vast majority of countries in Europe that have enjoyed positive globalization effects on the extensive margin between 2008 and 2018.

The second distinct group in Figure 2 includes countries with positive but relatively small estimates. Here, we find some of the other large economies in Europe, including Italy and the United Kingdom, along with some smaller economies such as Norway, Finland, and Greece. The absence of strong globalization effects on the extensive margin for the more developed economies in this group and the first group suggests that they had reached their extensive margin potential within Europe prior to the 2008-18 period. Combined with the decreased domestic intensive margin that we documented in Section 3 and Table 1, the moderate estimate for the United Kingdom suggests that the UK did not experience significant international extensive margin growth between 2008 and 2018. A possible explanation for the relatively small impact for Greece is that, as noted in Section 3 (see Table 1), Greece is among the countries that experienced the largest increase in the domestic extensive margin. Thus, the positive estimate for this country suggests very significant increase on the international extensive margin. In this group we also notice (i) some potential EU members,, e.g., Northern Macedonia and Bosnia and Herzegovina, that may have already gained access to the large European market in preparation for joining the EU; and (ii) some old members such as Ireland and Denmark.

The countries in the third group in Figure 2 have enjoyed significant positive effects from European integration. Almost all of these countries are recent EU members, e.g., Slovenia, Poland, Bulgaria, Slovakia, Estonia, and Croatia. Integration within the EU is the natural explanation for the large effects for the new members. The large gains for Iceland are probably due to the fact that this country is heavily integrated with the EU (through its membership in European Economic Area and the Schengen Agreement). Spain is also in this group, and we find the estimates for this country particularly interesting because, as discussed in the data description section, and as can be seen from Table 1, Spain is among the countries with the highest and the most stable domestic extensive margin indexes. Thus, in combination with the stable domestic extensive margin, the large and positive estimates for Spain can plausibly be interpreted as absolute positive effects of the impact of globalization on the extensive margin for this country.

Finally, the group of countries that have enjoyed the largest effects of globalization and integration within Europe includes some countries that joined the EU during, or close to the period of investigation (e.g., Romania, Latvia and Lithuania) as well as the Netherlands and Portugal. A tentative implication from Figure 2 is that the biggest winners from the impact of globalization on the extensive margin within Europe tend to be the smaller and poorer European economies, especially those that recently joined the European Union. The third group and even the second group also contain such economies, so more detailed analysis is required to explain the variation. Moreover, the fourth group contains the Netherlands, relatively small and rich, and a founding member of European integration. A full analysis requires more development of the structural implications of the theory. Nevertheless, this initial exploration yields an encouraging message with clear implications for the impact of European integration for development and inequality.

To test the robustness of our results, we reproduce the results from Table 3 with alternative estimators and with alternative samples. For clarity and simplicity of exposition, we do not report all results but, instead, we focus on one representative country from each of the

three groups that we identified in Figure 2, and we present our findings graphically. Specifically, we chose Sweden, Ireland, and Portugal. Figure 3 visualizes the estimates that we obtain with the four alternative estimators. Panel A presents our main estimates from Table 3, which are obtained with the *FLEX* estimator. The estimates in Panel B are obtained with the Tobit estimator. PPML estimates appear in Panel C. Finally, the results in Panel D are obtained with the OLS estimator. Based on the estimates in Figure 3, we conclude that our main findings about the (heterogeneous) impact of globalization on the extensive margin are robust to the use of alternative estimators.

The results in Figures 4 and 5 are obtained with alternative estimating samples. In particular, Panel A of Figure 4 visualizes estimates from our "Exporter-based sample" as described in the Data section. Panel B instead uses the "Importer-based sample" . The results in Panel C are based only on the positive observations in the main sample. The estimates in Panel D are obtained with three-year interval data. The estimates in Panel E use the "Conservative-product" sample. Finally, the results in Panel F are obtained with the "Extended-country" sample. Based on the estimates in Figure 4, we conclude that our main findings about the (heterogeneous) impact of globalization on the extensive margin are robust to the use of these alternative estimating samples.

Figure 5 reports estimates that are based on two samples with alternative definitions of the domestic extensive margin. Specifically, the estimates in Panel A of Figure 5 are obtained with our main sample, where the DEM is constructed directly from the raw PRODCOM data, while the results in Panel B are obtained from a sample where the DEM is constructed as the total number of products that are exported based on the COMEXT export, as described in Section 3. Two main findings stand out from Figure 5. First, we see that the estimates and their evolution over time is comparable, between the two panels, for Ireland and for Portugal. However, second, the evolution of the globalization estimates for Sweden is quite different. The natural explanation for this result is, of course, the difference in the construction of the domestic extensive margin. Comparison between the evolution of DEM for Sweden

depending on the construction method reveals that the number of products that Sweden produces has fallen in both cases. However, the decrease is almost three times larger (i.e., by 188 vs. 65 products) in the export-based DEM sample. This explains the difference between the two panels and points to the importance of proper measurement of the domestic extensive margin.

We conclude the empirical analysis with an investigation of the extensive margin effects of European integration on the new EU members. Given the period of investigation, we focus on three countries, including Bulgaria and Romania, which both joined in 2007 (the year before the start of our sample), and Croatia, which joined in 2013. Even though the sample of new EU members is small, we find the analysis instructive both from a methodological and from a policy perspective. In order to emphasize the some important aspects of our specifications and corresponding estimates, we develop the analysis sequentially, in three specifications. The estimates are presented in Table 4. Each of the three panels in Table 4 reports estimates from a different specification. The dependent variable is always the number of products sold from exporter i to importer j, including domestic sales, and all estimates are obtained with the Flex estimator of Santos Silva et al. (2014). All specifications include exporter-time, importer-time, and pair fixed effects, whose estimates are omitted for brevity. The difference between the three panels is in the set of covariates.

The results in Panel A correspond to the estimates from our main specification from column (1) of Table 2, which are obtained from equation (6). The difference between the results in Tables 2 and 4 is that, in addition to the common globalization effects from Table 2, we now introduce a set of border dummies for trade between the three new EU members and the old EU members. For brevity, we only report the estimates of the globalization effects on the extensive margin of trade between the new and the old EU members.[25] By construction, these estimates should be interpreted as deviations from the corresponding common globalization estimates for the same year. The main message from the results in

[25]The estimates of common globalization effects from Panel A of Table 2 are not statistically different from the corresponding estimates in column (1) of Table 2, and they are available by request.

Panel A is that we do not see stronger additional effects for trade between the new and the old EU members until in the last three years in our sample.

The results in Panel B Table 4 are obtained from the same specification as in Panel A, with the only difference that we allow for asymmetric/directional extensive margin effects for imports to the new from the old EU members ($Imp.EU$) vs. exports from the new EU to the old EU members ($Exp.EU$).

The results in Panel B tell a clear and interesting story. Specifically, we see that there was a significant increase in the extensive margin of trade from the old to the new EU members, but not the other way around. One implication, from a methodological perspective, is that the common/symmetric effects from Panel A are masking significant directional heterogeneity. The implication from a policy perspective is that the new EU members were not able to position their (possibly inferior) products very well in the developed West-European market, however, the new EU members welcomed the significant increase in varieties from Western Europe. We remind the reader that the negative estimates in column (3) should be interpreted as deviations from the common globalization effects. Thus, they imply that (i) there have still been some gains for the exports from the new to the old EU members, and (ii) that the effects on the exports from the new to the old EU members have been converging steadily toward the average common effects.

Finally, the results in Panel C replicate the directional borders specification from Panel B, but in addition we allow for country-specific effects for each of the three new EU members. For consistency, this specification allows for country-specific border effects as in Table 3. The results in Panel C confirm the main conclusions from Panel B, i.e., that for each country the effects on the extensive margin on the imports to the new from the old EU members are significantly larger than the average effects, while the effects on the exports from the new to the old EU members are small than the average but converging toward them.

Two interesting patterns are revealed in Panel C. First, while the estimates for the imports of Bulgaria and Romania are very similar, the effects on the extensive margin for Romanian

exports are more favorable. Second, in the first three years the estimates on EU exports to Croatia are actually negative and marginally insignificant. This is an important result because these are exactly the years when Croatia was not an EU member. After that, we see that the results for Croatia mimic the evolution of the estimates for Bulgaria and Romania. This confirms our implicit assumption that a significant fraction of the extensive margin effects for the new EU members have been triggered by their accession to the European Union.

5 Conclusion

This paper develops a short-run gravity theory of the extensive margin of production and trade and introduces the concept of the *domestic extensive margin (DEM)*. To demonstrate our methods, we utilize the domestic extensive margin to quantify the impact of globalization and European integration on the extensive margin of trade for 32 countries over the period 2008-2018. The new *DEM* concept and the accompanying analysis reveal a series of meaningful opportunities for future academic research and policy impact. We group this opportunities in four related areas, including: (i) theoretical contributions; (ii) new data development; (iii) the construction of new extensive margin and export diversification indexes; and (iv) a series of applications. We elaborate on each of these directions with some specific examples next.

On the theory front, we see potential to use *DEM* and its relation to the international extensive margin in order to challenge the standard assumption in the trade literature that, before exporting a given product, firms are already necessarily selling this product domestically. This idea is motivated by anecdotal evidence that points to alternative scenarios, e.g., where some products are simultaneously offered for sale on the domestic and on the foreign markets, or even when products are first exported and only then they are sold domestically. We believe that, in combination with theory, our new dataset that combines the international

and the domestic extensive margin can provide interesting insights in this direction.

To perform the empirical analysis we constructed a dataset covering the domestic (and international) extensive margin for the European economies. We see significant potential benefits from expanding the dataset to cover all possible countries in the world. For example, one clear advantage of such database would be that it will include the poorer and less-developed economies, where export diversification and the extensive margin are particularly important. We believe that the creation of such extended dataset is feasible and, in fact, significantly easier and more reliable as compared to a corresponding dataset on the intensive margin of trade. The reason is that in order to construct the *the domestic extensive margin*, all we need is an indicator on whether a given product is produced or not, and we do not need information on the actual volume of production (or trade), which is more problematic for various reasons and especially at the very disaggregated levels.

In addition to the *Domestic Extensive Margin Index* introduced in this paper (as the ratio between the number of domestically produced products and the total number of possible products that a country can produce), we see value in the construction of two related indexes. The first one is an *Export Diversification Index,* defined as the ratio between the number of exported products and the number of domestically produced products. We believe that this index will complement the existing Export Diversification (or Concentration) indexes, which are defined only based on export data and without taking into account the domestic extensive margin.[26] The second one is an *Extensive Margin Openness to Trade Index,* defined as the ratio between the sum of the number of exported products and the number of imported products divided by the number of domestically produced products. We see this index as the extensive margin counterpart of the standard Openness to Trade (OTT) index that is widely used in both the academic literature and for policy purposes. Consistent with our theory, each of these indexes can be constructed at the sectoral level.

Finally, our methods offer opportunities to evaluate a series of exciting applications.

[26]See https://wits.worldbank.org/wits/wits/witshelp/Content/Utilities/e1.trade_indicators.htm and https://www.imf.org/external/datamapper/datasets/SPRLU.

For example, the new DEM dataset calls for an analysis of the impact of the determinants of the domestic extensive margin. We believe that an important contribution in this area would be to use the data on the domestic extensive margin to perform a direct test for one of the main implications of the seminal theory of Melitz (2003), according to which trade liberalization leads to exit of the less productive firms. A descriptive look at our DEM data offers supportive preliminary evidence for the general validity Melitz's theory, but also points to potentially interesting heterogeneous effects.

In addition, our methods allow for an evaluation of the impact of non-discriminatory trade policies (e.g, export subsidies, export promotion, etc.) and country-specific characteristics (e.g., institutional quality, country-specific taxes, etc.) on export diversification and the extensive margin of trade. It is important to emphasize that without the domestic extensive margin one cannot identify the effects of any non-discriminatory trade policies and country-specific characteristics on the international extensive margin in a properly specified econometric model, i.e., with exporter(-time) and importer(-time) fixed effects that would control for the theory-motivated multilateral resistances.

References

Agnosteva, Delina E., James E. Anderson, and Yoto V. Yotov, "Intra-national Trade Costs: Assaying Regional Frictions," *European Economic Review*, 2019, *112*, 32–50.

Anderson, James E., "A Theoretical Foundation for the Gravity Equation," *American Economic Review*, 1979, *69* (1), 106–116.

_ **and Eric van Wincoop**, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 2003, *93* (1), 170–192.

Anderson, James E. and Yoto V. Yotov, "Short Run Gravity," *Journal of International Economics*, 2020, *126* (September).

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

_ **, Arnaud Costinot, and Andrés Rodríguez-Clare**, "New Trade Models, Same Old Gains?," *American Economic Review*, 2012, *102* (1), 94–130.

Besedes, Tibor and Thomas J. Prusa, "Product differentiation and duration of US import trade," *Journal of International Economics*, December 2006, *70* (2), 339–358.

Cadot, Olivier, Céline Carrère, and Vanessa Strauss-Kahn, "Export Diversification: What's behind the Hump?," *The Review of Economics and Statistics*, May 2011, *93* (2), 590–605.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 2008, *98* (4), 1707–1721.

_ , "The Network Structure of International Trade," *American Economic Review*, 2014, *104* (11), 3600–34.

Crucini, Mario J. and J. Scott Davis, "Distribution capital and the short- and long-run import demand elasticity," *Journal of International Economics*, 2016, *100* (C), 203–219.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Gourieroux, Christian, Alain Monfort, and Alain Trognon, "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, May 1984, *52* (3), 701–720.

Gurevich, Tamara and Peter Herman, "The Dynamic Gravity Dataset: 1948-2016," *USITC Working Paper 2018-02-A*, 2018.

Head, K., T. Mayer, and J. Ries, "The Erosion of Colonial Trade Linkages after Independence," *Journal of International Economics*, 2010, *81* (1), 1–14.

Heid, Benedikt, Mario Larch, and Yoto V. Yotov, "A Simple Method to Estimate the Effects of Non-discriminatory Trade Policy within Structural Gravity Models," *Canadian Journal of Economics*, forthcoming.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein, "Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 2008, *123* (2), 441–487.

_ , **Oleg Itskhoki, and Stephen Redding**, "Wages, Unemployment and Inequality with Heterogeneous Firms and Workers," *NBER Working Paper No. 14122*, 2008.

Hummels, D. and Peter J. Klenow, "The Variety and Quality of a Nation's Exports," *American Economic Review*, 2005, *95* (3), 704–723.

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

Mion, Giordano and Luca David Opromolla, "Managers' mobility, trade performance, and wages," *Journal of International Economics*, 2014, *94* (1), 85–101.

Papke, Leslie and Jeffrey Wooldridge, "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, 1996, *11* (6), 619–32.

Ramondo, Natalia, Andrés Rodríguez-Clare, and Milagro Saborio-Rodriguez, "Trade, Domestic Frictions, and Scale Effects," *American Economic Review*, 2016, *106* (10), 3159–3184.

Sampson, Thomas, "Dynamic Selection: An Idea Flows Theory of Entry, Trade, and Growth," *The Quarterly Journal of Economics*, 2016, *131* (1), 315–380.

Santos Silva, J. M. C., "A score test for non-nested hypotheses with applications to discrete data models," *Journal of Applied Econometrics*, 2001, *16* (5), 577–597.

Santos Silva, J.M.C. and Silvana Tenreyro, "The Log of Gravity," *Review of Economics and Statistics*, 2006, *88* (4), 641–658.

_ **and** _ , "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator," *Economics Letters*, 2011, *112* (2), 220–222.

_ , _ , **and Kehai Wei**, "Estimating the extensive margin of trade," *Journal of International Economics*, 2014, *93* (1), 67–75.

Yotov, Yoto V., "A Simple Solution to the Distance Puzzle in International Trade," *Economics Letters*, 2012, *117* (3), 794–798.

Table 1: The Domestic Extensive Margin, 2008-2018

Country Name	ISO	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	$\%\Delta DEM$	$\%\Delta IEM$
Austria	AT	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.52	0.52	0.52	1.69	-6.32
Bosnia and Herzegovina	BA	.	.	.	0.19	0.19	0.19	0.19	0.19	0.20	0.19	0.19	3.31	-1.05
Belgium	BE	0.49	0.49	0.48	0.48	0.49	0.49	0.49	0.49	0.49	0.50	0.50	2.63	-6.31
Bulgaria	BG	0.48	0.48	0.48	0.48	0.48	0.48	0.49	0.49	0.49	0.49	0.49	3.01	1.74
Czechia	CZ	0.65	0.65	0.65	0.65	0.65	0.66	0.65	0.66	0.66	0.66	0.66	2.05	-5.77
Germany	DE	0.90	0.89	0.89	0.89	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.69	-5.69
Denmark	DK	0.51	0.51	0.51	0.51	0.51	0.51	0.52	0.52	0.52	0.53	0.53	4.25	-5.56
Estonia	EE	0.32	0.32	0.33	0.34	0.34	0.34	0.34	0.34	0.33	0.34	0.34	5.47	2.84
Spain	ES	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.83	0.84	0.84	0.84	-0.21	-6.47
Finland	FI	0.53	0.52	0.52	0.52	0.51	0.51	0.51	0.51	0.50	0.51	0.51	-2.91	-4.39
France	FR	0.82	0.82	0.82	0.82	0.83	0.83	0.83	0.83	0.84	0.84	0.84	2.98	-6.84
United Kingdom	GB	0.82	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.80	0.80	-2.08	-6.47
Greece	GR	0.33	0.33	0.33	0.32	0.32	0.34	0.34	0.34	0.34	0.35	0.35	6.17	-2.98
Croatia	HR	0.42	0.42	0.41	0.41	.	0.40	0.41	0.40	0.39	0.40	0.39	-6.60	5.19
Hungary	HU	0.52	0.52	0.52	0.53	0.55	0.55	0.56	0.57	0.57	0.57	0.57	10.94	6.55
Ireland	IE	0.29	0.29	0.30	0.29	0.29	0.29	0.29	0.29	0.29	0.30	0.30	1.94	-1.93
Iceland	IS	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	1.25	2.63
Italy	IT	0.85	0.86	0.86	0.86	0.86	0.85	0.85	0.84	0.85	0.86	0.83	-2.13	-6.66
Lithuania	LT	0.37	0.38	0.36	0.37	0.38	0.38	0.37	0.39	0.38	0.39	0.40	8.00	0.79
Latvia	LV	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.21	0.21	0.21	2.12	3.47
Montenegro	ME	.	.	.	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	-15.58	53.23
North Macedonia	MK	.	.	.	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	1.40	13.15
Netherlands	NL	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.47	0.50	0.51	0.51	8.49	-6.33
Norway	NO	0.22	0.22	0.22	0.22	0.21	0.22	0.22	0.21	0.22	0.22	0.22	-0.47	-5.09
Poland	PL	0.74	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.76	0.75	2.23	-5.33
Portugal	PT	0.59	0.58	0.58	0.58	0.58	0.53	0.54	0.53	0.53	0.54	0.54	-9.57	-2.72
Romania	RO	0.42	0.42	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.44	0.44	2.88	2.35
Sweden	SE	0.46	0.47	0.46	0.47	0.47	0.47	0.47	0.47	0.48	0.48	0.48	2.99	-5.49
Slovenia	SI	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.37	0.38	6.52	-2.15
Slovakia	SK	0.31	0.31	0.31	0.31	0.31	0.32	0.31	0.31	0.31	0.33	0.32	3.23	2.58
Turkey	TR	0.44	0.44	0.45	.	.	.	.	.	.	.	.	0.95	-4.36
Serbia	XS	.	.	.	0.03	0.25	0.25	0.25	0.26	0.26	0.27	0.27	7.54	3.56
Total Number of Products	.	3,890	3,877	3,860	3,827	3,818	3,816	3,814	3,811	3,811	3,756	3,756	-3.44	.

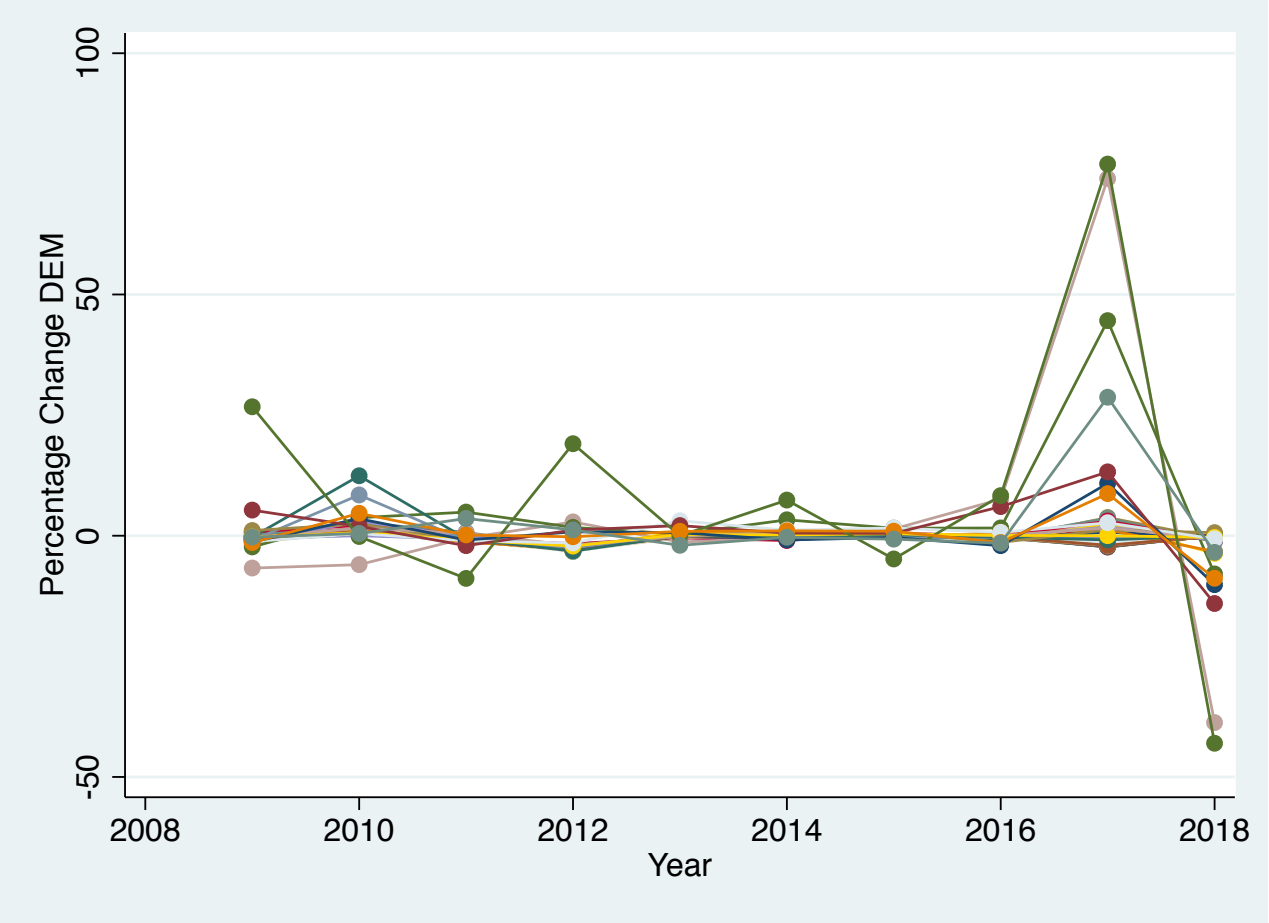
Notes: This table reports the domestic extensive margin (DEM) indexes for the countries in the PRODCOM database over the period 2008-2018. The indexes are constructed as the ratio between the number of products that are produced by a country in a given year, $D_{i,t}$, over the total number of possible products that could be produced in the same year, $N_{i,t}$: $DEM_{i,t} = D_{i,t}/N_{i,t}$. The total number of products are reported in the last row of the table. The second to last column of the table reports percentage changes in the DEM index between the first and the last year for which data were available. The exception is Serbia, for which the initial year for the percentage change in the past column is 2012. The last column of the table reports the percentage changes between 2008 and 2018 in the international extensive margin numbers, $\%\Delta IEM$, which are constructed as the total number of exported products for each year. See text for further details.

Table 2: Globalization/European Integration and the Extensive Margin of Trade, 2008-2018

	A. Main	B. Alternative Estimators			C. Alternative Samples							
	FLEX	TOBIT	PPML	OLS	EXPRT	IMPRT	PSTV	INTLRV	PROD	CNTRY	DEMEXP	NOOUTL
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
BRDR_2009	-0.009	-4.560	-0.000	-0.012	0.006	-0.026	-0.009	-0.027	-0.010	-0.011	0.017	0.017
	(0.009)	(7.514)	(0.006)	(0.011)	(0.011)	(0.008)**	(0.009)	(0.009)**	(0.009)	(0.011)	(0.011)	(0.011)
BRDR_2010	0.072	42.212	0.034	0.040	0.082	0.062	0.072		0.068	0.070	-0.046	-0.045
	(0.025)**	(14.843)**	(0.011)**	(0.018)*	(0.020)**	(0.035)$^{+}$	(0.025)**		(0.025)**	(0.022)**	(0.034)	(0.033)
BRDR_2011	0.117	71.679	0.051	0.067	0.133	0.104	0.117		0.122	0.106	0.064	0.065
	(0.030)**	(17.690)**	(0.014)**	(0.025)**	(0.025)**	(0.040)**	(0.030)**		(0.030)**	(0.027)**	(0.034)$^{+}$	(0.034)$^{+}$
BRDR_2012	0.154	102.959	0.067	0.076	0.175	0.141	0.154	0.146	0.171	0.142	0.292	0.289
	(0.034)**	(19.107)**	(0.016)**	(0.031)*	(0.033)**	(0.040)**	(0.034)**	(0.042)**	(0.033)**	(0.031)**	(0.043)**	(0.042)**
BRDR_2013	0.199	129.159	0.088	0.129	0.214	0.191	0.199		0.224	0.189	0.326	0.323
	(0.040)**	(22.612)**	(0.019)**	(0.035)**	(0.039)**	(0.045)**	(0.040)**		(0.039)**	(0.037)**	(0.044)**	(0.044)**
BRDR_2014	0.229	145.953	0.102	0.150	0.255	0.212	0.229		0.256	0.221	0.366	0.362
	(0.043)**	(24.266)**	(0.021)**	(0.036)**	(0.043)**	(0.047)**	(0.043)**		(0.043)**	(0.040)**	(0.048)**	(0.047)**
BRDR_2015	0.287	177.271	0.129	0.204	0.319	0.265	0.287	0.274	0.320	0.279	0.409	0.405
	(0.047)**	(26.322)**	(0.025)**	(0.040)**	(0.049)**	(0.051)**	(0.047)**	(0.052)**	(0.048)**	(0.044)**	(0.048)**	(0.047)**
BRDR_2016	0.325	198.697	0.147	0.255	0.367	0.296	0.325		0.370	0.318	0.483	0.477
	(0.052)**	(28.392)**	(0.029)**	(0.047)**	(0.055)**	(0.055)**	(0.052)**		(0.054)**	(0.049)**	(0.049)**	(0.048)**
BRDR_2017	0.429	254.843	0.186	0.314	0.456	0.422	0.429		0.495	0.422	0.399	0.492
	(0.072)**	(40.638)**	(0.037)**	(0.053)**	(0.066)**	(0.095)**	(0.072)**		(0.081)**	(0.071)**	(0.077)**	(0.066)**
BRDR_2018	0.456	268.345	0.200	0.339	0.479	0.454	0.456	0.466	0.525	0.449	0.579	0.574
	(0.065)**	(37.192)**	(0.038)**	(0.057)**	(0.065)**	(0.082)**	(0.065)**	(0.085)**	(0.072)**	(0.059)**	(0.063)**	(0.062)**
WTO	0.390	-16.522	0.396	0.389	0.391	0.373	0.390		0.364	0.379	-0.032	-0.026
	(0.054)**	(15.460)	(0.051)**	(0.070)**	(0.059)**	(0.050)**	(0.054)**		(0.051)**	(0.053)**	(0.093)	(0.073)
EIA	-0.067	-24.227	-0.063	-0.028	-0.131	0.004	-0.067	0.049	-0.051	-0.079	-0.087	-0.084
	(0.047)	(24.562)	(0.032)*	(0.093)	(0.063)*	(0.051)	(0.047)	(0.067)	(0.047)	(0.035)*	(0.031)**	(0.030)**
N	9656	9656	9656	9656	9656	9656	9656	4284	9641	22807	12984	12743
$R2$				0.988								

Notes: This table reports gravity estimates of the impact of globalization/European integration on the extensive margin of trade. The dependent variable in all columns, except column (4), is the number of products sold from exporter i to importer j, including domestic sales. All estimates are obtained with exporter-time, importer-time, and pair fixed effects, whose estimates are omitted for brevity. Column (1) reports estimates that are obtained with the Flex estimator of Santos Silva et al. (2014). The estimator in column (2) is Tobit. The results in column (3) are obtained with PPML. Column (4) employs the OLS estimator, and the dependent variable here is the log of the number of products. Columns (5) to (10) employ the Flex estimator, but rely on alternative estimating samples. The dependent variable in column (5) is based on the original export values, where we have replaced the missing export observations with import values, i.e., we employ the "Exporter-based sample" as described in the Data section. Column (6) instead uses the import values as the base to construct the depended variable and replaces the missing import values with export values. i.e., we employ the "Importer-based sample" . Column (7) reports estimates based only on the positive observations in the main sample. The estimates in column (8) are obtained with three-year interval data. The results in column (9) are obtained with the "Conservative-product" sample. The results in column (10) are obtained with the "Extended-country" sample. Finally, the estimates in columns (11)-(13) are obtained with samples where the domestic extensive margin is constructed as the total number of products exported by this country to any other country in the world. The estimates in columns (11) are based on the our main estimating sample with the alternative domestic extensive margin observations. The estimates in column (12) are the same as those in column (11) but without four outliers (Cyprus, Malta, Iceland, and Montenegro). The estimates in column (13) are from a sample with the new domestic margin observations and the "Conservative-product Sample". Standard errors are clustered by country pair and are reported in parentheses. $+$ $p < 0.10$, * $p < .05$, ** $p < .01$. See text for further details.

Figure 1: Export-based DEM Indexes. Percentage Changes, 2008-2018



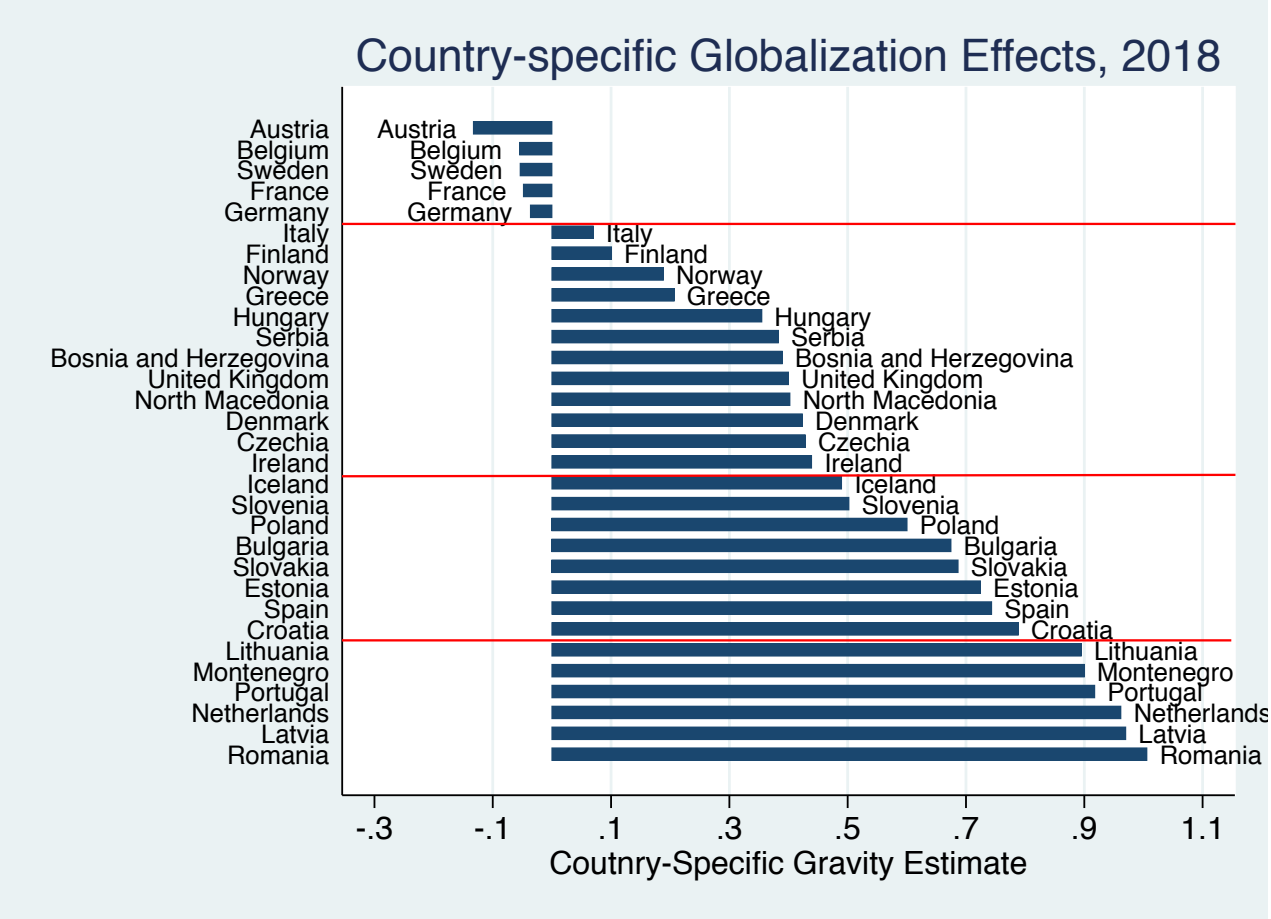
Notes: This figure visualizes the country-specific percentage changes in the domestic extensive margin indexes that are constructed based on export data. See text for discussion and further details.

Table 3: Country-specific Globalization Effects on the Extensive Margin of Trade, 2008-2018

ISO	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
AT	0.004	0.019	-0.018	-0.060	-0.057	-0.045	-0.055	-0.092	-0.133	-0.133
	(.013)	(.016)	(.021)	(.028)*	(.029)*	(.035)	(.036)	(.039)*	(.046)**	(.047)**
BA				0.149	0.199	0.301	0.300	0.288	0.470	0.389
				(.059)*	(.064)**	(.066)**	(.065)**	(.073)**	(.112)**	(.097)**
BE	-0.013	-0.029	-0.026	-0.056	-0.060	-0.019	-0.046	-0.015	-0.083	-0.055
	(.012)	(.019)	(.026)	(.034)	(.032)+	(.035)	(.038)	(.041)	(.052)	(.053)
BG	-0.071	0.044	0.262	0.350	0.333	0.391	0.474	0.510	0.587	0.675
	(.023)**	(.04)	(.047)**	(.049)**	(.047)**	(.05)**	(.05)**	(.048)**	(.055)**	(.061)**
CZ	0.001	-0.009	0.035	0.103	0.139	0.182	0.261	0.334	0.366	0.428
	(.014)	(.019)	(.024)	(.031)**	(.032)**	(.034)**	(.041)**	(.044)**	(.053)**	(.055)**
DE	0.361	0.297	0.017	0.152	0.197	0.168	0.085	0.009	-0.065	-0.036
	(.036)**	(.035)**	(.051)	(.053)**	(.055)**	(.058)**	(.054)	(.054)	(.063)	(.061)
DK	-0.007	0.518	0.541	0.450	0.429	0.450	0.474	0.408	0.367	0.423
	(.012)	(.065)**	(.065)**	(.067)**	(.064)**	(.066)**	(.064)**	(.057)**	(.064)**	(.063)**
EE	-0.057	0.031	0.146	0.222	0.329	0.313	0.425	0.550	0.641	0.725
	(.023)*	(.036)	(.039)**	(.052)**	(.061)**	(.065)**	(.077)**	(.081)**	(.092)**	(.102)**
ES	-0.031	0.019	0.155	0.259	0.370	0.465	0.548	0.624	0.676	0.743
	(.011)**	(.013)	(.032)**	(.041)**	(.048)**	(.054)**	(.06)**	(.066)**	(.074)**	(.079)**
FI	-0.035	-0.075	-0.057	-0.021	-0.027	-0.060	-0.023	0.043	0.060	0.101
	(.011)**	(.018)**	(.026)*	(.032)	(.034)	(.034)+	(.038)	(.037)	(.044)	(.048)*
FR	-0.067	-0.025	-0.081	-0.120	-0.122	-0.097	-0.068	-0.063	-0.061	-0.048
	(.012)**	(.026)	(.035)*	(.048)*	(.048)*	(.05)+	(.052)	(.055)	(.066)	(.067)
GB	-0.002	0.095	0.138	0.150	0.189	0.227	0.254	0.297	0.347	0.400
	(.015)	(.021)**	(.03)**	(.043)**	(.046)**	(.052)**	(.056)**	(.059)**	(.07)**	(.073)**
GR	-0.010	-0.033	-0.027	-0.042	-0.090	-0.042	-0.017	0.038	0.116	0.207
	(.017)	(.028)	(.038)	(.042)	(.045)*	(.049)	(.053)	(.054)	(.058)*	(.066)**
HR	-0.032	0.013	0.142		0.275	0.149	0.334	0.539	0.639	0.788
	(.019)+	(.031)	(.048)**		(.063)**	(.077)+	(.072)**	(.077)**	(.085)**	(.088)**
HU	0.050	0.244	0.224	0.199	0.242	0.237	0.278	0.326	0.297	0.355
	(.014)**	(.02)**	(.029)**	(.032)**	(.034)**	(.037)**	(.037)**	(.041)**	(.05)**	(.053)**
IE	-0.002	-0.049	0.004	0.123	0.188	0.179	0.242	0.289	0.341	0.439
	(.018)	(.032)	(.036)	(.036)**	(.036)**	(.037)**	(.038)**	(.043)**	(.051)**	(.048)**
IS	-0.196	-0.251	-0.145	-0.044	0.041	0.102	0.161	0.357	0.706	0.490
	(.055)**	(.06)**	(.091)	(.107)	(.106)	(.095)	(.104)	(.097)**	(.167)**	(.147)**
IT	-0.079	-0.001	-0.023	-0.021	-0.006	0.019	0.106	0.091	0.048	0.070
	(.013)**	(.022)	(.032)	(.039)	(.043)	(.046)	(.049)*	(.052)+	(.061)	(.064)
LT	-0.032	0.136	0.276	0.328	0.450	0.585	0.671	0.770	0.831	0.895
	(.02)	(.028)**	(.036)**	(.049)**	(.063)**	(.068)**	(.081)**	(.082)**	(.09)**	(.098)**
LV	0.007	0.189	0.344	0.445	0.507	0.541	0.649	0.706	0.817	0.970
	(.029)	(.038)**	(.053)**	(.079)**	(.081)**	(.083)**	(.089)**	(.093)**	(.106)**	(.114)**
ME				0.726	0.683	0.860	0.846	0.964	1.556	0.900
				(.149)**	(.087)**	(.125)**	(.1)**	(.116)**	(.283)**	(.114)**
MK				0.138	0.208	0.254	0.336	0.457	0.585	0.402
				(.059)*	(.076)**	(.089)**	(.091)**	(.094)**	(.097)**	(.12)**
NL	-0.019	-0.007	-0.026	-0.078	-0.099	-0.093	0.068	-0.047	1.652	0.962
	(.011)+	(.017)	(.025)	(.032)*	(.034)**	(.036)**	(.043)	(.043)	(.13)**	(.106)**
NO	-0.014	-0.006	0.079	0.149	0.184	0.263	0.255	0.262	0.263	0.188
	(.031)	(.035)	(.049)	(.062)*	(.061)**	(.067)**	(.061)**	(.065)**	(.069)**	(.075)*
PL	-0.012	0.057	0.141	0.203	0.262	0.378	0.473	0.481	0.484	0.600
	(.011)	(.016)**	(.022)**	(.031)**	(.035)**	(.042)**	(.052)**	(.055)**	(.062)**	(.067)**
PT	-0.020	0.113	0.189	0.294	0.539	0.532	0.659	0.755	0.870	0.918
	(.014)	(.021)**	(.029)**	(.038)**	(.045)**	(.048)**	(.055)**	(.057)**	(.069)**	(.068)**
RO	0.110	0.269	0.372	0.453	0.511	0.607	0.673	0.704	0.727	1.006
	(.025)**	(.037)**	(.042)**	(.048)**	(.048)**	(.052)**	(.056)**	(.058)**	(.064)**	(.071)**
SE	-0.063	-0.046	-0.048	-0.062	-0.071	-0.058	-0.065	-0.097	-0.106	-0.054
	(.01)**	(.018)*	(.025)+	(.032)+	(.033)*	(.035)+	(.037)+	(.039)*	(.045)*	(.046)
SI	0.003	0.068	0.173	0.236	0.398	0.360	0.421	0.465	0.513	0.502
	(.014)	(.021)**	(.038)**	(.038)**	(.047)**	(.04)**	(.038)**	(.041)**	(.049)**	(.053)**
SK	0.048	0.138	0.207	0.286	0.297	0.350	0.411	0.475	0.464	0.687
	(.018)**	(.02)**	(.028)**	(.034)**	(.037)**	(.044)**	(.048)**	(.049)**	(.055)**	(.067)**
TR	-0.026	0.112								
	(.02)	(.03)**								
XS					0.077	0.183	0.226	0.303	0.438	0.383
					(.031)*	(.035)**	(.041)**	(.051)**	(.072)**	(.075)**

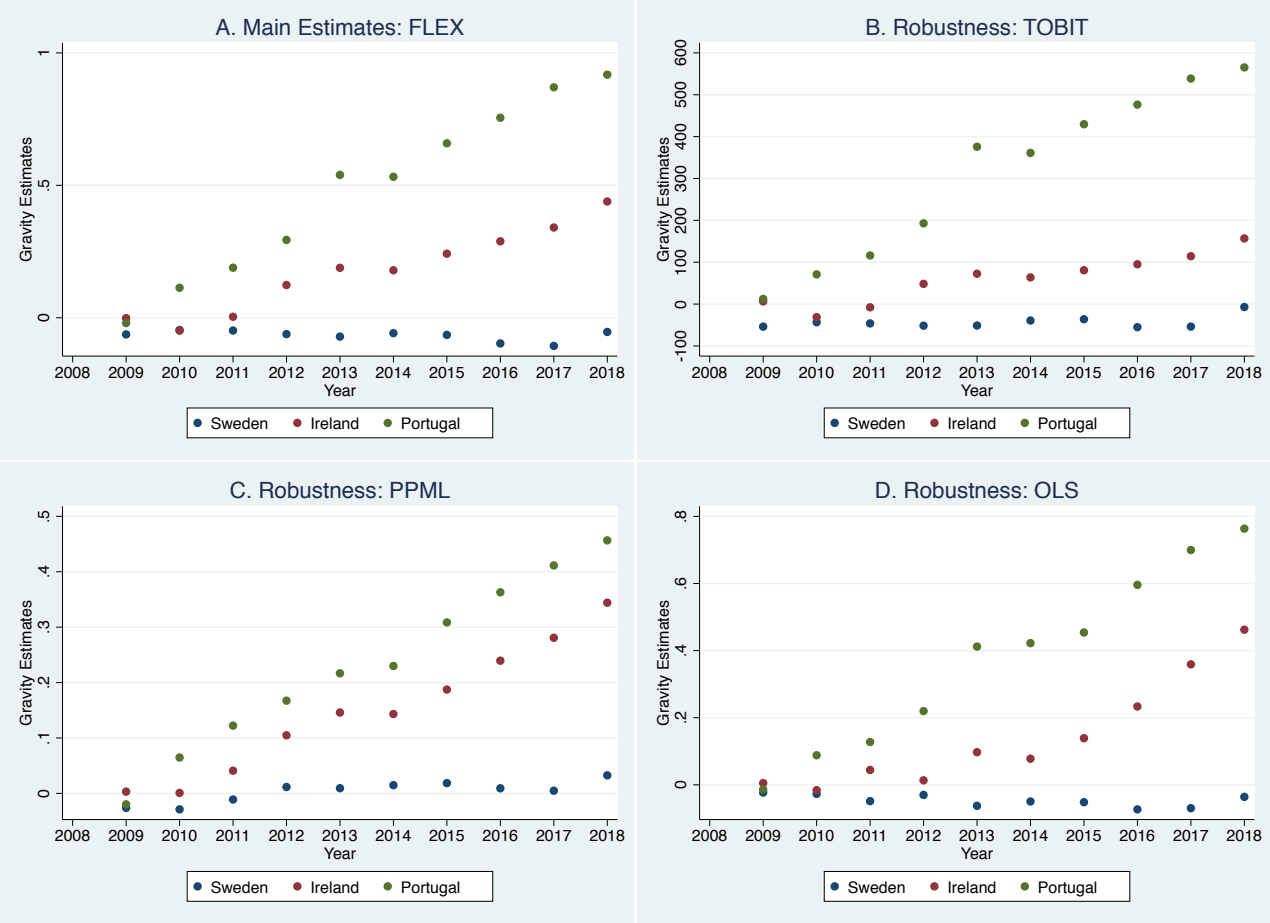
Notes: This table reports country-specific estimates of the impact of globalization/European integration on the extensive margin of trade. The estimates are obtained from equation (6). The dependent variable is the number of products sold from exporter i to importer j, including domestic sales and we use the Flex estimator of Santos Silva et al. (2014). The estimates are obtained with exporter-time, importer-time, and pair fixed effects, whose estimates are omitted for brevity. The two-letter country codes are listed in column (1) and the corresponding country names appear in column (1) of Table 1. Standard errors are clustered by country pair and are reported in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$. See text for further details.

Figure 2: Globalization and the Extensive Margin. Country-specific Effects, 2018



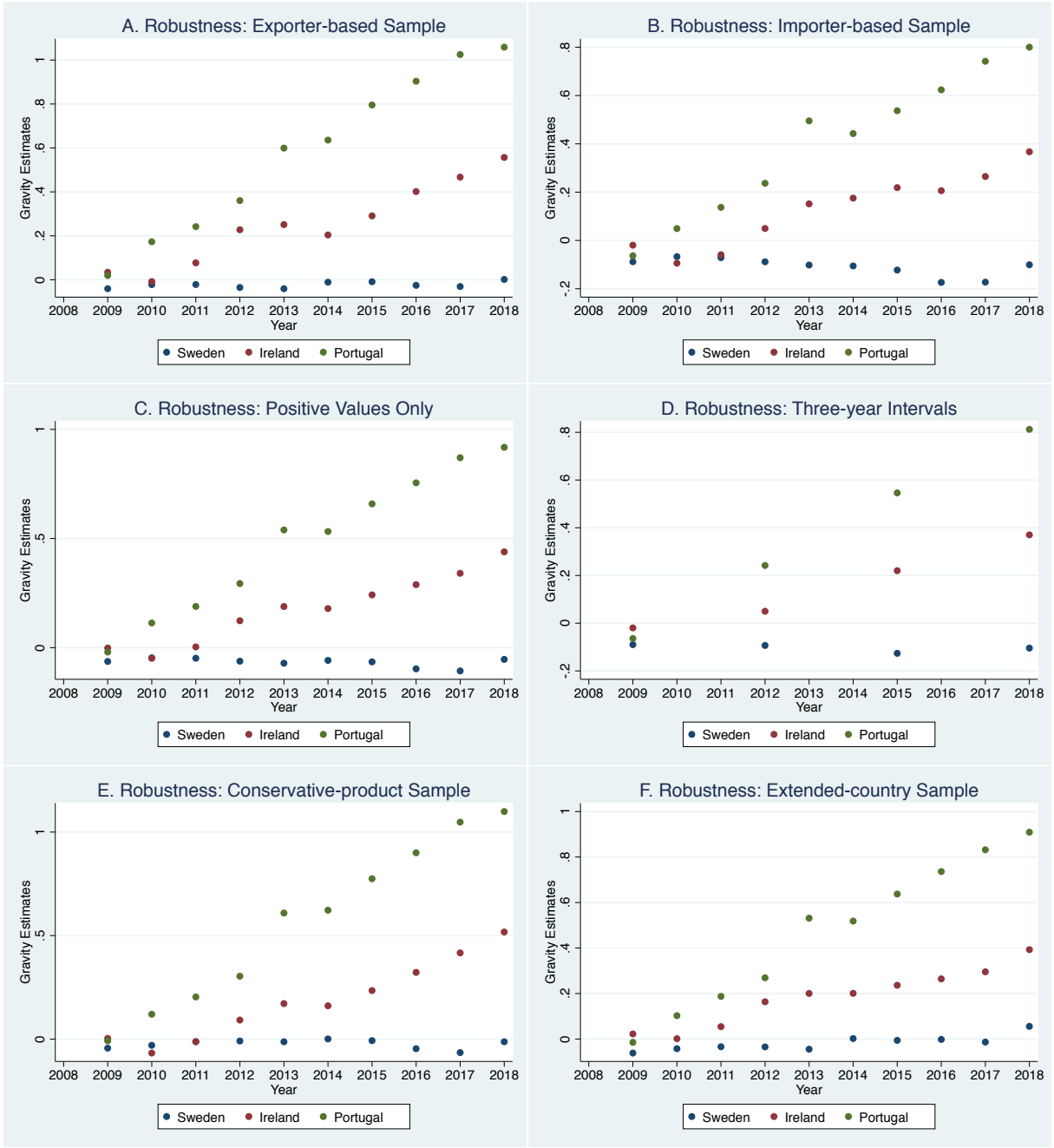
Notes: This figure visualizes the country-specific estimates of the globalization effects on the extensive margin in 2018 for all countries in our sample. The estimates are obtained from equation (6) and appear in the last column of Table 3. See text for discussion and further details.

Figure 3: Country-specific Estimates, Robustness: Alternative Estimators



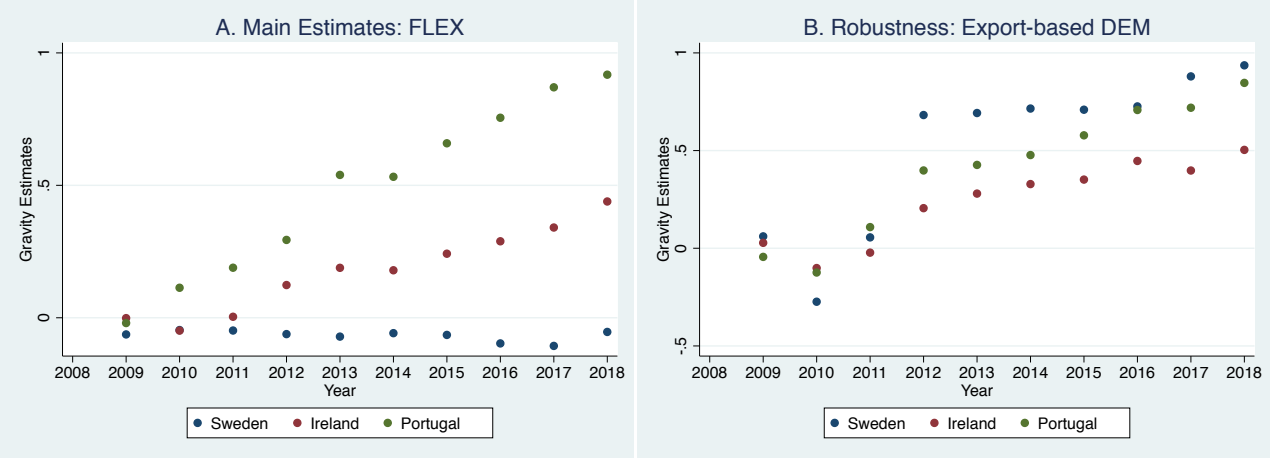
Notes: This figure visualizes the country-specific estimates of the globalization effects on the extensive margin for a selected group of countries including Sweden, Ireland, and Portugal. All estimates are obtained from equation (6), with exporter-time, importer-time, and pair fixed effects in a panel setting for all years and all countries in the sample. The difference between the four panels in the figure are due to the use of alternative estimators. Specifically, Panel A visualizes our main estimates, which are obtained with the Flex estimator of Santos Silva et al. (2014). The estimates in Panel A are in fact those from Table 3. The estimates in Panel B are obtained with the Tobit estimator. PPML estimates appear in Panel C. Finally, the results in Panel D are obtained with the OLS estimator. See text for further details.

Figure 4: Country-specific Estimates, Robustness: Alternative Samples



Notes: This figure visualizes the country-specific estimates of the globalization effects on the extensive margin for a selected group of countries including Sweden, Ireland, and Portugal. All estimates are obtained from equation (6), with exporter-time, importer-time, and pair fixed effects in a panel setting for all countries in the sample. The difference between the six panels of the figure are due to the use of alternative estimating samples. Specifically, Panel A visualizes estimates from our "Exporter-based sample" as described in the Data section. Panel B instead uses the "Importer-based sample" . The results in Panel C are based only on the positive observations in the main sample. The estimates in Panel D are obtained with three-year interval data. Panel E uses the "Conservative-product" sample. Finally, the results in Panel F are obtained with the "Extended-country" sample. See text for further details.

Figure 5: Country-specific Estimates, Robustness: DEM Definition



Notes: This figure visualizes the country-specific estimates of the globalization effects on the extensive margin for a selected group of countries including Sweden, Ireland, and Portugal. All estimates are obtained from equation (6), with exporter-time, importer-time, and pair fixed effects in a panel setting for all countries in the sample. The difference between the two panels of the figure are due to the definition/construction of the domestic extensive margin. Specifically, the estimates in Panel A are our main estimates, which are obtained with a DEM measure based on production data from PRODCOM, while the results in Panel B are obtained with a DEM measure that is constructed based on exports data, as described in Section 3. See text for further details.

Table 4: European Integration and the Extensive Margin for New EU Members

Year	A. Symmetric	B. Asymmetric		C. Country-specific					
				Bulgaria (2007)		Romania (2007)		Croatia (2013)	
		Imp.EU	Exp.EU	Imp.EU	Exp.EU	Imp.EU	Exp.EU	Imp.EU	Exp.EU
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2009	0.00200	-0.0270	0.0330	-0.00100	-0.192	0.0770	0.0560	-0.185	0.369
	(.015)	(.054)	(.059)	(.061)	(.099)+	(.069)	(.066)	(.088)*	(.095)**
2010	0.0260	-0.0200	0.0760	0.0400	-0.0230	0.0950	0.0370	-0.217	0.314
	(.019)	(.057)	(.058)	(.067)	(.094)	(.071)	(.076)	(.089)*	(.104)**
2011	0.0260	0.395	-0.339	0.499	-0.472	0.469	-0.273	0.168	-0.395
	(.024)	(.082)**	(.084)**	(.141)**	(.121)**	(.154)**	(.143)+	(.104)	(.141)**
2012	0.0290	0.378	-0.322	0.393	-0.407	0.332	-0.228		
	(.033)	(.069)**	(.077)**	(.107)**	(.113)**	(.115)**	(.127)+		
2013	0.0430	0.386	-0.297	0.407	-0.318	0.407	-0.197	0.188	-0.520
	(.026)	(.06)**	(.061)**	(.118)**	(.095)**	(.133)**	(.104)+	(.1)+	(.122)**
2014	-0.00100	0.355	-0.356	0.477	-0.286	0.306	-0.230	0.109	-0.707
	(.034)	(.062)**	(.067)**	(.102)**	(.092)**	(.102)**	(.115)*	(.107)	(.116)**
2015	0.0370	0.356	-0.282	0.401	-0.235	0.307	-0.198	0.175	-0.561
	(.034)	(.056)**	(.064)**	(.099)**	(.095)*	(.087)**	(.118)+	(.113)	(.112)**
2016	0.0740	0.401	-0.251	0.393	-0.226	0.324	-0.198	0.294	-0.466
	(.035)*	(.059)**	(.062)**	(.097)**	(.097)*	(.079)**	(.119)+	(.124)*	(.103)**
2017	0.136	0.428	-0.161	0.410	-0.120	0.311	-0.149	0.425	-0.305
	(.04)**	(.073)**	(.067)*	(.092)**	(.105)	(.1)**	(.137)	(.165)**	(.106)**
2018	0.160	0.424	-0.104	0.389	-0.0860	0.335	-0.0750	0.343	-0.295
	(.042)**	(.067)**	(.066)	(.103)**	(.103)	(.096)**	(.123)	(.137)*	(.11)**

Notes: This table reports estimates of the impact of European integration on the extensive margin of trade for the three most recent EU members in our sample, i.e. Bulgaria (2007), Romania (2007), and Croatia (2013). Each panel of the table reports estimates from a different specification. The dependent variable in each specification is the number of products sold from exporter i to importer j, including domestic sales, and all estimates are obtained with the Flex estimator of Santos Silva et al. (2014). All specifications include exporter-time, importer-time, and pair fixed effects, whose estimates are omitted for brevity. The difference between the three panels is in the set of covariates. Specifically, the results in Panel A are obtained from our main specification from column (1) of Table 2, i.e., based on equation (6), where in addition to the common border effects we have introduced a set of common (non-directional) border dummies that capture the EU effects on the three new EU members. For brevity we omit the common border effects, which are not statistically different from the estimates in column (1) of Table 2. The results in Panel B are based on the same specification but allow for asymmetric/directional effects for exports from the old EU to the new EU members ($Imp.EU$) vs. exports from the new EU to the old EU members ($Exp.EU$). Once again, we omit the common border estimates. Finally, the results in Panel C allow for directional and also country-specific effects for each of the three new EU members. For consistency, this specification allows for country-specific border effects as in Table 3. The country-specific border estimates are omitted for brevity. Standard errors are clustered by country pair and are reported in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$. See text for further details.

6 Appendix

This appendix includes details on the assumptions and derivations of our model.

6.1 Short-Run Gravity with Homogeneous Firms

Production and delivery from any origin i to any destination j) uses variable labor L_{ij} combined with fixed marketing capital K_{ij} in a Cobb-Douglas production function. Iceberg trade frictions $t_{ij} > 0$ melt all shipments such that 1 unit leaving origin i shrinks to $1/t_{ij}$ units at destination j. The short run gravity model that results is a geometric weighted average of long run gravity and a bilateral capacity variable λ_{ij} . Short run gravity trade flows are given by:

$$X_{ij} = \frac{Y_i E_j}{Y} \left[\frac{t_{ij}}{\Pi_i \widetilde{P}_j} \right]^{(1-\sigma)\rho} \lambda_{ij}^{1-\rho} \; \forall i, j; \tag{7}$$

where the multilateral resistances Π_i and $\widetilde{P}_j$ are interpreted as the sellers' and buyers' incidence of all trade costs respectively from origin i to the world market and from the world market to destination j. $\sigma > 1$ is the CES elasticity of substitution and $\rho \in (0, 1)$ is the buyers incidence elasticity of trade costs. ρ is a function of σ and supply elasticity parameters developed below.

The first term on the right hand side of (7) is the frictionless benchmark trade flow at given origin sales $\{Y_i\}$ and destination expenditure $\{E_j\}$. The middle term is the familiar effect of trade frictions in structural gravity, the ratio of bilateral to the product of buyers' and sellers' multilateral resistances. The short run trade elasticity falls from its usual (interpreted now as long run) absolute value $1 - \sigma$ to $(1 - \sigma)\rho$. The last term $\lambda_{ij}^{1-\rho}$ is the effect of fixed investment in capacity on link i, j.

Over time the allocation of destination specific capital $\{\lambda_{ij}\}$ presumably moves toward the long run efficient level. Anderson and Yotov (2020) show that the efficient allocation matches the long run equilibrium demand pattern, hence the level implied by the short run gravity equation at these long run λ_{ij}s equals its long run gravity value. The analysis is reviewed below and extended to treat the efficient extensive margin of trade in the sense of heterogeneous firms serving a particular destination and the extensive margin of exports from a particular origin. Since the heterogeneous firms model includes a fixed cost of serving a market, some origins do not serve a range of small

destination markets even in the long run. The model is then extended to multiple sectors with sectorally differing demand and supply parameters.

6.2 Basic Setup

Heterogeneous firms in an origin allocate capital to production and to distribution to a set of destinations. The capital becomes specific once allocated. Subsequently, the firms draw productivities from a Pareto distribution, and demand shocks are realized. The firms face iceberg trade frictions in distribution as well as the cost of resources needed to serve the destinations. The iceberg frictions include a fixed cost for each destination. The firms that can make operating profits hire labor from a national market and deploy it efficiently to production and to distribution to the various destinations. The description of the model is in terms of integrated production and distribution firms but it applies equally to arms length relations between production and distribution.

7 Heterogeneous Firms Case

Heterogeneous firms in an origin allocate capital to production and to distribution to a set of destinations. The capital becomes specific once allocated. Subsequently, the firms draw productivities from a Pareto distribution, and demand shocks are realized. The firms face iceberg trade frictions in distribution as well as the cost of resources needed to serve the destinations. The iceberg frictions include a fixed cost for each destination. The firms that can make operating profits hire labor from a national market and deploy it efficiently to production and to distribution to the various destinations. The description of the model is in terms of integrated production and distribution firms but it applies equally to arms length relations between production and distribution.

The firms use capital and labor for production and distribution with the Cobb-Douglas technology $K^{1-\alpha}L^{\alpha}$ where K and L without indexing denote any firm's capital and labor allocated to any activity. Competitive equilibrium requires that the value of sales net of distribution cost is equal to the net value of production.

Firms in origin i amplify their common technology by multiplicative productivity draws from

a Pareto distribution $G(\varrho) = 1 - (\varrho/\varrho_{\min})^{-}\theta, \varrho \geq \varrho_{\min} > 0$. Across origins i, the Pareto location parameter $\varrho_{\min}$ can vary to allow origin-specific differences in the productivity distribution. The firms face common iceberg frictions in distribution from origin i to destinations j, effectively reducing productivity in delivered goods by $1/t_{ij}, t_{ij} > 1$. The firms also face a fixed cost in terms of labor for each market served, $f_j = wa_j$ where a_j is the labor required to enter the market.[27] When sector differences are introduced (indexed by superscript k), they appear in the iceberg frictions t_{ij}^k, the Pareto location parameter $\underline{\varrho}_{\min}^k$ and the Pareto dispersion parameter $\theta_k > 0$. It is convenient to temporarily suppress the origin and sector notation, and to conduct the analysis with a continuum of firms.

An exogenously given mass M_j of firms in an implicit origin have previously committed capital k_j to each destination j. Distribution presupposes production denoted 'destination' 0. (Firms have identical per-firm capital k_j because prior to receiving their productivities, all firms are identical.) As the period of analysis opens each firm draws a Hicks-neutral productivity scalar $\varrho \geq \varrho_{\min} > 0$ from a Pareto distribution $G(\varrho) = 1 - (\varrho/\varrho_{\min})^{-\theta}$. Simplify by setting $\rho_{\min} = 1$. After the productivities are drawn, firms hire labor at wage rate w to produce and distribute the good, equating w to the value of marginal product of labor for production and for distribution to each destination.

Index the firms by their productivity draws ϱ. The profit of firm ϱ on sales to j using variable labor $L_j(\varrho)$ is

$$\varrho \frac{p_j}{t_j} L_j(\varrho)^{\alpha} [k_j]^{1-\alpha} - wL_j(\varrho).$$

Profit maximization on sales to j by a price taking firm ϱ implies that the value of marginal product of labor is equal to the wage, yielding demand for labor by firm ϱ:

$$L_j(\varrho) = k_j \left[\alpha \varrho \frac{p_j}{t_j} \frac{1}{w} \right]^{1/(1-\alpha)}. \tag{8}$$

[27] The case where the fixed cost is a fixed Cobb-Douglas function of capital and labor is essentially the same.

The resulting restricted profit function is $\varrho^{1/(1-\alpha)}\bar{R}_j$ where

$$\bar{R}_j = \left(\frac{p_j}{t_j}\right)^{1/(1-\alpha)} w^{-\alpha/(1-\alpha)} k_j [\alpha^{\alpha/(1-\alpha)} - \alpha^{1/(1-\alpha)}] \tag{9}$$

is the variable profit of the least productive active firm, equal to fixed cost.

The zero profit cutoff value of ϱ is

$$\underline{\varrho}_j = [wa_j/\bar{R}_j]^{1-\alpha}$$

where $f_j = wa_j,\ a_j > 0$ is the fixed cost of each firm in the implicit origin choosing to export to destination j. The proportion of firms with $\varrho \geq \underline{\varrho}$ is given by $1 - G(\varrho) = \int_{\underline{\varrho}_j}^{\infty} \theta \varrho^{-\theta-1} d\varrho$. The mass of firms choosing to serve destination j is

$$M_j U_j = M_j \int_{\underline{\varrho}}^{\infty} \theta \varrho^{1/(1-\alpha)-\theta-1} d\varrho = M_j \frac{\theta}{\theta - \eta} \left(\frac{wa_j}{\bar{R}_j}\right)^{(\eta-\theta)/\eta},\ \theta > \eta. \tag{10}$$

Here the right hand expression uses $\eta = 1/(1-\alpha)$ and U_j denotes the equilibrium utilization rate.

The aggregate value to sellers of trade shipped to destination j is given by integrating the value of marginal product of variable labor over firms ϱ. Use (9) in (10) and simplify to yield:

$$X_j = A k_j M_j \left(\frac{k_j}{a_j}\right)^{(\theta-\eta)/\eta} \left(\frac{p_j}{w t_j}\right)^{\theta-\eta}. \tag{11}$$

A is a constant function of the parameters α, η, θ.

Gravity combines this heterogeneous supply side with a CES demand side for many origin countries denoted by subscript i added to variables in (11). Expenditure on good i in destination j is given by

$$X_{ij} = \left(\frac{p_i t_{ij}}{P_j}\right)^{1-\sigma} E_j \tag{12}$$

where E_j is total expenditure on goods from all origins, $P_j = [\sum_i (p_i t_{ij})^{1-\sigma}]^{1/(1-\sigma)}$ is the CES price index and $\sigma > 1$ is the elasticity of substitution.

Market clearing bilaterally equates the right hand sides of (11) and (12). Solve this equation

for the bilateral market clearing price p_{ij}. This yields:

$$p_{ij} = \left[\frac{E_j P_j^{\sigma-1} w_i^{\theta-\eta} t_{ij}^{\theta-\eta}}{A k_{ij} M_{ij} (k_{ij}/a_{ij})^{(\theta-\eta)/\eta}} \right]^{1/(\theta-\eta+\sigma-1)} . \tag{13}$$

Substitute the right hand side of (13) for p_{ij} in the right hand side of the demand equation (12). After considerable simplification,[28] this yields short run gravity for the heterogeneous firms case:

$$X_{ij} = \left(\frac{t_{ij}}{P_j} \right)^{(1-\sigma)\rho} E_j^{\rho} w_i^{(1-\sigma)\rho} \left[A k_{ij} M_{ij} (k_{ij}/a_{ij})^{(\theta-\eta)/\eta} \right]^{1-\rho} \tag{14}$$

where $\rho \equiv (\theta - \eta)/(\theta - \eta + \sigma - 1)$. For purposes of adjustme nt on the external extensive margin it is helpful to consolidate the effect of k_{ij} into $k_{ij}^{(1-\rho)\theta/\eta}$

Comparing short run gravity in the heterogeneous case (14) with the homogeneous firms case, note that the short run trade elasticity remains the same $(1 - \sigma)\rho$. The origin and destination fixed effects differ slightly in interpretation. The analog to λ_{ij} in homogeneous short run gravity is the rightmost square bracketed term, Z_{ij}, a bilateral pair cost shifter with elasticity $1 - \rho$ as in the homogeneous case. Normalizing inside the square bracket to create the analog to the ex ante allocation shares λ_{ij} yields $\zeta_{ij} = Z_{ij}/Z_j$. Introduce an origin-specific factor $Z_i^{1-\rho} = \left[\sum_j A k_{ij} M_{ij} (k_{ij}/a_{ij})^{(\theta-\eta)/\eta} \right]^{1-\rho}$ to preserve the equation (14). The result is

$$X_{ij} = \left(\frac{t_{ij}}{P_j} \right)^{(1-\sigma)\rho} E_j^{\rho} w_i^{(1-\sigma)\rho} Z_i^{1-\rho} \zeta_{ij}^{1-\rho}; \ \forall i, j. \tag{15}$$

To obtain short run gravity, first replace $P_j^{(\sigma-1)\rho} E_j^{\rho}$ with $E_j \tilde{P}_j^{(\sigma-1)\rho}$ in (15). Second solve for $w_i^{(1-\sigma)\rho}$ from the market clearance condition. Summing (15) over j

$$w_i^{(1-\sigma)\rho} = \frac{Y_i/Y}{\sum_j (t_{ij}/\tilde{P}_j)^{(1-\sigma)\rho} E_j Z_i^{1-\rho} \zeta_{ij}} = \frac{Y_i/Y}{\Pi_i^{(1-\sigma)\rho} Z_i^{1-\rho}}$$

Substitute the right hand side for $w_i^{(1-\sigma)\rho}$ in (15). The result is short run gravity for the heterogeneous firms case.

$$X_{ij} = \frac{Y_i E_j}{Y} \left(\frac{t_{ij}}{\Pi_i \tilde{P}_j} \right)^{(1-\sigma)\rho} \zeta_{ij}^{1-\rho} \tag{16}$$

[28] The simplification uses $1 - \rho = -(1 - \sigma)/(\theta - \eta + \sigma - 1)$ in simplifying several complex exponents.

7.1 Returns to Capital

Now consider efficient investment. In this model the relative ex post return $r_{ij}/\bar{r}_i$ is given by differentiating (16) with respect to k_{ij} and placing that result relative to the effect of a uniform rise in all k_{ij}, equal to $\bar{r}_i$. The common exponents $1-\rho$ and $(\theta-\eta)/\eta$ cancel, hence

$$\frac{r_{ij}}{\bar{r}_i} = \frac{(s_{ij}/\zeta_{ij})(1-\zeta_{ij})/k_{ij}}{\sum_j (s_{ij}/\zeta_{ij})(1-\zeta_{ij})/k_{ij}}.$$

Alternatively and more simply, the denominator can be considered fixed at some proportion to an external $\bar{r}_i/(1-\rho)[\theta/\eta]$ where the deflator is the common exponent associated with k_{ij}.

Generally, investment will be inefficient because $\hat{s}_{ij} \neq s_{ij}$ and other adjustment frictions may impede adjustment toward efficient installed bilateral capacity. Short run gravity (16) obtains with ζ_{ij} shifting.

Deviation of relative returns from 1 is a signal for reallocation, provided it is systematic. Intuitively, the maximum profit capital allocation when costless reallocation is assumed should match anticipated sales shares.

Anderson and Yotov (2020) combine the production-*cum*-distribution model for each origin country i summarized in (**??**) with a standard Armington-CES demand system for all buyers in the multi-country world. Perfect spatial arbitrage subject to the trade costs (both iceberg and endogenous) yields the short run gravity model (7).

7.2 The Extensive Margin of Trade

There are two distinct forms of the extensive margin of trade, exclusive of the selection mechanism affecting the mass of firms in existing origin-destination trade. First is the *sectoral* margin whereby production and distribution of a product begins or ends. We call this the domestic extensive margin. Second is the external (cross-border) margin of trade whereby the production and distribution for export changes. There are two sources of time series variation in the two extensive margins. One is due to cyclic volatility of service in bilateral links. The other is secular change (growth or decline) in the number of markets served. Both forms of the extensive margin are described by the simple selection mechanism of the heterogeneous firm model that is embedded in the intensive margin of

trade. Both sources of action on the extensive margins are active and quantitatively important.[29] The analysis begins in Section 7.2.1 with the external margin of trade, followed by the domestic margin in Section 7.2.2.

7.2.1 External Margin of Trade

In the lens of the model, the switching off and on of bilateral exports is due to destination demand shocks large enough to alternately shut down and open up previously served markets. The fixed cost per period required to serve a small market can exceed the variable profit of even the most productive firm when the negative demand shock is large enough. (The equilibrium utilization rate U_j of destination-specific capital then drops to zero. For less drastic shocks, the market remains open with lower utilization of destination specific capital – $u_j \in (0,1)$.)

New markets are entered when demand in n becomes sufficiently large. A plausible feature of new markets that is distinct from re-entering formerly served markets is costly adjustment of capacity on the extensive margin of trade.[30]

Consider the efficient investment at the extensive margin, $j = n$. There is a one period cost of adjustment to be covered from a planned excess return $r_{in}/\bar{r}_i = \phi_i$. With perfect foresight the realized $s_{in} = \hat{s}_{in}$. Solving,

$$X_{in} = \frac{Y_n E_j}{Y} \left(\frac{t_{in}}{\Pi_i \tilde{P}_n} \right)^{1-\sigma} \phi_i^{(\rho-1)/\rho} \left[\bar{r}_i (1-\rho)(\theta/\eta) \frac{1-\zeta_{in}}{\zeta_{in} k_{in}} \right]^{1/\rho}.$$

The condition for entering under perfect foresight is based on

$$s_{in} \phi_i^{(\rho-1)/\rho} \left[\bar{r}_i (1-\rho)(\theta/\eta)(1-\zeta_{in})/\zeta_{in} k_{in} \right] \geq s_{i,n-1} \left[\bar{r}_i (1-\rho)(\theta/\eta)(1-\zeta_{i,n-1})/\zeta_{i,n-1} k_{i,n-1} \right].$$

Under perfect foresight the entry condition simplifies to

$$s_{in} \phi_i^{(\rho-1)/\rho} \left[(1-\zeta_{in})/\zeta_{in} k_{in} \right] \geq s_{i,n-1} \left[(1-\zeta_{i,n-1})/\zeta_{i,n-1} k_{i,n-1} \right]. \tag{17}$$

[29] Besedes and Prusa (2006) document the high volatility over time of 10 digit HS level bilateral US exports. Chaney (2014) describes the rich dynamics of French exporting firms focused on their entry into new bilateral markets.

[30] Anderson and Yotov (2020) treat adjustment on the intensive margin with simple ad hoc adjustment mechanisms and use this to infer values of the short run incidence parameter ρ.

Generally the anticipated share $\hat{s}_{in}$ is not equal to realized share s_{in} and the extensive margin is determined by

$$X_{in} = \frac{Y_n E_j}{Y} \left(\frac{t_{in}}{\Pi_i \tilde{P}_n} \right)^{(1-\sigma)\rho} \phi_i^{(\rho-1)} \left[\bar{r}_i (1-\rho)(\theta/\eta) \frac{\hat{s}_{in}}{\zeta_{in}} \frac{1-\zeta_{in}}{k_{in}} \right] \tag{18}$$

The entry condition (17) is amended with the $\hat{s}_{in}$ and $\hat{s}_{i,n-1}$ shares playing a role.

Notice that $k_{ij}/a_{ij} = \kappa_i$ implies a common utilization rate. The utilization rate of committed capital varies with $(k_{ij}/a_{ij})^{(\theta-\eta)/\eta}$, implying that $U_{ij}/\bar{U}_i \to 1$ as $k_{ij} \to \kappa_i a_{ij}$. Also, $Z_i^{1-\rho} \to (AK_i)^{1-\rho}$ where $K_i = \sum_j M_{ij} a_{ij} \kappa_i^{1+(\theta-\eta)/\eta} = \kappa_i^{\theta/\eta} \sum_j M_{ij} a_{ij}$. In this case $\zeta_{ij} = M_{ij} a_{ij} / \sum_j M_{ij} a_{ij}$. Then markets are efficiently served when $zeta_{ij} = s_{ij} \Rightarrow X_{ij} = (Y_i E_j / Y)(t_{ij}/\Pi_i P_j)^{1-\sigma}$. The entry condition here resembles the homogeneous firms case.

Generally the interior extensive margin $n < N$ is anticipated, where N is the maximum number of destinations. For destination markets with projected share below $\hat{s}_{in}$, the projected trade share remains latent.

The structure of (17) suggests a natural empirical approach to estimation of gravity with zeroes. The Tobit estimator can be applied, exploiting the close resemblance of cutoff share $\hat{s}_{in}$ in (**??**) to the share equation version of (7). The difference is that relative to (7), (17) is scaled by $\phi_i^{(\rho-1)/\rho}$. Imposing these structural restrictions on a Tobit estimator can potentially treat the evolution of export entry in the CES gravity framework.

This in principle permits a single Tobit regression equation to determine both positive and latent market shares. Fixed sunk entry cost is not necessary, though it is plausible.

The empirical model of this paper focuses on the extensive margin alone, abstracting from explaining the variation of market shares on the intensive margin. On the extensive margin, the count data is informative. The same forces determine the cutoff market, with the variation in the count reflecting the variation in (17).

7.2.2 The Domestic Margin of Trade

The model extends to allocation of labor and capital across multiple sectors when the Cobb-Douglas production function reasonably applies across sectors. The Cobb-Douglas labor share parameter

and the Pareto productivity distribution for firms' productivity draws can differ by sector, while the productivity distribution location parameter can differ across origins and sectors, and the CES elasticity of substitution can differ across sectors.

Suppress the origin notation and denote Y_k as the value of production in sector k at the implicit origin region or country. All previous steps are now indexed by the sector k and the equilibrium of production and distribution leads to

$$Y_k = K_k^{1-\alpha_k} L_k^{\alpha_k} \mathbf{P}_k$$

where $\mathbf{P}_k = \left[\sum_{j=0}^{N_k} (p_{jk}/t_{jk})^{1/(1-\alpha_k)}\right]^{1-\alpha_k}$ and $\lambda_{jk} = K_{jk}/K_k$. The residual rate of return on capital devoted to serving destination j relative to the average rate of return in sector k is $r_{jk}/\bar{r}_k = s_{jk}/\lambda_{jk}$ where s_{jk} is the share of total sales shipped to destination j.

The efficient allocation of labor across sectors and destinations implies $w = \alpha_k (K_k/L_k)^{1-\alpha_k} \mathbf{P}_k$ and hence

$$Y_k = \alpha_k^{\alpha_k/(1-\alpha_k)} K_k (\mathbf{P}_k/w)^{\alpha_k/(1-\alpha_k)}. \tag{19}$$

The aggregate value produced and distributed across all sectors is $Y = \sum_{k=1}^{M} Y_k$. Define $\Lambda_k \equiv K_k/K$. The share Y_k/Y is given by

$$S_k = \frac{a_k \Lambda_k (\mathbf{P}_k/w)^{\alpha_k(1-\alpha_k)}}{\sum_k a_k \Lambda_k (\mathbf{P}_k/w)^{\alpha_k(1-\alpha_k)}}, \tag{20}$$

where $a_k \equiv \alpha_k^{\alpha_k/(1-\alpha_k)}$.

The rate of return on capital specific to k relative to the average rate of return is

$$\frac{\bar{r}_k}{\bar{r}} = \frac{S_k/\Lambda_k}{\sum_k S_k/\Lambda_k}. \tag{21}$$

The extensive margin of destinations in sector k is denoted n_k and is determined in equilibrium by $\hat{s}_{n_k} = \phi_k \lambda_{nk}$. The same mechanism applied to the extensive margin of sectors yields $\hat{S}_M = \phi \Lambda_M$ where sectors are ordered by decreasing size of shares.

Notice that entry presupposes successful distribution – 'destination' 0 would not be entered

unless some of destinations $j > 0$ are also entered. The formal model implies that destinations served are ordered by decreasing market size. Thus the theoretical model allows for extensive margin of production for export only, or for distribution mainly for export. Our treatment in the empirical application assumes away production for export only.

Equations (20) and (21) together with the extensive margin condition $\widehat{S}_M/\Lambda_M = \phi$ are the elements of equilibrium selection of sectors. Characterizing this selection fully and carrying the characterization to data are beyond the scope of this paper. The reduced form model estimated below provides empirical regularities about the selection of sectors in a setting consistent with the theoretical model.

The empirical gravity model uses count data as the dependent variable, the count being interpreted as the marginal sector-destination served. Santos Silva et al. (2014) have a good discussion of the way in which their FLEX estimator applies to the essentially similar firm selection gravity model of Helpman et al. (2008b).