cesifo WORKING PAPERS

9822
2022

June 2022

Quantifying the Extensive Margin(s) of Trade: The Case of Uneven European Integration

James E. Anderson, Yoto V. Yotov

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9822*

Quantifying the Extensive Margin(s) of Trade: The Case of Uneven European Integration

Abstract

We propose a short-run theory of the extensive margins of trade, comprising the standard international extensive margin and a novel domestic extensive margin. The domestic extensive margin allows identification of globalization and specific policy effects not properly identified in previous literature. To apply our methods, we build a new dataset covering both the cross-border and domestic extensive margins for 35 countries, 1995-2014. We deploy it to quantify the extensive margins effects of globalization and European integration. We find strong positive effects of globalization and also significant but highly asymmetric effects of European integration in favor of more developed EU members.

JEL-Codes: F130, F140, F160.

Keywords: extensive margin, domestic extensive margin, globalization, EU, gravity.

James E. Anderson
Department of Economics
Boston College
USA - Chestnut Hill, MA 02467
anderson@bc.edu

Yoto V. Yotov
School of Economics
Drexel University
USA – Philadelphia, PA 19104
yotov@drexel.edu

May 4, 2022
We thank João Santos Silva for kindly sharing the original programming codes for the FLEX estimator. We thank Rebecca Freeman and Stephan-Alfons Nolte for very helpful advice and guidance to the data from Eurostat. We are indebted to Javier Flórez Mendoza for his tremendous effort with the construction and invaluable assistance with the interpretation of the PRODCOM-COMEXT, 1995-2014, concordance. We also are very grateful to Andrew Bernard, Ilke Van Beveren, and Hylke Vandenbussche for creating an earlier version of this concordance and for making the corresponding codes available. Finally, we thank the participants in the Winter 2020 NBER ITI meetings for useful comments, questions and suggestions.

"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 objective for both developed and developing countries. 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 literature on the extensive margin is large: 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 extensive margin literature (i) a simple model of adjustment on both domestic (range of products) and international (range of destinations) margins;[2] (ii) methods to identify policy effects that were not possible to obtain before, and to revisit

[1] 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).

[2] We could alternatively use 'product extensive margin' and 'destination extensive margin' labels, with more precision but less intuitive comprehension. Export logically requires production, but neither the model nor the empirical application require domestic sales of a product.

prior applications; (iii) a new dataset that covers both the domestic and the international extensive margins of trade; (iv) a novel export diversification index that complements existing indexes by leveling the field for the performance of smaller countries; and (v) a policy-relevant application that highlights our methodological contributions.

Building on three prominent strands of the literature,[3] we develop a short run structural gravity model focused on adjustments over time of fixed bilateral capacities by heterogeneous firms choosing destinations for given products as well as new product-*cum*-destinations (the domestic extensive margin). Capital is product-destination-specific. Investment on the extensive margins is selected when the expected return exceeds the product of the opportunity cost of investment and an adjustment cost factor. We derive a closed form solution of our model, which can be decomposed in four intuitive structural terms, including the standard long-run gravity model, a short-run gravity term, a capital utilization term due to selection of heterogeneous firms, and a term that captures action on the extensive margins. Importantly, the model applies to both the international and the domestic extensive margins.

A key implication of our model is that proper quantification of the international extensive margin (the set of partners any product is exported to) must also take into account the domestic margin (the set of products with positive production). The introduction of the domestic extensive margin allows for identification of the effects of a number of policies. Specifically, with data on the domestic extensive margin we can potentially identify the effects of (i) non-discriminatory export support policies, e.g., export subsidies, trade fairs, etc., (ii) non-discriminatory import protection policies, (iii) country-specific characteristics and policies, e.g. institutional quality, technical barriers to trade (TBT) etc., (iv) exchange rates, and (v) the effects of globalization on the extensive margin of trade. We show that this identification is impossible to obtain with data only on the external extensive margin. We

[3]The first one is the standard gravity literature, e.g., Anderson (1979), Eaton and Kortum (2002), Anderson and van Wincoop (2003), and Arkolakis et al. (2012). The second one is the literature on bilateral investment/dynamics, e.g., Arkolakis (2010), Head et al. (2010), Chaney (2014), Mion and Opromolla (2014), Sampson (2016), Crucini and Davis (2016), and Anderson and Yotov (2020). The third one is the literature on the extensive margin of trade, e.g., Melitz (2003), Helpman et al. (2008a), Chaney (2008), and Redding (2011).

also argue that the introduction of the domestic extensive margin may have implications for the estimates of bilateral trade policies, e.g., regional trade agreements (RTAs), membership in the World Trade Organization (WTO), etc.

Our empirical analysis is based on a novel data set that covers the extensive margins of trade in mining and manufacturing products for 35 European countries over the period 1995-2014. A key feature of our dataset is inclusion of the *domestic extensive margin*. The dataset combines two original sources. Production data are taken from Eurostat's *Production Communautaire* (PRODCOM) database. Trade data are from Eurostat's COMEXT data. The combination of PRODCOM and COMEXT allowed us to build an estimating sample with consistently constructed data on the external and domestic extensive margins for 35 European countries and about 3300 products, 1995-2014. We also experiment with several alternative estimating samples, which demonstrate the robustness of our main findings.

Variation in the domestic extensive margin is quantified in a novel index, the *Domestic Extensive Margin (DEM)*. DEM is defined as the ratio of the number of products actually produced by a given country in a given year to the total number of possible products that could have been produced by the same country and in the same year. Thus, DEM complements existing absolute measures by allowing for a consistent (relative) comparison of the export-diversification performance of smaller and less developed countries vs. large and advanced economies. DEM reveals several interesting patterns in its 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., Cyprus, Luxembourg, and Malta), while the countries with the largest indexes are large and rich economies (e.g., Germany, France, and Italy).[4] We also observe significant DEM variation over time. Most countries have experienced a decrease in DEM, which may be

[4]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.

interpreted as an indicator of specialization. The biggest decline is for smaller and relatively poor EU economies (e.g., Iceland, Hungary, and Ireland), while the rich European countries (e.g., Germany and France) experience a small decline in *DEM*. A small number of countries (e.g., Serbia, Bosnia and Herzegovina, and Turkey) have experienced an increase in DEM.

Three steps lead to specification of our econometric model. First, our theory extends the CES structural gravity model to a closed form that features fixed product-destination capital adjustment on both domestic and international extensive margins of trade. The highly stylized short run gravity model motivates our reduced-form empirical specification that identifies action on 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 model is estimated with the Santos Silva et al. (2014) FLEX estimator, which 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 conclusions to the use of alternatives estimators including Tobit, OLS, and the Poisson Pseudo Maximum Likelihood (PPML).

We highlight the use of our methods by quantifying the impact of globalization and European Union (EU) integration during the period 1995-2014. The application has three notable attributes. First, from a methodological perspective, it shows that the effects of globalization cannot be identified in a theory-consistent econometric specification without data on the domestic extensive margin. Second, from a practical perspective, the econometric specification consistent with the model allows us to obtain estimates of the globalization and EU integration effects within a simple, flexible, and robust econometric specification with fixed effects only.[5] The fixed effects treatment enables us to obtain a series of globalization and 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

[5]Technically, we do have controls for membership in the World Trade Organization (WTO) and in Economic Integration Agreements (EIAs). 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.

endogeneity concerns. Finally, from a policy perspective, the proposed application uncovers strong and positive globalization effects and also significant but very asymmetric effects of EU integration with potential implications for export diversification strategies.

The empirical analysis starts with a benchmark specification that imposes common globalization effects across all countries in the sample. Our preferred specification implies that, on average across the countries in our sample, the number of internationally traded products increased by about 511 relative to the number of domestically traded products during the period of investigation, or about 16 percent of the total number of possibly traded products in 2014. The lens of the model associates this increase with product-destination marketing capital investment due to EU membership initiation. Highly heterogeneous effects emerge when we allow for country-specific effects of EU integration. The estimates suggest that the extensive margin effects have been strongest for the recent and new EU members, while the large EU economies have experienced relatively small extensive margin gains. Focus on the new EU members extensive margins of trade reveals directional asymmetry. There are 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. An explanation is that the new EU members were less able to position their (possibly inferior) products well in the more developed old member West-European market. In contrast, the old (West European) EU members saw large opportunities to market their high quality varieties to new EU members.

Our work complements and extends two strands of the literature. Most closely related is the literature on the extensive margin of trade. Melitz (2003), Helpman et al. (2008a) and Chaney (2008) are prominent examples. Redding (2011) offers an excellent survey of the related theoretical literature, the empirical challenges related to this research, and its implications for the extensive margin of trade. 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 estimation 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.[6]

Another strand of related literature emphasizes the importance of proper accounting for domestic trade flows on the intensive margin of trade.[7] 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. (2021) 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 in relation to this literature is estimation of properly identified action on the extensive margins of trade, based on a tractable CES-type short run gravity model of product-destination capital adjustment. Our methods open avenues for many extensions and applications. 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,

[6]Thus, from a policy perspective, our contribution is related to a very large number of papers that study the impact of various determinants of the extensive margin of trade. Without an attempt to be exhaustive, for some excellent studies we refer the reader to Felbermayr and Kohler (2006), Berthou and Fontagne (2008), Cadot et al. (2011), Persson (2013), and Beverelli et al. (2015).

[7]Yotov (2022) surveys this literature.

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 yield a short-run gravity model of the extensive margin(s) of trade. 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 in marketing capital on the intensive margin, variously treated in papers by 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 the extensive margin of trade following Melitz (2003), Helpman et al. (2008a), Chaney (2008), and Redding (2011). Our key contributions in relation to the existing literature are modeling the *short-run extensive margin of international trade* and the introduction of adjustment on the *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 Theory

Heterogeneous small firms in each origin i use capital and labor for production-*cum*-distribution of each product with identical (up to a multiplicative productivity draw) Cobb-Douglas technology.[8] Firms in each origin *ex ante* commit capital to each product and destination j, and this capital becomes specific once allocated. For expositional simplicity, we temporarily suppress the product notation.

At the start of the period of analysis, demand shocks are realized and firms amplify

[8]The firms may be monopolistic or pure competitors. Also, the model applies equally to arms length relations between production and distribution.

their common *ex ante* technology[9] by multiplicative Hicks-neutral productivity draws from a Pareto distribution $G(\varrho) = 1-(\varrho/\varrho_{\min})^{-\theta}$, $\varrho \geq \varrho_{\min} > 0$. Across origins, 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, $w_i a_{ij}$, where w_i is the wage in i and a_{ij} is the labor required to enter market j.[10]

The firms that can make operating profits hire labor from a national market and deploy it efficiently to production and distribution to the various destinations, equating wages to the value of marginal product of labor for production and for distribution to each destination. Competitive equilibrium requires that the value of sales net of distribution cost is equal to the net cost of production.[11]

Index the firms in i selling to j by their productivity draws ϱ_{ij}. The operating profit of firm ϱ_{ij} on sales to j using variable labor $L_{ij}(\varrho_{ij})$ is

$$\varrho_{ij}\frac{p_{ij}}{t_{ij}}L_{ij}(\varrho_{ij})^{\alpha}[k_{ij}]^{1-\alpha} - w_i L_{ij}(\varrho_i).$$

For expositional ease, focus on the case of competitive firms.[12] Profit maximization on sales to j by a price-taking firm ϱ_{ij} implies that the value of marginal product of labor is equal

[9]Firms have committed identical per-firm capital k_j to j because prior to receiving their productivities, all firms are assumed to be identical. This is an inessential simplification. Differences across firms in their *ex ante* allocations act on their subsequent productivity like the random productivity shocks – an outside-the-model shifter treated as random.

[10]The case where the fixed cost is a common Cobb-Douglas function of capital and labor with labor share parameter α is essentially the same.

[11]Monopolistic competitive equilibrium has the same requirement.

[12]The case of monopolistic competitive firms with sufficiently small shares differs inessentially from (1)-(2) below in that marginal revenue $p_{ij}(\sigma-1)/\sigma$ replaces p_{ij} where $\sigma > 1$ is a constant elasticity of substitution facing all firms. $L_{ij}(\varrho_{ij})$ and $\bar{R}_{ij}$ pick up a constant term $(\sigma-1)/\sigma(1-\alpha)$. The small shares case is a standard simplification in the related literature. For discrete shares, the firms' *ex post* demand elasticities are functions of the shares and there is no closed form solution for the cutoff $\bar{R}_{ij}$.

to the wage, yielding demand for labor by firm ϱ_{ij}:

$$L_{ij}(\varrho_{ij}) = k_{ij}\left[\alpha\varrho_{ij}\frac{p_{ij}}{t_{ij}}\frac{1}{w_i}\right]^{1/(1-\alpha)}. \tag{1}$$

The resulting value function for (maximized) profits is $\varrho_{ij}^{1/(1-\alpha)}\bar{R}_{ij}$, where

$$\bar{R}_{ij} = \left(\frac{p_{ij}}{t_{ij}}\right)^{1/(1-\alpha)} w_i^{-\alpha/(1-\alpha)}k_{ij}[\alpha^{\alpha/(1-\alpha)} - \alpha^{1/(1-\alpha)}] \tag{2}$$

is the variable profit of the least productive active firm.

The proportion of active firms in i draw productivities above a cutoff $\varrho_{ij} \geq \underline{\varrho}_{ij}$. The active proportion is given by $1 - G(\underline{\varrho}_{ij}) = \int_{\underline{\varrho}_{ij}}^{\infty}\theta\varrho^{-\theta-1}d\varrho = \underline{\varrho}_{ij}^{-\theta}$. The zero profit cutoff value of ϱ_{ij}, $\underline{\varrho}_{ij}$, is solved from setting (2) equal to fixed cost. Thus

$$\underline{\varrho}_{ij} = [w_i a_{ij}/\bar{R}_{ij}]^{1-\alpha} = \frac{w_i t_{ij}}{p_{ij}}\left(\frac{a_{ij}}{k_{ij}}\right)^{1/\eta}[\alpha^{\alpha\eta} - \alpha^{\eta}]^{-1/\eta}, \tag{3}$$

using (2) in the second equation. For convenience we define $\eta \equiv 1/(1-\alpha)$. The proportion of active firms $\underline{\varrho}_{ij}^{-\theta}$ in i across destinations j is decreasing in the ratio of the fixed cost a_{ij} to the size of the bilateral capital k_{ij} and increasing in the ratio p_{ij}/t_{ij}, the sellers' net price of serving destination j.

The bilateral sales (at seller prices) of firm ρ_{ij} is given by plugging its equilibrium choice of labor (1) into its Cobb-Douglas production function:

$$Y_{ij}(\varrho_{ij}) = \frac{p_{ij}}{t_{ij}}k_{ij}\varrho_{ij}\left[\alpha\varrho_{ij}\frac{p_{ij}}{t_{ij}}\frac{1}{w_i}\right]^{\alpha/(1-\alpha)} = k_{ij}\varrho_{ij}^{\eta}(p_{ij}/w_i^{\alpha}t_{ij})^{\eta}\alpha^{\alpha\eta}.$$

The aggregate sales from origin i to destination j is given by multiplying the firm level sales $Y_{ij}(\varrho_{ij})$ by the mass of firms M_{ij} times the Pareto probability density of active firms $\theta\varrho_i^{-\theta-1}$ and integrating up from the zero cutoff productivity value $\underline{\varrho}_{ij} = (w_i a_{ij}/\bar{R}_{ij})^{1/\eta}$. The result is

$$Y_{ij} = M_{ij} k_{ij} \underline{\varrho}_{ij}^{\eta-\theta} \left(\frac{p_{ij}}{w_i^\alpha t_{ij}} \right)^\eta A = K_{ij} U_{ij} \underline{\varrho}_{ij}^{\eta} \left(\frac{p_{ij}}{w_i^\alpha t_{ij}} \right)^\eta A, \tag{4}$$

where $A = \theta/(\theta + 1 - \eta)\alpha^{\alpha\eta} > 0$ is a constant. In the second equation $K_{ij} \equiv M_{ij} k_{ij}$ is the aggregate marketing capital committed *ex ante* by origin i to serving market j, and $U_{ij} = \underline{\varrho}_{ij}^{-\theta}$ is the proportion of utilized capital (the proportion of *ex post* active firms). From the *ex ante* viewpoint this is the probability of a firm being active. Then (up to a scalar absorbed in A) $\underline{\varrho}_{ij}^{\eta}$ is recognized as the average productivity of active firms and $\underline{\varrho}_{ij}^{\eta}(p_{ij}/w_i^\alpha t_{ij})^\eta$ is the expected rent earned by an active unit of capital.

The utilization rate and the average productivity are determined by the cutoff productivity. $\underline{\varrho}_{ij}$ in turn varies in equilibrium with p_{ij} and w_i by equations (2)-(3). Building in this response, the resulting supply function has several cases depending on the sizes of θ and η. For the case $\eta = \theta$, the middle expression in (4) implies that aggregate sales is invariant to selection. The utilization rate is effectively constant. For the case $\eta > \theta$, rises in p_{ij} imply that adding *ex post* labor to active firms has higher payoff than adding labor to the below-cutoff productivity firms. Thus selection is inactive and supply is given by (4) with average productivity $\underline{\varrho}_{ij}^{\eta}$ constant.[13] For the case $\theta > \eta$, the opposite is true, adding more active firms dominates expanding on the intensive margin. The resulting short run sales equation is

$$Y_{ij} = K_{ij} \left(\frac{k_{ij}}{a_{ij}} \right)^{(\theta-\eta)/\eta} \left(\frac{p_{ij}}{w_i^\alpha t_{ij}} \right)^\theta A', \tag{5}$$

where $A' = A\left[\alpha^{\alpha\eta} - \alpha^\eta\right]^{-1}$. Comparing (4) to (5), the essential differences are that (i) $\eta =$ the 'intensive' margin supply elasticity without endogenous selection is replaced by $\theta =$ the 'interior extensive' margin of firms selection elasticity; and (ii) the constant utilization rate $U_{ij} = \underline{\varrho}_{ij}^{-\theta}$ is replaced by the price-independent proportion of the utilization rate $U'_{ij} = (k_{ij}/a_{ij})^{(\theta-\eta)/\eta}$.

Stepping back from the highly specialized model of selection with Pareto productivity

[13] For the case $\eta > \theta + 1$ the integral is unbounded, interpreted as a situation where the most productive firms dominate the market.

draws for a more realistic perspective, the expected productivity of active bilateral capital in the Cobb-Douglas production-*cum*-distribution model is due to a combination of productivity draws and *ex post* labor choice. Approximate this reality with a tractable log-linear model of supply change along both extensive and intensive margins of firms: $Y_{ij} = K_{ij} U^*_{ij} \left(p_{ij}/w_i^\alpha t_{ij}\right)^{\eta^*} B$. Here supply is the product of a constant $B > 0$, price-invariant portion of utilization rate U^*_{ij} a power function with supply elasticity η^* that combines intensive and extensive margin changes. From this perspective, supply equations (4) and (5) are special cases. For concreteness and without loss of generality, we use supply equation (4) in developing short run gravity.

The supply side of the model is closed by the labor market clearing condition at given prices $\{p_{ij}\}$ and labor endowment L_i for each country i:

$$L_i = \sum_j \int_{\underline{\varrho}_{ij}}^{\infty} L_{ij}(\varrho)\theta\varrho^{\eta-\theta-1}d\varrho \left[\alpha \frac{p_{ij}}{w_i t_{ij}}\right]^\eta K_{ij} = \alpha^\eta \frac{\theta}{\theta+1-\eta} \sum_j \underline{\varrho}_{ij}^{\eta-\theta} \left[\frac{p_{ij}}{w_i t_{ij}}\right]^\eta K_{ij}. \tag{6}$$

Solving equation (6) for w_i and using $\underline{\varrho}_{ij}^{-\theta} = U_{ij}$ yields:

$$w_i = \alpha \tilde{A} L_i^{-1/\eta} \left[\sum_j U_{ij} K_{ij} \left(\frac{p_{ij}}{t_{ij}/\underline{\varrho}_{ij}}\right)^\eta\right]^{1/\eta}, \tag{7}$$

where $\tilde{A} \equiv [\theta/(\theta+1-\eta)]^{1/\eta}$. The net value of sales to all destinations at sellers prices is given by $Y_i^S = w_i L_i/\alpha$ using the Cobb-Douglas property. Using equation (7), and noting that $L_i^{1-1/\eta} = L_i^\alpha$, this implies the CES joint product revenue function

$$Y_i^S = \tilde{A} L_i^\alpha K_i^{1-\alpha} U_i^{1-\alpha} \left[\sum_j u_{ij} \lambda_{ij} \left(\frac{p_{ij}}{t_{ij}/\underline{\varrho}_{ij}}\right)^\eta\right]^{1/\eta}, \tag{8}$$

where $u_{ij} = U_{ij}/U_i$, $U_i^{1-\alpha} = \left[\sum_j U_{ij}\right]^{1/\eta}$, $\lambda_{ij} = K_{ij}/K_i$ and $\left[\sum_j K_{ij}\right]^{1/\eta} = K_i^{1-\alpha}$. The square bracket term in (8) is the CES seller price index $\mathcal{P}_i$. Applying Hotelling's Lemma to

(8) implies that sales shares at seller prices from i to destinations j is given by

$$s_{ij} = \left(\frac{p_{ij}/(t_{ij}/\underline{\varrho}_{ij})}{\mathcal{P}_i} \right)^{\eta} u_{ij}\lambda_{ij}.$$

2.1.1 Short Run Gravity with Heterogeneous Firms

Gravity is based on spatial arbitrage that generates equilibrium buyer prices p_{ij} from each origin i to each destination j. Demand is characterized by CES expenditure on goods from many origin countries. Expenditure on good i in destination j is given by

$$X_{ij} = \left(\frac{p_{ij}}{P_j} \right)^{1-\sigma} E_j, \tag{9}$$

where E_j is total expenditure on goods from all origins, $P_j = [\sum_i (p_{ij})^{1-\sigma}]^{1/(1-\sigma)}$ is the CES price index, and $\sigma > 1$ is the elasticity of substitution.[14]

Market clearing in each origin-destination pair ij equates the right hand side of (9) with $t_{ij}s_{ij}Y_i^S$. This sets value at buyer prices on both sides of the equilibrium condition. Solve

$$\left(\frac{p_{ij}}{P_j} \right)^{1-\sigma} E_j = t_{ij} \left(\frac{p_{ij}}{t_{ij}/\underline{\varrho}_{ij}} \right)^{\eta} \mathcal{P}_i^{1-\eta} L_i^{\alpha} K_i^{1-\alpha} U_i^{1-\alpha} u_{ij}\lambda_{ij}\tilde{A}$$

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

$$p_{ij} = \left[\frac{E_j P_j^{\sigma-1} (t_{ij}^{\alpha}/\underline{\varrho}_{ij})^{\eta}}{\mathcal{P}_i^{1-\eta} R_i U_i^{1/\eta} u_{ij}\lambda_{ij}} \right]^{1/(\eta+\sigma-1)}, \tag{10}$$

where $R_i = L_i^{\alpha} K_i^{1-\alpha} \tilde{A}$. The cost term t_{ij}^{α} on the right hand side of (10) arises because the elasticity of sales $t_{ij}Y_{ij}^S$ with respect to t_{ij} is equal to $1 - \eta = -\alpha\eta$ using $\eta \equiv 1/(1-\alpha)$.[15]

Substitute the right hand side of (10) for p_{ij} in the right hand side of the demand equation

[14] Variation of preference weights by place of origin is absorbed in the t_{ij} frictions, as is well recognized in the literature.

[15] Equation (10) for the market-clearing bilateral price p_{ij} corrects the equivalent expression in Anderson and Yotov (2020), which omitted multiplying Y_{ij}^S by t_{ij}.

(9). After simplification this yields:

$$X_{ij} = \left(\frac{\tilde{t}_{ij}}{P_j}\right)^{(1-\sigma)\rho} E_j^{\rho}(R_i \mathcal{P}_i^{1-\eta} U_i^{1-\alpha})^{(1-\rho}(u_{ij}\lambda_{ij})^{1-\rho}, \tag{11}$$

where $\rho \equiv \eta/(\eta + \sigma - 1)$ and $\tilde{t}_{ij} \equiv (t_{ij})^{\alpha}/\underline{\varrho}_{ij}$, the effective bilateral iceberg friction affecting trade from i to j.

The next step in deriving the gravity equation is to solve the aggregate market clearing equation at buyers prices,[16] $Y_i = \sum_j X_{ij}$ for the sellers' price index $\mathcal{P}_i$ in combination with the exogenous sellers' mass terms. Define $Y \equiv \sum_i Y_i$ to scale i's sales relative to world sales at buyer prices. Then the market clearing equation for each origin i is

$$\frac{Y_i}{Y} = \left(\mathcal{P}_i^{1-\eta} R_i U_i^{1-\alpha}\right)^{(1-\sigma)\rho} \sum_j \frac{E_j^{\rho}}{Y} \left(\frac{\tilde{t}_{ij}}{P_j}\right)^{(1-\sigma)\rho} (u_{ij}\lambda_{ij})^{1-\rho}$$

The summation term on the right hand side is (a power transform of) the sellers' multilateral resistance

$$\tilde{\Pi}_i^{(1-\sigma)\rho} = \sum_j \frac{E_j}{Y} \left(\frac{\tilde{t}_{ij}}{\tilde{P}_j}\right)^{(1-\sigma)\rho} (u_{ij}\lambda_{ij})^{1-\rho}$$

where the (power transform of) buyers' multilateral resistance $\tilde{P}_j^{(1-\sigma)\rho} = E_j^{1-\rho} P_j^{(1-\sigma)\rho}$. Substitute in (11) $Y_i/Y\tilde{\Pi}_i^{(1-\sigma)\rho}$ for $\left(\mathcal{P}_i^{1-\eta} R_i U_i^{1-\alpha}\right)^{(1-\sigma)\rho}$ and substitute $E_j/\tilde{P}_j^{1-\sigma)\rho}$ for $E_j^{\rho}/P_j^{(1-\sigma)\rho}$ to yield (bringing in product index h notation)[17] :

Proposition 1: ***Short Run Gravity with Heterogeneous Firms.***

$$X_{ij}^h = \underbrace{\frac{(Y_i)^h E_j^h}{Y^h} \left(\frac{\tilde{t}_{ij}^h}{\tilde{\Pi}_i^h \tilde{P}_j^h}\right)^{(1-\sigma^h)\rho^h}}_{\textit{Structural Gravity}} \underbrace{\times\, (\lambda_{ij}^h)^{1-\rho^h}}_{\textit{Short Run}} \underbrace{\times\, (u_{ij}^h)^{1-\rho^h}}_{\textit{Heterogeneous Firms}}. \tag{12}$$

[16]The change in valuation implies a slight abuse of notation since $Y_i = \sum_i t_{ij} Y_{ij}$.

[17]Product differences indexed by superscript h appear in the iceberg frictions t_{ij}^h, the Pareto location parameter $\underline{\varrho}_{\min}^h$ and the Pareto dispersion parameter $\theta^h > 0$.

Equation (12) can be decomposed into three structural terms. 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. We label the second term in (12) *'Short Run'* because it reflects the 'short run' gravity model on the intensive margin of Anderson and Yotov (2020), where λ_{ij}^h is an *ex ante* bilateral capacity variable – the fraction of country i's capital in sector h allocated *ex ante* to destination j and ρ^h can be interpreted as the proportion by which the short run trade elasticity is reduced from the long run trade elasticity.[18] Finally, we label the third term *'Heterogeneous Firms'* because the relative utilization rate u_{ij}^h appears in equation (12) due to the presence of heterogeneous firms *a la* Melitz (2003) and Chaney (2008). Without heterogeneous firms, the term $(u_{ij}^h)^{1-\rho^h}$ will disappear and (12) will collapse to the short run gravity model of Anderson and Yotov (2020). Moreover, if we move to a hypothetical long run (i.e., when $\rho^h = 1$), the term $(\lambda_{ij}^h)^{1-\rho^h}$ will disappear as well, and (12) will become the standard long-run gravity equation.[19]

2.1.2 The Extensive Margins of Trade

The analysis in this section is greatly simplified by the extreme assumption of perfect foresight. This allows clean predictions to structure the empirical investigation of variation in the extensive margins of trade and production. In wider perspective, the model is a formal metaphor for the observed behavior of firms on the extensive margin. Imperfect information about future prospects and departures of the initial conditions from long run equilibrium efficiency complicate the entry/exit condition. Also, learning how to produce and serve new sector/destinations plausibly takes place over time, inducing partial adjustment and correcting for mistakes. The treatment here abstracts from all such dynamic considerations

[18] To see this intuition, note that ρ^h enters multiplicatively in the exponent of the trade cost element in the 'Structural Gravity' term, thus driving a wedge between the short-run and the long-run trade elasticity.

[19] Note that, when $\rho^h = 1$, the last two terms in (12) will disappear at the same time, even in the presence of heterogeneous firms. However, the presence of heterogenous firms would still be captured in the model through the definition of $\tilde{t}_{ij}^h$, i.e., (12) will collapse to a long run gravity model with heterogeneous firms *a la* Melitz (2003) and Chaney (2008).

to simplify focus on the essential static logic: entry requires a lower than eventually efficient capacity to raise the next period returns above the opportunity cost of capital. This logic rationalizes the widely observed regularity in the export behavior of firms over time – successful entry is followed by growth in export of the successful firms.

Consider the process for allocation of capital. A new market n would be allocated its capital when demand in n is expected to be sufficiently large. Suppose that entry into a new destination n requires an iceberg fixed entry cost $\phi_{in} \geq 1$. For new destination n toe be served by i, the creation of 1 unit of destination specific capital requires ϕ_{in} units of capital drawn from alternative uses. $\phi_{in} > 1$ represents a cost of learning when there are no incumbent origin i firms to learn from. Then the profitable creation of capacity for n must satisfy:

$$\frac{r_{in}}{r_i} \geq \phi_{in}, \quad \phi_{in} \geq 1, \tag{13}$$

where r_{ij} is the return in origin i on the specific capital for delivery to destination j, r_i is the corresponding average return on marketing capital in country i. Order the potential new destination markets in decreasing order of their relative return r_{in}/r_i. For the cutoff market entered, $j = n$, equality holds. For $j > n$, no entry occurs and for $j < n$ entry has previously occurred.

The *ex post* relative return to utilized marketing capital, r_{ij}/r_i, is obtained is obtained by differentiating (12) with respect to k_{ij} and placing that result relative to the effect of an equiproportionate rise in all k_{ij}, the weighted average of the r_{ij}s. This implies:

$$\frac{r_{ij}}{r_i} = \frac{s_{ij}}{\zeta_{ij}}, \tag{14}$$

where, s_{ij} is the share of sales from origin i to destination j, and $\zeta_{ij} \equiv \lambda_{ij} u_{ij}$ is the proportion of utilized marketing capital from i that is allocated to destination j.

Efficient allocation in terms of realized (*ex post*) returns implies equal returns on investment in each destination served, i.e., $s_{ij}/\zeta_{ij} = 1$. Utilization rates u_{ij} generally will differ across destinations given the efficient *ex post* allocation. *Ex ante* efficiency nevertheless holds with risk neutrality because the relative utilization rates $u_{ij} = U_{ij}/U_i$ are interpreted as the probabilities of an origin i firm's capital being utilized in destination j, hence the expected return per unit of *ex ante* firm capital is equalized across destinations. Higher realized payoffs are offset by lower probabilities of receiving them. Formally:

Proposition 2: *Ex ante and Ex post Efficiency.*

(a) *Ex post* efficiency is realized if and only if $s_{ij} = \zeta_{ij}, \ \forall i, j$;

(b) *Ex ante* efficiency under risk neutrality implies *ex post* efficiency.

With *ex ante* efficient capacity allocation, long run gravity obtains. In (12), fully efficient investment $\zeta_{ij} = X_{ij}/Y_i \Rightarrow$ implies a solution to (12) as if the exponent $\rho = 1$ and the multilateral resistances ($\tilde{\Pi}_i$ and $\tilde{P}_j$) revert to their interpretation in standard (long run) gravity. The fully efficient investment equilibrium implies that at the margin the opportunity cost of reallocating 1 unit of capital from the existing allocations is equal to 1.

Now consider investment. In combination, (13) and (14) imply the entry condition:

$$\frac{\hat{s}_{in}}{\zeta_{in}} \geq \phi_{in}, \quad \phi_{in} \geq 1. \tag{15}$$

where $\hat{s}_{in}$ is the expected sales share to the extensive margin destination. This implies setting $K_{in}/K_i = \lambda_{in}$ such that the expected relative utilization rate u_{ij} satisfies

$$\zeta_{in} = \frac{\hat{s}_{in}}{\phi_{in}}. \tag{16}$$

A useful simplification, due to perfect foresight, is the equivalence between expected and realized sales shares, i.e., $\hat{s}_{in} = s_{in}$. Using this relationship and (16) to substitute in equation (12) delivers:

Proposition 3: *Investment with Perfect Foresight.* *Investment with perfect foresight* *implies for origin i's extensive margin n of destinations*

$$\zeta_{in}^{\rho} = \frac{E_n}{Y\phi_{in}} \left(\frac{\tilde{t}_{in}}{\tilde{\Pi}_i \tilde{P}_n} \right)^{(1-\sigma)\rho}. \tag{17}$$

Following entry by firms from i in destination n, ζ_{in} is adjusted on the intensive margin to its efficient level ζ_{in}^*. The initial smaller allocation in equation (17) earns higher profit to offset the entry cost:

$$\zeta_{in} = \frac{\zeta_{in}^*}{\phi_{in}}. \tag{18}$$

This induces higher than long run efficient returns to offset the iceberg friction of entry. The opportunity cost of subsequently reallocating capital is equal to 1, $\phi_{in} = 1$ in equation (17) and thus

$$\zeta_{in}^* = \frac{E_n^*}{Y^*} \left(\frac{\tilde{t}_{in}}{\Pi_i^* P_n^*} \right)^{1-\sigma},$$

the long run efficient allocation of marketing capital. Effectively, it is as if in the long run $\rho = 1$.[20] Bringing in product index h notation and substituting ζ_{in} in (18) for $\lambda_{in} u_{in}$ in (12) delivers gravity including the extensive margin of trade at the time of entry:

Proposition 4: ***Short Run Gravity and the Extensive Margins of Trade.***

$$X_{in}^h = \underbrace{\frac{Y_i^h E_n^h}{Y^h} \left(\frac{\tilde{t}_{in}^h}{\tilde{\Pi}_i^h \tilde{P}_n^h} \right)^{(1-\sigma^h)\rho^h}}_{\textit{Structural Gravity}} \times \underbrace{(\lambda_{in}^{h,*})^{1-\rho^h}}_{\textit{Short Run}} \times \underbrace{(u_{in}^{h,*})^{1-\rho^h}}_{\textit{Heterogeneous Firms}} \times \underbrace{(\phi_{in}^h)^{\rho^h - 1}}_{\textit{Extensive Margin}}. \tag{19}$$

Equation (19) captures our two main theoretical contributions. First, it explicitly accounts for the extensive margin of trade in the rightmost term $(\phi_{in}^h)^{\rho^h-1}$ that temporarily reduces the capital allocation $\lambda_{in}^{h,*}$ below its long run efficient value. This term reflects the need for temporarily higher profit to offset the one-time fixed entry cost. This term winks off in the

[20]While it is simplest to think of *expanding* on the extensive margin, the extensive margin analysis above applies equally to exit on the assumption that exit costs are equal to entry costs. Exit costs that differ are analyzed by replacing ϕ_{in} with some $\phi_{in}' > 1$.

next period. Second, the term $(u_{in}^{h,*})^{1-\rho^h}$ accounts for the effect of heterogeneity of firms on the expected utilization rate.

Propositions 3 and 4 extend to characterize investment at the extensive margin of products by the device of interpreting ranges of destinations as products, maintaining all other formal elements. Entry of country i into extensive margin product $h = H_i$ requires entry into destination markets n in $\mathcal{M}_{H_i} = n \in [\underline{n}_{H_i}, \overline{n}_{H_i}]$ described by (18). $\mathcal{M}_{H_i}$ is the range of destinations for product H_i to which allocation of marketing capital $k_{in}^{H_i}$ is efficient, given production.

Assume there is a fixed cost $F_{H_i} > 1$ for entry to production of product line H_i. The entry logic of (16)-(18) applies to each destination n served in the aggregate sales share of product H_i (at sellers' prices):

$$\sum_{j \in \mathcal{M}_{H_i}} s_{in}^{H_i} \frac{Y_i^{S,h}}{\sum_{h=1}^{H_i} Y_i^{S,h}} = S_i^{H_i}.$$

A combined reduction $\phi_{in}^{H_i} F_{H_i}$ reduces the long run allocation $\zeta_{in}^{H_i,*}$ to each destination of the new product H_i by an amount sufficient to pay the entry cost to production and distribution. Each potential added destination pays its share of the production entry cost. In this sense, Proposition 3 applies.

Proposition 4 extends to the volume at the domestic extensive margin by aggregating (19) over the range of feasible destinations

$$\sum_{n \in \mathcal{M}_{H_i}} X_{in}^{H_i} = \frac{Y_i^{H_i}}{Y^{H_i}} \sum_{n \in \mathcal{M}_{H_i}} E_n^{H_i} \left(\frac{\tilde{t}_{in}^{H_i}}{\tilde{\Pi}_i^{H_i} \tilde{P}_n^{H_i}} \right)^{(1-\sigma^{H_i})\rho^{H_i}} (\zeta_{in}^{H_i,*})^{1-\rho^{H_i}} (F_{H_i} \phi_{in}^{H_i})^{\rho^{H_i}-1}.$$

The useful combined production-*cum*-distribution entry cost of H_i is implied by using $\sum_{n \in \mathcal{M}_{H_i}} X_{ij}^{H_i} = Y_i^{H_i}$ to divide both sides of the preceding equation:

$$1 = \sum_{n \in \mathcal{M}_{H_i}} \frac{E_n}{Y^{H_i}} \left(\frac{\tilde{t}_{in}^{H_i}}{\tilde{\Pi}_i^{H_i} \tilde{P}_n^{H_i}} \right)^{(1-\sigma^{H_i})\rho^{H_i}} (\zeta_{in}^{H_i,*})^{1-\rho^{H_i}} (F_{H_i} \phi_{in}^{H_i})^{\rho^{H_i}-1}.$$

Then:

$$\Phi_{H_i} = F_{H_i} \left[\sum_{n \in \mathcal{M}_{H_i}} \frac{E_n}{Y^{H_i}} \left(\frac{\tilde{t}_{in}^{H_i}}{\tilde{\Pi}_i^{H_i} \tilde{P}_n^{H_i}} \right)^{(1-\sigma^{H_i})\rho^{H_i}} (\zeta_{in}^{H_i,*})^{1-\rho^{H_i}} (\phi_{in}^{H_i})^{\rho^{H_i}-1} \right]^{1/(\rho^{H_i}-1)} = F_{H_i} \bar{\phi}_{H_i}.$$

The production-*cum*-entry decision requires the utilized capital committed to H_i to satisfy the analog of (18):

$$\mathcal{Z}_i^{H_i} = \frac{\widehat{S}_i^{H_i}}{\Phi_{H_i}}.$$

Note that $i \in \mathcal{M}_{H_i}$ is not formally necessary, some new products may export only. Although domestic sales often precede exports, this is not universally observed. In all cases, export logically requires production so we refer to the extensive margin of active products as the *'Domestic Extensive Margin' (DEM)*.[21]

Since Proposition 4 holds equally for international and domestic links (i.e., both $\forall\, i \neq n$ and $\forall\ i = n$), it captures two distinct forms of the extensive margin of trade. First is the standard external (cross-border) margin of trade on which the distribution to export destinations j changes. Second is the domestic margin $i = n$ of activity, i.e., the *'Domestic Extensive Margin' (DEM)*, where domestic sectoral production and/or consumption switches on or off. We demonstrate below that DEM impacts potentially both the internal and cross-border extensive margins of distribution. Moreover, we show in the empirical analysis 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 lens of the model remains sharp with small generalizations of the setup. The Cobb-Douglas production-*cum*-distribution function with sectoral labor share parameter

[21] Since our data would not allow us to identify such products, we will obtain our main estimates from a sample that only includes products that are produced domestically. However, in the sensitivity analysis we also allow for the external intensive margin to be defined more broadly. This does not have strong quantitative implications. More importantly, the definition of the external extensive margin has absolutely no impact on our methodological contribution regarding the importance of accounting for the domestic extensive margin.

variation makes no essential difference to the short run gravity structure except for an added source of supply elasticity variation. This follows because efficient national labor markets imply for each distinct sector k (extending the notation naturally) the common wage $w_i = \alpha_k Y^S_{i,k}/L_{i,k},\ \forall k$, while the common wage clears the labor market, $\sum_k L_{i,k} = L_i$. The model also readily extends to a Cobb-Douglas bundle of mobile factors that combine with destination-specific capital. Some production function forms more general than Cobb-Douglas may be well approximated by the Cobb-Douglas, though subject to blurred focus.[22] Similar approximation arguments apply to generalizing the productivity distribution beyond Pareto. Continued generalization eventually breaks the invertibility of derived labor (or composite mobile factors) demand that is essential for a gravity representation.

2.2 From Theory to Empirics: Estimating the Extensive Margin

To translate our structural model (19) into a corresponding estimating equation, we proceed in three steps that rely on three different strands of the literature. First, we transform 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).[23] 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 mapping the variables of our theory into observables to be used in an econometric model applied to panel data. First, we introduce notation for time periods in which the model applies to static equilibrium. Appealing to readers' familiarity with standard nomenclature, t_{ij} remains the iceberg cost factor applied to trade from origin i to destination j while subscript t refers to time period t. Let $N^h_{ij,t}$ be an indicator equal to one when

[22] The elasticity of labor demand with respect to the wage becomes endogenous outside the Cobb-Douglas case. This implies implicitly defined local supply elasticities.

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

at least one firm exports h from i to j at time t. In order for this to be the case, there should be at least one firm that finds it profitable to produce and export product h, i.e., $\pi^h_{ij,t}(\varrho) > 0$. This implies that the probability for a given product to be exported from origin i to destination j at time t is:

$$\Pr(N^h_{ij,t} = 1 | x_{ij,t}) = \Pr(\pi^h_{ij,t}(\varrho) > 0) = F^h({x'}_{ij,t}\beta). \tag{20}$$

Let $N_{ij,t} = \sum_h N^h_{ij,t}$ be the total number of products exported from i to j at time t. The previous expression implies:

$$E(N_{ij,t}|x_{ij,t}) = \sum_h \Pr(N^h_{ij,t} = 1|x_{ij,t}) = \sum_h F^h({x'}_{ij,t}\beta) = N_{i,t}F({x'}_{ij,t}\beta), \tag{21}$$

where $N_{i,t}$ is the total number of products available in origin i, and $F({x'}_{ij,t}\beta) = \left(F^h({x'}_{ij,t}\beta)\right)/N_{i,t}$ is interpreted as the probability that a randomly selected product h will be exported from country i to country j at time t.

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

$$F({x'}_{ij,t}\beta) = 1 - \left(1 + \omega \exp\left(x'_{ij,t}\beta\right)\right)^{-\frac{1}{\omega}}, \ \omega > 0.$$

This functional form has two 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. (Under the restriction $\omega > 0$, $F({x'}_{ij,t}\beta)$ is bounded between 0 and 1, which implies $[0; N_{i,t}]$ bounds for $N_{i,t}F({x'}_{ij,t}\beta)$.) 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 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. We will estimate the model by Bernoulli pseudo-maximum likelihood, which is easy to implement and it is consistent under very general conditions, c.f., Santos Silva et al. (2014) and Papke and Wooldridge (1996).[24]

The robustness of our main results to the FLEX estimator specification is investigated with three alternative estimators. First is a double-bounded Tobit estimator. Second, following the best current practices in the intensive margin gravity literature is the PPML estimator, which has the attractive property of being a count multiplicative model.[25] Finally, despite its limitations in the current setting (i.e., inability to capture the behavior of the distribution at its bounds because the partial OLS effects are assumed to be constant), we also obtain robustness estimates with the OLS estimator. As demonstrated in the sensitivity analysis, our main results and the conclusions that we draw in relation to our methodological contributions are robust to the use of alternative estimators.

Selection of the covariates in the model, the third step, completes the econometric specification. We combine the received empirical literature on the intensive and extensive margins of trade with the domestic extensive margin, our key contribution. We start by defining:

$$\exp\left(x'_{ij,t}\beta\right) = \exp\left(\pi_{i,t} + \chi_{j,t} + \gamma_{ij} + BIPOL_{ij,t}\beta_1\right), \quad \forall i \neq j. \tag{22}$$

Equation (22) 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

[24]We refer the reader to Gourieroux et al. (1984) and Papke and Wooldridge (1996) for a discussion.

[25]PPML established itself as the leading gravity estimator 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, and to Berthou and Fontagne (2008) for an application to the extensive margin of trade.

time-invariant bilateral determinants of the extensive margin of trade. Finally, $BIPOL_{ij,t}$ is a vector of time-varying bilateral policy determinants of trade, e.g., RTAs, tariffs, etc.

A common feature of all empirical papers on the extensive margin of trade, as illustrated by equation (22), is that, without exception, they use international trade data only ($i \neq j$). Thus they cannot account for effects of variation in the domestic extensive margin.[26] Once the domestic extensive margin is introduced, equation (22) becomes:

$$\exp\left(x'_{ij,t}\beta\right) = \exp\left(\pi_{i,t} + \chi_{j,t} + \gamma_{ij} + BIPOL_{ij,t}\beta_1 + EXS_{i,t}{\times}BRDR_{ij}\beta_2 + IMP_{j,t}{\times}BRDR_{ij}\beta_3\right) \times \exp(CNTRY_{j,t}{\times}BRDR_{ij}\beta_4 + EXR_{ij,t}{\times}BRDR_{ij}\beta_5 + \sum_{i,t}\beta_{i,t}GLOB_{i,t}\times BRDR_{ij}). \tag{23}$$

The introduction of the domestic extensive margin allows for identification the effects of five new terms, which appear in equation (23) but could not be identified from specification (22).[27] When the extensive margin is defined based on cross-border observations only, the effects of any non-discriminatory export policies are absorbed by the exporter-time fixed effects. In contrast, once the domestic extensive margin observations are introduced, the impact of any non-discriminatory export support policies can be identified in the presence of the exporter-time fixed effects because the export support policies apply only to international and not to domestic trade. Specifically, $EXS_{i,t}$ is a vector of non-discriminatory export support policies, e.g., export subsidies, trade fairs, etc. We interact $EXS_{i,t}$ with $BRDR_{ij}$, which is an indicator variable for cross-border trade, equal to 0 for domestic trade. Thus, the resulting interaction, $EXS_{i,t}{\times}BRDR_{ij}$, is time-varying and bilateral and, therefore, it can be identified in the presence of all fixed effects from (23).

The second new term in (23) 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-

[26] In our review of the voluminous empirical literatures on export diversification and on the extensive margin of trade we did not come across a single paper that took into account the domestic extensive margin.

[27] Thus our theory-consistent domestic extensive margin has potentially significant implications for estimating the impact of numerous determinants of the extensive margin of trade.

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 (23) 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 (TBT) 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.

The fourth new term in specification (23) is the exchange rate between i and j at t, $EXR_{ij,t}$. Even though exchange rates are bilateral their impact cannot be identified in gravity specifications with international trade data only due to perfect collinearity with the exporter time and importer time fixed effects. Once the domestic extensive margin is introduced, we can obtain estimates of the nonuniform/discriminatory impact of exchange rates on the external relative to the domestic extensive margin, because exchange rates do not vary domestically.

The fifth new term in specification (23) is $\sum_{i,t} \beta_{i,t} GLOB_{i,t} \times BRDR_{ij}$, which denotes a set of time-invariant cross-border dummies $BRDR_{ij}$ interacted with origin-time globalization dummies $GLOB_{i,t}$. The motivation for the inclusion and emphasis on this term is threefold. First, from a methodological perspective, the inclusion of the globalization dummies will highlight our key contribution. Specifically, we will demonstrate the these globalization effects cannot be identified without proper account for the domestic extensive margin. Second, from a policy perspective, the inclusion of the time-varying border indicators would enable us to resolve the 'the missing globalization puzzle', c.f., Coe et al. (2002), on the extensive margin of trade.[28] Third, from a practical perspective, the inclusion of the globalization

[28] In the empirical analysis we demonstrate that the effects of globalization are present and can be identified in our setting both when we constrain them to be common across countries and also when they are country-

dummies will enable us to address the challenge that we do not have data on some of the structural variables in (19), i.e., ${\zeta^h_{in,t}}^{1-\rho^h} \equiv (\lambda^h_{ij,t} u^h_{ij,t})^{1-\rho^h}$. Thus, the country-time specific globalization estimates that we will obtain in the empirical analysis will offer a flexible and comprehensive/all-inclusive account for the dynamic evolution of the international bilateral links relative to the domestic extensive margin.

Finally, we note that the introduction of the domestic extensive margin has two potentially 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.[29] 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 of more products, 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, i.e., to check whether joining the WTO has lead to changes in the overall trade of individual members.[30] 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.

specific. Importantly, as with all other new terms that appear in (23), neither the common globalization trends nor the country-specific globalization effects could be identified without the domestic extensive margin.

[29] 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.

[30] Alternatively, the introduction of the domestic extensive margin would allow for simultaneous identification of the effects for WTO membership on the extensive margin of trade with other members vs. trade with non-members for each WTO member cuntry.

3 Data: Construction and Sources

The empirical analysis uses our construction of a dataset that covers the extensive margin for about 3,313 mining and manufacturing goods for 35 countries, 1995-2014, a period which includes the main waves of EU enlargement.[31] The novel 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 experiment with several alternative constructed samples by extending the data coverage across the product and country dimensions. We construct the estimating sample(s) in steps described subsections 3.1-3.2. Added robustness checks use data described in subsection 3.3

3.1 The Domestic Extensive Margin

The original data source that we use to construct the *domestic extensive margin* is PRODCOM, a database developed, maintained, and hosted by Eurostat.[32] PRODCOM includes 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 3-letter ISO alpha codes, can be found in the first two columns of Table 1. Data for about half of the countries in PRODCOM are complete and balanced throughout the period of investigation. The countries for which data are not balanced include Bosnia and Herzegovina, Northern Macedonia, Montenegro, and Serbia (with data after 2010), Bulgaria, Croatia, Czech Republic, Hungary, Latvia, and Slovenia (with data after 2000), Cyprus and Malta (with data from 2003 to 2005), Estonia, Portugal, Lithuania, and Romania (with data after 1999), Luxembourg (with data from 1999 to 2005), Norway (with data after 2001), Slovakia (with data after 1998), and Turkey (with data from

[31]Specifically, Austria, Finland, and Sweden joined in 1995, Cyprus, Czechia, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, and Slovenia joined in 2004, Bulgaria and Romania in 2007, finally, Croatia in 2013.

[32]The title 'PRODCOM' comes from the French "PRODuction COMmunautaire" (Community Production). The description of PRODCOM can be found at https://ec.europa.eu/eurostat/web/prodcom.

2006 to 2010). In sum, in terms of country coverage, PRODCOM delivers an unbalanced sample of 35 countries over the period 1995-2014.

Turning to the product dimension, which is more important for our analysis, PRODCOM offers production data for the broad categories of mining, quarrying, and manufacturing (with the exception of military products and some energy products). On average, PRODCOM covers about 3800 products. However, due to the invention of new products, discontinued production of others, and changes in the classification, the number of categories covered in PRODCOM varies across years. Thus, a consistent classification over time within the PRODCOM database is needed for our purposes. Moreover, our analysis would require matching the PRODCOM data with data on international trade, i.e., the external extensive margin, which come from another source (COMEXT). With this in mind, we take three steps to construct the domestic extensive margin.

In the first step, we ensure consistency over time within the PRODCOM database, and consistency between the data from PRODCOM and the international trade data from COMEXT. To this end, we rely on Bradley et al. (2022), who develop a concordance between the PRODCOM and the COMEXT databases, which also ensures time consistency within each of them. We offer further details about this concordance in the next subsection, where we describe the matching between the domestic extensive margin and the external extensive margin. For now, we just note that ensuring internal consistency within each of the two databases over time as well as between the two datasets resulted in a total of 3,313 product categories in our estimating sample. The maximum number of products that can be produced and traded in each year of our sample appear in the bottom row of Table 1.

In the second step, we take advantage of three sets of additional variables from the raw PRODCOM data in order to improve the product coverage. Specifically, in addition to actual reported values of production, the original PRODCOM database includes: (i) data on production quantities; (ii) flags (Confidential (:C), Estimated (:E), or Confidential/Estimated (:CE)) for some missing observations; and (iii) data on total export values and export quantities for each country-product-year combination.

We start by creating an indicator variable for the presence of domestic production (of dimension country-product-year), which equals one if the PRODCOM database reports a corresponding positive production value number. Then, we utilize the additional data on production quantities. In most cases, the data on production values match or are better (i.e., with less missing values) than the data on production quantities. However, there are cases with non-missing production quantities but missing production values. For such cases, we set the missing production indicator variable to be equal to one. Note that the additional information on production quantities would not have been useful to us if we were interested in constructing a dataset on the *intensive margin* of trade. However, this procedure is valid for our purposes, where the focus is on the *extensive margin* and all we need to know is whether there is production, which must be the case if the quantity produced was positive.

Next, we utilize the information from the flag variable in PRODCOM. We assign a value of one to the missing production indicator variable when the corresponding observation in the flag variable is Confidential (:C) or Confidential/Estimated (:CE). The assumption is that if the data were classified as confidential, then there was at least some production. Once again, the flag data would not have been useful for constructing data on the *intensive margin*, however, it is useful for our purpose. We also use the flags with values Estimated (:E). In most cases, the production value observations that correspond to a flag Estimated (:E) are positive. However, there are some instances where the Estimated (:E) flag corresponds to a missing production value. For our main analysis, we assume that the production value is positive for such cases. However, we also construct and experiment with a sample that treats the missing production values with Estimated (:E) flag as missings.

We also take advantage of the export value and export quantity variables from PRODCOM. For the overwhelming majority of cases, the observations with non-missing export data (value or quantity) correspond to non-missing production value observations. However, there are cases with non-missing export data that correspond to missing production value observations. In such cases, our assumption is that if a product is exported by a country, then it must be produced by it, and we set the missing production indicator variable to one. Closer inspection of data reveals that in some cases, production values that are explicitly set to zero in PRODCOM correspond to positive export numbers. While relatively small, the number of such cases is significant and points to the possibility that positive exports may not necessarily correspond to positive production.[33] Therefore, for the main analysis, we still set the missing production indicator to one then there are exports, however, we also experiment with a more conservative sample, where we leave the production indicator as missing even if exports are positive.

The last step in the construction of the domestic extensive margin is to sum the positive values of the production indicator variable 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, $N_{i,t}$:[34]

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

The domestic extensive margin indexes for all countries in our sample appear in Table 1. For brevity, we only include the indexes for the odd years and for the last year in the sample (2014). 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 exceptions are Bulgaria, Croatia, Finland,

[33]Based official Eurostat documentation, it is possible for the data to include zero production and positive exports. Specifically, according to a note provided by the PRODCOM team when "... there is no production and there are no imports of a product, but an enterprise exports it from stock to a subsidiary or a subcontractor (i.e. with no sale taking place), there will be zero production, zero imports and non-zero exports, resulting in negative apparent consumption." (p.2).

[34]We realize that in reality some countries would/could never produce some products. However, this issue is mitigated by the fact that the vast majority of products in our sample are in fact manufacturing goods, which could, in principle, be produced by any country (i.e., they are not subject to exogenous constraints such as the weather.)

Luxembourg, Serbia, and Slovakia, for which we have dropped some years at the beginning or at the end of the reporting period due to suspicion of misreporting. For example, the percentage change for Bulgaria is based on data between 2004 and 2014, i.e., we do not take into account the first three years of reported data due to the suspicious increase in DEM from 2003 and 2004. We note that in each case, the few problematic (very different) observations are either in the beginning or at the end of the coverage period, which includes relatively stable indexes for the rest of the years. This adjustment resulted in eliminating 12 observations (2.2 percent of the domestic extensive margin) from our estimating sample.

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. 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., Bosnia and Herzegovina, Cyprus, Iceland, Luxembourg, Malta, Montenegro, and North Macedonia), while the countries with the largest indexes are large and rich economies (e.g., Germany, France, Italy, Spain, 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, the most notable message from Table 1 is that most countries in our sample have experienced a decrease in the domestic extensive margin during the period of investigation. This may be interpreted as an indicator of specialization. Second, within the group of countries that experience a decline in the *DEM* index, we see that the biggest decline is for smaller and relatively poor EU economies (e.g., Iceland, Hungary, Ireland, Belgium, and Slovakia), while the rich European countries (e.g., Spain, Germany, France) experience a small decline in *DEM*. Finally, we see that a small number of countries have experienced an increase on the domestic extensive margin. Some of these countries are non-EU members (e.g., Serbia, Bosnia and Herzegovina, and Turkey), while others are small EU members (e.g. Estonia and Malta).

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

The *international extensive margin* of trade for our analysis relies on the COMEXT database. 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.[35] We follow the standard method to construct the international 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 some challenges and several opportunities to construct and experiment with alternative estimating samples. We describe those

[35]Information about the Comext data can be found at https://ec.europa.eu/eurostat/web/international-trade-in-goods/data/focus-on-comext.

challenges, opportunities, and our choices next.

A consistent correspondence between the domestic and international extensive margins benefits from the fact that the two main underlying databases (PRODCOM and COMEXT) 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."* Eurostat provides concordances between the PRODCOM and the CN classifications on its RAMON site at: https://ec.europa.eu/eurostat/ramon/relations.

While the intended matching between PRODCOM and CN was 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. Finally, within each of the two classifications, there have been changes over time due to the exit of products, emergence of new products, and changes in the classifications.

To perform the analysis, we needed a consistent match between the PRODCOM and COMEXT that goes beyond the official concordance files provided by Eurostat. Creating such a concordance requires significant time and effort. We capitalized on the investment of Bradley et al. (2022), who follow and build on the methods of Van Beveren et al. (2012), Pierce and Schott (2012b), and Pierce and Schott (2012a), to construct the concordance we used in three broad steps. First, they create a classification that ensures internal consistency over time within the PRODCOM data. Then, they create a classification that ensures internal consistency over time within the COMEXT data. Finally, they create a concordance that ensures consistency between PRODCOM and COMEXT. For further details on the methods to construct these concordances, we refer the reader to Bradley et al. (2022).[36] Naturally, the required aggregations needed to ensure internal consistency and consistent matching between PRODCOM and COMEXT resulted in a smaller number of products in the final estimating sample. Nevertheless, the new database covers 3,313 products. We view this level of granularity as sufficient for our purposes.

Even though COMEXT is based on data reported by European Union members only (i.e., the maximum number of declarants in COMEXT is 28), the database allowed us to construct the international extensive margin for a very wide number of countries due to the fact that each EU declarant reported 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 international extensive margin of trade for the seven non-EU countries (e.g., Bosnia and Herzegovina, Iceland, Montenegro, Northern Macedonia, Norway, Serbia and Turkey) from the PRODCOM database, which were not declarants in COMEXT. This determined the 35×35 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 capitalize on the extensive country coverage of COMEXT to construct and experiment with an alternative *'Extended Country'* sample, which includes domestic extensive margin for the 35 PRODCOM countries, as well as the (international) extensive margin of trade between the 35 PRODCOM countries and 40 additional importers from COMEXT, thus ending up with a sample covering 75 importers and 75 exporters.[37] Finally,

[36] We are indebted to Javier Flórez Mendoza for his tremendous effort with the construction and invaluable assistance with the interpretation of the PRODCOM-COMEXT, 1995-2014, concordance.

[37] The COMEXT dataset includes information for more than 200 importers. However, (i) the wide importer coverage does not add to the domestic extensive margin, which is the focus of our study, and (ii) 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

we also construct and experiment with a *'Conservative Country'* sample, which only includes the 28 EU-members that appeared as declarants in both PRODCOM and COMEXT.

We also capitalized on the fact that, for each reporting country, COMEXT includes separate data on exports and on imports, i.e., for the pair Germany-France, Germany would report exports to and imports from France, and France would report imports from and exports to Germany. In the majority of cases, trade flows in each direction are reported. However, there is a significant number of instances, where one country reports imports from another country without the second country reporting exports to the first, e.g., Germany reports imports from France, but France does not report exports to Germany. In such instances, we assume that there were indeed French exports to Germany and, accordingly, we would take this information into account when we construct the international extensive margin.

Finally, inspection of the external extensive margin of trade reveals that, for many countries, the number of exported products in COMEXT exceeded the number of domestically produced products in PRODCOM. We assume that this discrepancy is due to the presence of re-exports. Accordingly, in order to ensure consistency in our main estimating sample, we restrict the possible set of exported products for a given country and in a given year to be the set of the products that are produced by this country in the same year. In addition, however, we also construct and experiment with an *'Extended Product'* sample, which constructs the external extensive margin without any restrictions on the number of products that are exported by any country and at any point of time.

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 35 exporters and 35 importers for which there are consistently constructed data on

selected the 75 importers as follows. First, we identified the 60 countries with the largest GDPs during the period of ovenstigation. Together they account for about 96% of world GDP during the period of investigation. 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"*.

the domestic extensive margin and on the international extensive margin of trade; (ii) It is based on all reported positive production value and production quantity observations, plus the observations that include any of the three flags, plus the observations for which there are positive exports from the PRODCOM database; (iii) Its international extensive margin is based on mirorred export and import trade flows.

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 *'More Conservative Product'* sample, which is based on all reported positive production value and production quantity observations plus the observations that include any of the three flags, i.e., this sample still treats as missing the observations for which there are positive export values; (ii) A *'Most Conservative Product'* sample, which is based on all reported positive production value and production quantity observations plus the observations that include flags Confidential (:C) and Confidential/Estimated (:CE), i.e., this sample still treats as missing the observations with flags Estimated (:E) and those missing production value observations for which there are positive export values; (iii) An *'Extended Country'* sample with 75 exporters and 75 importers, which takes advantage of the extended importer coverage in COMEXT; and (iv) A *'Conservative Country'* sample with 28 exporters and 28 importers, which appear as declarants both in PRODCOM and COMEXT.

3.3 Additional Data and Sources

Our main analysis uses a specification with a rich set of fixed effects. Exporter-time and importer-time fixed effects control for all possible country-specific determinants of the extensive margin on the exporter and importer sides respectively.Symmetric or asymmetric country-pair fixed effects absorb and control for all possible time-invariant bilateral determinants of the extensive margin. Finally, we 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). Despite the specifics of our sample (i.e., covering only European economies) and the use of country-pair fixed effects, it is possible to identify the effects of WTO membership due to the fact that there is a sufficient number of countries (e.g., Bosnia and Herzegovina, Bulgaria, Croatia, Estonia, Latvia, Lithuania, and Northern Macedonia) that joined the WTO after 1995. There is also sufficient variation in the EIA indicator, e.g., due to a number of trade agreements between EU members and non-members.

4 Globalization and the Extensive Margin of Trade

We apply the data sample(s) of Section 3 to empirical specification (23) to quantify the impact of globalization and EU membership on the extensive margin of trade in Europe. This application has three advantages for our purposes. First, from a methods perspective, the application highlights our key contribution that the effects of globalization on the extensive margin on trade can be identified in a theory-motivated econometric model such as structural model (19) with origin-time and destination-time fixed effects to control for the multilateral resistances.Second, from a policy perspective, the application answers an important quantitative question: What was the impact of European integration on the extensive margin of trade? Finally, from an econometric practice perspective, the desired globalization effects are captured within a simple and robust reduced form econometric specification with fixed effects only. The flexible econometric specification enables us to obtain a series of globalization estimates across time and for individual countries, while the rich fixed effects structure of our model will diminish omitted variable and, more broadly, endogeneity concerns. The lens of the model links the various globalization fixed effects to extensive margin changes in cross-border marketing capital and, for the EU members, reductions in the fixed cost of entry to EU partner markets.

For the application to the impact of globalization in the EU, the econometric setup (23) in Section 2.2 is simplified to the following estimating equation:

$$\frac{N_{ij,t}}{N_{i,t}} = 1-\left(1+\omega\exp\left(\pi_{i,t}+\chi_{j,t}+\gamma_{ij}+\delta_1 WTO_{ij,t}+\delta_2 EIA_{ij,t}+\sum_{i,t}\beta_{i,t}GLOB_{i,t}\times BRDR_{ij}\right)\right)^{-\frac{1}{\omega}}+\epsilon_{ij,t}. \tag{24}$$

The estimator, the dependent variable, and the fixed effects for origin-time, destination-time and time-invariant pairs in specification (24) were defined in (22). Controls for pairwise time variation include $WTO_{ij,t}$ as a fixed effect for membership in the World Trade Organization and $EIA_{ij,t}$ as a fixed effect for economic integration agreements (EIAs). Finally, the covariate $\sum_{i,t}\beta_{i,t}GLOB_{i,t}\times BRDR_{ij}$ comprises time-invariant cross-border dummies $BRDR_{ij}$ interacted with origin-time globalization dummies $GLOB_{i,t}$.

In the lens of theoretical model (19), suppressing the sector notation, the $\{GLOB_{i,t}BRDR_{ij}\}$ dummies control for the effect on extensive margins of the movement of ${\zeta_{ij,t}}^{1-\rho}$ induced by steps in European integration, 1995-2014. From a broader perspective, the globalization dummies $\{GLOB_{i,t}BRDR_{ij}\}$ absorb and control for any non-discriminatory policy or country-specific characteristic that may 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 overall bilateral marketing capacity effects of the powerful integration processes at work in Europe, 1995-2014. 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.

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, 1995. Thus, the globalization estimates that we obtain should be interpreted as deviations from the corresponding border effects in 1995.

We report the empirical results in three stages. Section 4.1 reports the effects of a common globalization across all countries for each year in our sample. Section 4.2 reports country-specific effects of globalization for each of the 35 countries in our sample. Finally, in Section 4.3, we zoom in on the effects of EU accession with reports on heterogeneous estimates for the impact of EU membership on the exports vs. imports of new members.

4.1 Benchmark Results: Common Globalization Effects

We benchmark the uneven impact of European integration on the extensive margins of trade by imposing a common globalization effect across all countries for each year in our sample. That is, we constrain the country-specific globalization effects to be common, $GLOB_t = \sum_i GLOB_{i,t}$ and $\beta_{i,t} = \beta_t$. The results are reported in column (1) of Table 2 and visualized in Figure 1.

Two findings stand out: (i) the globalization estimates are all positive; and (ii) they are increasing over time. In combination, the results imply that the impact of international borders on the international relative to the domestic extensive margin has fallen significantly between 1995 and 2014 for the countries in our sample. In other words, both on average and conditional on the number of products produced domestically, the countries in our sample have traded internationally within Europe a larger fraction products. In the lens of the model we interpret this as due to bilateral marketing capacity investments stimulated by the Intensive integration processes within Europe, 1995-2014.

In addition to our main findings, Figure 1 captures two slow-downs in the effects of globalization; one in 2000-2002 and another one around 2007-2009. Possible explanations for these patterns are the corresponding economic recessions. In the lens of the model, recession reduces the expected future sales share to a potential entrant market, hence lowers the expected return on marketing capital investment and discourages extensive margin investment. Compared to corresponding estimates on the intensive margin of trade, the impact of the recessions on the extensive margin of trade seems milder. The lens of the theoretical model suggests an explanation. A uniform fall in expected sales shares to foreign destinations lowers the expected rate of return by a common rate. Most destinations continue to be served, though with lower sales, and existing bilateral capacities may be depreciated. Marginal destinations will not be entered. Finally, the positive and statistically significant estimates on WTO and EIA suggest that joining the World Trade Organization and the Economic Integration Agreements in our sample have had positive impact on the extensive margin of trade.

The economic interpretation of our estimates is given by the marginal effect of the globalization estimate in 2014. By construction, it captures the total impact of globalization during the period of investigation. The marginal effect is 511.46 (std.err. 47.35), which means that on average the number of internationally traded products increased by about 511 relative to the number of domestically traded products during the period of investigation. This is about 16.2 percent of the total number of possibly traded products in 2014. This relatively large number suggests that globalization had strong effects on the extensive margin of trade in Europe during the period 1995-2014.

From a methodological perspective, our findings demonstrate that our version of structural gravity is well suited to capture the extensive margin effects of globalization.[38] What

[38]This result contrasts with the "missing globalization puzzle", according to which the gravity model cannot capture the impact of globalization on the intensive margin of trade, c.f., Coe et al. (2002). Our

enables us to identify these effects in our setting is the proper account of the *domestic extensive margin*. Without domestic extensive margin observations, the $GLOB_t$ variables would have been perfectly collinear with and, therefore, absorbed by the country-time fixed effects. Note also that identification does not come from any variation over time – identification of border effects is possible in the presence of exporter and importer fixed effects even in a cross section setting so long as the dataset includes domestic trade.

We demonstrate the importance of the domestic extensive margin in column (2) of Table 2, where we replicate our main specification from column (1) but on a sample that does not include observations for the domestic extensive margin. Consistent with our main contribution and argument about the crucial role of the theory-consistent domestic extensive margin of trade, the estimates in column (2) reveal that without the *DEM* observations we cannot identify the globalization effects that we are after. Another notable change between columns (1) and (2) is that the estimate of the impact of WTO in column (2) is no longer statistically significant. This result contributes to the debate on the effects of the WTO on international trade, e.g., Rose (2004). Comparison between the estimates from columns (1) and (2) suggests that (i) on average and conditional on the number of products produced domestically, the new WTO members in our sample have traded a larger fraction of products internationally; and (ii) gravity studies that do not account for the domestic extensive margin may underestimate the impact of WTO on the extensive margin of trade.

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 and econometric specifications, which are reported in panel B of Table 2; and (ii) alternative estimating samples, which are reported in panel C of Table 2.

The results in columns (3) to (6) of panel B are obtained with the Tobit, the PPML, the OLS estimators, and with the FLEX estimator with asymmetric pair fixed effects, re-

conclusion is that the effects of globalization were actually always in the structural gravity model, but 'hidden' in the exporter-time and/or the importer-time fixed effect.

spectively. Overall, the estimates in panel B confirm our main finding about a strong and positive impact of globalization on the extensive margin of trade. However, we also observe some differences. For example, we see that some of the PPML and OLS estimates before 2006 are not statistically significant. These results underscore the importance of selecting a proper estimator. In addition, we note that the globalization estimates with the directional/asymmetric pair fixed effects in column (7) are larger than those in column (1). A possible implication of this finding is that the effects of globalization could be asymmetric. We explore this hypothesis is Section 4.3, where we study the uneven effects of EU membership.

We conclude the analysis with the estimates in panel C of Table 2, where we experiment with the following alternative estimating samples: (i) the main sample but with 3-year interval data, in column (7); (ii) our *'Most Conservative Product'* sample, as described in the data section, in column (8); (iii) the *'More Conservative Product'* sample, in column (9); (iv) the *'Extended Product'* sample, in column (10); (v) the *'Conservative Country'* sample, in column (11); and (vii) the *'Extended Country'* sample, in column (12). 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 estimating samples. However, we also see some patterns. For example, based on the results from the alternative product samples, we see that the globalization effects are smaller the more conservative the sample becomes. The explanation for this could be simply mechanical, i.e., that a smaller number of domestically produced products may leave smaller room for improvements on the external extensive margin.

4.2 Country-specific Globalization Effects

Consistent with our theoretical model, the main specification allows for differential, country-specific effects of globalization. Thus, 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. 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 the first year for which data are available. Thus, all other country-specific globalization estimates should be interpreted as deviations from the corresponding border effect for the same country in the first year for which data are available. Also, by construction, the estimates for the last year for which data are available for a given country would capture the total (cumulated) effects during the period of investigation. The results appear in Table 3, where, for brevity, we only report the estimates for the even years in our sample.

The main implications of the estimates in Table 3 are as follows. First, globalization had positive effects on almost all of the European economies during the period 1995-2014. This is supported by the fact that all estimates in the last column of Table 3 are positive, sizable, and statistically significant. The only three countries for which we do not obtain significant globalization estimates are Cyprus, Malta, and Turkey.[39] Limited data may explain this finding. Second, we see that the estimated globalization effects are very heterogeneous across the countries in our sample. To facilitate discussion, in Figure 2 we plot the cumulative globalization effects for 2014 from the last column of Table 3. We do not include the insignificant estimates for Cyprus, Malta, and Turkey, and use the 2004 estimate for Luxembourg. Figure 2 enabled us to group the countries in our sample in four categories.

The first group includes ten countries with small gains from globalization, of which: four countries (Finland, Germany, France, Luxembourg, and Sweden) are old EU members,[40] including two of the biggest EU economies (Germany and France); four are small countries from Eastern Europe (Bosnia and Herzegovina, Croatia, Northern Macedonia, and Serbia) of which only Croatia joined the EU in 2013; and, finally, Norway is not an EU member. The

[39]Due to the dropping of the odd years, Table 3 does not include the insignificant estimates for Cyprus (0.155. std.err. 0.126) and Malta (-0.139, std.err. 0.085) in 2005, which is the last year for which we have data for these countries.

[40]We treat Austria, Finland, and Sweden as old members, even though they formally joined the EU in 1995. The reason is that, due to pre-accession processes, these countries were already relatively deeply integrated within the EU by 1995. Moreover, since 1995 is the first year in our sample, there is no time-variation for these countries that is useful for our identification purposes.

second distinct group in Figure 2 includes ten countries with positive but average globalization estimates. Four of these countries are old EU members (Austria, Italy, Netherlands, and UK), four countries are new EU members (Bulgaria, Czech Republic, Estonia, and Slovenia), and two are not EU countries (Montenegro and Iceland). The countries in the third group have enjoyed significant positive effects from European integration. Two of these countries are recent EU members (Poland and Slovakia), and three of them are small and relatively poor old EU countries (Greece, Ireland, and Portugal). Finally, the group of countries that have enjoyed the largest gains from globalization and integration within Europe includes four new EU members (Hungary, Latvia, Lithuania, and Romania) and three smaller old EU members (Belgium, Denmark, and Spain).

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 EU economies, especially those that recently joined the EU, while the large old EU countries have gained relatively little. The absence of strong globalization effects on the extensive margin for the larger and more developed economies suggests that they may have reached their extensive margin potential within Europe by exhausting the big gains from trade with other large EU economies, while the gains from trade with new EU members are relatively small. The small gains for the non-EU members are self-explanatory. The large gains for the old but smaller EU members are probably due to the free access to trade with the largest European economies, which are also part of the EU. Finally, integration within the EU is the natural explanation for the large effects for the new EU members. We also note that the gains for these countries are relatively large despite the fact that the data for each of them do not start in the beginning of our sample. Overall, the analysis in this section delivers an encouraging message for the impact of European integration for development and inequality. The next section narrows the focus to the impact of *changes in* EU membership.

4.3 On the Uneven Impact of Changes in EU Membership

The sample period suits the purpose of examining the extensive margin impact of changes in membership because it covers the accession of thirteen new EU members.[41] To highlight some important aspects of our specifications and corresponding estimates, we develop the analysis sequentially, in five steps/specifications. The estimates from the first four models appear in Table 4, and each of the four panels in this table reports estimates from a single 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. All specifications include exporter-time, importer-time, and directional pair fixed effects, whose estimates are omitted for brevity. The use of directional pair fixed effects, which would capture all time-invariant asymmetries in the trade costs between the countries in our sample, is important for the current purposes because some of our specifications would allow for asymmetric (i.e., on exports vs. imports) effects of EU membership. Thus, if we do not allow for asymmetric time-invariant trade costs, our policy estimates could capture asymmetries that should not be attributed to them. The difference between the four panels in Table 4 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, based on equation (24). The difference between the results in Tables 2 and 4 is that to obtain the new estimates, we have restricted the estimating sample to only include the countries and years for which there are data on the domestic extensive margin. For example, Poland will appear in the current sample only after 2003. We made this choice to ensure that the globalization estimates that we obtain are consistent over time. Specifically, keeping the focus on Poland, we could have estimated the impact of globalization on Poland's trade with the EU also for the years prior to 2004 because we

[41] The countries that joined the EU during the period of investigation include Bulgaria (2007), the Czech Republic (2004), Cyprus (2004), Estonia (2004), Croatia (2013), Hungary (2004), Latvia (2004), Lithuania (2004), Malta (2004), Poland (2004), Romania (2007), Slovakia (2004), and Slovenia (2004). As noted before, we treat Austria, Finland, and Sweden, which all joined the EU in 1995, as old EU members.

have data on Poland's international trade for earlier years. However, these estimates would have been obtained only relative to the domestic extensive margin of the old EU members. In contrast, when data on the domestic extensive margin for Poland are also available, i.e., for the years post 2003, the corresponding border effects for Poland would be identified relative to a convex combination of the domestic extensive margins of Poland and its trading partners. As can be seen from Panel A of Table 4, however, the estimates of the effects of globalization are very similar to our main findings from Table 2.

The estimates in Panel B of Table 4 are obtained after introducing a set of border dummies for trade between the EU members in our sample, in addition to the common globalization effects. By construction, the new EU estimates should be interpreted as deviations from the corresponding common globalization effects for the same year. Of course, we could have identified the new estimates in levels, i.e., by subtracting the new EU dummies from the corresponding common border variables. However, the current specification will enable us to gauge directly whether the effects of EU membership are different from the common globalization effects in our sample. Based on the estimates for the later years in column (3) of Table 4, we conclude that, overall, EU membership contributed to stronger positive impact on the extensive margin of trade. However, the negative estimates for the early years in column (3) suggest that the EU effects were weaker than the average for our sample. A possible explanation for the change from negative to positive estimates in column (3) is the accession of new EU members. This motivates our next specification.

In Panel C of Table 4, we go a step further by isolating the impact of EU membership on the extensive margin of trade between the old EU members and the new joiners. This is captured by the estimates of the new set of border dummies in column (6), which we label $O - N$. As before, the estimates of the new globalization variables should be interpreted as deviations, but this time from the corresponding EU effects that are reported in column (5) of panel C. Two findings stand out. First, the negative and significant estimates from column (6) suggest that the impact of EU membership on the extensive margin of trade between the old and the new EU members were smaller than the average EU effects. However, based on the positive and statistically significant sum (0.207, std.err. 0.096) between the estimates from columns (5) and (6), we conclude that the effects of EU membership on the extensive margin of trade were still stronger than the average impact of globalization in our sample.[42]

The results in Panel D of Table 4 are obtained from the same specification as in Panel C, with the only difference that we now allow for asymmetric/directional extensive margin effects for the exports of the old EU members to the new EU members ($O \rightarrow N$) vs. the imports of the old EU members from the new EU members ($O \leftarrow N$). The estimates in Panel D reveal significant asymmetries between the effects on the exports vs. imports of old to new EU members. Specifically, the estimates of the effects on the extensive margin of trade from the old to the new EU members are not statistically different from the average EU effects from column (1). However, the estimates of the effects on the exports of new to old EU members are negative, statistically significant, and comparable (in absolute value) to the average EU estimates from column (8), thus suggesting no gains on the extensive margin for the exports of the new EU members. The symmetric effects from Panel C masked a significant directional heterogeneity. An interpretation is that the new EU members were not able to position their (possibly inferior) products very well in the developed West-European market, while the new EU members benefitted from the significant increase in varieties from Western Europe. We remind the reader that the negative estimates in column (11) should be interpreted as deviations from the common EU effects in column (9), which themselves are deviations from the common globalization effects in column (8).

We conclude the analysis by providing country-specific estimates for the impact of EU membership on the extensive margin of trade for the 13 new members in our sample. The estimates in Table 5 are obtained from the same specification from Panel D of Table 4 with the

[42]To see this, consider the set of cumulative estimates for 2014 at the bottom of panel B. The impact of EU membership on trade between old and new EU members can be constructed by summing the three estimates for 2014. The result is a positive and statistically significant estimate of 1.253 (std.err. 0.150), which is also statistically different from the corresponding effect of globalization on the extensive margin for non-EU members.

exception that we have replaced the common directional border dummies with corresponding country-specific indicators. Panel A of Table 5 reports our findings for the imports of the new joiners from old EU members and panel B reports the effects on their exports. The estimates in panel A are almost exclusively small and not statistically significant, so the no-effect finding for new member exports reported in Panel D of Table 4 did not mask heterogeneity.

The results in panel B of Table 5 tell a different story. As with the common effects results in Panel D of Table 4, all estimates in panel B are negative, and all estimates for 2014, except the one for Bulgaria, are also statistically significant. In contrast, these negative effects are quite heterogeneous across the countries in our sample, ranging from not statistically significant (for Bulgaria) to very large and statistically significant for Latvia (e.g., -0.610, std.err. 0.150). In combination with the decrease in the DEM indexes for the new EU members, which we described in Section 3.1 and which are in the reference group for identification purposes, the negative estimates from panel B of Table 5 imply that the *relative* decrease in the export extensive margin for the new EU members was indeed strong.

In sum, the analysis in this section demonstrates that EU membership has stimulated trade on the extensive margin of trade on average. However, the effects have been very asymmetric and in favor of the exports of old EU members. Drawing possible policy implications (motivated by the opening quotes) from our results is premature. A relative fall or even absolute decline in the number of active products or in the number of destinations served by any product may well be an efficient response of product-*cum*-destination capital investment to the deep integration with large and small West European countries offered to new members by joining the EU. In contrast, larger than average adjustment cost barriers ϕ_{in}^h may represent inefficient financial constraints. Our model is an initial foundation for further investigation of these issues.

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 35 countries over the period 1995-2014. 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.[43] 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 and re-evaluate a series of applications. 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 character-

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

istics (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.

Berthou, A. and L. Fontagne, "The Euro effects on the firm and product-level trade margins: evidence from France," *Working Paper # 2008-21, CEPII*, 2008.

Beverelli, Cosimo, Simon Neumueller, and Robert Teh, "Export Diversification Effects of the WTO Trade Facilitation Agreement," *World Development*, 2015, *76* (C), 293–310.

Beveren, Ilke Van, Andrew B. Bernard, and Hylke Vandenbussche, "Concording EU trade and production data over time," *National Bureau of Economic Research*, 2012, *Working Paper* (18604).

Bradley, Sebastien, Javier Flórez, Mario Larch, and Yoto V. Yotov, "The Granular Trade and Production Activities (GRANTPA) Database," *Manuscript*, 2022.

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.

Coe, D., A. Subramanian, N. Tamirisa, and R. Bhavnani, "The Missing Globalization Puzzle," *IMF Working Paper No. 171*, 2002.

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.

Felbermayr, Gabriel and Wilhelm Kohler, "Exploring the Intensive and Extensive Margins of World Trade," *Review of World Economics*, 2006, *142* (4), 642–674.

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, "Estimating the Effects of Non-discriminatory Trade Policies within Structural Gravity Models," *Canadian Journal of Economics*, 2021, *54* (1), 376–409.

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.

Persson, Maria, "Trade facilitation and the extensive margin," *The Journal of International Trade & Economic Development*, August 2013, *22* (5), 658–693.

Pierce, Justin R. and Peter K. Schott, "A concordance between ten-digit US Harmonized System codes and SIC/NAICS product classes and industries," *Journal of Economic and Social Measurement*, 2012, *37* (1-2), 61–96.

_ and _ , "Concording US Harmonized System Categories over time," *Journal of Official Statistics*, 2012, *28* (1), 53–68.

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

Redding, Stephen J., "Theories of Heterogeneous Firms and Trade," *Annual Review of Economics*, September 2011, *3* (1), 77–105.

Rose, Andrew K., "Do We Really Know That the WTO Increases Trade?," *American Economic Review*, 2004, *94* (1), 98–114.

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. 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.

_ , "On the role of domestic trade flows for estimating the gravity model of trade," *Contemporary Economic Policy*, 2022, *DOI: 10.1111/coep.12567.*

Table 1: The Domestic Extensive Margin, 1995-2014

Country Name	ISO	1995	1997	1999	2001	2003	2005	2007	2009	2011	2013	2014	%ΔDEM
Austria	AUT	0.70	0.70	0.68	0.68	0.68	0.67	0.61	0.59	0.58	0.59	0.59	-16.05
Belgium	BEL	0.76	0.73	0.73	0.74	0.73	0.73	0.69	0.63	0.61	0.59	0.58	-23.62
Bosnia and Herzegovina	BIH	.	.	.	.	.	.	.	.	0.26	0.27	0.28	7.96
Bulgaria	BGR	.	.	.	0.19	0.23	0.66	0.62	0.59	0.57	0.58	0.58	-10.99
Croatia	HRV	.	.	.	0.32	0.48	0.48	0.43	0.43	0.42	0.41	0.42	-12.08
Cyprus	CYP	.	.	.	.	0.03	0.02	.	.	.	.	.	-8.17
Czech Republic	CZE	.	.	.	0.81	0.82	0.80	0.74	0.76	0.76	0.76	0.75	-6.55
Denmark	DNK	0.66	0.64	0.64	0.65	0.66	0.62	0.52	0.57	0.57	0.56	0.57	-13.61
Estonia	EST	.	.	.	0.37	0.39	0.40	0.35	0.33	0.34	0.34	0.36	1.63
Finland	FIN	0.98	0.65	0.65	0.66	0.63	0.60	0.55	0.54	0.53	0.53	0.53	-19.10
France	FRA	0.92	0.92	0.90	0.89	0.88	0.87	0.85	0.89	0.90	0.91	0.91	-1.04
Germany	DEU	0.96	0.95	0.95	0.95	0.95	0.94	0.95	0.95	0.95	0.94	0.94	-2.26
Greece	GRC	0.59	0.58	0.57	0.54	0.52	0.51	0.42	0.40	0.40	0.44	0.46	-21.47
Hungary	HUN	.	.	.	0.88	0.66	0.65	0.61	0.59	0.61	0.61	0.62	-28.92
Iceland	ISL	.	0.13	0.11	0.10	0.10	0.10	0.09	0.09	0.09	0.09	0.09	-35.57
Ireland	IRL	0.55	0.57	0.58	0.57	0.54	0.53	0.48	0.45	0.43	0.42	0.41	-26.18
Italy	ITA	0.99	0.94	0.95	0.92	0.92	0.92	0.89	0.90	0.90	0.90	0.90	-9.33
Latvia	LVA	.	.	.	0.40	0.39	0.41	0.32	0.30	0.31	0.32	0.35	-10.74
Lithuania	LTU	.	.	.	0.42	0.44	0.48	0.40	0.39	0.38	0.38	0.38	-10.77
Luxembourg	LUX	.	.	0.15	0.17	0.13	0.07	.	.	.	.	.	-6.19
Macedonia	MKD	.	.	.	.	.	.	.	.	0.15	0.15	0.15	-0.63
Malta	MLT	.	.	.	.	0.03	0.03	.	.	.	.	.	2.52
Montenegro	MNE	.	.	.	.	.	.	.	.	0.03	0.02	0.02	-22.21
Netherlands	NLD	0.72	0.72	0.70	0.69	0.68	0.66	0.61	0.60	0.59	0.58	0.57	-20.75
Norway	NOR	0.37	0.36	0.34	0.34	0.32	0.33	0.35	0.34	0.33	0.33	0.32	-14.25
Poland	POL	.	.	.	.	0.83	0.86	0.85	0.84	0.83	0.83	0.82	-6.76
Portugal	PRT	0.63	0.66	0.67	0.67	0.67	0.68	0.62	0.61	0.61	0.55	0.55	-11.43
Romania	ROU	.	.	.	0.65	0.94	0.57	0.58	0.56	0.53	0.53	0.53	-6.87
Serbia	SRB	.	.	.	.	.	.	.	.	0.03	0.44	0.43	0.48
Slovakia	SVK	.	.	0.93	0.95	0.56	0.51	0.45	0.43	0.42	0.42	0.41	-23.30
Slovenia	SVN	.	.	.	0.57	0.57	0.57	0.50	0.48	0.47	0.46	0.46	-18.92
Spain	ESP	0.92	0.92	0.93	0.94	0.94	0.94	0.92	0.92	0.91	0.90	0.89	-2.68
Sweden	SWE	0.70	0.68	0.68	0.67	0.67	0.66	0.60	0.59	0.59	0.58	0.59	-16.75
Turkey	TUR	.	.	.	.	.	.	0.71	0.73	.	.	.	8.17
United Kingdom	GBR	0.93	0.93	0.89	0.95	0.92	0.91	0.88	0.87	0.86	0.85	0.85	-8.22
Total Number of Products		3,035.00	3,038.00	3,058.00	3,065.00	3,150.00	3,176.00	3,181.00	3,159.00	3,157.00	3,150.00	3,151.00	3.82

Notes: This table reports the domestic extensive margin (DEM) indexes for the countries in the PRODCOM database over the period 1995-2014. For brevity, we report data for only the odd years plus 2014. 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 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 exceptions are Bulgaria, Croatia, Finland, Luxembourg, Serbia, and Slovakia, for which we have dropped some years at the beginning or at the end of the reporting period due to suspicion of misreporting. For example, the percentage change for Bulgaria is between 2004 and 2014, i.e., we do not take into account the first three years of reported data due to the suspicious increase in DEM from 2003 and 2004. See text for further details.

Table 2: Globalization/European Integration and the Extensive Margin of Trade, 1995-2014

	A. Main Results		B. Alternative Estimators				C. Alternative Samples					
	FLEX	NO DEM	TOBIT	PPML	OLS	ASYMM	INTLRV	PROD1	PROD2	PROD3	CNTRY1	CNTRY2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$GLOB_{1996}$	0.084	.	45.067	0.022	0.025	0.111		-0.046	0.062	0.084	0.096	0.094
	(0.031)**	.	(18.018)*	(0.007)**	(0.019)	(0.042)**		(0.057)	(0.023)**	(0.031)**	(0.037)**	(0.035)**
$GLOB_{1997}$	0.137	.	86.425	0.027	0.014	0.210		0.029	0.104	0.138	0.173	0.145
	(0.032)**	.	(23.700)**	(0.015)+	(0.051)	(0.056)**		(0.065)	(0.027)**	(0.033)**	(0.046)**	(0.033)**
$GLOB_{1998}$	0.167	.	100.604	0.036	0.029	0.253		0.061	0.129	0.167	0.213	0.185
	(0.034)**	.	(26.410)**	(0.017)*	(0.050)	(0.063)**		(0.065)	(0.029)**	(0.034)**	(0.047)**	(0.038)**
$GLOB_{1999}$	0.220	.	117.887	0.049	0.071	0.348	0.138	0.105	0.174	0.220	0.249	0.247
	(0.033)**	.	(43.898)**	(0.022)*	(0.064)	(0.059)**	(0.033)**	(0.067)	(0.032)**	(0.033)**	(0.044)**	(0.039)**
$GLOB_{2000}$	0.228	.	137.283	0.038	0.009	0.385		0.081	0.196	0.228	0.282	0.258
	(0.041)**	.	(41.066)**	(0.031)	(0.096)	(0.072)**		(0.078)	(0.039)**	(0.041)**	(0.050)**	(0.046)**
$GLOB_{2001}$	0.173	.	99.524	0.005	0.009	0.311		0.169	0.172	0.173	0.238	0.224
	(0.063)**	.	(48.506)*	(0.047)	(0.110)	(0.110)**		(0.100)+	(0.059)**	(0.063)**	(0.069)**	(0.068)**
$GLOB_{2002}$	0.197	.	100.835	0.010	0.030	0.337	0.113	0.132	0.193	0.197	0.273	0.252
	(0.081)*	.	(57.295)+	(0.056)	(0.118)	(0.135)*	(0.079)	(0.099)	(0.072)**	(0.082)*	(0.092)**	(0.089)*
$GLOB_{2003}$	0.287	.	158.360	0.051	0.095	0.456		0.167	0.281	0.282	0.376	0.354
	(0.083)**	.	(59.555)**	(0.059)	(0.125)	(0.143)**		(0.096)+	(0.074)**	(0.084)**	(0.094)**	(0.094)**
$GLOB_{2004}$	0.373	.	202.141	0.083	0.131	0.572		0.220	0.352	0.368	0.454	0.441
	(0.070)**	.	(54.398)**	(0.059)	(0.125)	(0.133)**		(0.100)*	(0.065)**	(0.070)**	(0.080)**	(0.081)**
$GLOB_{2005}$	0.385	.	211.359	0.086	0.111	0.579	0.297	0.198	0.354	0.380	0.467	0.450
	(0.078)**	.	(58.831)**	(0.060)	(0.127)	(0.147)**	(0.075)**	(0.104)+	(0.070)**	(0.078)**	(0.089)**	(0.095)**
$GLOB_{2006}$	0.646	.	386.837	0.183	0.237	0.973		0.209	0.362	0.641	0.795	0.746
	(0.083)**	.	(62.109)**	(0.066)**	(0.132)+	(0.169)**		(0.107)+	(0.075)**	(0.084)**	(0.088)**	(0.100)**
$GLOB_{2007}$	0.707	.	422.066	0.206	0.273	1.059		0.270	0.419	0.702	0.856	0.813
	(0.087)**	.	(64.827)**	(0.069)**	(0.135)*	(0.174)**		(0.110)*	(0.078)**	(0.087)**	(0.093)**	(0.102)**
$GLOB_{2008}$	0.736	.	439.034	0.218	0.290	1.106	0.643	0.300	0.446	0.732	0.895	0.849
	(0.090)**	.	(65.480)**	(0.071)**	(0.136)*	(0.179)**	(0.085)**	(0.113)**	(0.081)**	(0.090)**	(0.096)**	(0.107)**
$GLOB_{2009}$	0.751	.	453.566	0.225	0.285	1.125		0.313	0.460	0.746	0.920	0.859
	(0.091)**	.	(64.854)**	(0.070)**	(0.136)*	(0.181)**		(0.116)**	(0.082)**	(0.091)**	(0.099)**	(0.110)**
$GLOB_{2010}$	0.844	.	506.663	0.265	0.334	1.252		0.402	0.548	0.839	1.041	0.961
	(0.100)**	.	(71.149)**	(0.076)**	(0.141)*	(0.193)**		(0.122)**	(0.089)**	(0.100)**	(0.109)**	(0.116)**
$GLOB_{2011}$	0.900	.	548.422	0.288	0.360	1.339	0.809	0.455	0.600	0.895	1.100	1.016
	(0.101)**	.	(71.916)**	(0.080)**	(0.146)*	(0.190)**	(0.099)**	(0.124)**	(0.091)**	(0.101)**	(0.112)**	(0.114)**
$GLOB_{2012}$	0.954	.	578.661	0.312	0.401	1.419		0.505	0.652	0.950	1.163	1.075
	(0.098)**	.	(69.571)**	(0.080)**	(0.145)**	(0.186)**		(0.122)**	(0.087)**	(0.098)**	(0.109)**	(0.110)**
$GLOB_{2013}$	0.992	.	597.443	0.329	0.439	1.471		0.541	0.686	0.987	1.208	1.113
	(0.098)**	.	(66.735)**	(0.082)**	(0.151)**	(0.185)**		(0.121)**	(0.086)**	(0.098)**	(0.109)**	(0.109)**
$GLOB_{2014}$	1.025	.	613.847	0.345	0.462	1.509	0.929	0.575	0.718	1.020	1.232	1.151
	(0.100)**	.	(68.479)**	(0.084)**	(0.152)**	(0.186)**	(0.099)**	(0.123)**	(0.089)**	(0.100)**	(0.113)**	(0.111)**
WTO	0.327	-0.533	88.297	0.448	0.029	0.281	0.654	-0.121	0.173	0.332	0.144	0.081
	(0.100)**	(0.445)	(55.271)	(0.105)**	(0.327)	(0.149)+	(0.119)**	(0.100)	(0.096)+	(0.100)**	(0.099)	(0.079)
EIA	0.052	0.044	20.514	0.004	0.035	0.081	0.067	0.085	0.095	0.051	0.139	0.122
	(0.025)*	(0.024)+	(14.215)	(0.020)	(0.035)	(0.034)*	(0.030)*	(0.025)**	(0.025)**	(0.025)*	(0.033)**	(0.024)**
Omega	0.713	0.701				1.928	0.686	0.589	0.592	0.716	1.062	1.008
	(0.063)**	(0.063)**				(0.137)**	(0.062)**	(0.076)**	(0.076)**	(0.063)**	(0.086)**	(0.055)**
Sigma			165.084									
			(3.442)**									
N	20394	19896	20394	20394	20394	20394	7176	20387	20387	20394	14508	58196
R^2					0.964							

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 estimates in column (2) are obtained from a sample that does not include observations on the domestic extensive margin. The estimator in column (3) is Tobit. The results in column (4) are obtained with PPML. Column (5) employs the OLS estimator, and the dependent variable here is the log of the number of products. Columns (6) to (11) employ the Flex estimator. In column (6) we use asymmetric pair fixed effects. The estimates in column (7) are obtained with three-year interval data. The results in columns (8)-(10) are obtained with alternative samples depending on product coverage. Moving from the most conservative sample (PROD1) to the most extended sample (PROD3). The results in columns (11) are (12) are obtained with a 'Conservative' and an 'Extended' country sample, respectively. See text for details on the alternative samples depending on product and country coverage. Standard errors are clustered by country pair and are reported in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$.

Figure 1: Globalization and the Extensive Margin of Trade, 1995-2014.

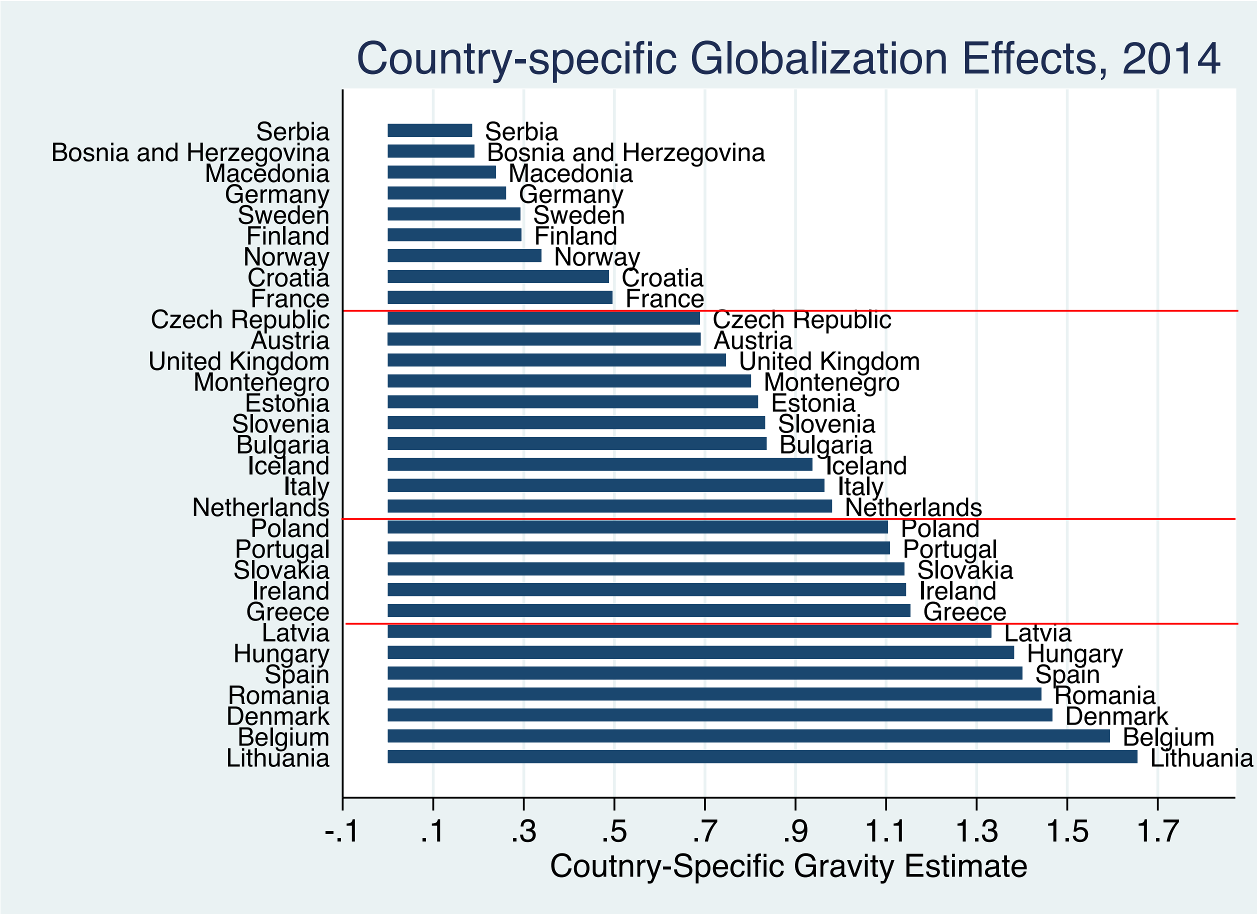
Note: This figure visualizes the estimates of the coefficients on the time-varying border/globalization variables that are obtained from our main specification in column (1) of Table 2. All estimates are relative to the baseline year – 1995. See text for further details.

Table 3: Country-specific Globalization & the Extensive Margin of Trade, 1995-2014

ISO3	1996	1998	2000	2002	2004	2006	2008	2010	2012	2014
AUT	-0.024	0.134	0.089	0.051	0.306	0.643	0.683	0.805	0.709	0.690
	(.019)	(.03)**	(.047)+	(.065)	(.059)**	(.072)**	(.073)**	(.077)**	(.076)**	(.075)**
BEL	0.119	0.218	0.128	0.238	0.598	0.863	1.281	1.323	1.461	1.594
	(.017)**	(.021)**	(.035)**	(.04)**	(.055)**	(.061)**	(.072)**	(.08)**	(.082)**	(.087)**
BGR						0.476	0.446	0.568	0.776	0.836
						(.039)**	(.063)**	(.076)**	(.084)**	(.088)**
BIH									0.095	0.190
									(.048)*	(.061)**
CYP					0.261					
					(.082)**					
CZE				-0.082	0.038	0.383	0.473	0.493	0.573	0.689
				(.017)**	(.026)	(.039)**	(.04)**	(.039)**	(.039)**	(.041)**
DEU	0.015	0.053	-0.016	-0.108	-0.058	-0.131	-0.084	0.032	0.240	0.260
	(.018)	(.026)*	(.042)	(.056)+	(.067)	(.078)+	(.075)	(.077)	(.081)**	(.083)**
DNK	0.030	0.061	0.152	0.309	0.466	0.834	0.768	1.565	1.457	1.467
	(.014)*	(.02)**	(.034)**	(.043)**	(.065)**	(.07)**	(.067)**	(.112)**	(.113)**	(.112)**
ESP	-0.036	0.115	0.177	0.470	0.373	0.624	0.705	0.770	1.108	1.401
	(.024)	(.023)**	(.04)**	(.066)**	(.067)**	(.064)**	(.068)**	(.074)**	(.077)**	(.084)**
EST				0.098	0.202	0.298	0.511	0.521	0.716	0.817
				(.065)	(.066)**	(.085)**	(.076)**	(.092)**	(.091)**	(.101)**
FIN		0.033	-0.046	-0.021	0.067	0.348	0.370	0.288	0.341	0.294
		(.018)+	(.058)	(.069)	(.067)	(.069)**	(.072)**	(.08)**	(.084)**	(.086)**
FRA	0.207	0.297	0.414	0.427	0.714	0.844	0.682	0.668	0.566	0.495
	(.021)**	(.025)**	(.042)**	(.054)**	(.065)**	(.071)**	(.074)**	(.072)**	(.071)**	(.071)**
GBR	0.067	0.003	0.042	0.026	0.028	0.318	0.421	0.571	0.698	0.746
	(.016)**	(.027)	(.052)	(.064)	(.054)	(.059)**	(.064)**	(.072)**	(.084)**	(.09)**
GRC	0.143	0.275	0.383	0.547	0.721	1.189	1.426	1.332	1.342	1.153
	(.06)*	(.066)**	(.085)**	(.096)**	(.103)**	(.108)**	(.093)**	(.095)**	(.099)**	(.101)**
HRV					0.076	0.363	0.520	0.469	0.685	0.487
					(.027)**	(.084)**	(.05)**	(.055)**	(.063)**	(.058)**
HUN				0.984	0.807	1.061	1.148	1.374	1.293	1.382
				(.04)**	(.045)**	(.07)**	(.068)**	(.073)**	(.072)**	(.075)**
IRL	0.075	0.040	0.305	0.575	0.614	0.833	0.879	0.916	1.018	1.144
	(.018)**	(.023)+	(.045)**	(.075)**	(.077)**	(.07)**	(.077)**	(.076)**	(.079)**	(.078)**
ISL		0.065	0.673	0.605	0.744	0.936	0.970	0.634	0.796	0.937
		(.031)*	(.095)**	(.09)**	(.112)**	(.1)**	(.151)**	(.138)**	(.147)**	(.146)**
ITA	0.735	0.493	0.667	0.624	0.650	0.979	0.961	0.935	0.921	0.963
	(.047)**	(.045)**	(.063)**	(.074)**	(.073)**	(.089)**	(.093)**	(.097)**	(.097)**	(.101)**
LTU				0.108	0.293	0.842	1.044	1.183	1.382	1.655
				(.076)	(.073)**	(.073)**	(.073)**	(.076)**	(.085)**	(.096)**
LUX			-0.026	0.076	0.437					
			(.023)	(.049)	(.067)**					
LVA				0.095	0.304	0.706	0.850	1.126	1.398	1.332
				(.028)**	(.064)**	(.079)**	(.088)**	(.087)**	(.092)**	(.092)**
MKD									0.082	0.237
									(.055)	(.082)**
MLT					0.025					
					(.061)					
MNE									0.468	0.801
									(.078)**	(.113)**
NLD	0.081	0.286	0.357	0.313	0.584	0.919	0.832	0.904	0.996	0.980
	(.022)**	(.034)**	(.034)**	(.041)**	(.054)**	(.067)**	(.064)**	(.069)**	(.071)**	(.067)**
NOR	0.028	0.092	0.281	0.081	0.132	-0.007	0.018	0.027	0.260	0.338
	(.025)	(.033)**	(.083)**	(.075)	(.093)	(.097)	(.1)	(.1)	(.093)**	(.094)**
POL					0.177	0.286	0.485	0.714	0.914	1.103
					(.029)**	(.05)**	(.051)**	(.055)**	(.059)**	(.064)**
PRT	0.023	0.090	0.219	0.272	0.247	0.513	0.538	0.664	0.821	1.108
	(.019)	(.023)**	(.046)**	(.074)**	(.069)**	(.068)**	(.061)**	(.062)**	(.074)**	(.077)**
ROU				-1.473	0.154	0.655	0.780	1.126	1.321	1.442
				(.082)**	(.061)*	(.075)**	(.075)**	(.086)**	(.09)**	(.093)**
SRB										0.185
										(.034)**
SVK					0.148	0.464	0.631	0.895	1.031	1.140
					(.044)**	(.062)**	(.059)**	(.062)**	(.069)**	(.066)**
SVN				-0.005	0.389	0.350	0.503	0.613	0.657	0.833
				(.021)	(.049)**	(.054)**	(.052)**	(.055)**	(.057)**	(.06)**
SWE	0.007	0.085	0.050	0.002	0.056	0.263	0.269	0.232	0.311	0.292
	(.018)	(.029)**	(.05)	(.063)	(.066)	(.073)**	(.074)**	(.078)**	(.077)**	(.078)**
TUR							0.130	-0.005		
							(.029)**	(.047)		

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 (24). 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, and we control for the impact of WTO and EIAs. For brevity, we only report the country-specific effects of interest. The three-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. Total Country-specific Effects



Notes: This figure visualizes the total country-specific estimates of the globalization effects on the extensive margin for all countries in our sample. The estimates are obtained from equation (24) and appear in the last column of Table 3. See text for discussion and further details.

Table 4: EU membership and the Extensive Margin of Trade

Year	A. Common	B. EU Membership		C. OLD-NEW EU Members			D. Asymmetric OLD-NEW EU			
		NoEU	EU	NoEU	EU	O-N	NoEU	EU	$O \to N$	$N \to O$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1996	0.0950	0.305	-0.265	0.192	-0.121		0.183	-0.110		
	(.037)*	(.109)**	(.091)**	(.093)*	(.092)		(.091)*	(.09)		
1997	0.146	0.298	-0.197	0.181	-0.0490		0.172	-0.0400		
	(.037)**	(.105)**	(.107)+	(.107)+	(.114)		(.107)	(.113)		
1998	0.179	0.306	-0.162	0.189	-0.0150		0.180	-0.00700		
	(.044)**	(.1)**	(.092)+	(.099)+	(.099)		(.098)+	(.099)		
1999	0.246	0.413	-0.206	0.296	-0.0580		0.286	-0.0500		
	(.05)**	(.104)**	(.071)**	(.09)**	(.076)		(.088)**	(.075)		
2000	0.276	0.380	-0.144	0.233	0.0440		0.224	0.0480		
	(.067)**	(.1)**	(.064)*	(.092)*	(.077)		(.091)*	(.077)		
2001	0.220	0.276	-0.0620	0.125	0.160		0.116	0.165		
	(.079)**	(.094)**	(.048)	(.088)	(.068)*		(.087)	(.067)*		
2002	0.252	0.309	-0.0630	0.161	0.170		0.151	0.174		
	(.097)**	(.115)**	(.055)	(.117)	(.08)*		(.115)	(.079)*		
2003	0.358	0.391	0.0350	0.246	0.265		0.235	0.267		
	(.104)**	(.117)**	(.044)	(.118)*	(.073)**		(.116)*	(.072)**		
2004	0.451	0.488	0.0120	0.375	0.118	-0.130	0.364	0.120	-0.0740	-0.188
	(.095)**	(.109)**	(.047)	(.099)**	(.053)*	(.038)**	(.096)**	(.051)*	(.048)	(.053)**
2005	0.456	0.525	-0.0420	0.417	0.104	-0.216	0.406	0.106	-0.111	-0.322
	(.104)**	(.127)**	(.061)	(.109)**	(.059)+	(.04)**	(.105)**	(.056)+	(.052)*	(.054)**
2006	0.796	0.856	0.00600	0.745	0.159	-0.234	0.726	0.162	-0.0830	-0.381
	(.133)**	(.165)**	(.07)	(.142)**	(.072)*	(.04)**	(.138)**	(.069)*	(.049)+	(.053)**
2007	0.867	0.844	0.127	0.730	0.268	-0.212	0.714	0.266	-0.0630	-0.359
	(.139)**	(.181)**	(.098)	(.16)**	(.103)**	(.037)**	(.156)**	(.101)**	(.049)	(.05)**
2008	0.907	0.875	0.145	0.761	0.274	-0.190	0.744	0.271	-0.0110	-0.366
	(.144)**	(.186)**	(.104)	(.164)**	(.109)*	(.038)**	(.16)**	(.107)*	(.049)	(.054)**
2009	0.926	0.867	0.182	0.753	0.315	-0.197	0.735	0.312	-0.00200	-0.388
	(.146)**	(.181)**	(.1)+	(.16)**	(.105)**	(.039)**	(.156)**	(.103)**	(.051)	(.055)**
2010	1.041	0.916	0.281	0.802	0.398	-0.172	0.785	0.393	-0.00900	-0.333
	(.156)**	(.182)**	(.103)**	(.162)**	(.109)**	(.039)**	(.158)**	(.106)**	(.052)	(.056)**
2011	1.120	1.025	0.236	0.918	0.337	-0.158	0.896	0.336	0.0410	-0.353
	(.158)**	(.178)**	(.088)**	(.158)**	(.094)**	(.039)**	(.154)**	(.091)**	(.054)	(.058)**
2012	1.188	1.111	0.209	1.003	0.299	-0.142	0.982	0.296	0.0430	-0.325
	(.159)**	(.179)**	(.09)*	(.159)**	(.095)**	(.039)**	(.155)**	(.093)**	(.056)	(.059)**
2013	1.232	1.090	0.290	0.980	0.383	-0.141	0.960	0.376	0.0470	-0.326
	(.16)**	(.173)**	(.093)**	(.153)**	(.099)**	(.037)**	(.149)**	(.097)**	(.058)	(.062)**
2014	1.271	1.157	0.249	1.046	0.337	-0.130	1.026	0.330	0.0370	-0.295
	(.161)**	(.173)**	(.096)**	(.152)**	(.101)**	(.038)**	(.149)**	(.099)**	(.06)	(.063)**

Notes: This table reports estimates of the impact of EU integration on the extensive margin of trade. The results in each panel are obtained from the same specification. The results in Panel A are obtained from our main specification from column (1) of Table 2 but after limiting the sample only to the countries and years for which we have data on the domestic extensive margin. The results in Panel B are based on the same specification but, in addition to the common border effects ($NoEU$), we have introduced a set of border dummies (EU) that capture the common impact of EU membership (EU). The estimates in Panel C further isolate the effects of EU membership between old and new members ($O - N$). Finally, the specification in Panel D allows for asymmetric/directional effects for exports from the old EU to the new EU members ($O \to N$) vs. exports from the new EU to the old EU members ($N \to O$). The dependent variable in each specification is the number of products sold from exporter i to importer j, including domestic sales. All estimates are obtained with the Flex estimator of Santos Silva et al. (2014). Each specification includes includes exporter-time, importer-time, and pair fixed effects, and we control for the impact of WTO and EIAs. For brevity, we only report the globalization estimates of interest. Standard errors are clustered by country pair and are reported in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$. See text for further details.

Table 5: EU Membership and the Extensive Margin of Trade

Year	BGR	CYP	CZE	EST	HRV	HUN	LTU	LVA	MLT	POL	ROU	SVK	SVN
A. Imports from Old to New EU Members													
2004		0.210	0	-0.0900		-0.150	-0.100	0.0300	0.100	-0.0400		-0.270	0.210
		(.11)+	(.08)	(.08)		(.09)	(.13)	(.08)	(.1)	(.08)		(.12)*	(.12)+
2005		0.430	0.0200	-0.100		-0.260	-0.120	-0.0400	0.290	-0.110		-0.410	-0.100
		(.13)**	(.09)	(.11)		(.11)*	(.12)	(.1)	(.12)*	(.1)		(.13)**	(.12)
2006			0.0400	-0.0500		-0.260	0.0300	0.0300		-0.0800		-0.270	0.0400
			(.09)	(.11)		(.12)*	(.13)	(.09)		(.1)		(.12)*	(.12)
2007	-0.0700		0.0300	-0.0500		-0.230	0.0100	0.0300		-0.110	-0.0600	-0.210	0.0100
	(.09)		(.08)	(.12)		(.13)+	(.13)	(.1)		(.1)	(.09)	(.11)+	(.13)
2008	-0.0100		0.0400	-0.0100		-0.220	0.0900	0.110		-0.0400	0.0200	-0.150	0.0600
	(.1)		(.09)	(.11)		(.13)+	(.13)	(.1)		(.1)	(.1)	(.11)	(.12)
2009	-0.0400		0.0600	-0.0600		-0.200	0.0600	0.230		-0.0100	-0.0300	-0.0700	0.0600
	(.09)		(.09)	(.11)		(.13)	(.13)	(.1)*		(.1)	(.12)	(.11)	(.12)
2010	-0.0100		0.0500	-0.110		-0.200	0.0700	0.230		-0.0800	-0.0300	-0.0800	0.0700
	(.1)		(.09)	(.12)		(.14)	(.13)	(.1)*		(.11)	(.12)	(.12)	(.12)
2011	0.0300		0.120	-0.0300		-0.220	0.170	0.280		-0.0400	-0.0200	-0.0400	0.120
	(.11)		(.1)	(.11)		(.13)+	(.13)	(.12)*		(.11)	(.13)	(.13)	(.13)
2012	0.0500		0.0800	0		-0.260	0.180	0.330		-0.0300	0.0600	-0.0800	0.0600
	(.12)		(.1)	(.11)		(.14)+	(.12)	(.11)**		(.11)	(.13)	(.13)	(.13)
2013	0.0600		0.0300	0.0100	-0.0100	-0.200	0.180	0.290		0.0100	0.0700	-0.0400	0.0400
	(.12)		(.1)	(.11)	(.06)	(.14)	(.13)	(.12)*		(.11)	(.13)	(.13)	(.14)
2014	0.100		0.0700	0.0100	-0.110	-0.240	0.250	0.280		-0.0300	0.0400	-0.0800	0.0500
	(.12)		(.1)	(.11)	(.09)	(.14)+	(.13)+	(.12)*		(.11)	(.14)	(.14)	(.13)
B. Exports from New to Old EU Members													
2004		-0.550	-0.250	-0.250		-0.410	-0.400	-0.500	-0.180	-0.370		-0.280	-0.0500
		(.23)*	(.08)**	(.1)**		(.1)**	(.11)**	(.13)**	(.15)	(.08)**		(.09)**	(.08)
2005		-0.600	-0.360	-0.360		-0.560	-0.570	-0.520	-0.0900	-0.570		-0.380	-0.350
		(.29)*	(.08)**	(.12)**		(.1)**	(.11)**	(.13)**	(.23)	(.09)**		(.1)**	(.1)**
2006			-0.420	-0.450		-0.670	-0.640	-0.560		-0.610		-0.360	-0.390
			(.08)**	(.13)**		(.11)**	(.13)**	(.15)**		(.1)**		(.11)**	(.1)**
2007	-0.210		-0.420	-0.400		-0.640	-0.610	-0.630		-0.530	-0.230	-0.410	-0.320
	(.11)*		(.08)**	(.12)**		(.11)**	(.13)**	(.15)**		(.11)**	(.08)**	(.11)**	(.1)**
2008	-0.270		-0.410	-0.420		-0.590	-0.590	-0.650		-0.520	-0.290	-0.400	-0.340
	(.12)*		(.09)**	(.13)**		(.11)**	(.12)**	(.15)**		(.1)**	(.09)**	(.12)**	(.1)**
2009	-0.360		-0.400	-0.510		-0.590	-0.600	-0.800		-0.500	-0.280	-0.370	-0.350
	(.13)**		(.09)**	(.15)**		(.12)**	(.13)**	(.14)**		(.1)**	(.1)**	(.12)**	(.1)**
2010	-0.180		-0.400	-0.430		-0.550	-0.540	-0.670		-0.480	-0.160	-0.390	-0.370
	(.14)		(.09)**	(.15)**		(.13)**	(.13)**	(.15)**		(.11)**	(.12)	(.12)**	(.11)**
2011	-0.180		-0.430	-0.400		-0.540	-0.600	-0.700		-0.480	-0.210	-0.400	-0.390
	(.14)		(.09)**	(.14)**		(.13)**	(.13)**	(.15)**		(.11)**	(.13)+	(.13)**	(.11)**
2012	-0.190		-0.410	-0.430		-0.490	-0.550	-0.590		-0.460	-0.250	-0.310	-0.360
	(.15)		(.09)**	(.16)**		(.12)**	(.13)**	(.15)**		(.12)**	(.13)+	(.13)*	(.1)**
2013	-0.210		-0.370	-0.480	-0.180	-0.470	-0.580	-0.680		-0.430	-0.290	-0.350	-0.320
	(.16)		(.1)**	(.17)**	(.09)*	(.12)**	(.13)**	(.15)**		(.12)**	(.12)*	(.13)**	(.1)**
2014	-0.170		-0.330	-0.380	-0.440	-0.370	-0.560	-0.610		-0.390	-0.220	-0.300	-0.260
	(.16)		(.09)**	(.17)*	(.1)**	(.13)**	(.14)**	(.15)**		(.12)**	(.13)+	(.13)*	(.1)**

Notes: This table reports country-specific estimates of the impact of European integration on the extensive margin of trade for the thirteen EU joiners in our sample. Panel A reports the impact on EU exports to the new members, and Panel B reports the impact of EU imports from the new members. All estimates are obtained from the same specification. The dependent variable is the number of products sold from exporter i to importer j, including domestic sales. The estimates are obtained with the Flex estimator of Santos Silva et al. (2014). The specification includes includes exporter-time, importer-time, pair fixed effects, and common globalization effects, and we control for the impact of WTO and EIAs. For brevity, we only report the country-specific estimates of interest. Standard errors are clustered by country pair and are reported in parentheses. + $p < 0.10$, * $p < .05$, ** $p < .01$. See text for further details.