Export Entry and Network Interactions

Evidence from the Belgian Production Network

[*]Emmanuel Dhyne[◊], Philipp Ludwig[△] and Hylke Vandenbussche[§]

August 2022

Abstract

Low export participation of firms across countries is typically related to high entry costs allowing only the most productive firms to serve foreign markets. In this paper, we move beyond individual firm characteristics to explain export participation and investigate whether firms' domestic network linkages can facilitate export entry. Firms receive information from business interactions with experienced exporters which lowers sunk entry costs and allows them to enter the foreign market. Using rich data of buyer-seller linkages in the Belgian production network, we find that network heterogeneity is a key determinant of the extensive margin of trade. Each additional export signal received via network linkages increases the entry probability by 0.4 – 1.5 percentage points, giving firms with suitable networks a key advantage in accessing foreign markets. The marginal impact of network effects decreases in network size which we attribute to negative assortative matching in the underlying network formation process.

Keywords: Export entry, buyer-seller network, information frictions, trade barriers, heterogeneous firms

JEL Classification: F12, F13, L14, D83

[*]We thank Pol Antras, Paola Conconi, Ronald Davies, Xiang Ding, Sharat Ganapati, Keith Head, Beata Javorcik, Amit Khandelwal, Isabelle Méjean, Ralph Ossa, Christian Proebsting, Gonzague Vannoorenberghe, Jaume Ventura and participants of the CES-LCIS-UCD 2021 PhD seminar, the Leuven-Louvain trade workshop, NBB internal seminar, 21^{st} RIEF Doctoral Meeting, EOS conference, SMYE 2022, 4^{th} QMUL PhD Workshop and WIEN Workshop 2022 for valuable comments and discussions. Philipp Ludwig gratefully acknowledges funding from the Research Foundation Flanders (FWO) and thanks the National Bank of Belgium for their hospitality during a research stay where this paper developed.

[◊]National Bank of Belgium and University of Mons. Emmanuel.dhyne@nbb.be

[△]Department of Economics, KU Leuven. philipp.ludwig@kuleuven.be

[§]Department of Economics, KU Leuven. hylke.vandenbussche@kuleuven.be

1 Introduction

Despite continued trade liberalization export participation remains an exclusive phenomenon with only 4-5% of firms directly engaged in cross-border trade (Bernard et al., 2007; Dhyne et al., 2015). This concentration of economic activity at the extensive margin of trade has far-reaching consequences. Low export participation not only weakens competition in domestic markets by allowing a small number of exporters to consolidate market power (De Loecker and Warzynski, 2012) but also severely restricts aggregate export growth which crucially depends on the entry of new firms (Eaton et al., 2009a). The key question is therefore to understand why only a handful of firms can overcome entry barriers and access foreign markets.

This paper aims to contribute to this debate by investigating which factors ultimately determine a firm's entry decision. Traditionally, low export participation has been related to the presence of sunk entry costs allowing only a select number of firms to enter and remain profitable in a foreign market. This insight sparked an influential literature of heterogeneous firm trade models which directly link export entry decisions to individual firm characteristics. Prominent attributes determining export participation are firm productivity (Melitz, 2003), access to finance (Manova, 2013) or previous experience in similar markets (Morales et al., 2019). What all these examples share is that they relate export participation to firm-level characteristics. Firms however do not operate in isolation but constantly interact with other firms in their production network. These business transactions represent an potentially important determinant of export participation because they expose firms to the export experience of network peers which might lower entry barriers to foreign markets. In this setting, each firm's individual network serves as a conduit of export-related information creating additional heterogeneity in entry behavior beyond firm-level characteristics.

Our paper allows for this network-dimension of firm heterogeneity and empirically investigates whether export participation is facilitated by buyer-seller interactions in domestic production networks. Using detailed data from the universe of Belgian firms, we capture each firm's entire domestic production network and export behavior for the years 2002 − 2014. To trace the diffusion of export-related information along buyer-seller linkages, we combine export and network data to detect changes in firms' export experience. Each time a firm starts to export to a new destination, it gathers entry-related information which diffuses through business interactions to the firm's immediate production network. At the receiving end, these export signals reduce market access costs and increase the probability of the receiver to start exporting to the same foreign market in the following year. Our approach therefore exploits changes in the export behavior of network peers to detect export market signals and investigates whether these signals increase the entry probability of connected firms.

To formalize this idea, we introduce network interactions to a stylized model of export entry in which firms are heterogeneous along both a firm-dimension, capturing individual characteristics like productivity, and a network-dimension, capturing their linkages to other firms. Export decisions are based on a simple trade-off between the costs and benefits of entering a foreign market. The key novelty is to express sunk entry costs as a function of network linkages using functional form assumptions inspired by the literature of social networks (Bramoullé et al., 2009) and spatial economics (Anselin et al., 2008). As a consequence, differences in network linkages directly translate into different market access costs creating additional variation in entry decisions beyond firm productivity. To achieve this, we treat network effects as an externalitiy which implies that network formation is not strategic and conditionally exogenous. This assumption is relaxed at a later stage. While restrictive, the stylized setting allows us to study both dimensions of firm heterogeneity at the same time and empirically investigate the relevance of network linkages for export participation.

The resulting estimation equation takes the form of a time-space recursive model which relates export entry decisions to individual firm characteristics and incoming export signals. Since not every network interaction contributes to the diffusion of valuable export signals we make two important adjustments. First, we distinguish between matching and non-matching signals based on whether the signal origin coincides with the chosen export destination of the receiver. Only matching signals are expected to lower market access costs and facilitate entry, meaning an export start to China should not benefit from signals originating in the US market. Second, we consider two separate model specifications to account for the fact that network interactions expose firms to both valuable export signals and unrelated network noise. While a signal intensity specification ignores any negative impact of network noise, a signal clarity specification discounts the value of incoming export signals by the amount of noise in the network. In both cases, a precise estimation of network effects relies on a combination of high-dimensional fixed effects and network instruments to ensure that any correlation in firm responses is caused by network interactions and not driven by unobserved common shocks. While network linkages are initially considered exogenous, we account for endogenous network formation in an extension by introducing a network selection model as in Arduini et al. (2015) and Qu et al. (2017) to our benchmark setting. This allows us to control for any selection bias in estimated network effects from correlated linkage and entry decisions. Using detailed balance sheet, trade transaction and network data from the universe of Belgian firms, we then estimate our augmented entry model for the years 2004-2014.

Our results indicate that firm networks are an important determinant of export participation decisions of Belgian firms. Each additional matching signal increases the entry probability between 0.4–1.5 percentage points while non-matching signals do not have any effect. As firms can receive multiple signals in any given period, the economic impact of this channel can be considerable which emphasizes that network heterogeneity needs to be considered alongside firm productivity to understand firm behavior at the extensive margin of trade. At the same time, network effects show distinct differences to traditional sources of firm heterogeneity. We find that the marginal effect of network externalities decreases significantly in network size, to the extent that export signals have a 3-4 times larger impact on small compared to large firms. We show that these stark differences can be traced back to negative assortative matching in the underlying network formation process. Large firms face a disproportionate amount of network noise which systematically attenuates the beneficial impact of export signals. This size penalty means that network effects are unlikely to contribute to the observed concentration at the extensive margin of trade. Instead, it suggests that network effects might be a particularly important tool for policy makers to connect small firms to foreign markets.

While our findings emphasize that network interactions are an important factor in understanding firm behavior at the extensive margin of trade, networks and entry decisions have typically been considered as separate phenomena in the international trade literature. Studies investigating the determinants of export participation mostly focus on firm-level characteristics. Traditionally, firm productivity and firm size have received most attention to explain the sorting of firms into exporters and non-exporters in accordance with the canonical heterogeneous firm trade model of Melitz (2003). Recent contributions also highlight the role of experience generated from a firm's past export activity. This helps to account for correlated export patterns across time (Albornoz et al., 2012) and space (Morales et al., 2019) as entering foreign markets occurs more frequently if firms are already active in the region. Similarly, Arkolakis et al. (2021) find that firm experience can occur in form of scope economies at the product level as market-access costs fall with the number of products the firm already exports to the destination. We directly contribute to this literature by highlighting that expertise generated from past export activity not only benefits firms' own export expansion but also diffuses through network linkages and thereby becomes an important determinant for export participation of others. Other insights from previous studies are directly mirrored in our estimated network effects. Non-matching export signals for example still have a mildly postive impact on export entry if they originate in markets with close geographic proximity to the actual entry destination.

A separate literature in international trade has started to formally integrate buyer-seller linkages into existing trade models to provide a theoretical foundation for observed firm behavior at the extensive margin of trade. Chaney (2014) explains spatially correlated entry patterns of firms with a search mechanism that allows firms to use their existing contacts in foreign markets to expand to nearby destinations. A further export expansion of the firm is therefore facilitated by existing cross-border linkages. Eaton et al. (2016) and Bernard et al. (2018) look at the microfoundation of the extensive margin of trade by directly modeling the formation of international buyer-seller linkages. In contrast to these papers, we study domestic buyer-seller linkages and allow them to facilitate the creation of cross-border linkages in form of export starts. As export starts are only captured at the firm-destination level, we abstract from the identity of the importing firm in the foreign market. Nevertheless, we contribute to the literature on international buyer-seller linkages in an important way. By highlighting that diffusion in domestic networks facilitates export entry, we show that domestic networks actively contribute to the formation of international networks.

Closest to this paper is a third strand of empirical literature which directly relates export decisions to activity in domestic networks. A clear point of distinction among papers in this group is the way authors try to capture domestic firm-to-firm interactions. A common approach in the export spillover literature (Fernandes and Tang, 2014; Koenig, 2009) is to extrapolate unobserved network linkages from spatial proximity which implicitly assumes that interactions are limited to the specified geographic unit and occur between all firms within it. Our production network approach in contrast is based on observed buyer-seller transactions which mitigates concerns of network misspecification common to geographic proxies. Recent alternatives have been presented by Patault and Lenoir (2021) who study if hiring employees from exporters increases the likelihood to start serving the same foreign clients and by Connell et al. (2019) who explore if indirect exporting through wholesalers facilitates subsequent export entry. While human capital movements and learning through wholesalers are important mechanisms to access export-related information, they are far more restrictive than our approach which allows information to diffuse in absence of labor movements and through any business interaction.

Finally, this paper is related to a vast literature on peer effects in networks (Advani and Malde, 2018; Bramoullé et al., 2020). Depending on the setting, the relationship between agent and network behavior is either characterized by homophily, meaning agents seek to conform to average network behavior or strategic complementarity, where agent and network behavior is mutually reinforcing. Introducing these ideas to study export participation leads to two important methodological contributions. First, we show that in an information diffusion setting, choosing between strategic complementarity and homophily translates to ignoring or accounting for network noise. Second, this distinction is particularly important in production networks where negative assortative matching exposes firms to a disproportionate amount of network noise. This emphasizes that researchers need to be mindful of the underlying network formation process when deciding how to model agent behavior.

This paper has 7 sections. Section 2 presents a stylized model of export entry, shows how we introduce network interactions and discusses strengths and weaknesses of the approach. Section 3 describes our data sources and sample selection. Section 4 discusses identification and estimation of our augmented entry model. Our results are presented in section 5 and discussed in section 6. The last section concludes.

2 Theoretical framework

Our goal is to assess whether foreign market access not only depends on each firm's individual characteristics but can also be facilitated by information externalities originating from a firm's domestic production network. We therefore need a theoretical framework which relates export entry decisions to two dimensions of firm heterogeneity: firm productivity and network interactions. While the former is a standard component in most trade models since the seminal contribution of Melitz (2003), heterogeneity in firm networks only recently attracted attention in the trade literature[1] and has commonly not been considered as a direct determinant of export participation[2].

We therefore proceed in two steps. First, we use the model of Koenig (2009) to illustrate the key determinants of firm export entry behavior in a standard setting which only allows firms to differ in productivity. More productive firms can expect higher profits in foreign markets and are therefore more likely to enter. In a second step, we then introduce network interactions to the model by allowing entry barriers in foreign markets to change in response to export-related information received from network peers. In this setting, firms that receive more information from their network are more likely to enter because a part of the informational cost burden related to entering the foreign market is offset by the network. We outline the limitations and strengths of this novel modeling approach and derive our estimation equation. An extended discussion of model identification is deferred to section 4.1.

[1] For an overview of the role of networks to trade see Chaney (2016).

[2] Notable exceptions are Connell et al. (2019) who relate export entry decisions to interactions with wholesalers and Patault and Lenoir (2021) who use labor movements between firms to explain subsequent export decisions of the hiring firm.

2.1 Stylized entry framework

We follow the framework of Koenig (2009) to study the decision of firm i to enter foreign market d. In this stylized setting, firms have absolute certainty about their expected profits $\Pi_{id,t}$ but face sunk entry costs f_d whenever entering a foreign market. A firm will start to export if the present value of profits (assuming constant discount factor r) exceeds the costs of entry. The probability to enter market d is thus

$$Pr(y_{id,t} = 1) = Pr\left(\frac{\Pi_{id,t}}{r} > f_d\right) \tag{1}$$

Firms therefore face a trade-off between costs and benefits of exporting to the foreign market and their ultimate decision will rest on the relative strength of both elements. Suppressing time subscript t, firm profits in market d can be described as

$$\Pi_{id} = p_{id}q_{id} - a_i w_i q_{id}$$

The first term on the right-hand side represents firm sales as the product of price p_{id} and demand q_{id} in the foreign market whereas the second describes the production costs which per unit of demanded quantity q_{id} amount to $w_i a_i$, the product of nominal wages w_i and inverse productivity a_i[3].

The model relies on a canonical setting where single product firms operate under monopolistic competition and consumers have CES utility which means the demand for firm i's products in market d is given by $q_{id} = p_{id}^{-\sigma} P_d^{\sigma-1} \mu_d E_d$ where $P_d^{\sigma-1} = \left[\int_l p_{ld}^{1-\sigma} dl\right]^{\frac{1}{1-\sigma}}$ represents the price index in market d, σ is the elasticity of substitution, μ_d is the expenditure share devoted to the representative industry and E_d denotes the level of income in d. The optimal mill price charged by firm i in this setting is $p_i = \frac{\sigma}{\sigma-1} a_i w_i$ as a constant markup over marginal costs $a_i w_i$. The final price faced by foreign consumers is $p_{id} = p_i \tau_d$ where τ_d represents ad-valorem iceberg-type trade costs related to shipping goods to market d. The profit of firm i in foreign market d is therefore

$$\Pi_{id} = \left[\frac{a_i w_i \tau_d}{(\sigma-1)P_d^{1-\sigma}}\right]^{1-\sigma} \mu_d E_d \tag{2}$$

Plugging Π_{id} into 1 and assuming (for now) that entry barriers f_d are common to all prospective entrants we can express a firm's entry decision as

$$Pr(y_{id,t} = 1) = Pr\left(\left[\frac{a_{i,t} w_{i,t} \tau_{d,t}}{(\sigma-1)P_{d,t}^{1-\sigma}}\right]^{1-\sigma} \frac{\mu_{d,t} E_{d,t}}{r} - f_d > 0\right) \tag{3}$$

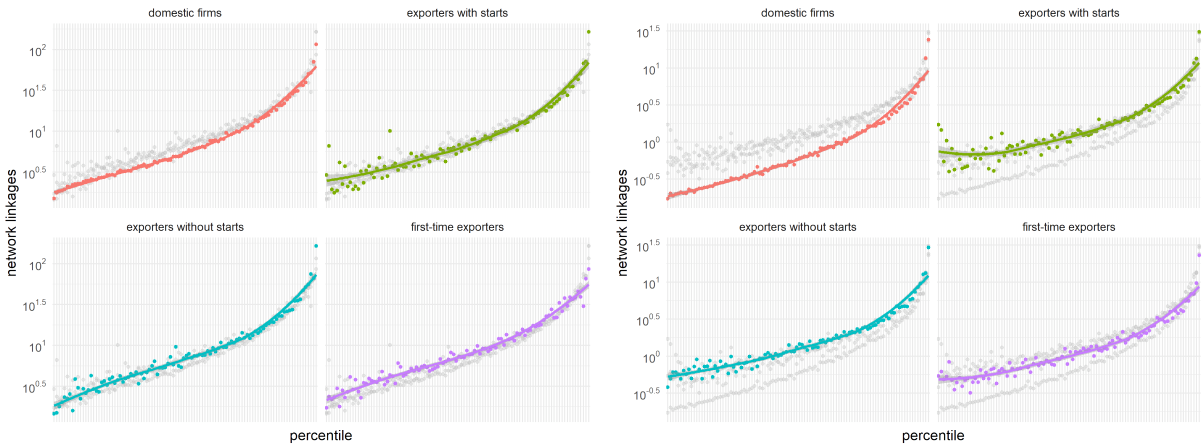
There are several things worth noting at this point. First, equation 3 illustrates that in the canonical setting all sources of firm heterogeneity are generated from firm-level characteristics, namely wages $w_{i,t}$ and (inverse) productivity $a_{i,t}$. This is emblematic of Melitz (2003) type trade models in which firm productivity is the key determinant in explaining the sorting of firms into exporters and non-exporters. Similarly, the Koenig (2009) model predicts that firms with higher productivity (lower $a_{i,t}$) are more likely to start exporting in presence of common entry costs f_d. Second, it highlights the role of gravity for the extensive margin of trade. Destinations with

[3] The units of labor needed to produce one unit of q_i.

higher income levels $E_{d,t}$ or lower trade costs $\tau_{d,t}$ will attract more exporters conditional on firm characteristics.

Our goal is to extend this framework by a new source of firm heterogeneity that accounts for the fact that each individual firm has a unique network which may serve as a conduit for export-related information. Before formally introducing networks to the model in the next section, we provide some descriptive evidence to emphasize that network heterogeneity is not simply a primitive of firm productivity but generates additional variation that might be directly related to a firm's export participation.

Figure 1: Seller linkages by TFP percentile



(a) all linkages

(b) linkages to exporters

This figure shows shows the average number of buyers for a seller in a given productivity percentile. Seller productivity is computed using the approach of Levinsohn and Petrin (2003). Sellers are separated into four types: Non-exporters (red), exporters with export starts (green), exporters without starts (blue) and first-time exporters (purple). Figure 1a plots the average linkages to any buyer, while figure 1b plots the average linkages to buyers that export. The figure uses production network data of Belgian firms explained in detail in section 3.1.

Figure 1 relies on Belgian firm-level data explained in detail in section 3.1 and plots the average number of domestic buyer-seller linkages for sellers in a given productivity percentile. Both consist of four panels, each plotting the number of linkages for one seller type in color and the corresponding linkages of the other three seller types in grey. The only difference between the two graphs is the type of network linkages plotted on the y-axis. Figure 1a considers linkages to all buyers while figure 1b only considers the subset of linkages which involve exporting buyers.

Focusing on figure 1a, we observe two patterns. Firstly, across seller types, the total number of linkages seems to increase in seller productivity. This pattern is common to production networks (Zi and Bernard, 2021). High productivity sellers are likely to attract more buyers because they can charge lower prices than their competitors. Secondly, when comparing colored and grey dots in each panel, we see that conditional on seller productivity network sizes hardly differ across seller types. Comparing the network size of domestic (red) and exporting sellers (green, blue and purple) across productivity percentiles, we see that the network size premium of exporting sellers amounts to 28-37%. This indicates that exporting sellers overall seem to have more network interactions than domestic sellers even after controlling for firm productivity.

This difference increases dramatically in figure 1b where we exclusively focus on seller linkages to exporting buyers. Network interactions with exporters are potentially important for a firm's own export entry decision because they capture to what degree a firm is exposed to the export experience of other firms . What we see in the graph is that domestic sellers interact with fewer exporting buyers across productivity percentiles and that the network size premium of exporting sellers increases to 75-144%. This significant wedge in network interactions between domestic and exporting sellers cannot be explained by seller productivity (which we condition on) or total network size (as seen in figure 1a) and holds for incumbent and first-time exporting sellers alike. This strong correlation between seller's export status and their access to export experience through network linkages has important implications. First, it highlights that network interactions represent a new dimension of firm heterogeneity and are not simply a primitive of firm productivity. Exporting and non-exporting sellers with equal productivity interact with a different number of firms and this difference is especially pronounced when considering interactions with exporting buyers. Secondly, it introduces the possibility that networks are causing sellers to become exporters in the first place. To investigate this channel, we will formally introduce network interactions to the standard export entry framework outlined above and discuss how this additional dimension of firm heterogeneity can affect firm's entry behavior.

2.2 Augmented entry framework with network interactions

We treat each firm i's domestic production network as a pool of export-related information the firm can draw from. Each time a firm j in the network starts to export to a new destination d for the first time, it needs to pay sunk entry cost f_d which includes expenses to assess local demand preferences, search costs to identify foreign retailers and administrative costs related to obtaining mandatory certification for imported products or other non-tariff barriers. Each of these cost components involve a non-negligible share of information which the firm gathers upon entry and given sufficient interaction strength diffuses along network linkages. We define the diffusion triggered by each individual export start of firm j as an export signal y_{jd} which contains entry-related information for one particular market d. Information is assumed to diffuse immediately[4] upon entry of firm j. We formally model the impact of network interactions as an externality which enters firm i's entry equation 1, expressed in logs

$$Pr(y_{id,t} = 1) = Pr(\ln \Pi_{id,t} > \ln r + \ln f_d) \tag{4}$$

via entry cost term $\ln f_d$. This is achieved by making a functional form assumption which is inspired by the literature on social networks (Bramoullé et al., 2009; Calvó-Armengol et al., 2009) and explicitly expresses entry costs as a function of firm i's individual network

$$\ln f_{id,t} = \alpha_d - \delta \sum_j \bar{s}_{ij,t} x_{j,t} - \beta \sum_j s_{ij,t} y_{jd,t} - \psi_{S_{i,t}} + \varepsilon_{id,t} \tag{5}$$

where α_d denote sunk entry costs incurred in market d, $x_{j,t}$ represent time-varying peer characteristics, $y_{jd,t}$ are export start decisions of network peers, $s_{ij,t}$ ($\bar{s}_{ij,t}$) are elements of a (row-normalized) binary interaction matrix S_t[5] which captures all domestic firm-to-firm interactions in the economy in year t, $\psi_{S_{i,t}}$ denote common shocks to all firms in firm i's network S and $\varepsilon_{id,t}$

[4] Our empirical specification relies on annual firm-to-firm interactions. Immediate diffusion in this context means signals reach a connected firm i within the same calendar year as the export start of firm j.

[5] We will describe interaction matrix S_t in more detail in section 3.1.

is an idiosyncratic error term.
In this setting, network interactions can affect firm i's entry decision in two distinct ways. First, we allow entry costs f_{id} to directly respond to the average characteristics x_j of firms in the network[6]. This channel controls for network effects unrelated to the diffusion of export signals such as productivity spillovers. Second, we allow entry costs f_{id} to directly respond to the total number of export signals y_{jd} received from the network. Each incoming export signal is assumed to lower entry barriers of firm i to destination d because the obtained information offsets a part of the informational cost burden otherwise incurred during export entry. An important distinction between both network externalities is the observational unit they operate at. Network characteristics δ vary at the firm level and can therefore only impact entry costs across export destinations (f_i). Network signals β however vary at the firm-destination level which means signals only affect entry costs of the market they originate from (f_{id}). Plugging equation 5 and 2 into equation 4 then yields our final entry equation

$$Pr(y_{id,t}=1) = Pr\left(\boldsymbol{x_{id,t}}\gamma + \delta\sum_j \bar{s}_{ij,t}x_{j,t} + \beta\sum_j s_{ij,t}y_{jd,t} + \psi_{S_{i,t}} - \alpha_d - \varepsilon_{id,t} > 0\right) \tag{6}$$

where $\boldsymbol{x_{id,t}}$ captures firm and destination variables[7] related to firm profits in foreign market d. This equation closely resembles models from the peer effects (Manski, 1993; Bramoullé et al., 2009) and spatial economics (Anselin et al., 2008; Qu and Lee, 2015) literature where it is commonly referred to as a spatial autoregressive (SAR) model with panel data. Both strands guide our identification strategy which we defer to section 4.1. For the trade literature this explicit introduction of network interactions to a baseline model of export entry represents a novel modeling approach which warrants a further discussion regarding its strengths and weaknesses.

2.3 Concept limitations

The augmented entry framework described above introduces a new approach to allow network-transmitted information to affect export participation decisions of firms. Both the design of information diffusion and the choice to model network effects as an externality to market access costs carry important implications which will be discussed in turn.
Our diffusion mechanism relies on variation in export activity of network peers. An export entry decision of a firm in the network triggers the transmission of export signals which immediately propagate to connected firms. The type of diffusion mechanism assumed here is therefore based on changes in network experience rather than the stock of accumulated export information available in the network. Our choice to focus on changes rather than accumulated export information has several advantages. First, it offers a clear timing for the diffusion mechanism by linking both the creation and propagation of export information to the moment of export starts. In a model where information is accumulated over time the exact moment of diffusion is unclear. Secondly, limiting attention to new entry decisions ensures that the transmitted information remains relevant for

[6] We follow the standard convention of the social network literature and assume sunk entry costs respond to average rather than aggregate characteristics in firm i's network. This is achieved by row-normalizing entries in interaction matrix S_t such that $\sum_j \bar{s}_{ij} = 1$. As we are assuming all $s_{ij} \in \{0,1\}$ in our empirical setting, $\sum_j \bar{s}_{ij}x_j$ simply represents an unweighted average of characteristics of all firms j in the network of firm i.

[7] $\boldsymbol{x_{id,t}} = (1-\sigma)(\ln a_{i,t} + \ln w_{i,t} + \ln \tau_{d,t} - (\sigma-1)\ln P_{d,t}^{1-\sigma}) + \ln \mu_{d,t} + \ln E_{d,t}$.

firms at the receiving end. Entry information accumulated in previous periods would require additional assumptions regarding information depreciation rates to differentiate between the value of old and recent experiences. Our approach avoids this issue[8] by assuming that relevant information diffuses quickly along network linkages which is reasonable in our empirical setting[9]. A drawback of our baseline approach is that it treats incoming export signals as uniform and therefore abstracts from weighting them based on linkage or peer characteristics. We refrain from *exact* weighting in our model in lack of a theory-consistent weighting scheme but do explore signal heterogeneity empirically in section 5.3 by studying the impact of different signal *groups* which are based on peer and linkage characteristics. Another important limitation of our diffusion approach is that it does not grant insights into the relevance of individual information cost components. Entry-related information costs can be related to the assessment of demand preferences of foreign consumers, the search for local distribution partners or compliance with import requirements and product standards. Our model is not equipped to investigate the relative importance of each of these cost components which likely differ across products and destinations and instead focuses on their combined impact on export participation.

A second important aspect of our augmented framework concerns the way network interactions are allowed to influence firm entry decisions. Choosing to introduce network effects in form of an externality to market access costs f implies that accessing export information is not a first-order concern in the the underlying formation process and network interactions are therefore not strategic. Instead, any benefit in form of export signals is assumed to be unexpected and not allowed to alter the profit optimization in equation 2. Firms do not account for network effects when setting optimal prices but can benefit from network interactions in form of lower entry costs f. Our framework therefore abstracts from models with network games (König et al., 2019) where optimal firm and network behavior is interdependent due to the presence of strategic complementarities or models with strategic network formation (Badev, 2021; Hsieh et al., 2020) where firms anticipate network effects when choosing which agents to interact with. One concern related to network formation however remains. While firms may not choose business partners strategically with the intent to acquire information about foreign markets, linkages could still be endogenous due to unobserved shocks which simultaneously affect network formation and export entry. We discuss the identification challenges arising from this form of network endogeneity in section 4.3.

2.4 Concept advantages

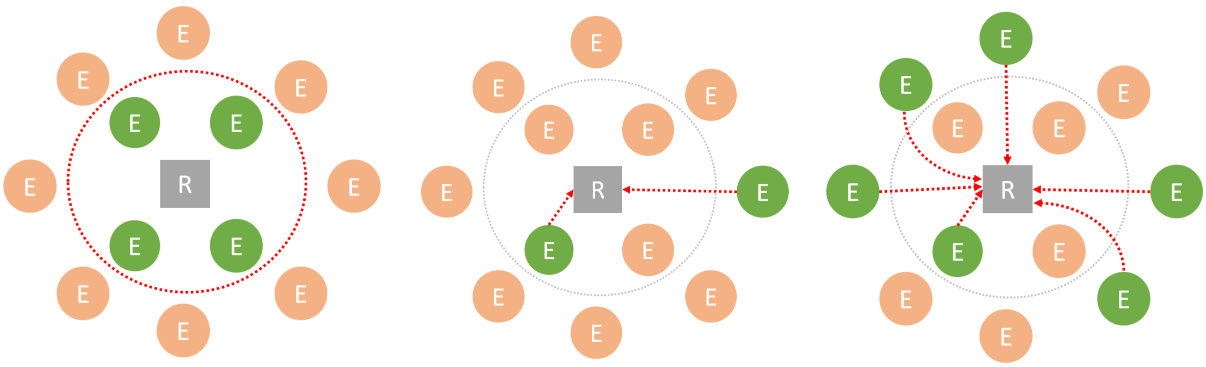
Taken together, the diffusion mechanism and characterization of network effects as externalities augment the standard entry framework of Koenig (2009) by a new dimension of firm heterogeneity which offers several key advantages.

First, linking export participation to two distinct dimensions of firm heterogeneity offers new explanations for observed patterns at the extensive margin of trade. An important example is

[8]Our approach does not allow for the value of information to depreciate before emission. Information depreciation however can still play a role after emission, if the receiving firm takes time to act upon the information received from its network.

[9]We capture network interactions and export starts at an annual level. This means our model assumes that export signals reach firms in the same year as they are emitted.

the existence of small exporters across export destinations documented by Eaton et al. (2011). A Melitz (2003) model with destination-specific entry costs f_d, constant marginal costs and productivity as the only dimension of firm heterogeneity typically struggles to explain the presence of small exporters if their productivity falls below model implied minimum thresholds required for export participation. Arkolakis (2010) addresses this shortcoming by presenting a model in which firms need to pay marketing costs to reach foreign buyers. Market access costs increase for each additional buyer the firm wants to reach, which explains the presence of small exporters since reaching a few customers remains relatively inexpensive. Firm-specific network externalities studied in this paper offer an alternative explanation without assuming increasing market access costs. Firms at the margin that just fall below the minimum productivity threshold might still choose to export if they receive sufficient information from their network to bring the cost of entry below expected profits[10]. Allowing for both types of firm heterogeneity therefore captures additional variation which might be especially important to understand export participation decisions of small exporters.



(a) geographic proxy (b) labor movements (c) production network

Figure 2: Comparison of network approaches

Second, the augmented framework describes between-firm effects in a true network context which adds a substantial amount of empirical precision compared to the preceding export spillover literature. Prior studies like Koenig (2009) or Fernandes and Tang (2014) typically relate firms' export participation to the degree of export activity within the same geographic unit. This type of geographic proxy is depicted in figure 2a and suffers from several important shortcomings as it implicitly assumes that all firms within the same geographic unit interact and that no interactions occur across units. These requirements directly contradict each other. To credibly capture firm-to-firm interactions in physical space, geographic units often need to be very small as spillovers may not even extend to firms across the street Bisztray et al. (2018). At the same time, a small radius increases the likelihood of missing interactions which occur across geographic units, making the optimal unit size an important but completely arbitrary determinant of diffusion. A

[10]Increasing market access cost in the Arkolakis (2010) model not only help to rationalize the existence of small exporters in presence of high entry costs but also match important export growth dynamics following episodes of trade liberalization. We do not investigate the relationship between network externalities and exporter dynamics post-entry in this paper. The comparison made therefore only seeks to illustrate how multi-dimensional firm heterogeneity could lead to new explanations for the presence of small exporters.

related issue is the inability to capture different spillover effects for firms within the same unit as all of them by assumption share the same network. Both issues, the definition of unit size and lack of within-unit heterogeneity, similarly occur when capturing interactions via input-output tables[11] but are resolved in our augmented framework. The production network approach shown in figure 2c exploits all domestic firm-to-firm linkages revealing each firm's true network which is unique to each firm and free from any geographic constraints. These key characteristics allow us to move from the geographic group to the individual firm and express networks as a firm-specific conduit of information.

A related approach exploits labor movements to capture firm-to-firm interactions and is depicted in figure 2b. Choquette and Meinen (2015) and Patault and Lenoir (2021) both investigate whether hiring employees from firms with export experience increases the likelihood to start exporting. While highly intuitive, this approach differs from ours in several important aspects. Hiring is more likely to be a strategic decision to access specialized information as firms pay particular attention to the skill set of applicants. Information therefore diffuses endogenously which stands in sharp contrast to the exogenous diffusion process in our setting. Moreover, the alternative approach is far more restrictive as it rules out information transfers between firms that do not hire from each other. In contrast, we simply require firms to have business interactions which results in a much wider interaction network involving many more linkages and hence opportunities for information to flow between firms. To the extent that labor movements and business interactions overlap, we therefore generalize previous approaches by capturing a wider diffusion mechanism of export-related information[12].

A third important advantage of our augmented framework is the ability to nest two distinct network specifications which differ in their treatment of network noise. We define network noise as any firm-to-firm interaction that does not yield an export signal. From the perspective of export participation these linkages represent a burden as they do not lower entry barriers to foreign markets but still take up time and resources at the firm. Our framework can easily accommodate settings with and without network noise by making minimal adjustments to our function form assumption which relates market access costs f to network linkages. In equation 5 we effectively ignore network noise by simply defining network externalities as the number on export signals firm i receives, a measure we call *signal intensity*. If we instead want to account for network noise, we need to express network externalities as share of total network interactions that yield export signals, a measure we call *signal clarity*. For any fixed number of export signals

[11] Approaches using input-output data assume the same network for all firms operating in the same industry as interaction strength only varies at the sectoral level. This masks any network heterogeneity within sectors.

[12] Labor movements and business interactions will likely not overlap perfectly. While this means that our production network approach will likely miss certain linkages, it captures a much wider extent of firm interactions than previous approaches. Patault and Lenoir (2021) for example use labor movements of sales managers as a proxy for firm-to-firm interactions. The implicit network size of their approach is therefore equivalent to the total number of competitors a firm hires from in any given year. As this number is not reported in their paper, we resort to an indirect comparison. In the extreme, the largest number of interactions from labor movements occurs if all French companies were to hire their entire staff from a distinct set of competitors every year. In this case, the average number of employees per firm serves as a proxy for the maximum network size as every employee is hired from a different competitor. Based on business demographics from Eurostat, the average number of employees per firm in France in 2019 is 5. When excluding firms with less than 10 employees, the average increases to 79. In contrast, The average number of buyers a seller interacts with each year in the Belgian production network is 20 for all sellers and 416 for sellers with at least 10 employees. Hence, even when taking labor movement-based network size to the extreme, production networks still capture a much larger number of firm-to-firm interactions.

larger levels of network noise will lead to a decrease signal clarity which attenuates the positive impact of the network on export participation. In practice, this changes our functional form assumption for entry costs f to

$$\ln f'_{id,t} = \alpha_d - \delta \sum_j \bar{s}_{ij,t} x_{j,t} - \beta \sum_j \bar{s}_{ij,t} y_{jd,t} - \psi_{S_{i,t}} + \varepsilon_{id,t} \tag{7}$$

Equation 7 is identical to equation 5 apart from a single change in the network information term β which now involves row-normalized interaction matrix elements $\bar{s}_{ij,t}$ which now to relate market access costs to the share of export signals $y_{jd,t}$ out of all network interactions of the firm[13]. The distinction between signal intensity and signal clarity is a key contribution of this paper and will become important to understand diverging network effects for firms with small and large networks which are explained in detail in section 5.4.

3 Data, empirical setting and descriptive statistics

To empirically investigate the role of network externalities for the extensive margin of trade we rely on detailed firm-level data from Belgium. As a small open economy Belgium represents an ideal setting for our analysis since it relies heavily on trade making export participation a key concern for firms and policy makers alike. In this section we first describe our main data sources and link them to the augmented framework derived above. We then present descriptive statistics to illustrate how network heterogeneity shapes the diffusion of export information among Belgian firms.

3.1 Data sources and sample selection

At the center of our analysis are three administrative datasets which are linked via unique firm identifiers and capture characteristics, export behavior and network interactions of Belgian firms for the years 2002 – 2014. Firstly, we use the Annual Account Filings database (National Bank of Belgium, 2002–2014a) which collects balance sheet information such as sales, revenues, input costs (labor, capital, material), 4-digit industry codes (NACE), zip code and ownership information from mandatory annual account filings of all firms operating in Belgium. We complement firm characteristics with annual import and export transaction data at HS6 product-level from the International Trade Dataset (National Bank of Belgium, 2002–2014b) which combines information from customs records and intra-EU trade declarations[14]. Together, balance sheet and trade data provide a detailed picture of performance and export activity of Belgian firms but do not grant any insights into firm-to-firm interactions. To fill this gap, we use the Business-to-Business Transactions Dataset (National Bank of Belgium, 2002–2014c) which records any buyer-seller transaction of firms operating in Belgium, provided the annual transaction value amounts to at least 250 €[15]. Belgian firms are required by law to file a breakdown of their an-

[13] To see this, recall that $s_{ij,t} \in \{0, 1\}$ and row-normalized $\bar{s}_{ij,t} = \frac{s_{ij,t}}{\sum_j s_{ij,t}}$ such that $\sum_j \bar{s}_{ij,t} = 1$. The term $\sum_j \bar{s}_{ij,t} y_{jd,t}$ hence equals the average number of network peers which emit export signals at time t. If no firms in the network emit signals it equals 0, it all firms in the network emit signals it equals 1.

[14] Intra- and extra-EU transactions have different reporting thresholds which are explained in appendix B.1

[15] For a detailed description of the dataset we refer to Dhyne et al. (2015)

nual sales by each individual buyer which allows us to identify individual firms involved in each transaction and thereby capture virtually all firm-to-firm interactions at an annual interval. To handle the vast amount of information contained in the combined dataset we implement important sample restrictions along firm, destination and network dimensions.

At the firm level, we follow the sample selection procedure of Dhyne et al. (2021) which significantly reduces the sample size while remaining very close to aggregate national statistics. In a first step this involves exploiting ownership information to single out observations which have unique identifiers but ultimately relate to the same firm. Identifiers in the data are constructed from value-added tax (VAT) numbers and some firms choose to use multiple VAT numbers for tax or accounting purposes. We aggregate these entries to the level of the firm which reduces the number of observations by around 4%. The second step of the selection procedure was originally introduced by De Loecker et al. (2014) and restricts our sample to firms with at least one full-time employee, more than 100 € of tangible assets, positive total assets in at least one reported year and positive labor costs and output. This step alone excludes more than 80% of the remaining observations as many firms in the original data are one-person companies[16]. The remaining sample is identical to the one used in Dhyne et al. (2021), includes between 90k-100k firms per year and remains very close to aggregate statistics in terms of value added, gross output, exports, and imports[17].

At the destination level, we only consider market access decisions for destinations outside the European Economic Area (EEA) as information frictions are expected to represent a much larger barrier to entry compared to highly integrated EEA countries[18]. Non-EEA destinations on average account for roughly two-thirds of all export starts of Belgian firms which means our sample still captures the majority of activity at the extensive margin of trade. We follow Koenig (2009) and define an export start as a transaction to a destination which has not been served by the firm in the previous two years. Resuming exports to a foreign market after a single year of inactivity therefore are not treated as export starts[19]. This ensures that sufficient time has passed for market conditions to change such that information costs again become a relevant barrier to entry[20]. For our sample this implies that all observations of the first two years are dropped reducing the sample timeframe to 2004-2014. Further, we only consider firm-destination pairs with at least one export start across years to facilitate comparability across different estimation approaches[21].

At the network level, we start by characterizing the main network components. A network is defined as a collection of nodes and edges which in our case are represented by firms and their

[16] In 2012 there are 750,100 firms reporting less than 1 full-time employee.

[17] For a detailed comparison with aggregate statistics we refer to Table 1 in Dhyne et al. (2021)

[18] The list of EEA countries includes Greece, Lithuania, Portugal, Bulgaria, Spain, Luxembourg, Romania, Czech Republic, Hungary, Slovenia, Denmark, Croatia, Malta, Slovakia, Germany, Italy, the Netherlands, Finland, Estonia, Cyprus, Austria, Sweden, Ireland, Latvia, Poland, the United Kingdom, Norway, Liechtenstein, Switzerland, Iceland. We disregard all export transactions of Belgian firms to any of these countries for all sample years.

[19] Note that this allows for restarts within firm-destination pairs. In practice only 11% of entries are restarts.

[20] We assume that firms gather entry-related information upon entry. Firms that reenter after a single year of inactivity still possess very recent entry information and could benefit from their previous experience. By enforcing a 2-year period of inactivity we assume that entry requirements, consumer preference and non-tariff barriers in the destination have sufficiently changed such that information again presents a barrier to entry.

[21] Logistic regressions require variation in the outcome variable. To facilitate a comparison with results from linear probability model, we require at least one export start within each firm-destination pair which ensures sufficient variation for logistic regressions and allow to use the same sample for both estimation methods.

business transactions. Transactions (edges) therefore link firms (nodes) to each other and the transaction value (edge weight) gives an indication about the respective strength of each network interaction. In production networks edges are always directed because each firm involved in a transaction either acts as a buyer or a seller. In our setting, we need to distinguish between two distinct types of direction. First, the flow of goods and services from sellers to buyers along the supply chain which we define as a forward linkage. Second, the flow of money for goods and services sent from buyers to sellers which we define as a backward linkage. This distinction is important because network externalities in principle could go in either direction. In this paper we focus on information diffusion along backward linkages meaning sellers learn from their buyers. This direction has been identified as the most relevant one for information diffusion by the preceding export (Choquette and Meinen, 2015) and productivity (Javorcik and Spatareanu, 2011) spillover literature and is favored in our empirical setting. Sellers typically do not care which buyers they sell to which creates little incentive for them to communicate export-related information along forward linkages. Buyers on the other hand care about their suppliers as their own performance depends on the quality of sourced inputs. Information in our empirical setting mainly diffuses from buyers to sellers but we offer additional results for alternative diffusion directions in section 9.

While this clearly denotes which firms emit and receive export signals, in practice it is unlikely that all buyer-seller interactions meaningfully contribute to the diffusion of export signals. Suppliers which only account for a small share of total buyer sourcing may receive no information because the small transaction size does not necessitate any communication with buyers or indicates a low level of importance attached to the sourced input. We therefore need to distinguish between relevant and non-relevant network linkages and exclude those which too small to play any meaningful role for the diffusion of export signals. To do so we compute the share of total buyer sourcing accounted for by individual suppliers as

$$\nu_{ij,t} = \frac{\kappa_{ij,t}}{\sum_j \kappa_{ij,t}}$$

where $\kappa_{ij,t}$ represents the value of annual transactions between seller i and buyer j in year t taken from transaction value matrix K_t. An interaction is defined as relevant for diffusion if suppliers account for at least 1% of buyer sourcing. Interactions that account for less than 1% of buyer sourcing[22] are treated as irrelevant for information diffusion and are excluded from the sample[23]. Applying this rule to all entries of transaction value matrix K_t leads to a binary interaction matrix S_t[24] with elements

$$s_{ij,t} = \begin{cases} 1, & \nu_{ij,t} \geq 1\% \\ 0, & \text{otherwise} \end{cases}$$

Each row of matrix S_t contains linkages of seller i and the row sum indicates the number of buyers j a seller interacts with each year. As customary, self-links are not allowed which means

[22] Our empirical results are robust to alternative thresholds as demonstrated in section 5.2.

[23] Our network sample is also subject to the firm-level restrictions described above which exclude 52% of network linkages from the sample. Of the remaining interactions, non-relevant linkages account for 85% in number but only account for 8% of total buyer sourcing. The network restriction therefore retains the majority of sourcing value $\nu_{ij,t}$ which is our key indicator of diffusion probability and greatly facilitates the analysis by reducing the sample size.

[24] Our baseline model does not differentiate between transactions beyond the 1% threshold. To learn more about the role of interaction strength for network externalities, please see section 5.3.

all diagonal elements s_{ii} are set to zero.

3.2 Descriptive statistics

After implementing firm, destination and network restrictions our final sample contains characteristics of around 98,000 firms, 25,000 export starts to 188 non-EEA destinations and more than one million firm-to-firm interactions per year. The combined data allows us to trace the diffusion of export signals along network linkages and relate it to the entry behavior of Belgian firms. To understand how each data source contributes to this analysis we present descriptive evidence about firm behavior at the extensive margin of trade, the prevalence of signal diffusion and the role of network structure for the diffusion process.

3.2.1 Extensive margin of trade

Figure 3 shows the geographic dispersion of non-EEA export starts of Belgian firms between 2004-2014. Export decisions follow the rules of gravity and mainly occur in markets that are attractive due to their large size or limited distance to Belgium. One exception is the concentration of export entry in the Democratic Republic of the Congo. As a former colony the country retains strong ties to Belgium which potentially facilitates market access for Belgian exporters. Another important pattern shown in appendix A.1 is distribution of export entry across geographic regions. While large countries like the US individually still account for the largest number of export starts, appendix A.1 shows that more than two-thirds of non-EEA new market entries occur in Africa and Asia. As these blocks comprise a large number of countries with different import regulations, consumer preferences and local supply networks, we expect entry-related information costs for these destinations to be high. This emphasizes the role of network externalities as many Belgian firms might want to reach the large consumers base in these emerging markets but lack the ability to overcome entry barriers.

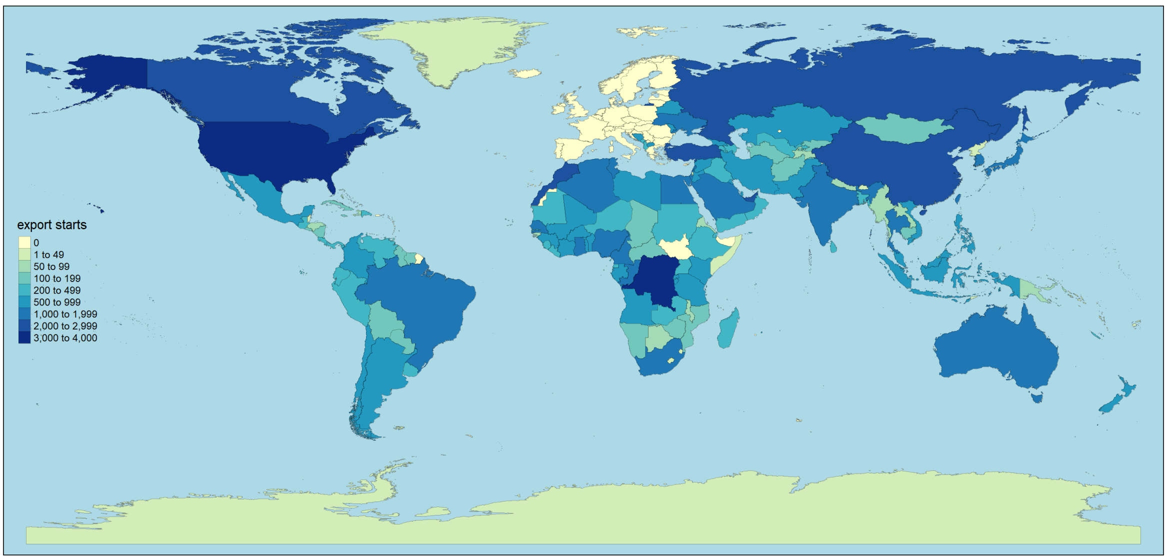


Figure 3: Geographic distribution of export starts (2004-2014)

3.2.2 Prevalence of export signals

Our augmented framework allows sellers to benefit from the export experience of their buyers in form of export signals that are emitted every time a connected buyer starts to export to a new destination. Our novel approach allows us to identify over 728,000 export signals received by sellers between 2004-2014 which highlights the vast scope of export information diffusing between firms. We distinguish between matching and non-matching export signals signals to indicate whether the origin of the incoming signal matches the destination the seller starts exporting to. Matching signals therefore represent valuable information information related to a firm's own export entry whereas non-matching signals capture the overall scope of the diffusion process.
Table 1 illustrates the prevalence of both signal types in each year and over the entire sample period. A first insight is that despite the large number of signals identified, a majority of sellers do not receive any export signals. Each year only 5.5% of sellers benefit from matching export signals which emphasizes that many entry decisions are still taken in absence of network externalities. This means there remains a large amount of cross-sectional variation we can exploit for our empirical analysis[25]. Second, the distribution of firms receiving export signals appears to be highly skewed. Over the entire sample period roughly one half of all sellers do not receive any signals while a quarter of them receive more than 5[26]. This concentration of information diffusion among a small number of sellers is related to firm and network characteristics and further explored below.

Table 1: Share of firms receiving export signals

	per year		2004-2014	
signals	any signal	matching signal	any signal	matching signal
0	0.777	0.945	0.555	0.740
1	0.077	0.041	0.057	0.074
2	0.041	0.008	0.040	0.035
3	0.026	0.003	0.029	0.023
4	0.017	0.001	0.024	0.017
5	0.012	0.001	0.020	0.012
more than 5	0.045	0	0.248	0.081

This table indicates the share of firms that receive export signals in a single year and over the whole sample period. Matching signals represent the subset of total signals that originate in the same market as the subsequent export entry.

3.2.3 Network descriptives

We start with presenting a first set of characterisitcs related to network heterogeneity. Seller networks on average consist of 14 different buyers including 2 exporters and 1 export starter.

[25] If most sellers received export signals in every period, identification of network externalities would mostly rely on within-firm variation in incoming export signals over time. Table 1 shows that our analysis can rely on both within- and between-firm variation when estimating network effects.

[26] The full distribution of export signals is shown in appendix A.2.

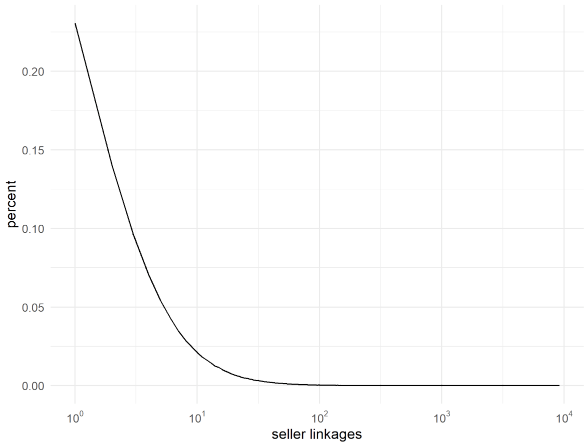


Figure 4: Distribution of seller linkages

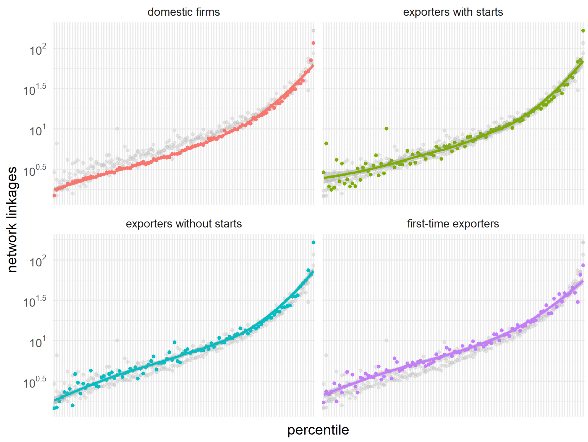


Figure 5: Seller linkages by sales percentile

These numbers mask a large amount of heterogeneity in firm networks as the distribution of seller linkages is highly skewed[27]. Figure 4 shows that while 25% of sellers only maintain a single network interaction, sellers in the top decile on average count over 1000 linkages to buyers each year. These vast differences in network size are closely related to seller size as shown in figure 5. Sellers in higher sales percentiles (a proxy for firm size) on average interact with more buyers which holds for exporting and non-exporting sellers alike. This pattern is common to production networks (Zi and Bernard, 2021) and typically explained by superior firm performance. Large sellers can offer products at better quality or lower prices and thereby attract a larger number of buyers.

This correlation between seller and network size has direct implications for the export participation pattern predicted by our augmented framework. If larger networks are associated with more export information and lower market access costs f_{id}, then network heterogeneity will simply reinforce selection patterns of the standard framework[28] by increasing the existing advantage of large sellers. In that case firm and network heterogeneity reinforce each other and our main contribution would be to estimate the relative importance of each channel. Conversely, if larger networks do generate superior access to export information, then firm and network heterogeneity might work in opposite directions with each channel benefiting firms of different sizes.

The latter case is of particular interest in our setting due to a second set of network characteristics related to negative degree assortativity. While larger sellers typically have larger networks, the average performance of their buyers is worse. This surprising fact can be seen as a curse of large sellers' own success. Being able to offer products at lower prices than competitors implies that even underperforming firms can afford these products which increases the total number of buyers but lowers average buyer performance. Small sellers in comparison tend to be less productive, offer products at relatively high prices which means only very performant firms are able to buy them. Figure 6 illustrates this pattern for seller networks in 2014 and shows how the share of exporting buyers changes with network size. The average buyer in a large network is significantly less likely to be an exporter compared to the average buyer in a small network. This

[27] The distribution of linkages to exporters and export starters is equally skewed as shown in appendix A.3.

[28] In absence of network interactions, larger (more productive) firms are already more likely to access foreign markets due to their superior firm performance (Melitz, 2003). This fact receives large empirical support as exporters tend to be larger and more productive than non-exporters (Bernard et al., 2003).

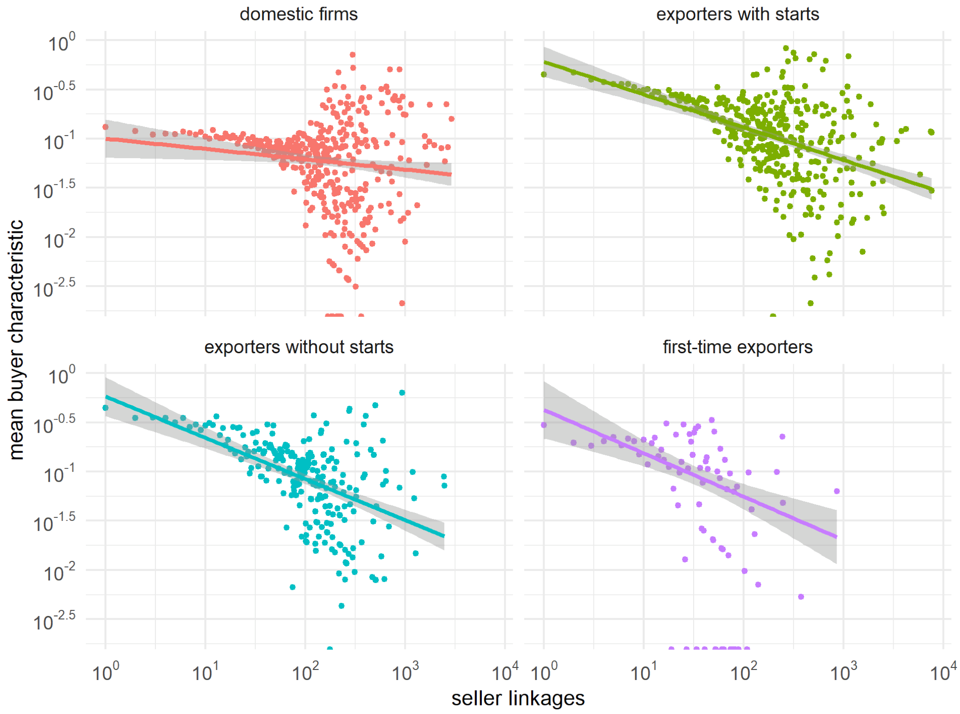


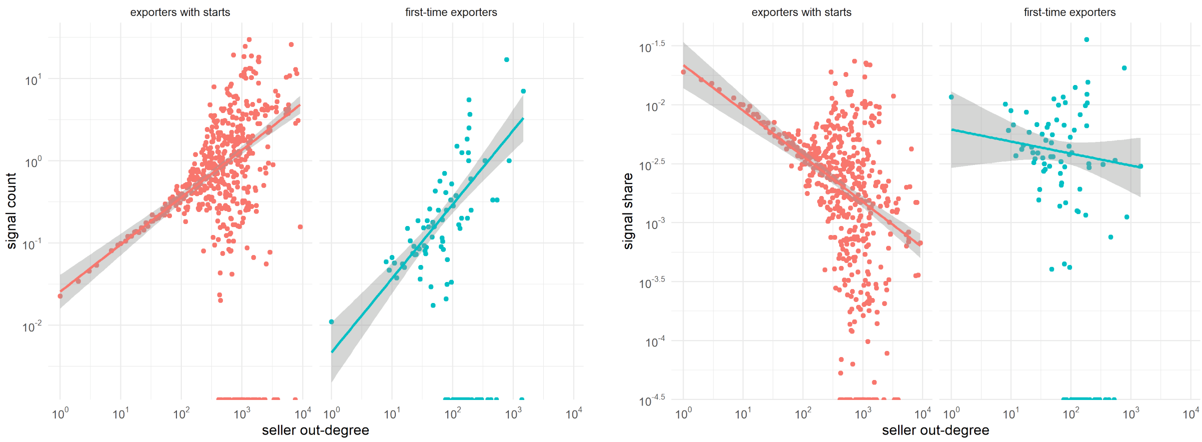
Figure 6: Seller network size and mean buyer export probability in 2014

relationship holds across years, seller types and alternative buyer characteristics such as sales, employment, productivity or export starts as shown in appendix A.4. If information diffusion depends on average rather than aggregate buyer behavior in the network, negative degree assortativity therefore reverses the impact of network heterogeneity on export entry. In this setting, small firms are expected to receive more export information than large firms as the average buyer is more likely to emit export signals.

3.2.4 Network externalities

A key takeaway from the previous section is that network externalities do not necessarily favor sellers with larger networks. Negative degree assortativity in production networks introduce a penalty that increases in network size and becomes important if the average network behavior is of interest. The decision to relate outcomes to aggregate or average network activity therefore has first-order importance for model predictions.

This finding is directly linked to the treatment of network noise discussed in section 2.4. If we assume network noise can be ignored, then the aggregate number of export signals matters (signal intensity). If we instead beleive that network noise needs to be accounted for, then we care about the relative number of export signals (signal clarity). Figure 7 shows how signal intensity and clarity change according to seller network size. As expected, we see a diverging pattern across measures. While signal intensity indicates that large firms receive more export-related information than small firms, the opposite is the case for signal clarity where network noise penalizes sellers with large networks. This indicates that network effects may not linearly increase in network size.



(a) signal intensity

(b) signal clarity

Figure 7: Network externalities in 2014

4 Econometric framework

Introducing network interactions to the standard model of export entry creates several econometric challenges. Export decisions now not only depend on sellers' own characteristics but also on the characteristics and export experience of their buyers. We first outline the new identification challenges created in this augmented setting and then describe how we approach them in our empirical application.

4.1 Identification

As a starting point we rewrite our entry equation 6 in matrix notation. Using S_t and $\bar{S}_t$ to indicate standard and row-normalized binary interaction matrices we get

$$Pr(y_{d,t} = 1) = Pr(\boldsymbol{x_t}\gamma + \bar{S}_t\boldsymbol{x_t}\delta + \beta S_t y_{d,t} + \psi_{S_t} - \alpha_d + \varepsilon_t > 0) \tag{8a}$$

and

$$Pr(y_{d,t} = 1) = Pr(\boldsymbol{x_t}\gamma + \bar{S}_t\boldsymbol{x_t}\delta + \beta \bar{S}_t y_{d,t} + \psi_{S_t} - \alpha_d + \varepsilon_t > 0) \tag{8b}$$

for signal intensity and clarity respectively. Following the terminology of the peer effects literature these are commonly known as local-aggregate and local-average models. Both allow networks to affect seller outcomes in three distinct ways. First, in form of contextual peer effects which relate seller outcomes to buyer characteristics. These capture general externalities unrelated to export information such as productivity spillovers. Second, in form of endogenous peer effects β which represent signal intensity or clarity. Third, in form of network fixed effect ψ_{S_t} which expose sellers to common shocks occurring in their immediate network. While this creates interesting interdependencies in firm behavior, the models cannot be readily taken to the data before dealing with several econometric challenges. To facilitate the discussion of each challenge, we take an intermediary step and assume $E(\varepsilon_t|\boldsymbol{x_t}, S_t) = 0)$ which means networks form exogenously after conditioning on observable firm characteristics. We relax this assumption in section 4.3 where we introduce a network selection model to control for the potential selection bias created by endogenous network formation.

4.1.1 The reflection problem

A first challenge arises from the joint determination of buyer and seller outcomes y_t. If both terms enter the equations 8a and 8b contemporaneously, common shocks can lead to a simultaneity of buyer and seller export behavior. This well-known *reflection problem* (Manski, 1993) can prevent the separate identification of contextual and endogenous peer effects δ and β if individual firm networks do not sufficiently overlap[29]. In that case firm-to-firm linkages create separated network clusters in which firms only interact with members of the same cluster but have no linkages with firms in other clusters. If firms then experience a common shock, contextual and endogenous network effects become perfectly collinear as all firms within the same cluster act simultaneously and there is no variation from cross-cluster linkages to separately determine the impact each network channel.

The separate identification of both network effects is of particular importance in our setting, as we want to ensure that our main coefficient of interest β does not capture general spillover effects unrelated to export information. Bramoullé et al. (2009) and (Liu et al., 2014) show how this can be achieved in network settings for local-average and local-aggregate models respectively. In a local-average model contextual and endogenous network effects are identified if identity matrix I, and interaction matrices S, and S^2 are linearly independent. In a local-aggregate model separate identification requires the rowsums of S to be non-constant and linear independence between I, S, $\bar{S}$ and $S\bar{S}$. Both sets of conditions are met in our setting as linkages in production networks are typically unidirectional which ensures linear independence of network matrices due to the presence of intransitive triads[30] and the fact that each seller interacts with a different number of of buyers leading to a non-constant rowsum of S.

Despite meeting the general conditions to identify network effects in a contemporaneous setting, we take a different approach because sellers are expected to respond to incoming export signals with delay. Assuming a temporal lag between signal reception and response is more realistic in our setting as sellers may take time to process information and adjust their production processes before entering a foreign market. Empirically, a delayed response also mitigates concerns related to the timing of buyer and seller export starts within the same year by ensuring that all sellers have sufficient time to react irrespective of the exact time a signal was received[31]. Lagging network effects changes our equations for signal intensity and signal clarity to

$$Pr(y_{d,t}=1) = Pr(\boldsymbol{x_t}\gamma + \bar{S}_{t-1}\boldsymbol{x_{t-1}}\delta + \beta S_{t-1}y_{d,t-1} + \psi_{S_{t-1}} - \alpha_d + \varepsilon_t > 0) \tag{9a}$$

and

$$Pr(y_{d,t}=1) = Pr(\boldsymbol{x_t}\gamma + \bar{S}_{t-1}\boldsymbol{x_{t-1}}\delta + \beta \bar{S}_{t-1}y_{d,t-1} + \psi_{S_{t-1}} - \alpha_d + \varepsilon_t > 0) \tag{9b}$$

Buyer and seller export starts now no longer occur simultaneously which resolves issues related to Manski's reflection problem. Conceptually, the change also marks a departure from the local-aggregate and local-average models of peer effects literature. Instead, it brings our approach

[29] In our setting, firm networks do not overlap if sellers act as exclusive suppliers for all buyers in their network and buyers themselves source but do not sell (= have positive indegree but zero outdegree).

[30] An intransitive triad describes a network structure where firm A interacts with firm B, B interacts with firm C. It is called intransitive if there is no direct interaction between A and C.

[31] Sellers receive export signals at different points of the year. If responding to signals takes time, then sellers receiving a signal towards the end of the year are disadvantaged which may introduce a downward bias to the estimation of endogenous peer effects.

closer to time-space recursive models[32] studied in spatial economics (Anselin et al., 2008; Halleck Vega and Elhorst, 2017) where current outcomes y_t are related to past network outcomes $\bar{S}_{t-1}y_{t-1}$. An important difference to time-space recursive models is that we do not consider lagged seller outcomes y_{t-1} as additional controls. This type of autocorrelation cannot occur in our setting due to the definition of export starts[33] which rules out entries to the same destination in two consecutive years.

A delayed response to network externalities facilitates model identification but requires additional assumptions regarding the timing of the underlying diffusion process. First, incoming export signals are not allowed to affect market access costs f_d in the same period. This assumption is compatible with a setting where networks form at the beginning of each period, firms choose whether to export and only then receive export signals. This order of events implies that information received in the current period only facilitates foreign market entry in next one. Second, to ensure conditional network exogeneity holds we need to assume that export signals received in period t do not influence network formation at the beginning of period $t + 1$. This rules out that sellers form linkages strategically with the intent to reduce market access costs in the next period. Choosing to ignore export information received from previous linkages when forming new ones is restrictive but compatible with scenarios where the value of past signals is not yet realized at the beginning of the current period. Relaxing this assumption would require a formal model of strategic network formation as in (Badev, 2021) which is beyond the scope of this paper.

4.1.2 Correlated effects

A second challenge in our setting is to demonstrate that estimated network coefficients δ and β capture a causal relationship between network behavior and seller outcomes instead of a mere correlation driven by unobserved shocks. The latter are typically referred to as *correlated effects* which arise naturally in our study as buyers and sellers face various domestic and foreign shocks that can alter their export participation decision but remain unobserved by the econometrician. Failing to account for correlated effects will introduce a bias to estimated network coefficients and cast doubts on the relevance of network effects for export entry.

To understand what type of shocks might cause concerns when estimating equations 9a and 9b, it is key to consider the timing assumptions and different levels of observation at which our network effects operate. First, assuming a lagged seller response to network effects rules out most correlated effects from temporary shocks as buyer and seller actions no longer occur in the same period. Second, while contextual peer effects δ operate at the firm-year level, endogenous peer effects β operate at the firm-destination-year level. This opens up the opportunity to employ high-dimensional fixed effects (FE) to account for a vast array of correlated effects from unobserved supply chain disruptions (firm-year FE), foreign demand shocks (destination-year

[32]Time-space here refers to two different types of lag from perspective of the dependent variable y_t. A spatial lag $S_t y_t$ indicating the relationship to network outcomes and a temporal lag $t - 1$.

[33]An export start requires inactivity in the foreign market in the previous two periods. This implies that the two periods after an export start are excluded from the sample. A firm starting to export in year 3 and stopping in year 5 therefore is only faces entry decisions in periods 1,2,3 and 6. Years 4 and 5 are dropped from the sample.

FE) or export specialization patterns within networks (firm-destination FE)[34].

The main concern regarding correlated effects in our setting therefore does not emanate from unobserved time-varying shocks per se, but instead arises from common shocks which buyers and sellers may respond to at different points in time. To illustrate this point, assume Chinese customs officials unexpectedly relax import requirements in period $t-1$ resulting in a decreased sunk entry cost α_d for all Belgian firms. If buyers in network immediately respond to the shock and start exporting to China but sellers only react to the shock in period t, the delayed shock response of sellers would be observationally equivalent to the network effect we try to capture. To isolate buyer-seller entry variation induced by network effects, we need an instrument that can absorb any correlation in export behavior that is driven by delayed response times to common destination-specific shocks.

In our setting, this means that we need to treat export starts of connected buyers $y_{jd,t-1}$ as endogenous and find a suitable instrument that is correlated with buyers' entry behavior in $t-1$ (relevance) but uncorrelated with sellers' entry decision in period t (exclusion restriction). An instrument that meets these requirements are export starts of firms which are directly linked to buyers j in period $t-1$ but have no direct link to the seller i as illustrated in figure 8.

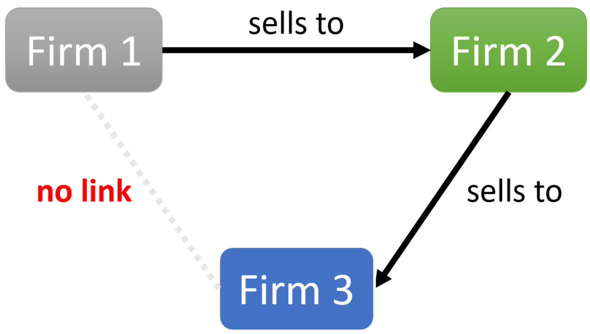


Figure 8: Intransitive triad

A destination-specific shock like the unexpected Chinese import liberalization discussed above is expected to lead to correlated export responses of direct (firm 2) and indirect buyers (firm 3) in $t-1$ due to their buyer-seller relationship. If indirect buyers' only network link to seller i goes via direct buyer j, we can use the export starts of indirect buyers to control for correlated export behavior caused by the shock in $t-1$. After controlling for the impact of destination-specific shocks in $t-1$ with our instrument, any remaining correlation between direct buyers j and sellers i must be driven by network effects.

The validity of this instrument crucially relies on the assumption that indirect buyers are only linked to sellers via buyers j and do not affect seller outcomes directly. To corroborate the credibility of this exclusion restriction, we exploit the full network structure[35] and exclude all indirect buyers (firm 3) which are connected to sellers (firm 1) directly or via higher-order linkages

[34] Firm-year FE are a special case in this context. While including them absorbs contextual network effect δ and many firm-level controls, it allows us to control for *any* unobserved shock at the firm level. This includes important examples like supply chain disruptions, efficiency spillovers within networks or changes in local infrastructure and policy. Given they also absorb many variables of interest from the stylized entry model, we do not include them in our benchmark regression and instead only use them in robustness checks shown in section 5.2

[35] A natural concern in this setting is that our network sample does not accurately capture all relevant linkages of each seller. While this is likely the case in practice, we expect our approach to perform reasonably well as most

which do not involve immediate buyers (firm 2). In other words, we exclude all firms from the set of indirect buyers that have a first-, third-, forth- or fifth-order linkage to sellers i. To curb the influence of unobserved firm linkages created outside production networks, we only consider indirect buyers located outside the province of the seller. The results of this instrumentation strategy are presented in section 5.2.2.

4.2 Estimation

We now present our empirical framework. Under the assumptions discussed in the previous section and assuming networks form exogenously, we only need to make a distributional assumption for error terms $\varepsilon_{id,t}$ to take models 9a and 9b to the data. As a starting point, our benchmark estimation uses a linear probability model with fixed effects (LPM-FE) which assumes that errors $\varepsilon_{id,t}$ are $i.i.d$ and follow a normal distribution. A key advantage of the LPM-FE is the ability to easily accommodate high-dimensional fixed effects which allows us to capture important unobservable factors in domestic and foreign markets that otherwise may give rise to correlated effects. At the same time, assumed linearity limits the accuracy of predicted probabilities which can exceed the {0,1} interval.

Non-linear alternatives such as Probit and Logit models restrict predicted probabilities to the unit interval and therefore deliver more precise estimates for extreme values but typically suffer from an incidental parameter problem (IPP) when featuring high-dimensional fixed effects (Neyman and Scott, 1948). If the number of parameters that need to be estimated increases with sample size, maximum likelihood asymptotics no longer converge resulting in inconsistent parameter estimates. Our empirical setting is prone to this issue as the analysis considers export decisions at the firm-destination level which involves a large number of unobserved characteristics that need to be estimated. To evaluate benchmark estimates of LPM-FE model, we therefore contrast them with the fixed-effects logit estimator of Fernández-Val and Weidner (2016) and the fixed effects probit estimator of Hinz et al. (2021) which both feature a bias correction for the IPP while remaining directly comparable[36] to the LPM-FE via average partial effects.

Under normally distributed errors, we estimate the following reduced-form equation based on our time-space recursive lag model:

$$Pr(y_{id,t}=1) = Pr\left(\boldsymbol{x_{id,t}} + \sum_j \bar{s}_{ij,t-1}\boldsymbol{x_{j,t-1}}\delta + \beta\sum_j s_{ij,t-1}y_{jd,t-1} + \psi + \varepsilon_{id,t} > 0\right) \tag{10}$$

Seller export starts $y_{id,t}$ are related to their own characteristics $\boldsymbol{x_{id,t}}$, network effects in form of buyer characteristics $\boldsymbol{x_{j,t-1}}$ and export signals $y_{jd,t-1}$ and a set of fixed effects ψ. We summarize the variables contained in each component below and present additional details in appendix B.2.

i. Seller characteristics $\boldsymbol{x_{id,t}}$ capture determinants that affect seller export decisions in absence of any network effects. These include firm-level controls such as total factor produc-

social networks are extremely sparse. Missing or misspecified network linkages should therefore only represent a small fraction of total linkages when compared to the correctly identified absence of linkages between most firms.

[36] A common approach that avoids the IPP overall is the conditional logit model suggested by Chamberlain (1980). While delivering consistent parameter estimates it is not able to estimate average partial effects and therefore cannot be directly compared to the other methods.

tivity (TFP), estimated using the procedure of Levinsohn and Petrin (2003), seller wages and employment-based seller size. Higher levels of TFP, wages and size are typically associated with increased export probability Bernard et al. (2003). Complementary to these firm-level controls, we exploit available data about trade transactions to construct additional variables at the firm-destination level. First, we identify the products underlying a seller's export start and use this information to construct a firm-specific measure of import demand in each foreign market. This variable controls for export decisions as a direct response to foreign demand shocks. Second, we control for sellers' experience in a foreign market prior to their export start. Even without network linkages, sellers might accumulate expertise about destinations from other activities. We therefore add dummy variables to control for seller experience from importing, exporting to bordering destinations or destinations with historic ties[37]. Lastly, we control for a seller's overall export expertise via the share of export sales in total sales.

ii. Buyer characteristics $\boldsymbol{x_{j,t-1}}$ capture general network spillovers that affect seller entry across export destinations. Their main function is to ensure that export signal coefficient β is identified from destination-specific variation in the network. We include buyer sales and TFP to control for general spillovers unrelated to entry information.

iii. Lastly, we employ two distinct fixed-effect (FE) specifications to control for correlated effects. In the benchmark case, ψ contains firm and destination-year FE. This allows to control for unobserved differences in firm performance and time-varying demand shocks in foreign markets. A second and more stringent specification extends this to firm-year FE. In this case, fixed effects absorb *any* time-varying characteristic at the firm-level which includes most variables of the standard entry framework as well as network effects from buyer characteristics. To remain as close as possible to the theoretical framework, we therefore rely on the weaker FE specification for the benchmark case and use the more stringent specification as a robustness check.

Table 2: Regression sample (firm-years)

Statistic	N	Min	Pctl(25)	Median	Mean	Pctl(75)	Max
employees	89,120	1.00	4.50	12.60	76.45	36.90	59,691.68
wage (k)	89,120	0.70	39.51	48.59	52.76	60.30	574.71
TFP (log)	89,120	3.39	12.80	13.56	13.69	14.47	21.05
border dummy	89,120	0	0	0	0.15	0.2	1
history dummy	89,120	0	0	0.6	0.52	0.9	1
import dummy	89,120	0	0	0	0.09	0.1	1
export sales share	89,120	0.00	0.00	0.05	0.21	0.38	1.00
export demand (mn)	89,120	0.00	0.01	0.05	6.02	0.30	11,328.10
mean buyer sales (mn)	89,120	0.00	1.14	2.62	30.56	7.85	47,125.25
mean buyer TFP (log)	89,120	2.78	12.00	12.49	12.66	13.15	20.97

This table shows firm characteristics of our final regression sample. All variables have been aggregated to the firm-year level to facilitate readability. The reported number of observations therefore differs from the regression tables which capture entry decisions at the firm-destination-year level.

[37]The sequence of entry decisions is not random. Firms tend to enter markets that are similar to previous destinations (Morales et al., 2019) creating spatially correlated entry patterns (Albornoz et al., 2012)

4.3 Endogenous network linkages

The preceding analysis relies on a conditional exogeneity assumption for network formation. As long as $E(\varepsilon_{id,t}|x_{id,t}, S_{i,t-1}) = 0$, interaction matrix elements $s_{ij,t-1}$ remain uncorrelated with outcome error $\varepsilon_{id,t}$ and network effects δ and β can be accurately estimated. Conditional exogeneity of network linkages, however, is unlikely to hold in practice because firms' ability to sell their products domestically may be systematically correlated with their likelihood of conducting business transactions across borders in form of exporting. New employees who were originally hired to assess the firm's domestic product appeal might for example develop methodologies that can be employed to foreign markets as well and thereby facilitate the firm's expansion abroad. These firm-level shocks which both affect domestic and foreign link formation (= exporting) are problematic because they render network linkages s_{ij} endogenous and introduce bias to estimated network effects.

To account for endogenous network formation, we introduce the network selection model of Arduini et al. (2015) and Qu et al. (2017) to our estimation procedure. This allows us to express network endogeneity as an unobserved shock to domestic production network S and export behavior y_{id} and correct for the selection bias resulting from correlated linkage formation and entry decisions[38]. Formally, network formation is expressed by equation 11. Firms trade off the value of being linked to other firms and form linkages if

$$V(S_{ij,t} = 1) - V(S_{ij,t} = 0) > 0 = U_{ij,t} + \xi_{ij,t} \tag{11}$$

where $U_{ij,t}$ represents the linkage surplus and ξ is a random error term. The surplus is typically expressed as

$$U_{ij,t}(\theta) = \theta_0 + \boldsymbol{z_{i,t}\theta_1} + \boldsymbol{z_{j,t}\theta_2} + \boldsymbol{z_{ij,t}\theta_3} + \theta_4 A_{ij,t-1} \tag{12}$$

where coefficients $\boldsymbol{\theta_1}$, $\boldsymbol{\theta_2}$, $\boldsymbol{\theta_3}$ and θ_4 capture the impact of individual characteristics of firm i and j, characteristics of the dyad ij and their past relationship status. Dyad characteristics $z_{ij,t}$ are important in this setting because they control for key matching determinants like bilateral distance or common language which by definition cannot affect export decisions of individual firms directly. Controlling for past relationship status is important because firm linkages in production networks tend to be very persistent (Martin et al., 2020) due to high search costs involved in the matching process.

If we assume that the random surplus component ξ is i.i.d and follows a logistic distribution, we can write the linkage probability $S_{ij,t}$ as

$$P(S_{ij,t=1}) = P(U_{ij,t}(\theta) + \xi_{ij,t} > 0) = \frac{e^{U_{ij,t}(\theta)}}{1 + e^{U_{ij,t}(\theta)}} \tag{13}$$

[38] Our approach introduces network endogeneity in form of a correlation between network formation and market access error terms. We believe this modeling choice is appropriate given the similarity of both processes. If exporting is considered as the search for foreign buyers, we can interpret network formation and exporting as domestic and foreign search processes which are likely affected by common unobserved shocks. Arduini et al. (2015) show that this form of endogenous network formation can be controlled for with a standard selection correction term which otherwise leaves the structure of the outcome estimator unchanged. For an empirical application of productivity spillovers see Iyoha (2021). Alternative modeling approaches which link outcome errors to unobserved variables in the formation process (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016) require Bayesian methods to estimate the likelihood functions. To keep the estimation parsimonious, we abstract from these alternatives.

To arrive at this expression, we assume that the conditional probability of i forming a link with j is independent from the decision to interact with another firm k. This implies that there is no strategic behavior in the network formation process which may characterize linkage formation in practice. Given the large network size considered our setting[39] we believe this assumption is appropriate to render the problem computationally tractable. The resulting formation process still includes many important characteristics of production networks. Seller i can interact with multiple buyers j at the same time, demonstrate persistence in their linkage decision and attach value to having business partners in close proximity.

While endogeneity in this context arises from unobserved shocks to network formation and export entry error terms ξ and ε, the timing of events in outcome equation 9a implies that we are mainly concerned about common shocks which have an immediate impact on domestic matching but only alter export decisions in the next period. Continuing the example from above, this would mean that the unexpected change in firm capacity is first employed in the domestic market, before being rolled out to prospective foreign destinations. Shocks affecting network formation and market access contemporaneously do not need to be considered as we only allow (endogenous) network effects to change firms' export decisions with a lag.

Before formalizing the correlation between formation and outcome errors, it is important to underline the dimensions at which both error terms operate. Network formation only considers firm-level characteristics of i and j whereas export decisions occur at the firm-destination level. This implies that our selection correction approach will only be able to capture correlated network and export behavior at the firm-level as formation errors $\xi_{ij,t-1}$ of seller i do not vary across seller export destinations d.

We start by collecting all network formation errors of seller i from dyadic regression 13 in row vector $\xi'_{i,t-1} = \{\xi_{ij,t-1}\}_{j\neq i}$. To relate formation errors to destination-specific outcome errors $\varepsilon_{id,t}$, each block of seller-specific error terms $\xi'_{i,t-1}$ is then repeated for each export destination seller i serves in year t. We denote the extended vector of formation errors as $\xi'_{i\{d\},t-1}$ where subscript d indicates that original formation errors have been repeated d-times for each seller i[40].

The correlation in error terms can then be expressed as $(\varepsilon_{id,t}, \xi'_{i\{d\},t-1}) \sim i.i.d.(0, \Sigma_{\varepsilon\xi})$ where $\Sigma_{\varepsilon\xi} = \begin{pmatrix} \sigma^2_\varepsilon & \sigma'_{\varepsilon\xi} \\ \sigma_{\varepsilon\xi} & \Sigma_\xi \end{pmatrix}$, σ^2_ε is a scalar, $\sigma'_{\varepsilon\xi}$ and $\sigma_{\varepsilon\xi}$ are $(n_{ijd}-1)$-dimensional row and column covariance vectors and Σ_ξ a $(n_{ijd}-1)$ by $(n_{ijd}-1)$ diagonal matrix with scalars σ^2_ξ on the diagonal. If we stack all row vectors of extended formation errors in a matrix:

$$\Xi_{t-1} = \begin{bmatrix} \xi'_{1\{d\},t-1} \\ \xi'_{2\{d\},t-1} \\ \vdots \\ \xi'_{n\{d\},t-1} \end{bmatrix}$$

we can decompose the outcome error as:

$$\varepsilon_t = \boldsymbol{\Xi_{t-1}}\eta + \upsilon_t$$

[39] There are around 100k firms in them Belgian production network. Further steps to reduce the dimension of the formation process are discussed in section XYZ.

[40] Assume there are two sellers A and B, each forming linkages with buyers 1 and 2 but differing in the number of export destination destinations they serve. Seller A exports to China and India, seller B only to India. The destination-extended vector of formation errors for all sellers would thus be $\xi = (\underbrace{\xi_{A1}, \xi_{A2}}_{\text{China}}, \underbrace{\xi_{A1}, \xi_{A2}}_{\text{India}}, \underbrace{\xi_{B1}, \xi_{B2}}_{\text{India}})$.

where $\eta = \Sigma_{\xi}^{-1}\sigma_{\varepsilon\xi}$, $\sigma_{\upsilon}^{2} = \sigma_{\xi}^{2} - \sigma_{\varepsilon\xi}'\Sigma_{\xi}^{-1}\sigma_{\varepsilon\xi}$ and υ_t is independent of formation error ξ_{t-1}. Plugging the decomposed outcome error into equation 9a then yields

$$Pr(y_{d,t} = 1) = Pr(\boldsymbol{x_t}\gamma + \bar{S}_{t-1}\boldsymbol{x_{t-1}}\delta + \beta S_{t-1}y_{d,t-1} + \psi_{S_{t-1}} - \alpha_d + \boldsymbol{\Xi_{t-1}}\eta + \upsilon_t > 0) \tag{14}$$

where $\boldsymbol{\Xi_{t-1}}\eta$ describes the selection bias induced by endogenous network formation. If $\sigma_{\varepsilon\xi} \neq 0$, seller networks S_{t-1} become endogenous and network effects δ and β will be be biased unless we control for $\boldsymbol{\Xi_{t-1}}$.

To construct the selection correction term, we follow Arduini et al. (2015) and assume that outcome error $\varepsilon_{id,t}$ is normally distributed. This allows us to use predicted linkage probabilities $p = P(S_{ij,t-1} = 1) = \frac{e^{U_{ij,t}(\theta)}}{1+e^{U_{ij,t}(\theta)}}$ from equation 13 and construct the selection correction term using a Heckman-type mills ratio:

$$\hat{\Xi}_{i,t-1} = \sum_{j \neq i} s_{ij,t-1}\frac{\phi(\Phi^{-1}(p)}{p} + (1 - s_{ij,t-1})\frac{\phi(\Phi^{-1}(p)}{1-p} \tag{15}$$

where ϕ and Φ represent probability and cumulative density functions of a standard normal variable. The estimated selection correction term can then be used as an additional regressor in equation 14 to purge the outcome error of unwanted correlation from endogenous network formation. A side effect of implementing this selection correction approach is the ability to directly test whether linkage endogeneity represents a concern when studying export entry decisions. If coefficient η is significantly different from zero, this would suggest that network formation is endogenous and controlling for selection important to recover accurate estimates of network effects.

5 Results

In this section we use the empirical framework outlined above to assess how network externalities shape export participation decisions of Belgian firms. We proceed in two steps.

Our first goal is to test whether network effects have any impact on export entry that goes beyond conventional forms of firm heterogeneity like firm productivity. We therefore bring the augmented entry framework to the data to learn how firms' own productivity, productivity spillovers from network peers and export signals affect individual entry decisions. Our benchmark estimates rely on a LPM-FE, remain close to the theoretical framework and assume no correlated effects from network signals and linkages. To assess their credibility, we then run a battery of robustness checks which involve more stringent model specifications and control for endogeneity via network instruments and a dyadic network selection model. Lastly, we shed light into the vast heterogeneity of network effects which points towards some of the underlying forces at work.

Our second goal is to investigate how network effects contribute to the observed concentration at the extensive margin of trade. To this end, we study the relationship between the strength of externalities and network size to investigate whether network effects solidify the dominant position of large firms in export markets.

5.1 Benchmark results

The main mechanism of our augmented entry framework is the ability of firm networks to provide access to specialized market information in form of export signals. Receiving an export signal lowers entry barriers for that particular market and induces firms to enter. This represents a key difference to conventional sources of firm heterogeneity like productivity which simultaneously boost entry across export destinations. A natural way to assess the relevance of network effects is therefore to sort incoming export signals by destination and test if signals indeed only facilitate access to the same market they originate from (matching signal) or can also promote entry in other markets (non-matching signal).

Table 3 summarizes the marginal effect of receiving an additional export signal by signal type. The full table is available in appendix C.1. Matching export signals (column 1) appear to be an important determinant for export entry even after controlling for sellers'own productivity, their experience in the foreign markets and network spillovers related to peer characteristics. On average each additional incoming matching signal increases the probability to start exporting by 0.43 percentage points. While we have seen in table 1 that the chance of receiving a matching signal is small, conditional on receiving one the economic significance of this network effect can become quite sizeable if the number of received signals is sufficiently large. Moreover, the fact that matching signals turn out to be statistically significant even after controlling for seller productivity underlines that both firm and network heterogeneity matter to explain export

Table 3: Benchmark results - signal type

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
matching signals	0.0043***					
	(0.0011)					
non-matching signals		3.44×10^{-5}				
		(7.77×10^{-5})				
total signals			4.14×10^{-5}			
			(7.79×10^{-5})			
EEA signals				9.03×10^{-6}		
				(0.0001)		
border signals					0.0003	
					(0.0004)	
history signals						0.0002***
						(5.93×10^{-5})
Peer characteristics	yes	yes	yes	yes	yes	yes
Firm destination experience	yes	yes	yes	yes	yes	yes
Firm characteristics	yes	yes	yes	yes	yes	yes
firm FE	yes	yes	yes	yes	yes	yes
destination-year FE	yes	yes	yes	yes	yes	yes
R^2	0.127	0.127	0.127	0.127	0.127	0.127
Observations	474,830	474,830	474,830	474,830	474,830	474,830

This table shows regression results of estimating equation 10 with a LPM-FE. Each column shows the marginal effect of receiving a different type of export signal on a seller's probability to start exporting. Standard errors in parentheses are clustered at the firm level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

participation patters of Belgian firms. In contrast, non-matching (column 2) and total export signals (column 3) seem to have no impact on entry behavior. This finding is important, as it shows that our diffusion mechanism is not picking up more general spillover effects that facilitate market access across destinations. Instead, entry barriers appear to be different in each market and as expected can only be reduced by matching signals.

In the remaining columns we disaggregate non-matching signals into three separate groups to corroborate these findings. Column 4 performs a placebo test by investigating the impact of incoming EEA signals. These destinations have been excluded from the sample in a previous step and signals should therefore under no circumstances have any impact on export decisions which reassuringly is not the case. On the contrary, signals originating in markets with a close relation to the actual export destination could facilitate entry if entry barriers or local demand preferences are correlated in space. Columns 5 and 6 illustrate that there is mild evidence of these indirect channels. This result is interesting, as it shows that previous findings of spatially correlated entry patterns of firms' own export expansion (Albornoz et al., 2012; Morales et al., 2019) carries over to information diffusion. While direct information in form of matching signals is most valuable, indirect information from related markets, albeit to a much lesser degree, also facilitates entry.

5.2 Benchmark robustness

To assess the validity of our benchmark results we perform several robustness checks which can be divided in two distinct camps. A first set investigates to what extent network effects depend on the underlying model specification. A second set is exclusively dedicated to endogeneity concerns arising from endogenous export signals and network linkages.

5.2.1 Model specification

To begin with, we need to ensure that our results are not contingent on the chosen estimation method. Benchmark estimates are obtained from a LPM-FE whose linearity assumption implies that the marginal impact of matching signals is common to all firms and constant for each additional signal received. Non-linear alternatives such as Logit and Probit models relax this assumption and allow the marginal effects of export signals to vary with characteristics of each seller. At the same time, the inclusion of high dimensional fixed effects presents a real challenge for non-linear models as the asymptotics of the underlying ML estimator break down due to the incidental parameter problem (IPP). As the inclusion of fixed effects is essential to rule out unobserved heterogeneity that would otherwise plague our network estimates, we employ the bias-adjusted non-linear logit and probit models developed by Fernández-Val and Weidner (2016) and Hinz et al. (2021) which mitigate concerns related to the IPP and remain comparable to our linear estimates via APEs. Appendix C.2.1 shows the results from the direct comparison between linear and non-linear models for signal intensity and clarity. While coefficients naturally differ across models, APEs remain remarkably close to each other irrespective of which model is preferred. The comparison also reiterates the importance of network heterogeneity for export entry as matching signals remain statistically significant throughout all specifications.

Next, we compare our benchmark results to estimates obtained from more stringent fixed effect specifications presented in appendix C.2.2. These mark a departure from the baseline model, as the additional firm-year FE and firm-destination FE absorb most controls of the standard entry model. In turn, this allows us to control for any unobserved time-varying influence at the firm level and firm's unobserved proclivity towards certain destinations. Results from alternative FE specifications remain very close to our benchmark estimates, indicating that these additional sources of correlated effects are not responsible for the observed network effects.

In appendix C.2.3 we repeat the benchmark estimates using an alternative network cutoff value $\nu_{ij,t} = 5\%$ to define relevant network linkages. Compared to the benchmark case, this significantly reduces the number of network linkages and thus the number of export signals diffusing through the network. While this reduces the chance of receiving export signals, specifying the network in this conservative manner still singles out matching export signals as the dominant source of network effects. The chosen network cutoff therefore does not determine the general mechanism at play but changes the marginal effect size of export signals which seem to increase when diffusion is constrained by a more conservative linkage cutoff.

Lastly, we restrict the sample to first-time exporters to ensure that network effects not only act as a catalyst for firms which already have a presence in foreign markets but also facilitate export transactions of firms that in the past have only served domestic customers. For this purpose, we define first-time exporters as firms that have not exported anywhere in the first 1, 5 and 10 years of the sample. Estimates obtained from running the benchmark model for these different subsets of firms are shown in appendix C.2.4. Compared to the baseline results, network effects not only remain significant but generally appear to be stronger for first-time exporters. This suggests that having access to the export experience of one's peers via network linkages is particularly important for first-time exporters as these firms have not yet gained any export experience of their own.

5.2.2 Model endogeneity

Up to this point, network effects were obtained under the assumption that export decisions of network peers $y_{jd,t-1}$ and network linkages $s_{ij,t-1}$ are exogenous. To relax this assumption, we employ the two-stage least squares (2SLS) approach and network selection model outlined in sections 4.1.2 and 4.3 to investigate if network effects persist once we allow both signals and linkages to become endogenous.

To rule out that any observed correlation of buyer and seller export decisions is driven by a common shock, we instrument export decisions of direct buyers with export decisions of indirect buyers that exclusively interact with sellers through intermediate buyers[41]. Results of this instrumentation strategy are reported in table 4. As expected, first-stage estimates in columns 1 and 3 reveal that the direct B2B relationship between first- and second-order buyers indeed results in a strong correlation of export decisions in period t-1, underlining the relevance of the proposed instrument. Controlling for this source of correlated effects in our main outcome equation (columns 2 and 4) leads to network effects which remain statistically significant but the marginal effect of export signals exceeds our benchmark estimates by a factor of 1.5-2.5. This

[41] As explained in section 4.1.2, we ensure that second-order buyers are not linked to sellers via higher-order linkages or located in close proximity to each other.

Table 4: Endogenous export signals - 2SLS

Dependent variable:	matching signals	export starts	matching signals	export starts
IV stages:	First	Second	First	Second
Variables				
second-order signals	0.3848***		0.4204***	
	(0.0669)		(0.0752)	
matching signals		0.0100**		0.0063**
		(0.0042)		(0.0025)
Peer characteristics	yes	yes		
Firm destination experience	yes	yes	yes	yes
Firm characteristics	yes	yes		
firm FE	yes	yes		
destination-year FE	yes	yes	yes	yes
firm-year FE			yes	yes
R^2	0.595	0.127	0.678	0.306
Observations	474,676	474,676	904,896	904,896

This table shows results of a 2SLS regression of equation 10. Endogenous export signals are instrumented by second-order signals. Columns 1 and 3 show first-stage results, columns 2 and 4 second-stage results. Standard errors in parentheses are clustered at the firm level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

not only indicates that accounting for unobserved shocks which create spurious correlation in seller and network export behavior is important but also reveals that true network effects might be much larger than the initial benchmark estimates suggest.

Next, we contrast benchmark estimates to a setting where firms choose linkages according to the dyadic network formation model presented in equation 12. Endogeneity concerns in this setting arise whenever unobserved shocks simultaneously affect domestic (Belgian B2B) and foreign (exporting) business transactions. We follow Arduini et al. (2015) and Qu et al. (2017) and construct a selection correction term which accounts for the resulting selection bias from correlated outcome and formation error terms ε and ξ. To render the required estimation of the dyadic formation model feasible for all 100k firms in our sample, we take several steps to reduce the dimension of the linkage formation process which are detailed in appendix C.3. One of these steps is to impose restrictions on the number of candidates a firm considers when forming domestic linkages. As the true number of candidates a firm considered but eventually did not decide to interact with is unobserved, we compute selection correction terms for different candidate set sizes which are expressed as the number of candidates per actual match.

Results of estimating equation 14 which includes selection correction terms as an additional regressors are presented in table 5. Despite the fact that selection correction terms are statistically significant, indicating that the underlying formation process is indeed endogenous, matching signals remain virtually identical to the benchmark estimates. This suggests that while a selection bias from endogenous network formation is present, it does not appear to be a major concern in our setting.

Table 5: Endogenous network linkages - selection correction

Candidates per match:	baseline	n=1	n=5	n=10	n=20
Variables					
matching signals	0.0042^{***}	0.0043^{***}	0.0043^{***}	0.0043^{***}	0.0043^{***}
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
selection correction		$2 \times 10^{-5**}$	$1.95 \times 10^{-5**}$	$1.93 \times 10^{-5**}$	$1.9 \times 10^{-5**}$
		(8.44×10^{-6})	(8.25×10^{-6})	(8.17×10^{-6})	(8.14×10^{-6})
Peer characteristics	yes	yes	yes	yes	yes
Firm destination experience	yes	yes	yes	yes	yes
Firm characteristics	yes	yes	yes	yes	yes
firm FE	yes	yes	yes	yes	yes
destination-year FE	yes	yes	yes	yes	yes
R^2	0.127	0.127	0.127	0.127	0.127
Observations	450,243	450,243	450,243	450,243	450,243

This table shows results of estimating equation 14 which accounts for endogenous network formation via a selection correction term. The selection correction term is based on the dyadic network formation model outlined in equation 12 and calculated for several buyer candidate sets which differ in size. Standard errors in parentheses are clustered at the firm level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

5.3 Signal heterogeneity

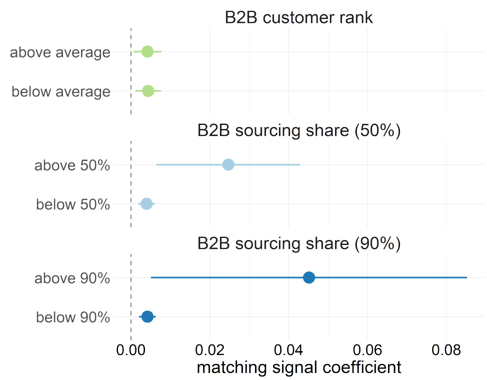
The baseline findings presented in section 5.1 point towards a clear export inducing effect of domestic B2B networks that appears to be robust to alternative model specifications and can accommodate common endogeneity concerns as discussed in section 5.2. These benchmark results however likely mask an important differences in network effects given the vast heterogeneity of the underlying network characteristics. In this section we therefore disaggregate network effects along linkage and peer characteristics to investigate under which conditions export signals are most conducive for export entry. Results are summarized in figure 9, further details available in appendix C.4.

We start by comparing how the strength of B2B linkages shapes the impact of export signals. To this end, panel (a) plots the estimated coefficients of matching signals that have been received via different linkage types. While the seller-specific rank of individual buyers does not seem to matter for the strength of network effects, receiving a signal from buyers which rely on a single seller for a majority of their sourcing (B2B sourcing shares above 50%) appears to have a stronger impact on subsequent entry decisions of sellers. While signals received from buyers with more diversified sourcing strategies still facilitate entry, this suggests that linkage dependency amplifies the impact of export signals.

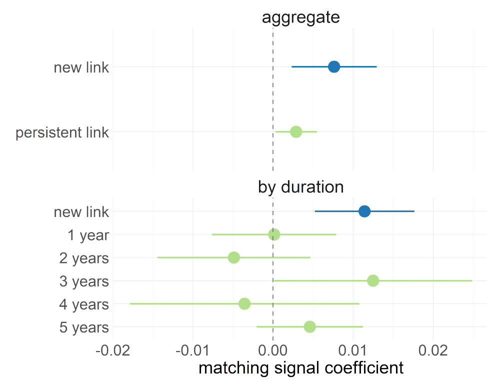
In panel (b) we study how the duration of the B2B relationship affects network effects. Although network linkages are often sticky given the non-negligible fixed costs involved in identifying suitable business partners (Martin et al., 2020), we find that both new and persistent linkages facilitate export entry. This suggests that sellers not only respond to signals received from trusted sources but also remain open to insights from new business partners.

A last linkage characteristic which we investigate in panel (c) is the direction of the underlying supply chain interaction. While our analysis focuses on backward linkages where sellers receive signals from their buyers, network effects might also arise from forward (the network of suppliers) or mixed linkages where firm i simultaneously acts as a buyer and a supplier for peers j. Coeffi-

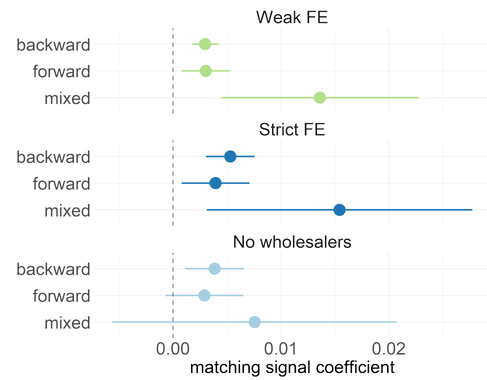
Figure 9: Matching signal heterogeneity



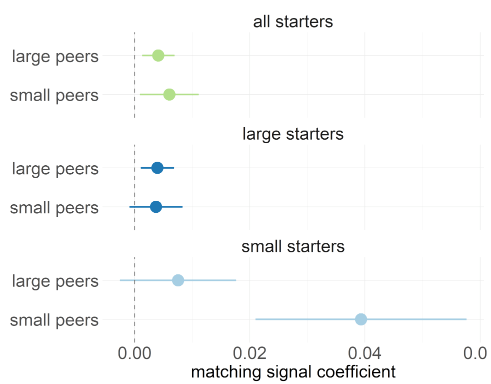
(a) linkage dependency



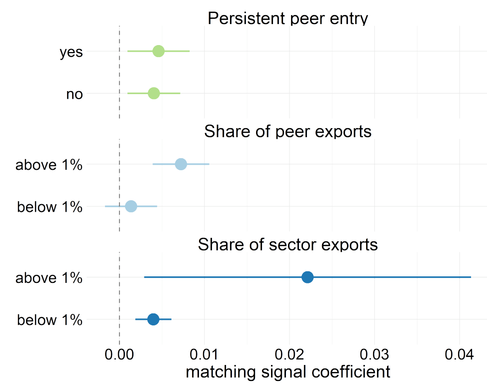
(b) linkage persistence



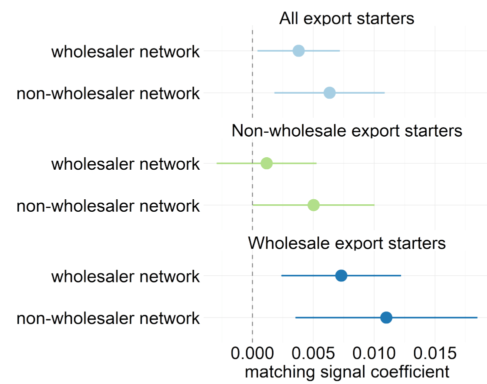
(c) linkage direction



(d) peer size



(e) peer credibility



(f) peer sector

cients obtained from our benchmark equation 10 with weak and strict FE suggest that all three linkage directions facilitate export entry. A striking results is the large effect of signals received from mixed linkages which are almost three times larger than backward and forward effects. A closer inspection of mixed linkages reveals that most of them involve large wholesalers which naturally hold a dual role as both buyers and sellers in the domestic production network. To ensure that our findings do not simply reflect entry decisions of wholesalers, we re-estimate our strict FE specification excluding export starts of firms i which operate in NACE sectors 45, 46, and 47. Our results show that non-wholesalers only benefit from signals received from backward linkages while forward and mixed linkages no longer promote export entry. Wholesalers which hold special position in domestic production networks therefore seem to benefit from a wider exposure to specialized export information along all three linkage directions, whereas all other firms exclusively benefit from signals received through backward linkages. As our study set out to study network heterogeneity for the entire population of firms, a focus on backward linkages appears to be reasonable. Nevertheless, we take the uncovered sectoral heterogeneity of network effects into account and present additional results for wholesalers in panel (f).

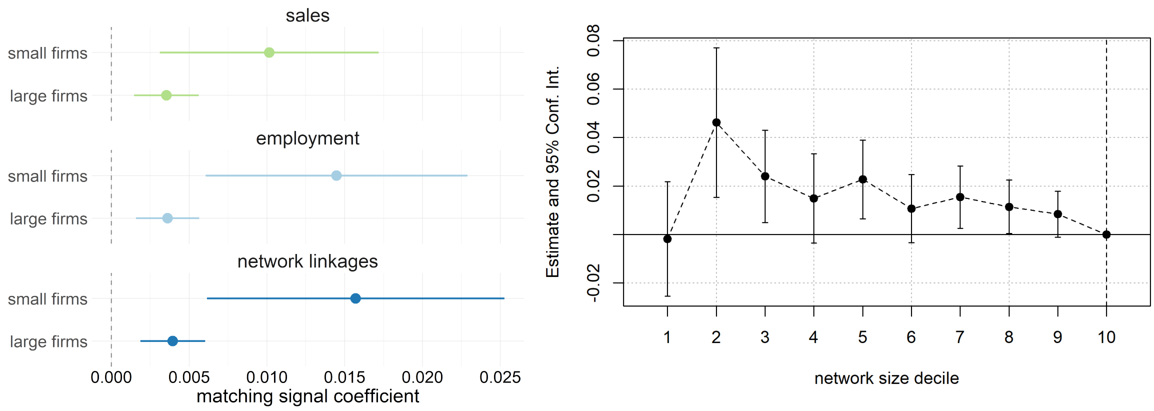
In panels (d), (e) and (f) we study the impact of peer heterogeneity on network effects. Panel (d) starts by investigating the role of peer size. Our findings show that export signals originating from small and large firms both matter for the observed conducive impact on export entry, but also reveal a substantial degree of homophily in the underlying network effects. While foreign market access large firms is disproportionately driven by interactions involving other large firms, the opposite is true for small firms. Our methodology is not equipped to uncover the underlying mechanism at play, but clearly suggests an unequal response of firms to interactions with peers of different size.

Next, we investigate whether the credibility of export signals shapes the entry behavior of sellers. Results are presented in panel (e). Our first disaggregation accounts for the export behavior of network peers one year after they emitted an export signal. Cases in which peers immediately leave the foreign market after their initial entry could indicate a bad experience which nullifies the positive impact of the emitted signal for the receiver. Our results do not corroborate this claim suggesting that peers' post-entry behavior has no impact on the strength network effects. Conversely, we do find that signals which are more credible because they account for a significant share in the peer's own or sectoral exports have a larger impact on the receiver. While our benchmark analysis treats each incoming signal equally, firms do seem to differentiate between signals they receive which likely leads to an underestimation of the true network effect.

Lastly, we return the role of wholesalers in panel (e). Trade intermediaries have been shown to play an important role for firms to access foreign markets by initially allowing them to circumvent high entry barriers via indirect exporting before ultimately entering the market directly (Connell et al., 2019). Our mechanism generalizes this idea by considering each firm's entire network as a source of promoting export entry. This allows us to separately account for the role of wholesalers and non-wholesalers in diffusion of export signals. Our results reveal that this distinction is crucial for understanding aggregate network effects. While export signals from wholesalers do contribute to foreign market access of other wholesalers, this is not true to the entire population of firms. Instead, the majority of network effects seems to be driven by non-wholesalers which underlines the importance of considering the entire production network when estimating network effects.

5.4 Network heterogeneity and the extensive margin of trade

After demonstrating that networks shape market access decisions of Belgian firms, we now turn to the consequences of network heterogeneity. As documented by Mayer and Ottaviano (2008), export sectors tend to be dominated by a small number of large firms which also happen to have the largest networks as shown in figure 5, which means they will on average receive more export signals. They key question we want to investigate is therefore whether network externalities increase in firm size and thereby contribute to the observed dominance of large firms at the extensive margin of trade.

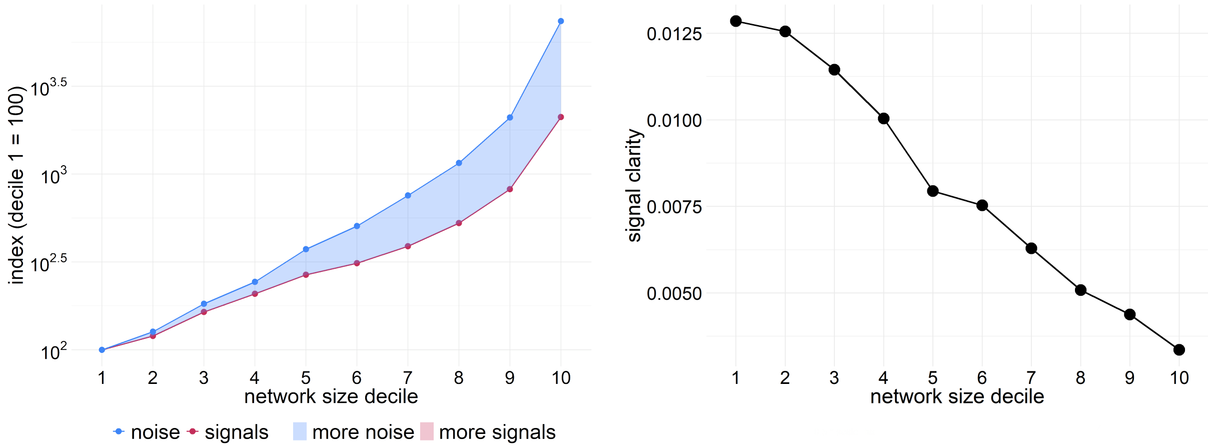


(a) subsample regressions (b) full sample with interaction term

Figure 10: Network effects by firm size

Figure 10 shows results of two distinct exercises to illustrate which firms benefit most from network effects. The left panel plots matching signal coefficients obtained from of six regressions in which the full sample has been divided into small and large firms based on median sales, employment and the number of network linkages. Irrespective of which firm size dimension is considered, we see that the marginal impact of export signals is significantly larger for small firms. For these firms, a single export signal increases the probability to start exporting by 1.0-1.6 percentage points, outpacing the impact on large firms by a factor of 2-4. To corroborate this large difference in network effects by firm size, we return to the full sample, group firms into network size deciles and adjust our benchmark equation by interacting network effects with firm size. The right panel plots the resulting matching signal coefficients by size decile which are expressed relative to firms with the largest networks (decile 10). Apart from firms with very small networks, marginal effects appear to almost linearly decrease in network size which indicates that network effects exhibit decreasing returns to network scale.

To rationalize this surprising result two explanations come to mind. First, repeated exposure to the same export signal might decrease marginal effects in large networks where the likelihood of receiving the same signal multiple times is higher. Unfortunately our data does not allow us to compare the content overlap of individual matching signals which means we cannot test to what extent signal redundancy can explain decreasing returns to network scale. Second, firms' ability to process export signals might be constraint by the presence of network noise. We define noise as the number of network interactions that do not yield matching export signals. As each maintained network linkage takes up time and resources at the firm irrespective of whether a linkage transmits export signals, firms with large networks might not benefit from the large number of export signals when simultaneously facing a disproportionate amount of network noise. The key question is therefore if firms with small and large networks are exposed to an unequal amount of valuable export signals *relative* to the amount of network noise.



(a) signal vs. noise growth rates

(b) signal clarity

Figure 11: Network noise by firm size

Figure 11a plots the average growth rates of export signals and network noise by size decile. Y-axis values of signals (red) and noise (blue) are expressed as an index which uses the first size decile as the reference category. As expected, both lines are monotonically increasing given that larger networks naturally include a larger number of valuable and non-valuable linkages. The main insight, however, is that both linkage types grow at different rates. While growth rates are comparable across small networks, noise appears to grow disproportionately in network size as indicated by the growing blue shaded area between both lines. Firms with large networks therefore receive more export signals in absolute terms, but at the same time are exposed to increasing levels of network noise. This trend is similarly depicted in figure 11b. Here we express network effects as the share of total linkages which yield matching export signals (signal clarity) as discussed in detail in section 2.4. High levels of network noise decrease signal clarity. This means that firms with large networks on average receive less export signals per linkage than firms with small networks.

This size penalty is closely related to the underlying network formation process. As shown in figure 6, the Belgian production network is characterized by negative assortative matching. Highly productive sellers can offer products at competitive prices which attracts a large number of buyers, many of which are not necessarily very performant themselves. While having a large number of unproductive buyers in the network may benefit the seller in other dimensions, it potentially mitigates the positive impact of the network for export participation as this group of buyers is unlikely to provide export experience to the seller. Instead, they emit network noise making it harder to respond to valuable export signals received from other network peers.

To see if our model can generate any support for detrimental impact of network noise, we re-estimate our benchmark equation using the signal clarity specification outlined in equation 9b. Network effects $\bar{S}_{t-1}y_{d,t-1}$ are again interacted with size deciles to see if firms' export behavior

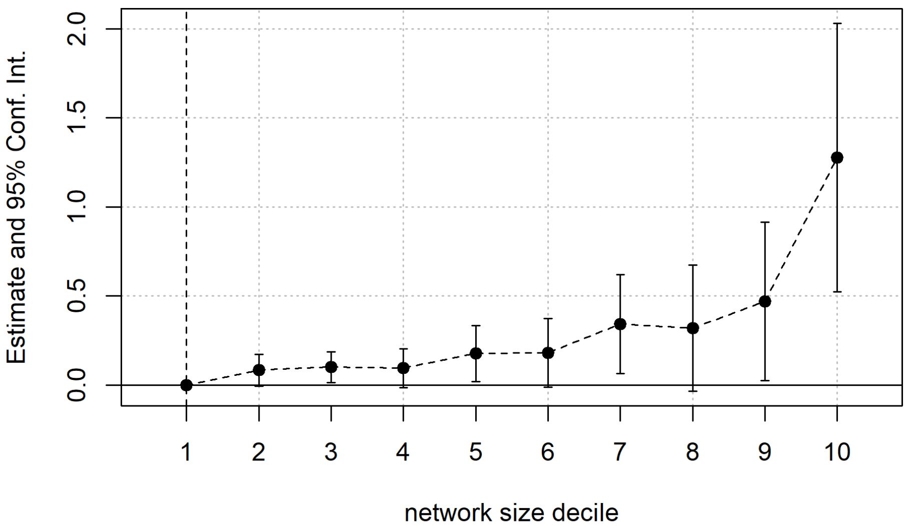


Figure 12: Signal clarity results by network size

changes when facing different levels of network noise. Figure 12 plots estimated signal clarity coefficients for each size group which indicate how sellers respond to incoming export signals which are received alongside network noise. A unit increase in signal clarity can be interpreted as a relative decrease in network noise, holding the absolute number of export signals constant. It therefore shows how firms would react to a change in their network configuration in which the relative share of exporting to non-exporting peers is higher. The results reveal two important findings. First, firms across size deciles benefit from lower levels of network noise in form of increased propensity to enter foreign markets. This indicates that network effects increase both in absolute (signal intensity) and relative (signal clarity) strength. Second, we see that a relative decrease in network noise is most beneficial for firms with large networks where this translates to the largest absolute reductions in network noise. Seeing that the ordering of the descriptive analysis presented above carries over to our benchmark model is important as it shows that network noise at least partially explains the observed decreasing returns to network scale. The alternative explanation of signal redundancy cannot generate this pattern, as it solely focuses on linkages which generate export signals and ignores any impact of network noise. We discuss the consequences of this finding in the next section.

6 Discussion

The results presented above demonstrate that networks act as an important determinant for export participation. Network linkages expose firms to the activity of others which can facilitate access to foreign markets if peers happen to be experienced exporters but also become a burden if not. These two central findings separate network effects from traditional sources of firm heterogeneity and carry important implications.

For trade theory, our results suggest that network linkages constitute an important dimension of firm heterogeneity that needs to be considered to understand firm behavior at the extensive margin of trade. Considering a setting with multidimensional firm heterogeneity relaxes the singular focus on firm productivity and offers a new perspective on export participation patterns that were hard to reconcile before. An example is the existence of small exporters which often fail to meet theory-implied productivity levels that would justify their participation in foreign markets. Network heterogeneity offers an explanation for this pattern as suitable network configurations might lower entry barriers to a particular market, allowing firms to enter despite insufficient levels of productivity.

Another important difference to traditional sources of firm heterogeneity is that network effects exhibit decreasing returns to scale. While productivity continues to increase export participation in our theoretical framework even at extreme levels, network effects dissipate as the network grows. This raises interesting questions regarding network efficiency. If network expansion adds noise to export signals but benefits the firm in other dimensions, there might be an optimal size which balances both effects. We are not convinced that learning about export markets is a first-order concern in the network formation process, which is why we disregard strategic network formation and treat network effects as an externality. Our work nevertheless highlights a potentially important but indirect cost of unrestricted network expansion that goes beyond direct search and matching costs typically associated with the network formation process.

Relatedly, we want to stress the importance of understanding the role of network assortativity in this context. Decreasing returns to network scale in our setting are closely linked to a distinct feature of production networks which are marked by a negative relationship between firm size and average peer performance. The opposite is often observed in social networks where agent and peer characteristics tend to be positively correlated. Realizing that the direction of assortativity varies across settings is important because a production network with positive assortative matching would have resulted in increasing returns to network scale. In this case, sellers would get access to more performant buyers as the network grows which decreases the average level of network noise and means that export participation monotonically increases in firm and network size. Any study mapping average network characteristics to outcomes should therefore be mindful of the underlying formation process and consider the possibility of decreasing returns to scale in settings marked by negative assortative matching.

On the policy side, our findings emphasize the role of information frictions in trade. Only matching signals stimulate foreign market access which indicates that informational cost barriers differ substantially across export destinations. Export promotion agencies often address this problem by investing considerable resources to provide a *select group of members* with up-to-date market information and organize costly matching events to connect domestic firms with foreign buyers. Our results suggest that domestic production networks can act as powerful tool to provide similar benefits to *all domestic firms* in a relatively cost-effective way. Instead of trying to directly link domestic firms to foreign buyers, policy makers could utilize the existing export experience in the network and facilitate the diffusion of specialized export information by creating new linkages among domestic firms. This could be an especially promising strategy to connect small and medium-sized enterprises (SMEs) to global markets as our results indicate that small firms stand to gain most from network externalities.

7 Conclusion

Export participation remains low across countries which causes concerns as it weakens competition in domestic markets (De Loecker and Warzynski, 2012) and restricts aggregate export growth (Eaton et al., 2009b). In this paper, we empirically investigate the determinants of export participation. A large preceding literature of heterogeneous firm trade models has emphasized the role of firm-level characteristics like productivity to rationalize observed entry patterns among firms.

We move beyond firm-level characteristics and investigate whether domestic production networks actively influence export entry decisions. Firms that interact with experienced exporters receive export-related information via network linkages which lowers sunk entry costs and thereby facilitates foreign market access. To formalize this mechanism, we introduce network interactions into a stylized model of export entry. Our augmented framework features firms which differ in both productivity and network linkages which allows us to assess the relevance of each dimension of firm heterogeneity for export participation.

To estimate the model, we rely on detailed data from the universe of Belgian firms which contains firm characteristics, imports, exports and domestic firm-to-firm transactions of every firm operating in Belgium between 2002-2014. Combined, these unique datasets allow us to observe each firm's individual network as well as the export behavior of network peers. Every time a firm starts to export to a new export destination, it emits an export signal to connected firms which contains valuable entry information and lowers the market access costs. The number of received export signals varies across firms as each network is unique. This creates a second dimension of firm heterogeneity beyond firm productivity.

Taking this model to the data reveals that network heterogeneity play a decisive role for export participation even after controlling for productivity. Each additional export signal received from the network increases the entry probability to a specific foreign market by 0.4-1.5 percentage points. While firms with large networks also receive the highest number of export signals, we find that the marginal effect of signals decreases in network size. We relate this size penalty to negative assortative matching in the underlying network formation process. Network expansions are associated with a disproportionate growth of network interactions that do not yield valuable export signals but still take up time and resources at the firm. We find evidence that this form of network noise mitigates the positive impact of network effects and is partially responsible for the observed decreasing returns to network scale.

Taken together, our findings demonstrate that network heterogeneity acts as an important new determinant of export participation but is unlikely to exacerbate the observed concentration at the extensive margin of trade. At the same time, they raise important questions regarding the strategic link between network formation and export participation and under which conditions networks should be considered efficient. These questions lie outside the scope of the current paper and require a more structural treatment of network linkages and firm outcomes. We believe our results provide important empirical evidence for this future avenue of research and will promote a stronger consideration of networks in international trade.

References

Advani, A. and Malde, B. (2018). Methods to identify linear network models: a review. *Swiss Journal of Economics and Statistics*, 154(1):1–16.

Albornoz, F., Calvo Pardo, H. F., Corcos, G., and Ornelas, E. (2012). Sequential exporting. *Journal of International Economics*, 88(1):17–31.

Anselin, L., Gallo, J. L., and Jayet, H. (2008). Spatial Panel Econometrics. In Mátyás, L. and Sevestre, P., editors, *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, Advanced Studies in Theoretical and Applied Econometrics, pages 625–660. Springer, Berlin, Heidelberg.

Arduini, T., Patacchini, E., and Rainone, E. (2015). Parametric and Semiparametric IV Estimation of Network Models with Selectivity. Technical Report 1509, Einaudi Institute for Economics and Finance (EIEF). EIEF Working Papers Series.

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

Arkolakis, C., Ganapati, S., and Muendler, M.-A. (2021). The Extensive Margin of Exporting Products: A Firm-Level Analysis. *American Economic Journal: Macroeconomics*, 13(4):182–245.

Badev, A. (2021). Nash Equilibria on (Un)Stable Networks. *Econometrica*, 89(3):1179–1206.

Bernard, A. B., Eaton, J., Jensen, J. B., and Kortum, S. (2003). Plants and Productivity in International Trade. *The American Economic Review*, 93(4):1268–1290.

Bernard, A. B., Jensen, J. B., Redding, S. J., and Schott, P. K. (2007). Firms in International Trade. *Journal of Economic Perspectives*, 21(3):105–130.

Bernard, A. B., Moxnes, A., and Ulltveit-Moe, K. H. (2018). Two-Sided Heterogeneity and Trade. *The Review of Economics and Statistics*, 100(3):424–439.

Bisztray, M., Koren, M., and Szeidl, A. (2018). Learning to import from your peers. *Journal of International Economics*, 115:242–258.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2020). Peer Effects in Networks: A Survey. *Annual Review of Economics*, 12(1):603–629.

Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009). Peer Effects and Social Networks in Education. *Review of Economic Studies*, 76(4):1239–1267.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies*, 47(1):225–238.

Chaney, T. (2016). Networks in International Trade. In Bramoullé, Y., Galeotti, A., and Rogers, B., editors, *The Oxford Handbook of the Economics of Networks*, pages 753–775. Oxford University Press.

Choquette, E. and Meinen, P. (2015). Export Spillovers: Opening the Black Box. *The World Economy*, 38(12):1912–1946.

Connell, W., Dhyne, E., and Vandenbussche, H. (2019). Learning about demand abroad from wholesalers: a B2B analysis. *National Bank of Belgium, Brussels*, NBB Working Paper(377).

De Loecker, J., Fuss, C., and Biesebroeck, J. V. (2014). International Competition and Firm Performance: Evidence from Belgium. *National Bank of Belgium, Brussels*, NBB Working Paper(269).

De Loecker, J. and Warzynski, F. (2012). Markups and Firm-Level Export Status. *American Economic Review*, 102(6):2437–2471.

Dhyne, E., Kikkawa, A. K., Mogstad, M., and Tintelnot, F. (2021). Trade and Domestic Production Networks. *The Review of Economic Studies*, 88(2):643–668.

Dhyne, E., Magerman, G., and Rubínova, S. (2015). The Belgian production network 2002-2012. *National Bank of Belgium, Brussels*, NBB Working Paper 288(288).

Eaton, J., Eslava, M., Kugler, M., and Tybout, J. R. (2009a). 8. Export Dynamics in Colombia: Firm-Level Evidence. In Helpman, E., Marin, D., and Verdier, T., editors, *The Organization of Firms in a Global Economy*, pages 231–272. Harvard University Press.

Eaton, J., Eslava, M., Kugler, M., and Tybout, J. R. (2009b). Export Dynamics in Colombia: Firm-Level Evidence. In *The Organization of Firms in a Global Economy*, pages 231–272. Harvard University Press.

Eaton, J., Jinkins, D., Tybout, J., and Xu, D. (2016). Two-sided Search in International Markets. *mimeo*.

Eaton, J., Kortum, S., and Kramarz, F. (2011). An Anatomy of International Trade: Evidence From French Firms. *Econometrica*, 79(5):1453–1498.

Fernandes, A. P. and Tang, H. (2014). Learning to export from neighbors. *Journal of International Economics*, 94(1):67–84.

Fernández-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large N , T. *Journal of Econometrics*, 192(1):291–312.

Gaulier, G. and Zignago, S. (2010). Baci: International trade database at the product-level. the 1994-2007 version. Working Papers 2010-23, CEPII.

Goldsmith-Pinkham, P. and Imbens, G. W. (2013). Social Networks and the Identification of Peer Effects. *Journal of Business & Economic Statistics*, 31(3):253–264.

Halleck Vega, S. and Elhorst, J. P. (2017). Regional labour force participation across the European Union: a time–space recursive modelling approach with endogenous regressors. *Spatial Economic Analysis*, 12(2-3):138–160.

Hinz, J., Stammann, A., and Wanner, J. (2021). State Dependence and Unobserved Heterogeneity in the Extensive Margin of Trade. *arXiv:2004.12655 [econ]*.

Hsieh, C.-S. and Lee, L. F. (2016). A Social Interactions Model with Endogenous Friendship Formation and Selectivity. *Journal of Applied Econometrics*, 31(2):301–319. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.2426.

Hsieh, C.-S., Lee, L.-F., and Boucher, V. (2020). Specification and estimation of network formation and network interaction models with the exponential probability distribution. *Quantitative Economics*, 11(4):1349–1390.

Iyoha, E. (2021). Estimating Productivity in the Presence of Spillovers: Firm-level Evidence from the US Production Network. page 77.

Javorcik, B. S. and Spatareanu, M. (2011). Does it matter where you come from? Vertical spillovers from foreign direct investment and the origin of investors. *Journal of Development Economics*, 96(1):126–138.

King, G. and Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, 9(2):137–163.

Koenig, P. (2009). Agglomeration and the export decisions of French firms. *Journal of Urban Economics*, 66(3):186–195.

König, M. D., Liu, X., and Zenou, Y. (2019). R&D Networks: Theory, Empirics, and Policy Implications. *The Review of Economics and Statistics*, 101(3):476–491.

Levinsohn, J. and Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies*, 70(2):317–341.

Liu, X., Patacchini, E., and Zenou, Y. (2014). Endogenous peer effects: local aggregate or local average? *Journal of Economic Behavior & Organization*, 103:39–59.

Manova, K. (2013). Credit Constraints, Heterogeneous Firms, and International Trade. *The Review of Economic Studies*, 80(2):711–744.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542.

Martin, J., Mejean, I., and Parenti, M. (2020). Relationship stickiness and economic uncertainty. page 37.

Mayer, T. and Ottaviano, G. I. P. (2008). The Happy Few: The Internationalisation of European Firms: New Facts based on Firm-level Evidence. *Intereconomics*, 43(3):135–148.

Mayer, T. and Zignago, S. (2011). Notes on CEPII's Distances Measures: The GeoDist Database. *SSRN Electronic Journal*.

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

Morales, E., Sheu, G., and Zahler, A. (2019). Extended gravity. *The Review of economic studies*, 86(6):2668–2712.

National Bank of Belgium (2002-2014a). Annual Account Filings. Confidential Dataset.

National Bank of Belgium (2002-2014b). International Trade Dataset. Confidential Dataset.

National Bank of Belgium (2002-2014c). Business-to-Business Transactions Dataset. Confidential Dataset.

Neyman, J. and Scott, E. L. (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 16(1):1–32.

Patault, B. and Lenoir, C. (2021). How valuable are business networks? Evidence from sales managers in international markets. *mimeo*.

Qu, X. and Lee, L.-f. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2):209–232.

Qu, X., Lee, L.-f., and Yu, J. (2017). QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Journal of Econometrics*, 197(2):173–201.

Zi, Y. and Bernard, A. B. (2021). Sparse Production Networks. *mimeo*.

A Additional descriptives

A.1 Export starts

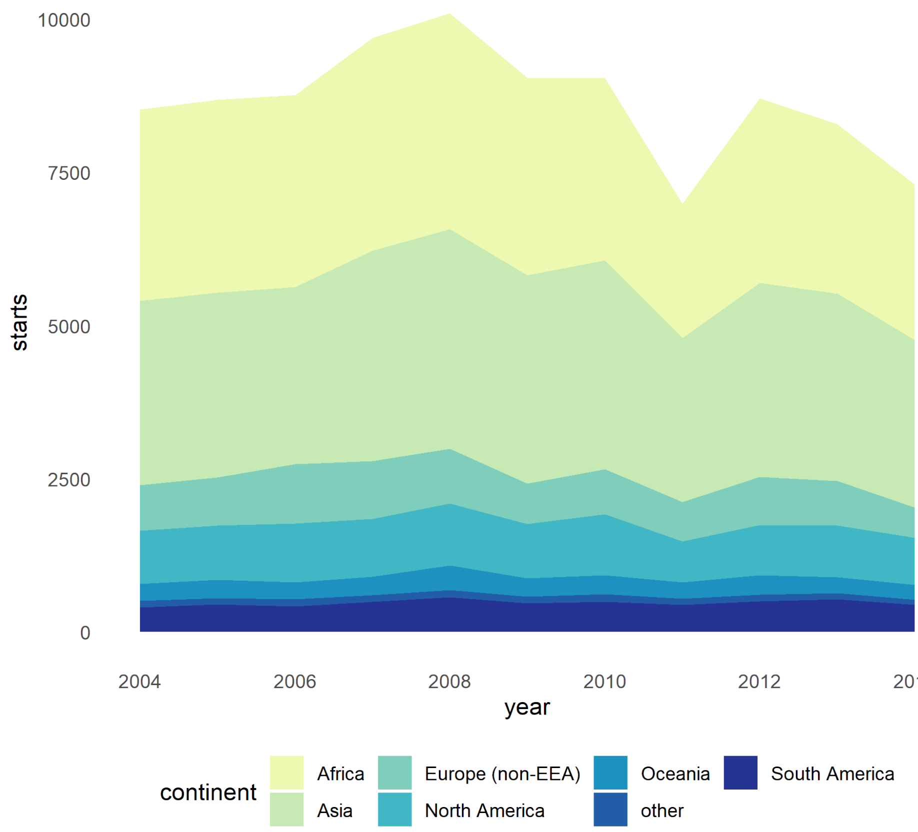
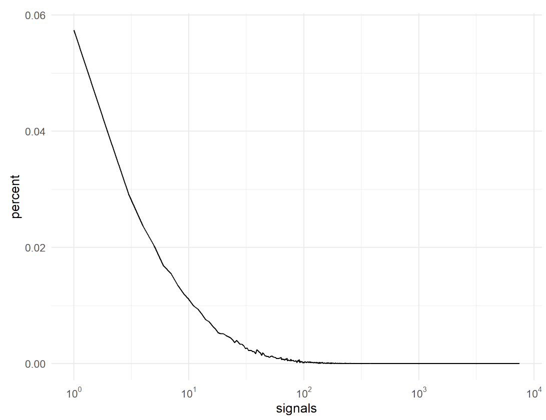
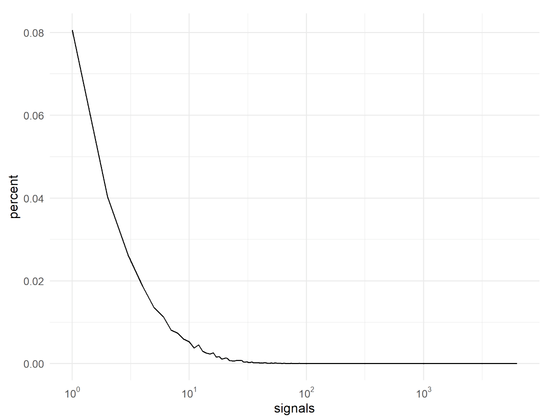


Figure 13: Share of non-EEA starts by region (2004-2014)

A.2 Export signals



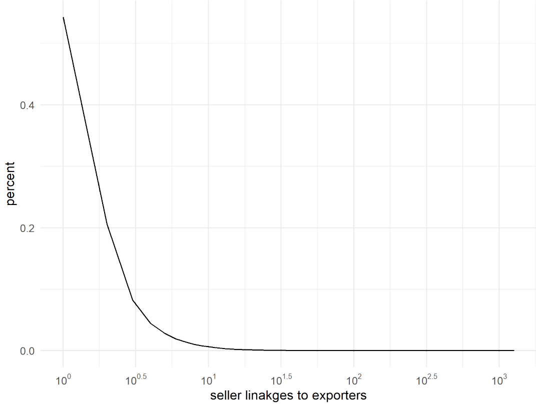
(a) any export signal



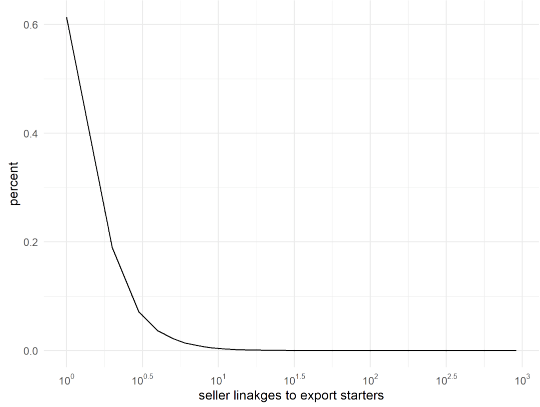
(b) matching export signal

Figure 14: Distribution of firms receiving export signals

A.3 Network linkages



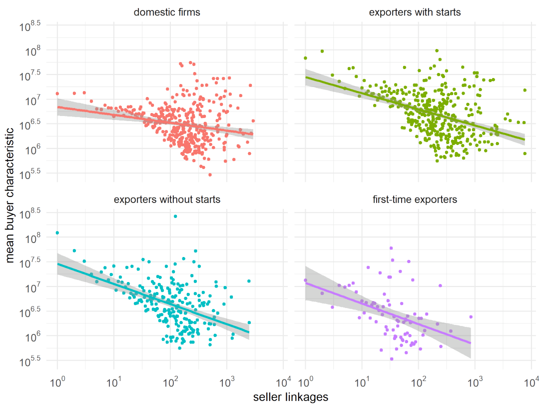
(a) linkages with exporters



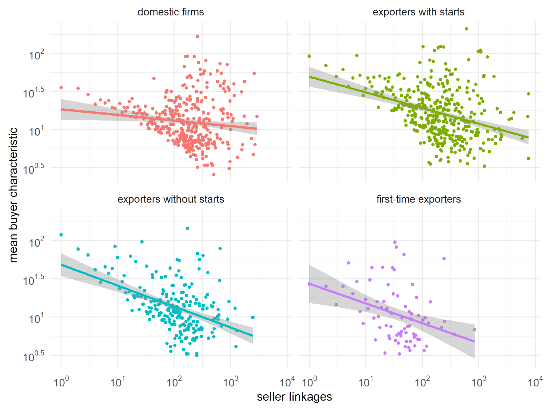
(b) linkages with export starters

Figure 15: Distribution of seller linkages

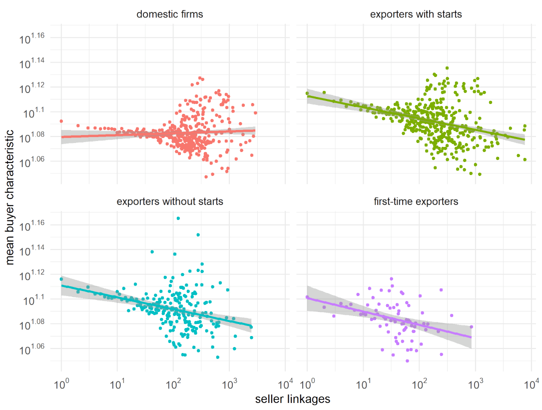
A.4 Network assortativity



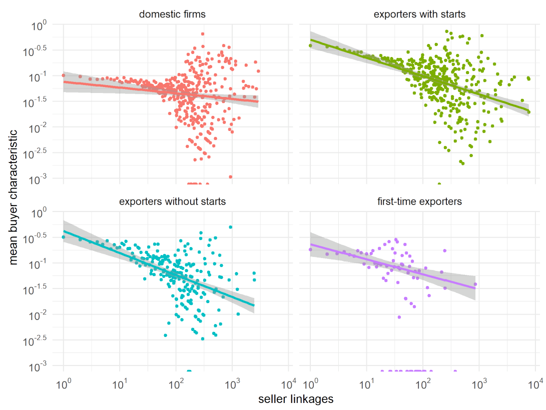
(a) average buyer sales



(b) average buyer employment



(c) average buyer productivity



(d) average buyer export start probability

Figure 16: Seller network size and mean buyer characteristics in 2014

B Additional dataset information

B.1 Reporting thresholds for trade transactions

The reporting thresholds differ across intra-EU and extra-EU trade transactions. Extra-EU export and import transactions follow a common reporting standard across all sample years. They are covered in the dataset if the transaction value exceeds 1,000 € or the volume is bigger than 1,000kg. In rare instances transactions below the minimum volume threshold are observed if the respective firm uses electronic reporting standards.
Intra-EU transaction thresholds are much higher and change over the sample period. Before 2006, they are reported if the combined import and export value of a firm exceeds 250,000 €. Between 2006-2010, the reporting threshold for imports was 400,000 € and 700,000 € for exports before both where harmonized to 700,000 € in 2010.
Our analysis mainly focuses on extra-EU transactions and therefore avoids measurement issues related to changing reporting thresholds or high threshold levels.

B.2 Construction of the regression sample

The variables used in our regression sample draw on the rich information contained in our merged dataset.

i. *Export starts* rely on detailed HS6 product-level export-transaction data which we aggregate to the firm-destination level. A firm with positive export transactions each year is counted as an exporter. An export start is defined as an export transaction to a destination that has not been served in the previous two periods. All observations within the two-year buffer period are dropped as firms by definition do not face an export decision. Likewise, non-starts are also only included in the data, if the firm has not been exporting in the past two years to ensure that a start could have occurred mean the firm faced a actual entry decision.

ii. Data on the number of *employees* and firm *wages* can be directly obtained from the available balance sheet data.

iii. *Total factor productivity* (TFP) is estimated using the approach of Levinsohn and Petrin (2003). The estimation requires data on firm sales, capital, labor and material inputs which are all available in the balance sheet data. Deflators for each input at 2-digit NACE codes are provided by the NBB based on internal price information. Our estimation is performed sector-by-sector and we only include sectors for which at least 50 non-missing observations are available.

iv. *Export experience dummies* rely on a combination of Belgian trade-transaction data for import and exports and the GeoDist database (Mayer and Zignago, 2011) freely available from CEPII's website. The latter includes information on bilateral relationships between all more than 200 countries including historic links and geographic borders. We merge this country relationship information with trade transaction data to create history and border dummies depending on the recorded relationship between Belgium and the respective trade partner. Import dummies on the other hand are only require the original trade transaction data and mark whether a seller has directly imported products from the future export destination. *Export sales shares* compare aggregate export values to sales information in the balance sheet records.

v. The idea for *export demand* is to capture the demand for the products underlying the export starts of Belgian firms in the foreign market prior to the actual export entry. To do so we proceed in several steps. First, we collect import data at HS6 product level for all destinations and sample years from the BACI database (Gaulier and Zignago, 2010) which we complement with WTO data for missing import information for Taiwan. Next, we identify the products underlying the export starts of each firm using the Belgian trade transaction database. For these products, we compute the export value at HS6 product-level in each destination originating in non-EEA countries. These non-EEA exports should capture changes in product demand in the destination without being correlated with Belgian exports due to common trade policy. For each firm, this gives us a proxy of how strongly their product was demanded in the destination prior to the export start. We then aggregate this export demand information to the firm-destination level and introduce it to the regression sample to control for the firm-specific export demand in each destination in each year t.

vi. Peer characteristics included in our regression sample are buyer *TFP* and buyer *sales* available from the Belgian balance sheet data. To relate buyer characteristics to sellers, we use row-normalized interaction matrices $\bar{S}_t$ and compute the average TFP and sales of buyers in a seller's network.

C Additional results

C.1 Benchmark regressions - full table

Table 6: Benchmark results - signal type

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
matching signals	0.0043***					
	(0.0011)					
non-matching signals		3.44×10^{-5}				
		(7.77×10^{-5})				
total signals			4.14×10^{-5}			
			(7.79×10^{-5})			
EEA signals				9.03×10^{-6}		
				(0.0001)		
border signals					0.0003	
					(0.0004)	
history signals						0.0002***
						(5.93×10^{-5})
log employment	0.0476***	0.0477***	0.0477***	0.0477***	0.0477***	0.0476***
	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)	(0.0041)
log wage	0.0127**	0.0127**	0.0127**	0.0127**	0.0127**	0.0127**
	(0.0060)	(0.0060)	(0.0060)	(0.0060)	(0.0060)	(0.0060)
log TFP	0.0325***	0.0325***	0.0325***	0.0325***	0.0325***	0.0326***
	(0.0040)	(0.0040)	(0.0040)	(0.0040)	(0.0040)	(0.0040)
log export demand	0.0179***	0.0179***	0.0179***	0.0179***	0.0179***	0.0179***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
border dummy	0.0693***	0.0694***	0.0694***	0.0694***	0.0694***	0.0694***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
history dummy	-0.0140***	-0.0140***	-0.0140***	-0.0140***	-0.0140***	-0.0140***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
export propensity	0.1702***	0.1701***	0.1701***	0.1701***	0.1701***	0.1701***
	(0.0115)	(0.0115)	(0.0115)	(0.0115)	(0.0115)	(0.0115)
import dummy	0.0980***	0.0982***	0.0982***	0.0982***	0.0982***	0.0982***
	(0.0034)	(0.0034)	(0.0034)	(0.0034)	(0.0034)	(0.0034)
log peer size	-0.0005	-0.0004	-0.0004	-0.0003	-0.0003	-0.0004
	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
peer TFP	0.0004	0.0003	0.0004	0.0003	0.0003	0.0004
	(0.0027)	(0.0027)	(0.0027)	(0.0028)	(0.0027)	(0.0027)
firm FE	yes	yes	yes	yes	yes	yes
destination-year FE	yes	yes	yes	yes	yes	yes
R^2	0.127	0.127	0.127	0.127	0.127	0.127
Observations	474,830	474,830	474,830	474,830	474,830	474,830

This table shows regression results of estimating equation 10 with a LPM-FE. Each column shows the marginal effect of receiving a different type of export signal on a seller's probability to start exporting. Standard errors in parentheses are clustered at the firm level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

C.2 Benchmark robustness

C.2.1 Comparison of estimation methods

Table 7: Robustness - Nonlinear models - signal intensity

Model	LPM-FE	Logit-FE	Logit-FE-IPP	Probit-FE	Probit-FE-IPP
Coefficient for *matching signal*	0.0043***	0.0348***	0.0346***	0.0187***	0.0186***
	(0.0011)	(0.0059)	(0.0060)	(0.0034)	(0.0034)
APE for *matching signal*	0.0043***	0.0050***	0.0051***	0.0047***	0.0049***
	(0.0011)	(0.00089)	(0.00090)	(0.00091)	(0.00091)
firm FE	yes	yes	yes	yes	yes
destination-year FE	yes	yes	yes	yes	yes
Observations	474,830	475,928	475,928	477,263	477,263

This table compares regression results of equation 10 for different linear and non-linear models. Standard errors in parentheses are clustered at the firm level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

C.2.2 Different fixed effect specifications

Table 8: Robustness - Fixed effects

Model:	(1)	(2)	(3)	(4)
Variables				
matching signals	0.0043***	0.0052***	0.0043***	0.0041***
	(0.0011)	(0.0008)	(0.0010)	(0.0010)
Peer characteristics	yes			
Firm destination experience	yes			yes
Firm characteristics	yes			
firm FE	yes			
destination-year FE	yes	yes	yes	yes
firm-year FE		yes	yes	yes
firm-destination FE			yes	yes
R^2	0.127	0.293	0.424	0.435
Observations	474,830	929,117	929,117	904,896

This table compares regression results of equation 10 for different fixed effect specifications. Standard errors in parentheses are clustered at the firm level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

C.2.3 Network threshold

Table 9: Robustness - 5% network threshold

Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
matching signals	0.0088***					
	(0.0022)					
non-matching signals		0.0001				
		(0.0002)				
total signals			0.0002			
			(0.0002)			
EEA signals				0.0001		
				(0.0005)		
border signals					0.0015*	
					(0.0009)	
history signals						0.0004*
						(0.0002)
Peer characteristics	yes	yes	yes	yes	yes	yes
Firm destination experience	yes	yes	yes	yes	yes	yes
Firm characteristics	yes	yes	yes	yes	yes	yes
firm FE	yes	yes	yes	yes	yes	yes
destination-year FE	yes	yes	yes	yes	yes	yes
R^2	0.124	0.124	0.124	0.124	0.124	0.124
Observations	363,607	363,607	363,607	363,607	363,607	363,607

This table compares regression results of equation 10 using a 5% buyer sourcing threshold to define relevant network linkages. Standard errors in parentheses are clustered at the firm level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

C.2.4 First-time exporters

Table 10: Robustness - First-time exporters

No export activity before:	baseline	2003	2006	2012
Variables				
matching signals	0.0043***	0.0092***	0.0079***	0.0112
	(0.0011)	(0.0028)	(0.0031)	(0.0077)
Peer characteristics	yes	yes	yes	yes
Firm destination experience	yes	yes	yes	yes
Firm characteristics	yes	yes	yes	yes
firm FE	yes	yes	yes	yes
destination-year FE	yes	yes	yes	yes
R^2	0.127	0.233	0.229	0.373
Observations	474,830	140,622	112,884	25,014

This table compares regression results of equation 10 focusing on sellers with no export experience at the firm-level up to the indicated year. Standard errors in parentheses are clustered at the firm level. Significance codes: ***: 0.01, **: 0.05, *: 0.1.

C.3 Endogenous network formation

This section provides additional details on how to obtain selection correction terms needed to estimate equation 14.

To begin with, we need to estimate the dyadic formation model outlined in equation 12 which poses two distinct challenges. A first challenge relates to the large sample size n. As the model requires us to estimate the linkage probability between any pair of firms operating in Belgium, each firm in theory considers all $n-1$ other firms as candidates for establishing a linkage. Including all $n * n - 1$ firm pairs in our dyadic formation model is not only computationally infeasible given our sample includes around 100k firms per year, but also highly unrealistic as firms are unlikely to consider the entire population of firms as matching candidates when searching for an individual business partner. A second challenge is that even if the size of candidate sets becomes computationally feasible, the exact candidate set a firm considered in the matching process remains unobserved. Both challenges require additional assumptions which we discuss in turn.

To reduce the dimension of the problem, we impose several restrictions on the n x n firm-to-firm interaction matrix. Instead of treating all $n - 1$ firms as potential matching candidates for an observed linkage $s_{ij,t} = 1$, we only consider firms as candidates if they operate in the same 4-digit NACE industry as the actual match and themselves have interacted with firms in the same 4-digit NACE sector of firm i. All candidates which do not meet these criteria are dropped from the candidate set which implies that they were never considered as potential business partners for the observed B2B linkage. This two-sided sector-specific restriction will create a distinct candidate set for each observed linkage and significantly reduces the size of the candidate set such that on average we are left with 50 candidates per observed match.

While this selection potentially introduces some error by potentially excluding individual candidates which firm i did consider as business partners in the matching process, we believe that restricting candidate sets to the sector of the actual match j is intuitive and expected to preserve the majority of true candidates.

As we do not know which of the 50 candidates firm i actually considered in the matching process, we restrict the number of candidates a firm considers per match to $n^{random} = \{1, 5, 10, 20\}$, take a random sample and estimate the dyadic formation model with a logit model for a given draw of n^{random} candidates plus the actual match. As shown in figure XXX, the estimated coefficients $\boldsymbol{\theta}$ of the network formation model which controls for individual firm employment and productivity, bilateral distance, common language and past linkage status hardly vary across random samples within a particular candidate set size n^{random}. This indicates that selection correction terms does not depend on the variation created from random sampling *within* given set size, but may still lead to different results due to differences *between* imposed set sizes n^{random}. To mitigate concerns that our results might be driven by the chosen candidate set size, we employ the rare events correction introduced by King and Zeng (2001) which accounts for the different ratios of events ($s_{ij,t} = 1$) to non-events ($s_{ij,t} = 0$) created by the choosing a distinct candidate set size n^{random}.

We then use the estimated formation coefficients $\boldsymbol{\theta}$ to predict linkage probabilities $p = P(S_{ij,t-1} = 1) = \frac{e^{U_{ij,t}(\theta)}}{1+e^{U_{ij,t}(\theta)}}$ and compute selection correction terms $\hat{\Xi}_{i,t-1}$ as shown in equation 15 in the main text.

C.4 Signal heterogeneity regressions

In this section we provide additional details for the signal heterogeneity results presented in 9. All panels show matching signal coefficients obtained from estimating equation 10 via a LPM-FE. The only difference to our benchmark results is that matching signal counts are disaggregated by linkage or peer type to assess the impact of underlying network heterogeneity. The disaggregation exercise in each panel uses the following definitions:

- Panel (a) uses three approaches to separate linkages into strongly and weakly dependent. First, by ranking ranking all buyers j based to their sourcing share from seller i. Buyers above the median rank are then defined as more dependent and vice versa. As an alternative, we use the observed sourcing shares to define strongly dependent buyers as those that source at least 50% (90%) of domestic inputs from seller i.
- Panel (b) defines linkage persistence as the number of consecutive years a seller-buyer pair ij have interacted with. Incoming export signals are then assigned according to the maturity of linkage ij in year t. To ensure all linkage maturities can be observed in our sample, regressions only consider entry decisions after 2006.
- Panel (c) considers export signals received from backward, forward and mixed linkages. The three linkage types reflect the relationship of buyers and sellers in the production network. Backward linkages capture signals received from buyers, forward linkages capture signals received from suppliers and mixed linkages capture signals received from firms which simultaneously act as buyers and sellers for firm i. The third regression in this panel excludes export starts of wholesalers by droppings operating in NACE sectors 45, 46 and 47 from the sample.
- Panel (d) disaggregates incoming export signal by peer size. We define large an small firms based on sales and use median sales as the cutoff value.
- Panel (e) studies the credibility of export signals in three distinct ways. First, by investigating whether peer entries are persistent or immediately leave the market in $t+1$. Second, by checking whether exports to the new destination account for more than 1% of total peer exports in that year. Third, by examining whether exports to the new market account for more than 1% of total exports in the firm's 4-digit NACE sector.
- Panel (f) finally uses NACE sectors 45, 46 and 47 to identify wholesalers and separetly counts export signals originating from wholesaler and non-wholesaler networks.