Sparse Production Networks*

Andrew B. Bernard† Yuan Zi‡

March 2022

Abstract

Firm-to-firm connections in domestic and international production networks are increasingly viewed as playing a fundamental role in economic outcomes. Firm heterogeneity and the sparse nature of firm-to-firm connections implicitly discipline network structure. We find that a large group of well-established statistical relationships are not useful in improving our understanding of production networks. We propose an "elementary" model for a production network based on random matching and firm heterogeneity and characterize the families of statistics and data generating processes that may raise underidentification concerns in more complex models. The elementary model is a useful benchmark in developing "instructive" statistics and informing model construction and selection.

JEL Classification: F11, F14

Keywords: firm-to-firm networks, model selection, balls-and-bins, buyer-seller matching, underidentification

*We thank Costas Arkolakis, Kirill Borusyak, Julia Cajal, Arnaud Costinot, Dave Donaldson, Alessandro Ferrari, Simon Galle, Pawel Gola, Edwin Leuven, Andreas Moxnes, Ralph Ossa, Kjetil Storesletten, Zi Wang, Zhihong Yu and participants at various seminars and conferences for valuable comments and discussions. Yuan Zi gratefully acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 715147, 819394).

†Tuck School at Dartmouth, NBER & CEPR, *email*: andrew.b.bernard@tuck.dartmouth.edu

‡The Graduate Institute of International and Development Studies (IHEID) & University of Oslo & CEPR, *email*: yuanzi.economics@gmail.com

1 Introduction

From the propagation of microeconomic shocks into the aggregate economy to the determinants of international trade flows to the sources of heterogeneity in firm size and performance, firm-to-firm connections in production networks are increasingly viewed as playing a fundamental role in economic outcomes. Recent research has taken advantage of the proliferation of data on international trade transactions and firm-to-firm connections in domestic production networks to establish a set of common empirical regularities on buyer-seller relationships and the evolution of the production network. These regularities, or stylized facts, have in turn disciplined approaches to modeling and testing the nature of firm-to-firm connections and their consequences.

In production networks there is a fundamental sparsity: buyers do not purchase from every seller of a given product. This fundamental sparsity, coupled with firm heterogeneity, implicitly disciplines the structure of production network. This paper shows that most of the established stylized facts on firm-to-firm networks are, in fact, not informative in developing associated models on how buyer-seller relationships are established, nor guiding related empirical tests. Conditional on buyer and seller heterogeneity, a simple allocation model with purchases from buyers randomly assigned to sellers based solely on buyer and seller size is equally capable of generating a sparse production network and the additional empirical facts used to motivate, and often verify, more complex models of firm-to-firm connections.

This simple allocation model of a production network, henceforth the "elementary" model, generates a rich set of statistical relationships that match those from numerous existing datasets. A key point is that as the data becomes more disaggregated, the more sparse, granular, and stochastic the data becomes. A discrete, stochastic model would thus naturally provide a more realistic characterization of data than a continuous, deterministic model. Continuity itself has implications for network structures, necessitating the introduction of additional mechanisms to explain certain network features. Forgoing continuity, an assumption generally made for modeling simplicity, firm heterogeneity can result in complex network structures that match the data well.

In addition to matching existing stylized facts and empirical results, we develop the general properties of the elementary model, characterizing the *families* of statistics it generates. Rather than having to take every empirical prediction of a model to the data, the researcher can easily pin down the corresponding the prediction of the elementary model for *any* empirical fact in a given family. This paper focuses on three issues confronting researchers deciding how to present empirical characteristics of their data: (1) What statistical transformation should be applied? (2) What level of aggregation should be considered? (3) What variables of interest should be examined and how should they be linked?

We present the connection of the elementary model to three general classes of statistical transformations: monotonic statistics such as intensive and extensive margins, order statistics such as percentiles, and higher-order relationship statistics such as assortativity. Then, we show that all results from the elementary model that hold at a disaggregated level are maintained with aggregation. Finally, we show that if two variables of interest, such as productivity and size, are first-order stochastically monotone in the realization of any variable that the model directly generates, e.g., pairwise flows between buyers and sellers, their correlation can be pinned down. These findings suggest that, on the one hand, the elementary model provides a natural benchmark for characterizing real-world production networks due to its simple yet rich properties; on the other hand, a large group of statistics may not be informative in providing insights on firm-to-firm relationships beyond random matching.

Our findings suggest that the elementary model can be a powerful tool to select informative statistics and benchmark more elaborated theories. Researchers can use the elementary model to identify statistics that are "instructive" for the development of more elaborate models of production networks. "Instructive" statistics can take two broad forms. First, those statistics which cannot be produced by the elementary model, i.e. they have a different sign than predicted by the elementary model. Second, related to a suggestion by Armenter and Koren (2014), While the signs of some statistical relationships may not be informative, the magnitudes may be if they differ substantially from the predictions of the elementary model. We use Norwegian transaction level trade data to illustrate both ideas with examples.

In addition, we propose to use the elementary model as a benchmark to identify necessary building blocks for applied theories. This is a slightly different thought experiment: instead of starting from data and the elementary model with minimum structure, we start from a candidate model and compare its explanatory power to the elementary model. For example, the matching probability between buyers and sellers (p_{ij}) across different buyer-seller bins in the data is close to what is predicted by the elementary model. In contrast, models emphasizing negative degree assortativity in matching perform much worse in every buyer-seller bin. This poor performance may appear puzzling at first glance as these models are developed to understand firm-to-firm networks. However, while models as such are intended to explain how buyers match sellers – the data counterpart is the *distribution* of p_{ij} – what they actually match are certain statistics of p_{ij}, such as the sign of assortativity. To match these statistics, strong assumptions are often imposed, which generate greater prediction errors in untargeted statistics and in the distribution of p_{ij}.

Finally, in the last part of the paper, we approach theory comparison from a structural perspective and introduce the idea of Bayesian model selection. We show that Bayes' theorem gives a natural characterization of how we move from empirical evidence to theory. It enables us to think about preferences for a certain theory as a product of its ability to explain the data and prior beliefs compared to alternatives. We show that the elementary model is so far the best baseline alternative, for selecting richer structural models of production networks.

The primary concern of random-allocation models, such as Armenter and Koren (2014), is that they are statistical models lacking economics. However, we show that if the allocation is designed with the correct economic counterpart, it can be viewed as an intermediate step of a discrete choice model. The elementary model can be derived from a competitive environment, with Eaton and Kortum (2002, henceforth, EK) being a special case. This view further suggests that when linking a model to data, the allocation problem needs to be designed with the correct data counterparts. Otherwise, implicitly, we are imposing additional assumptions (often inconsistent with empirical context or the model itself) on the distribution of firm-to-firm sales. We use balls-and-bins models designed by Bernard et al. (2018a) and Armenter and Koren (2014) as examples to illustrate this point. These bins and balls models fail to match relevant moments in the data either because the balls-and-bins counterpart is misspecified or because of additional assumptions made in the simulations.

The elementary model can also be used to understand network adjustments to shocks. We show a rich set of partner switching patterns between Mexican exporters and US importers in response the end of the Multifiber Arrangement (MFA) documented in Sugita et al. (forthcoming) can be explained by the elementary model, i.e., the changes in the existing network in response to shocks may not be sufficient to identify specific, more complex matching mechanisms between buyers and sellers. As transactions represent the foundation of any aggregate relationship, the findings also raise concerns regarding whether existing empirical regularities at more aggregate levels can distinguish between competing economic models. We illustrate this point by focusing on firms, showing that heterogeneous responses of firms to improved market access or exchange rate changes can be generated by heterogeneous firms randomly reshuffling buyer-seller links rather than firm-varying pricing-to-market adjustments, e.g., Berman et al. (2012).

Thus far, we have emphasized the importance and potential benefits of our findings. It should be clear that our work also has important limitations. Our approach is not useful in uncovering the microfoundations of firm heterogeneity. In most of the paper, we implicitly treat buyer and seller size as a function of some fundamental primitives, the identification of which would require additional data and research design. Therefore, our work should be viewed as complementary to studies that examine the sources of firm heterogeneity (e.g., Atkin et al., 2017; Cai and Szeidl, 2018; Alfaro-Urena et al., 2020).

The elementary model is a balls-and-bins model which is part of a set of classic problems in probability theory with wide applications in statistics and computer science.[1] Our paper is related to Armenter and Koren (2014), who first brought the balls-and-bins to international trade, and explained why the categorical nature of trade data may provide challenges in distinguishing sample zeros from fundamental zeros. In production networks there is a fundamental sparsity: buyers do not purchase from every seller of a given product. This fundamental sparsity, coupled with

[1] See Mitzenmacher and Upfal (2017) and Raab and Steger (1998) for a brief overview. Additionally, Roch (2015) provides an excellent reference for the modern treatment of discrete probability.

firm heterogeneity, implicitly disciplines the structure of production network. Unlike Armenter and Koren (2014), we consider a general version of the allocation problem, in which the number of transactions differs across buyers, and buyers and sellers are allowed to overlap. We characterize the model's statistical features under general classes of statistical transformations, where the extensive margin is one special case. It is important to note that the goal of this paper is not to propose any particular model or to explain any particular set of facts. Rather, we focus on the basic structures required for a model to adequately explain empirical findings on production networks, acknowledging the discreteness and sparseness nature of firm data at a very disaggregated level.[2]

Our analysis relies on a significant methodological contribution by introducing coupling methods to the production network literature and demonstrating that they are powerful tools for stochastic comparative statics analysis and in characterizing real-world firm networks. In spirit, our emphasis on sufficient model structure in explaining rich empirical facts is relate to Chaney (2018), who provides sufficient conditions for trade being inversely proportional to distance, and Costinot (2009), who offers a unifying perspective on the fundamental forces that shape comparative advantage.

Our work also adds to a large literature on model selection.[3] Even though model-selection methods have become increasingly important in other scientific disciplines (Hooten and Hobbs, 2015; Ding et al., 2018; Madigan and Raftery, 1994), model-selection methods are not commonplace in research on international trade or production networks. We show the problem of model selection generally exists in empirical production network studies, and we propose to use the elementary model to select theory and benchmark empirical results.

In terms of results, our paper is related research on how random allocation may lead to seemingly nonrandom outcomes. Ellison and Glaeser (1997) study whether the observed levels of geographic concentration of industries are greater than would be expected to arise randomly. Dingel and Tintelnot (2020) note that granularity could lead to confusion between sample zeros and fundamental zeros in spatial studies. Borusyak and Hull (2020) show how random shocks may lead to nonrandom exposure and bias estimation results. While the abovementioned papers focus on geography, our focus is on production networks.

The difficulty of describing production networks lies in their complex structure. In this paper we show the topology of production networks are primarily governed by simple yet robust organizing principles. In this regard, our paper contributes to the literature on statistical mechanics of complex networks (Barabási and Albert, 1999; Albert and Barabási, 2002; Onnela et al., 2007). The networks studied in this literature (e.g., world wide web, social network, nervous system) however exhibit very distinct topological and dynamic features, and hence are explained by different network formation

[2] In a recent paper, Herkenhoff et al. (2021) explain a similar set of existing stylized facts on exporters and importers in international trade by re-interpreting the Krugman (1980) model with randomly bundled varieties and statistical reporting thresholds for cross-border transactions. Their approach follows much of the literature by starting with a continuum of heterogeneous firms and adding constraints to match the sparse nature of the data.

[3] Gelman et al. (2013) provides an introduction to the topic.

principles (Barabási and Albert, 1999; Bollobás et al., 2001; Albert and Barabási, 2000).[4] Moreover, none of these papers provides any analytical results on comparative statics.

Overall, the contribution of this paper is threefold. First, we find that a large group of well-established statistical relationships may not be useful in improving our understanding of production networks. Second, we propose an elementary model to characterize the organizational principles of production networks and describe the general statistical families that may subject a more elaborate model to underidentification concerns. We show how coupling methods and the statistical dominance theorem can be useful when working with discrete stochastic models. Lastly, we demonstrate that our framework is a useful benchmark to identify statistics for model construction. As micro-level data from firms becomes increasingly available, the number of relationships or variable combinations that can be examined will rapidly grow. Our work will serve as a useful guide for connecting data to theory.

The remainder of the paper is organized as follows. In Section 2, we present a set of existing stylized facts on domestic and international trade networks. Section 3 presents the model and derives theoretical counterparts to the stylized facts. Section 4 considers the general properties of the elementary model. Section 5 shows that the elementary model can be derived from a competitive environment. Section 6 shows that the model also matches a rich set of data responses to shocks. Section 7 introduces the concept of "instructive" statistics as well as structural methods to guide more elaborate modeling of the production network. Section 8 concludes.

2 Stylized Facts on Production Networks

We first briefly review the existing empirical literature on firm-to-firm production networks, both international and domestic. We highlight a set of facts common to empirical work on a variety of countries. These facts have been the basis for many of the models of firm-level production networks.

Fact 1. *Production networks, both domestic and international, are sparse.*

In every production network examined to date, most buyers and sellers are not connected. In Japan (Bernard et al., 2019), fewer than 1 in 130,000 potential buyer-seller connections are active in the domestic production network, while in Belgium, the figure is 1 in 23,000 (Dhyne et al., 2015).

Even conditioning on firm participation in a particular international market, buyer-seller connections are sparse. For Colombia, Bernard et al. (2018b) find that fewer than 1 in 15,000 connections exist between Colombian firms that import and foreign firms that export to Colombia. Comparable

[4] One distinct feature of networks studied in these papers is positive degree assortativity (instead of the negative degree assortativity found in production networks). This among other things is explained by the growth and the preferential attachment principles. Generally speaking, these papers are more interested in the development (expansion) of networks. In spirit, they are closer to Chaney (2014), which studies the dynamic formation of an international network of exporters. A more detailed discussion of this model feature can be found in Chaney (2016).

numbers are 1 in 12500 for French exporters and importers from France and 1 in 9000 for Norwegian exporters and firms that import from Norway (Kramarz et al. 2020 and Bernard et al. 2018a).

Fact 2. *Firms in a production network are heterogeneous in the number of links and the value per link.*

Both the out-degree (number of buyers per seller) and in-degree (sellers per buyer) show substantial dispersion across firms. Large exporters (90th percentile) have 6.5 to 11 times as many foreign customers as the median exporter in Costa Rica, Uruguay and Ecuador (Carballo et al., 2018). Similar dispersion of connectivity is found in Norwegian exporters (Bernard et al., 2018a), Colombian importers (Bernard et al., 2018b) and in US, Japanese and Belgian domestic production networks (Atalay et al., 2011; Bernard and Moxnes, 2018; Bernard et al., forthcoming).

The value of trade between pairs of firms is also highly skewed. For the Belgian domestic network (Dhyne et al., 2015) and for Costa Rican, Ecuadorian and Uruguayan exporters (Carballo et al., 2018), the ratio of the sales of the 90th percentile to the 50th percentile pair is greater than 12.

Fact 3. *The largest firms in terms of sales have the most buyers and suppliers and reach the largest number of markets (cities or countries).*

The largest firms in terms of sales account for a large majority of economic activity in the production network. They connect to the most (domestic and foriegn) markets and have the greatest number of buyers and suppliers, e.g. Japan (Bernard and Moxnes, 2018), Norway (Bernard et al., 2018a), Chile (Blum et al., 2012), and Belgium (Dhyne et al., 2015).

Fact 4. *Negative degree assortativity: Firms with many connections, on average, trade with firms that are less well connected.*

Production networks typically exhibit significant negative degree assortativity: highly connected firms sell to more customers, but their average customer purchases from a relatively small number of suppliers. The same is true on the purchasing side, highly connected buyers, on average, buy from less-connected sellers. This relationship has also been confirmed using production-network data from various countries, time periods, and for both within- and cross-country networks (e.g., Lim 2020 for US Compustat data; Bernard et al. 2019 for Japan; Bernard et al. forthcoming for Belgium; Bernard et al. 2018b for Colombian imports). Negative assortativity has often been used to motivate relationship-specific fixed costs. In such instances, only the best sellers find it profitable to incur fixed costs and reach small buyers (e.g. Bernard et al., 2018a; Bernard and Moxnes, 2018; Lim, 2020; Sugita et al., forthcoming; and Serrano and Boguná (country level evidence), 2003).

Fact 5. *Hierarchy: Well-connected firms trade with a range of partners from the best-connected to the least. Firms with few connections match with well-connected partners.*

Notably, negative degree assortativity does not mean that well-connected sellers only sell to poorly connected buyers. In data, the best connected (most productive) firms often sell to a wider range of buyers, from the best connected (most productive) down to the least connected (least productive). Less connected (low-productivity) sellers, however, will only be able to connect to high-productivity partners. The hierarchy nature of production networks was presented explicitly in Bernard et al. (2019) and Blum et al. (2012), but it is a common feature shared by many datasets and has been implicitly or explicitly incorporated in most of the related production network models.

While these facts do not cover the entirety of the empirical work on firm-to-firm production networks, they are often used to motivate the need for more elaborate models of buyer-seller, or importer-exporter interactions. In the next section, we show that a parsimonious model of buyer-seller transactions based on random matching and firm heterogeneity can match these facts and then we generalize the results to broad classes of statistics.

3 A Random Allocation Model

We model the assignment of transactions from buyers to sellers as balls falling into bins. By doing so, we capture the discrete nature inherent in the micro-level data, as a firm's revenue consists of a finite number of orders placed by its customers, each of which is a discrete unit of observation and is the microfoundation of all aggregate statistics at, but not limited to the firm, buyer-seller, country, and origin-destination levels.

3.1 Basic Environment

Consider an economy with $\mathcal{M} \in \mathbb{N}$ buyers, indexed by subscript $j \in \{1, 2, ..., \mathcal{M}\}$. Each buyer makes $b_j \geq 0$ purchases and buyers are ordered such that $b_j \leq b_{j+1}$. By one purchase we mean one-unit of a product. To fix ideas, we first consider the case in which all purchases are identical in value (we relax this assumption in Section 4.5).[5] Similarly, let $\mathcal{N} \in \mathbb{N}$ be the number of sellers, indexed by subscript $i \in \{1, 2, ..., \mathcal{N}\}$. The probability that any purchase lands on seller i is characterized by s_i, with $s_i \in [0, 1)$ and $\sum_i s_i = 1$. Sellers are ordered such that s_i is weakly increasing in i.[6] As in equilibrium the expected size of a seller will be positively associated with s, we sometimes refer to s and b as seller size and buyer size, respectively.

The setup is as simple as it appears. It is based on only two natural and verifiable assumptions: (1) buyers have different purchasing powers and sellers have different abilities to attract buyers; and (2) who matches with whom is not deterministic. We refer to our formulation in terms of s, b and the allocation rule as the *"elementary model structure"*: any two models with the same formulation

[5] A sufficient assumption is that firms' total sales are positively related to total number of transactions, which is supported by all transaction-level data we have worked with thus far.

[6] Unless otherwise specified, we refer to a function as "increasing" ("decreasing") if it is strictly increasing (decreasing).

will give rise to the same network structure regardless of their micro-foundations. Correspondingly, we refer to our model as the *"elementary model"* of production networks.

It's worth noting that we make no assumptions on how sellers and buyers overlap. A node in the network can be a buyer and a seller at the same time, with sales and purchases being characterized by s and b, respectively. If a node has $b > 0$ and $s = 0$, it represents a final consumer; if $b = 0$ and $s > 0$, it represents the most upstream producer; and if the realized assignment for some nodes with $s > 0$ is zero, those nodes indicate sellers who have had no sales in a given period.

In the remainder of this section, we illustrate how the model matches the stylized facts shown in Section 2. Our main theoretical contribution, the generalized results, is introduced in Section 4. As our model belongs to the family of balls-into-bins models, sometimes when it is convenient, we will use "balls" and "bins" to explain the intuition of the results.[7]

3.2 Extensive Margin

Given the environment, the probability that buyer j matches seller i equals the probability that the seller receives at least one purchase from j:

$$p_{ij} = 1 - (1 - s_i)^{b_j}. \tag{1}$$

The probability p_{ij} is an increasing function of both seller and buyer size: larger sellers (buyers) are more likely to match partners of a given size than small sellers (buyers). In data, the size distribution of firms is highly skewed, meaning that the majority of firms have s_i close to zero and b_j being a small number. As a result, p_{ij} is close to zero for most of the firm-firm connections. The model therefore generates network sparseness (Fact 1).

The expected number of buyers that a seller matches is then given by the sum of p_{ij} over all buyers:

$$N_i^s = \sum_j p_{ij} = \sum_j (1 - (1 - s_i)^{b_j}). \tag{2}$$

Similarly, the expected number of sellers that a buyer matches is given by

$$N_j^b = \sum_i p_{ij} = \sum_i (1 - (1 - s_i)^{b_j}). \tag{3}$$

It is easy to verify that N_i^s increases in s_i and N_i^b increases in b_j. Thus, on average, larger buyers match more sellers, and larger sellers match more buyers. Intuitively, buyers who make more purchases have a greater chance of connecting with any seller; hence they are on average more likely to match with more sellers. Sellers with greater s are more likely to receive purchases from

[7] We can view sellers as bins of difference sizes, and buyers making b_j purchases as buyers holding b_j numbers of balls. None of our results depend on the assumption that buyers throw balls and sellers are bins, as will become clear in the subsequent sections.

any buyer and thus are more likely to match with more buyers on average (Fact 2).

Next, let us consider $\mathcal{C} \in \mathbb{N}$ locations indexed by $c \in \{1, 2, ..., \mathcal{C}\}$. The probability that seller i exports to location c equals the probability that the seller matches at least one buyer in c:

$$\mathbb{P}[D_{ic} = 1] = 1 - \Pi_{j \in c}(1 - s_i)^{b_j} = 1 - (1 - s_i)^{\sum_{j \in c} b_j}, \tag{4}$$

where D_{ic} is an indicator variable that equals one if seller i sells to c. We use $j \in c$ to indicate buyer j at location c. Therefore, the expected number of locations that a seller reaches equals

$$N_i^{sc} = \sum_c (1 - (1 - s_i)^{\sum_{j \in c} b_j}). \tag{5}$$

Again, N_i^{sc} increases in s_i because each element in the sum is an increasing function of s_i. This analysis is equivalent to treating all buyers at location c as a single representative buyer with $\sum_{j \in c} b_j$ purchases, hence the same intuition follows. Similarly, one can show that the number of locations that a buyer reaches also increases with its size. Thus the model matches Fact 3.

3.3 Intensive Margin

Since purchases are of equal size, the total purchases of buyer j from seller i is captured by the number of orders from j to i, which we denote by x_{ij}. The probability that $x_{ij} = x$ is given by

$$\mathbb{P}[x_{ij} = x] = C_{b_j}^x s_i^x (1 - s_i)^{b_j - x}, \tag{6}$$

where $C_{b_j}^x$ denotes the combination formula. Thus, x_{ij} follows the binomial distribution $Bin(b_j, s_i)$. Let X_{ij} denote the expected purchase amount from j to i; we then have

$$X_{ij} = b_j s_i. \tag{7}$$

Thus, the model predicts that large sellers on average sell more to a given buyer and larger buyers purchase more from a given seller. Perhaps more interestingly, the model suggests that expected value of pairwise flows between buyers and sellers follows gravity rules (without the distance component), which in general ensures a model's good fit to data.

Moreover, the binomial distribution has the following property:

$$b_j \geq b_{j'},\ s_i \geq s_{i'} \ \Rightarrow\ Bin(b_j, s_i) \succ Bin(b_{j'}, s_{i'}). \tag{8}$$

In words, total sales tend to be bigger between larger seller and buyers. Combined with the ranking theorem, it is immediate that for any variable y_{ij} that is a nondecreasing function of x_{ij}, we will have

$$b_j \geq b_{j'},\ s_i \geq s_{i'} \ \Rightarrow\ E(y_{ij}) \geq E(y_{i'j'}), \tag{9}$$

which holds with strict inequality if $(b_j - b_{j'})(s_i - s_{i'}) \neq 0$. This implies that for any relationship-specific variable that is increasing in sales, such as relationship-specific investment or price discount, its expected value will be greater for larger buyers (sellers). Similarly, if a variable is negatively associated with x, the model will predict that, on average, the value of that variable is smaller for larger buyers (sellers). We will come back to this point with a general treatment in Section 4.4.

Equations (2), (3), and (7) suggest that in our model, heterogeneity in the size distribution of firms naturally translates to network heterogeneity in the number of links and the value per link (Fact 2).

3.4 Degree Assortativity

Assortativity examines the extent to which a network's nodes attach to others that are similar in some way. Although the specific measure of similarity may vary, existing studies of production networks often examine assortativity in terms of a node's degree. Unlike many other types of networks, production networks exhibit significant negative degree assortativity. This data feature has often been used to motivate the modeling of relationship-specific fixed costs. For example, only the best sellers find it optimal to incur fixed costs to reach small buyers. However, negative degree assortativity naturally emerges from the elementary model.

The in-degree of a buyer is nothing but the number of sellers the buyer matches. Its expected value is given by equation (3) and increases in b_j. The average connections of sellers that a buyer matches, by the law of iterated expectation, can be written as the sum of each seller's expected total connection times its relative probability of matching that buyer:

$$\bar{N}^s(j) = \frac{\sum_i N_i^s p_{ij}}{\sum_i p_{ij}} := \sum_i N_i^s g(i;j),$$

where $\bar{N}^s(j)$ denotes the expected average out-degree of sellers that buyer j matches, and the probability mass function $g(i;j) := \frac{p_{ij}}{\sum_i p_{ij}}$ gives the relative probability of i matching j – in other words, $g(i;j)$ characterizes the distribution of sellers conditional on matching buyer j. Intuitively, if a buyer makes one random purchase, she has a greater probability to match with bigger sellers; but if a buyer makes an infinite number of random purchases, she matches with every seller almost surely. In other words, $g(i;j)$ is more left-skewed for small buyers, i.e., smaller buyers have *relatively* greater probability of matching with big sellers. Thus, the expected average out-degree of sellers that a buyer matches is adversely correlated with her size and, as a result, with her connection. (Fact 4).[8]

[8]The formal proof can be found in Appendix A. In the next section, we also show that this correlation persists

3.5 Hierarchy

As emphasized in the literature, negative degree assortativity does not mean that well-connected sellers only sell to poorly connected buyers. In data, the most productive firms typically sell to a broader range of buyers, from the most productive (best connected) down to the least productive (less connected). Low-productivity sellers, on the other hand, only connect with high-productivity buyers (Bernard and Moxnes, 2018). At first glance, this appears to be strong support for fixed-cost hierarchy models, in which only firms with significant economic scale can overcome fixed costs and engage with small partners.

However, this finding also arises naturally from the elementary model. The best way to demonstrate this is to construct a coupling, which we do in the next section, so we will focus on the intuition for now. The key is to understand that, in comparison to a smaller seller, while a seller with greater s is more likely to match *any* buyer, it is *relatively* more likely to match small buyers. The intuition is essentially the same as that underpins negative degree assortativity. For example, consider two buyers with $b_1 = 1, b_2 = 100$, as well as a seller with $s = 0.01$. The seller has 1% chance of matching with the first buyer, and $1 - 0.99^{100} \approx 100\%$ chance of matching with the second. In contrast, when $s \to 1$, the matching probability for both buyers is approximately equal to one. While both sellers are quite likely to match with the large buyer in this example, the large seller has a significantly higher probability of matching with the smaller buyer.

4 Generalization

The previous section introduced the elementary model and showed how it matched existing stylized facts on production networks. We now introduce the general properties of the elementary model, characterizing the *families* of statistics it produces. Data on production networks are usually quite detailed, hence the number of relationships or variable combinations studied by researchers might quickly grow. Instead of exhausting all combinations in data and checking their corresponding elementary-model prediction one at a time, it is useful to characterize general predictions of the model. By doing so, researchers can easily employ our results to compare new empirical regularities from their data with those generated by the elementary model.

We consider three choices confronting researchers deciding how to present empirical characteristics of their data. (1) What statistical transformation should be applied? In particular, we examine three classes of statistical transformations: monotone statistics, order statistics and higher order relationships. (2) What level of aggregation should be considered? Depending on the question, the researcher might be interested in transactions, firm-level variables or even country-level aggregates. (3) What variables of interest should be examined and how should they be linked? The fundamental

with aggregation, matching the finding of Serrano and Boguná (2003) that the partners of well-connected countries are, on average, less-well connected themselves.

of the elementary model is a firm-to-firm transaction but studies might be interested in characterizing the network based on firm productivity or other variables. We explore the general properties of the elementary model along these three dimensions. Finally, we also provide robustness checks of our results by considering several model extensions.

4.1 Ranking and Stochastic Dominance Theorems

To facilitate the analysis, we first introduce two theorems:

Ranking Theorem. *Cumulative G^X first-order stochastically dominates (FSD) cumulative G^Y if and only if for any increasing and piecewise differentiable function $u(z)$,*

$$E_{G^X}[u(z)] > E_{G^Y}[u(z)].$$

Let $\succ$ denote first-order stochastic dominance. If there are two random variables such that $X \succ Y$, we also refer to X and Y as *statistically ordered in FSD*. Intuitively, ranking theorem tells us the stochastic ordering (in the first-order sense) of two random variables preserves with monotonically increasing transformations.

Stochastic Dominance Theorem. *The random variable $X \succ Y$ if and only if there is a coupling $(\hat{X}, \hat{Y})$ of X and Y such that*

$$\mathbb{P}[\hat{X} \geq \hat{Y}] = 1,$$

which holds with inequality for some values. We refer to $(\hat{X}, \hat{Y})$ as a monotone coupling of X and Y.

To compare two probability measures X and Y, sometimes it is useful to construct a joint probability space, i.e. the coupling $(\hat{X}, \hat{Y})$, with its marginal distributions being X and Y. The right-to-left direction of stochastic dominance theorem says that if we can find such a probability space, with $\hat{X} \geq \hat{Y}$ being a sure event, then we know that $X \succ Y$. The left-to-right direction of the theorem says that if $X \succ Y$, there always exist a coupling $(\hat{X}, \hat{Y})$ such that $\hat{X} \geq \hat{Y}$ surely.

When dealing with probabilistic models, the ranking theorem and stochastic dominance theory are useful tools. Both theorems, in a broad sense, allow us to compare random variables in the same way that we compare deterministic variables. When we use the notion of first order stochastic dominance to compare two random variables, we are anchoring *values* and comparing *probabilities*. The stochastic dominance theorem, on the other hand, allows us to translate the problem to anchoring *probabilities* and comparing *values*. Economic models usually apply functional transformations on the value side, hence when using the stochastic dominance theorem, we don't need to alter our anchoring every time when making comparisons. As a result, we can perform comparative statics for probabilistic models in the same way that we do for deterministic models.

4.2 General Statistical Transformations

4.2.1 Monotone Statistics

Denote the assignment result of seller i by $\mathbf{x}_i^s = (x_{i1}, ..., x_{i\mathcal{M}}) \in \mathbb{N}^{\mathcal{M}}$, and let $t : \mathbb{R}^{\mathcal{M}} \to \mathbb{R}$ be some statistical function of $\mathbf{x}_i^s$.[9] Consider two sellers i and i', with $s_i > s_{i'}$. We can couple $\mathbf{x}_i^s$ and $\mathbf{x}_{i'}^s$ in the following way: first, set $\hat{\mathbf{x}}_{i'}^s = \mathbf{x}_{i'}^s$. Then, set $\hat{\mathbf{x}}_i^s$ as the sum of two random vectors: the first random vector equals $\hat{\mathbf{x}}_{i'}^s$ with probability one, and the second random vector is generated independently according to the distribution of $\mathbf{x}_{i''}^s$ with $s_{i''} = s_i - s_{i'}$.[10] That is, $\hat{\mathbf{x}}_i^s = \mathbf{x}_{i'}^s + \mathbf{x}_{i''}^s$. Intuitively, the realization of purchases to seller (bin) i can always be viewed as the union of the realizations in two smaller-sized bins; one with size $s_{i'}$ and the other with size $s_i - s_{i'}$.

Then, instead of directly comparing the statistics of $\mathbf{x}_i^s$ and $\mathbf{x}_{i'}^s$, we only need to compare the statistics of their coupling. Note that by construction, $(\hat{\mathbf{x}}_i^s, \hat{\mathbf{x}}_{i'}^s)$ is a monotone coupling of $\mathbf{x}_i^s$ and $\mathbf{x}_{i'}^s$. Therefore, for any monotonic t, $(t(\hat{\mathbf{x}}_i^s), t(\hat{\mathbf{x}}_{i'}^s))$ is also a monotone coupling of $t(\mathbf{x}_i^s)$ and $t(\mathbf{x}_{i'}^s)$. This greatly simplifies our analysis because instead of computing the exact distribution for every statistic t that we are interested in, we only need to examine the monotonicity of t. Then, we can directly apply the stochastic dominance theorem to determine the stochastic ordering of $\{t(\mathbf{x}_i^s), t(\mathbf{x}_{i'}^s)\}$.

For example, consider the number of buyers that seller i matches with, which we denote by n_i^s. The random variable n_i^s is generated from the following statistical transformation: $t_1(\mathbf{x}_i^s) = \sum_j 1_{x_{ij}>0}$. It is trivial to show that $t_1(\hat{\mathbf{x}}_{i'}^s + \hat{\mathbf{x}}_{i''}^s) \geq t_1(\hat{\mathbf{x}}_{i'}^s)$ because the realized number of balls from each buyer can only weakly increase when we add the realization of $\mathbf{x}_{i''}^s$. Therefore by construction,

$$\mathbb{P}[\hat{n}_i^s \geq \hat{n}_{i'}^s] = 1.$$

By the stochastic dominance theorem, we then have $n_i^s \succ n_{i'}^s$. The ranking theorem subsequently implies that $N_i^s > N_{i'}^s$.

Formally, given two random vectors $\mathbf{X}$, $\mathbf{Y} \geq \mathbf{0}$,[11] a statistical transformation t is said to be *monotonically increasing* if $t(\mathbf{X} + \mathbf{Y}) \geq t(\mathbf{X})$ for any realization of $\mathbf{X}$, $\mathbf{Y}$, which holds with strict inequality for some realizations. Similarly, t is said to be *monotonically decreasing* if $t(\mathbf{X} + \mathbf{Y}) \leq t(\mathbf{X})$. We then summarize the result in the following proposition.

PROPOSITION 1. *For $s_i > s_{i'}$, $t(\mathbf{x}_i^s) \succ t(\mathbf{x}_{i'}^s)$ if t is a monotonically increasing transformation; $t(\mathbf{x}_{i'}^s) \succ t(\mathbf{x}_i^s)$ if t is a monotonically decreasing transformation.*

[9] To save notation, throughout the paper, we use t to refer a general statistical transformation and $t1, t2, ...$ to denote specific statistical transformations (functional forms are specified in the paper) regardless the dimension of the input vector.

[10] Two random vectors are said to be *equal* if they are equal as functions on their probability space. If two random vectors $\mathbf{X}$ and $\mathbf{Y}$ are equal in distribution and $corr(\mathbf{X}, \mathbf{Y}) = 1$, we say that $\mathbf{X}$ and $\mathbf{Y}$ are *equal with probability one*, or $\mathbf{X} = \mathbf{Y}$.

[11] We follow the convention to express vector inequality in the following way: any two vectors $\mathbf{X} \leqq (\geqq)\mathbf{Y}$ means that each component of vector X is less (greater) than or equal to each corresponding component of vector Y. If $\mathbf{X} \leq (\geq)\mathbf{Y}$, it means that $\mathbf{X} \leqq (\geqq)\mathbf{Y}$ and $\mathbf{X}! = \mathbf{Y}$.

Applying proposition 1, it is immediate that the size of the largest buyer that seller i matches with FSD that with which seller i' matches. In contrast, the size of the smallest buyer that seller i' matches with FSD that with which seller i matches. As a result, we will always find that larger sellers match with more dispersed buyers. Moreover, the above arguments are true regardless of the metric considered. For instance, the largest buyer can be defined by total purchases, the number of connections, or the most productive seller with which it matches.

We construct the coupling on the buyer side in a similar way. Denote the assignment result of buyer j by $\mathbf{x}_j^b = (x_{1j}, ..., x_{\mathcal{N}j}) \in \mathbb{N}^{\mathcal{N}}$. Consider two buyers, j and j', with $b_j > b_{j'}$. First set $\hat{\mathbf{x}}_{j'}^b = \mathbf{x}_{j'}^b$. Then, let $\hat{\mathbf{x}}_j^b$ equal the sum of two random vectors: one equals $\hat{\mathbf{x}}_{i'}^s$ with probability one, and the other is generated independently according to the distribution of $\mathbf{x}_{i''}^b$ with $b_{i''} = b_i - b_{i'}$. Intuitively, as purchases are not ordered, making b_j purchases can always be viewed as one first make $b_{j'}$ and then the other $b_i - b_{i'}$ purchases. Then, the following proposition is immediate:

PROPOSITION 2. *For $b_j > b_{j'}$, $t(\mathbf{x}_j^b) \succ t(\mathbf{x}_{j'}^b)$ if t is a monotonically increasing transformation; $t(\mathbf{x}_{j'}^b) \succ t(\mathbf{x}_j^b)$ if t is a monotonically decreasing transformation.*

4.2.2 Order Statistics

Next, we consider another class of commonly used statistical transformations: order statistics. Most order-statistic functions, such as quantiles or percentiles, are not monotone. Nevertheless, given our environment, the sample average of these statistics can be easily ordered. We use percentiles to illustrate the point. One can construct similar proofs to examine other order statistics and the associated rank statistics.

Without loss of generality, consider the τ_{th} percentile buyer of seller i. Let $O(\tau; i)$ denote its expected rank. Obviously, $O(\tau; i)$ is an increasing function of τ: higher the percentile, by definition we refer to the larger-valued buyer in the matched list. Denote $\xi_i(z) := \frac{\sum_{j=1}^{z} p_{ij}}{\sum_j p_{ij}}$ as the expected fraction of buyers whose rank is smaller than z. Then, by definition, $O(\xi_i; i) = z$, and the inverse function $O^{-1}(z; i)$ is also an increasing function.

Let $g(j; i) := \frac{p_{ij}}{\sum_j p_{ij}}$. It is easy to verify that $g(j; i)$ has the monotone likelihood ratio (MLR) property, which is a sufficient condition for $g(j; i')$ FSD $g(j; i)$ when $s_i > s_{i'}$. Therefore, it is immediate that $\xi_{iz} \geq \xi_{i'z}$ for all z by the definition of FSD, i.e., the expected fraction of buyers that are below an arbitrary rank is always greater for large sellers. By definition $O^{-1}(z; i) := \xi_{iz}$, hence it is immediate that $O^{-1}(z; i) \geq O^{-1}(z; i')$. Combining this with the fact that $O^{-1}(z)$ increases in z, we conclude that $O(\tau; i) \leq O(\tau; i')$ for all τ. A similar proof can be constructed for buyers. □

PROPOSITION 3. *The average rank of the τ_{th} percentile buyer (seller) that a seller (buyer) matches with decreases in seller (buyer) size.*

Intuitively, if a seller increases its size, it becomes more likely to match with every buyer, but more importantly, it becomes relatively more likely to match with smaller buyers. This pushes up every existing buyer's rank, which in turn means for a given percentile, the expected rank of its associated buyer decreases.

4.2.3 Higher-order Relationships

Production network properties are often described by higher-order relationships. The assortativity discussed in the previous section is one example, various forms of upstreamness and downstreamness measures are another. In this subsection, we show how to pin down the elementary-model predictions on higher-order relationship statistics in general.

The trick is to note that a higher-order statistic can be viewed as the composition of a series of statistical transformations. Thus, we can construct a statistic generating function for k_{th}-order relationships. Let $\phi_t^k(\iota)$ describe the k_{th}-order statistics given a statistical function t. This function applies to both buyers and sellers with a slight abuse of notation: when $\iota = i$, the statistics are calculated for a seller, and when $\iota = j$, they are calculated for a buyer. We use $-\iota$ to denote partners on the other side of the relationship: when ι refers to a seller, then $-\iota$ refers to a buyer and vice versa. The variable $\mathbf{x}_\iota$ describes the assignment result. Then, the function $\phi_t^k(\iota)$ can be constructed as:

$$\phi_t^k(\iota) = \begin{cases} E(t(\mathbf{x}_\iota)) & \text{if } \ k = 1, \\ \frac{\sum_{-\iota} \phi_t^{k-1}(-\iota) p_{-\iota\iota}}{\sum_{-\iota} p_{-\iota\iota}} = \sum_{-\iota} \phi_t^{k-1}(-\iota) g(-\iota;\iota) & \text{if } \ k > 1. \end{cases}$$

For instance, let $t_1(\mathbf{x}_\iota) = \sum_\iota 1_{\mathbf{x}_\iota > 0}$, where $\mathbf{x}_\iota$ refers to the realized assignment for ι. Buyer j's expected connectivity can therefore be written as $\phi_{t_1}^1(j)$, and the expected average connectivity of its matched seller is $\phi_{t_1}^2(j)$. Generally speaking, assortativity is nothing but a correlation between one second-order and one first-order statistic.

Recall that the distribution $g(i;j')$ FSD $g(i;j)$ when $b_j > b_{j'}$, and the distribution $g(j;i')$ FSD $g(j;i)$ when $s_i > s_{i'}$. By iteration, we arrive at the following proposition.

PROPOSITION 4. *Let $\phi_t^k(\iota)$ and $\phi_{t'}^{k'}(\iota)$ be two higher-order relation statistics across buyers (or sellers). When both transformations t and t' are monotonically increasing or decreasing, $\phi_t^k(\iota)$ and $\phi_{t'}^{k'}(\iota)$ are positively associated if $(-1)^{k+k'} = 1$ and negatively associated if $(-1)^{k+k'} = -1$. When one of t and t' is monotonically increasing and the other is monotonically decreasing, $\phi_t^k(\iota)$ and $\phi_{t'}^{k'}(\iota)$ are positively associated if $(-1)^{k+k'} = -1$ and negatively associated if $(-1)^{k+k'} = 1$.*

For example, consider the out-degree of sellers against the expected average in-degree of buyers with which sellers match. The latter is given by:

$$\bar{N}^b(i) = \frac{\sum_j N_j^b p_{ij}}{\sum_j p_{ij}} := \sum_j N_j^b g(j;i) = E_{G_i}(N_j^b).$$

By definition $\bar{N}^b(i) := \phi_{t_1}^2(i)$ and $\bar{N}_i^s := \phi_{t_1}^1(i)$. Since $t1$ is monotonically increasing, and $k + k' = 3$, from proposition 4 we immediately know that sellers' connection also exhibits negative assortativity, i.e, the average customer of a highly connected seller purchases from a relatively small number of suppliers.

We can also view $\phi_t^k(\iota)$ as a $g(-\iota;\iota)$-weighted sum of $\phi_t^{k-1}(-\iota)$. In some cases, one may want to use weighting functions other than $g(-\iota;\iota)$. Similar to the above analysis, we can pin down the correlation between the higher-order statistics by examining the monotonicity of associated statistical transformations and the stochastic dominance property of the weighting function.

4.3 Arbitrary Aggregation

In the elementary model, any relationship that holds at the disaggregated level is maintained with aggregation. Because we can always view a set of buyers (denote the set by I) as one buyer with $\sum_{j\in I} b_j$ purchases and a set of sellers (denote the set by I') as one seller whose size given by $\sum_{i\in I'} s_i$. This yields our next proposition:

PROPOSITION 5. *Relationships that hold at a disaggregated level are preserved under aggregation in the elementary model.*

Consider for example the firm-level statistics at the intensive margin. As buyers are independent, the total number of purchases that fall into seller i also follows a binomial distribution:

$$\mathbb{P}[x_i = x] := \mathbb{P}[\sum_j x_{ij} = x] = C_{\sum_j b_j}^x s_i^x (1 - s_i)^{\sum_j b_j - x}. \quad (10)$$

The sum of independent binomial random variables is itself a binomial random variable if they share the same success probability. Therefore the probability of x purchases falling into i is identical to $\mathbb{P}[x_{iJ} = x]$, where the representative buyer J has $\sum_j b_j$ purchases. In words, the ordering of sales at the bilateral level preserves at the firm level, firms with higher s_i on average not only sell more to each buyer but also have greater total sales.

4.4 General Variables of Interest

The elementary model essentially generates only one outcome: the realization of purchase assignment $\{x_{ij}\}_{\mathcal{N}\times\mathcal{M}}$. Therefore, all statistical relationships that the model can directly match are necessary functions of $\{x_{ij}\}_{\mathcal{N}\times\mathcal{M}}$. However, we are sometimes interested in characterizing the network based on other firm features, such as productivity or R&D intensity. If we observe, for example, that in the data, a more productive buyer matches, on average, with less productive sellers, to what extent can this empirical regularity inform us about the underlying matching mechanism?

We formalize the question as follows: let the buyer-seller assignment follow the elementary model, and consider a firm-specific variable of interest, ψ. Under which data generating process can $\{\psi_\iota\}$ be FSD ordered in ι, so that the results of the previous subsections could be applied immediately?

Clearly, if (1) ψ is a monotone deterministic function of ι or (2) ψ is FSD generated in ι, that $\{\psi_\iota\}$ can be FSD ordered in ι is trivially true. Moreover, if (3) ψ is a monotone deterministic function of the realizations of t, where t can be any transformation of $\{x_{ij}\}_{\mathcal{N}\times\mathcal{M}}$ that has an FSD order or (4) ψ is FSD generated based on such a t, $\{\psi_\iota\}$ is also FSD ordered in ι (either in ascending or descending order). Take productivity as an example. One example of case (1) is that we write a model in which firm size is an increasing function of productivity. In case (2), the productivity distributions of small and large firms are allowed to overlap, but larger firms have a greater probability of being more productive. If we consider an environment where firms' productivity is a function of their realized sourcing outcomes $\mathbf{x}_j^b = (x_{1j}^b, ..., x_{nj}^b)$ and productivity is a nondecreasing transformation of $\mathbf{x}_j^b$, it belongs to case (3). Finally, if in case (3), a firm's productivity is not a deterministic function of its realized sourcing outcomes but a stochastic one following an FSD ordering, it corresponds to case (4). We can view the settings of Melitz (2003), Bernard et al. (2003), and Antras et al. (2017) as examples of (1), (2), and (3), respectively.

If productivity falls into one of the above four data generating processes, the elementary model predicts that more productive buyers will match with less *productive* sellers on average. Consequently, we won't be able to draw any conclusions about how firm matching behaviors differ depending on productivity. To summarize, buyer or seller attributes can be uninformative in assisting us in understanding how firms match as long as in data they are stochastically monotone in fundamental model parameters or in realized variables that follow a FSD ordering.

4.5 Robustness

In this section we discuss a few extensions of the basic model to show that the results are robust with respect to a number of modeling choices we made for convenience. In the interests of space, we present the results and their intuitions. Formal proofs and further discussions are relegated to the Appendix.

First, we discuss how our results may or may not differ when purchases are of unequal size. Two purchases of unequal size can be viewed as two purchases with the same price but different quantities (bundled purchase) or two purchases with the same quantity but different prices. But as long as we are interested in sales-related statistics, this distinction is not important: we loosely refer to the value of a purchase as purchase size. Introducing purchase-size heterogeneity is equivalent to adding more parameters to the model; hence, we can view purchase-size as "another variable of interest". Then, the results from Section 4.4 directly apply. For instance, when purchase size is weakly increasing in b_j, s_i, or x_{ij}, all results derived in our paper remain unchanged.[12]

Secondly, we consider the case in which assignments are not i.i.d events. Without loss of generality, let one purchase be randomly allocated at a time. Assume that at time one, the probability of a seller i receiving a purchase from any buyer is still s_i. At time $1+\zeta$, for any $\zeta > 0$, its probability of receiving a purchase from buyer j is given by:

$$s_{ij,1+\zeta} = \frac{\Omega\left(s_{ij,\zeta}, 1(x_{ij,\zeta} > 0)\right)}{\sum_{i'} \Omega\left(s_{i'j,\zeta}, 1(x_{i'j,\zeta} > 0)\right)}, \tag{11}$$

where Ω is an increasing function of both s and $1(x > 0)$. That is, the probability that a purchase is assigned to i at time $1+\zeta$ is positively associated with the seller's general attractiveness to j and their realized matching outcome in the previous period. The denominator reflects the fact that the unconditional probability needs to be normalized across all sellers. Given the setting, we show in the Appendix that $\{s_{ij,\zeta}\}$ can be FSD ordered ascending in i for all buyers and periods, hence all our results remain unchanged. Intuitively, equation (11) describes a "rich-get-richer" phenomenon: assignments are more likely to arrive at seller who the buyer matched previously, but the previous chance of matching is still positively associated with s_i to start with. One can also incorporate richer dynamics, as long as Ω and the stochastic ordering of the input vector of Ω are monotone.

Finally, in practice, all empirical correlations are calculated based on realized assignments instead of expected assignments. The sufficient condition for a correlation having the same sign when we compare firms based on realized assignments is that the variable of interest in the x-axis follows a monotonic probability ratio ordering (MPR), a condition that is stronger than first-order stochastic ordering but weaker than monotonic likelihood ordering (Eeckhoudt and Gollier, 1995). Reassuringly, under the elementary model structure, this actually covers all x-axis statistics that the literature has used to date.

5 Linking to Economics

One argument on the limitation of random-allocation models, such as the one proposed by Armenter and Koren (2014), is that they are statistical models lacking economics. In Section 5.1, we show that the elementary model can be derived from a competitive environment, where models with an EK structure are a special case. This finding in turn greatly improves our ability to map the elementary model to data to analyse or falsify economic problems, which we explain with examples

[12]Without loss of generality, we can always choose s_i such that for the given minimum purchases, the expected sales of a seller equal its observed sales. We maintain this view of normalization throughout our analysis.

in Section 5.2.

5.1 Elementary Model from a Competitive Environment

Consider an environment where buyers choose the seller that provides the highest utility for each of their purchase. Buyer preferences over seller i's product are drawn from a distribution H_i and denoted by ψ_i. The distributions $H_1, ..., H_{\mathcal{N}}$ are independent, admissible, and FSD ordered. We can view H as a function of some model primitives, such as various product characteristics, including price.

Intuitively, $H_i \succ H_{i'}$ implies that i always has a higher probability of receiving better utility draws; hence, its probability of being chosen is always higher in the competitive equilibrium. Given the setting, the probability that a purchase goes to seller i is given by:

$$s_i = \mathbb{P}[\psi_i > max(\psi_{-(i)})],$$

where $\psi_{-(i)}$ is the vector of $\mathcal{N}-1$ random variables excluding ψ_i. For simplicity, let us focus on the case in which H_i is continuous and differentiable for all i. Consider two sellers, $H_i \succ H_{i'}$, then we have

$$s_i := \mathbb{P}[\psi_i > max(\psi_{-(i)})] = \mathbb{P}[\psi_i > max(\psi_{-(i,i')})]\mathbb{P}[\psi_i > \psi_{i'}].$$

Note that $max(\psi_{-(i,i')})$ is also a random variable. Let $J_{-(i,i')}$ denote its cumulative distribution. Thus,

$$\begin{aligned}\frac{s_i}{s_{i'}} &= \frac{\mathbb{P}[\psi_i > max(\psi_{-(i,i')})]\mathbb{P}[\psi_i > \psi_{i'}]}{\mathbb{P}[\psi_{i'} > max(\psi_{-(i,i')})]\mathbb{P}[\psi_{i'} > \psi_i]} \\ &= \frac{\int_{-\infty}^{\infty} J_{-(i,i')}(z)dH_i(z)}{\int_{-\infty}^{\infty} J_{-(i,i')}(z)dH_{i'}(z)} \frac{\int_{-\infty}^{\infty} H_{i'}(z)dH_i(z)}{\int_{-\infty}^{\infty} H_i(z)dH_{i'}(z)} \\ &> \frac{\int_{-\infty}^{\infty} J_{-(i,i')}(z)dH_i(z)}{\int_{-\infty}^{\infty} J_{-(i,i')}(z)dH_{i'}(z)} \frac{\int_{-\infty}^{\infty} H_i(z)dH_i(z)}{\int_{-\infty}^{\infty} H_i(z)dH_{i'}(z)},\end{aligned}$$

where the inequality is ensured by $H_i(z) < H_{i'}(z)$. In addition, by $H_i \succ H_{i'}$ and both $J_{-(i,i')}(z)$ and $H_i(z)$ being increasing in z, the ranking theorem implies that both fractions on the right hand side (RHS) of the above expression are larger than 1. Therefore, $s_i > s_{i'}$. □

Note that the competitive-selection environment described above is the same as in EK. The Fréchet distribution is just one of many distributions that satisfy the FSD-ordering property: if one Fréchet distribution has the same shape parameter but a greater level parameter than another Fréchet, the former FSD the latter. When assuming that H_i follows the Fréchet distribution, the probability of a seller i matching a buyer j for a transaction in our model gives exactly the same expression as the probability that country i provides a good at the lowest price in country j in the EK model. With additional assumptions on the utility function and continuity, the model will collapse to EK, with sellers being isomorphic to exporting countries in the EK model.

5.2 Linking the Model to Data: Design Matters

As random-allocation can be viewed as an economic instead of a purely statistical process, this in turn disciplines how we should link the model to data when analyse certain economic problems. To illustrate this point we provide two examples with existing random matching (bins and balls) designs: the examination of negative degree assortativity in Bernard et al. (2018a) and the analysis of the fraction of exporting firms by Armenter and Koren (2014).

In their paper on two-sided heterogeneity and exporter-importer matching, Bernard et al. (2018a) ask if the negative degree assortativity among Norwegian exporters and Swedish importers can be captured by a simple balls and bins model. They conclude that a balls and bins model does not produce negative assortativity and use that result to argue in favor of a model with match-specific fixed costs. However, negative assortativity is obtained by treating each transaction in the Norwegian trade data as a ball and assigning these balls randomly according to the elementary model, as shown in Figure 1.[13]

The source of difference lies in their exercise design. Bernard et al. (2018a) consider $\mathcal{N} \times \mathcal{M}$ bins, with $\mathcal{N}$ being the number of exporters from Norway and $\mathcal{M}$ being the number of importers from Sweden. Each bin size is set as the product of the relative degree of the buyer and seller: $\hat{n}_i^s = \frac{n_i^s}{\sum_i n_i^s}$, $\hat{n}_j^b = \frac{n_j^b}{\sum_j n_j^b}$, and $s_i = \hat{n}_i^s \times \hat{n}_j^b$. The total number of balls from Sweden is set as the total number of existing connections: $n =: \sum_i n_i^s = \sum_j n_j^b$. The balls are then assumed to be assigned to bins at random, with the probability of any ball landing on a bin i being s_i.

There is an immediate problem with this setup. If a ball is a connection, once a ball has landed in a bin (an importer-exporter pair) that bin should be closed. Instead they allow bins to catch more than one ball. As a result, the model-simulated expected number of connections does not, by construction, match the sample mean, which renders inconsistency in the model design in the first place.[14] This example emphasizes the importance of a correctly designed balls and bins model in assessing empirical regularities of the data and highlights the benefits of the general results presented in Section 4.

Armenter and Koren (2014) is the first and most well-known example of a bins and balls model in international trade. The paper examined, among other things, how well such a model captures the fraction of US firms that export. Each US manufacturing firm was modeled as a seller throwing balls

[13] Bernard et al. (2018a) had access to annual import-exporter trade data rather than the underlying transaction data. We use the underlying transaction-level sales by Norwegian exporting firms to foreign importing firms. Sweden is the largest destination for Norwegian exports.

[14] To see this more clearly, if we plug the balls and bins' sizes into the statistical function for the extensive margin, it is immediate that the design implicitly requires that $E(n_i^s) = E(p_{ij}) = E(\sum_i((1-(1-\frac{n_i^s n_j^b}{n^2})^n))$, which does not hold in general.

(i.e., sales) into two bins (i.e., buyers), one domestic and one international. The number of balls of each firm was set to equal its total sales divided by the average shipment size, which is $36,000. The relative size of the two bins is set to equal their proportionate share of US manufacturer purchases. Armenter and Koren (2014) simulated 297,873 US firms with a size distribution approximated by a log-normal distribution with $\sigma = 2.66$ and $\mu = 13.2$. Given this setup, they find that 74% of US manufacturing firms export, in sharp contrast to the 18% in the US data.

Starting from the elementary model, we reexamine the question of how many US manufacturing firms export. We assume the same distribution of US firm sizes as Armenter and Koren (2014) and maintain their assumption that US firms throw balls at domestic and foreign customers.[15] However, we consider several adjustments to their setup. In addition to two aggregate buyers bins (domestic and foreign), we consider simulations with country-level bins and with firm-level bins. With country-level bins, each bin is proportional to the share of country purchases from the US. For the firm-level bins, each bin size should ideally be each firm's share of US purchases. We impute this information by assuming that the relative firm size distribution in the buying countries follows the size distribution of Chinese importing firms.[16] Linking to economics, the elementary model with both balls and bins' data counterparts being firms is our preferred design, as it is the positive transactions between firms that get recorded in customs – and hence are what trade statistics are based on.

There are two additional important choices to be made in designing the bins and balls framework. The first is the ball size and second is the presence and level of any cutoff which limits what transactions are recorded in the data. Armenter and Koren (2014) assumed that each ball represented $36000 of sales, the average transaction size in the data, and imposed no cutoff. The US data have a minimum transaction size of $2000, i.e. a cutoff below which export transactions are not systematically recorded. We consider balls of three sizes: $36000, $2000, and $100, where $100 is the minimum international transaction size observed in Norwegian customs data. Clearly, we prefer $100-sized balls with an export recording cutoff of $2,000, as this corresponds to the empirical context the best. We recognize that the "true" minimum export transaction size is likely larger than $100, but use this to demonstrate the importance of "hidden" assumptions in comparing a balls and bins model to data.

The results of this exercise are intuitive and important for the design of balls and bins models for particular data applications. As reported in panel (a) of Table 1, in our preferred design, the

[15]Simulation exercises assuming importers throw balls are reported in Table 1 panel (b). The results change very little.

[16]In the case of China, we take the actual distribution from Chinese customs data for the same year. Then each Chinese firm's bin size is s_{china}*import share of that firm, yielding n_{china} bins whose distribution is A. For other countries, we assume their importer bin distribution is the same as that of China. For example, if country j's total manufacturing imports from US are twice as large as China's, we randomly draw 2*n_{china} bins from A so that the sum over all foreign firms and countries equals the share of US manufacturing that is exported. Armenter and Koren (2014) did not have access to data on foreign firm (buyer) sizes or shares.

model predicts 11% of US manufacturing firms export, which is much closer to the 18% in data. With aggregate or country-level bins, the model performs poorly in matching the US exporter share regardless of ball size and the presence of a cutoff. This implies that in order to match the low fraction of US exporters, there must be firm heterogeneity in both buyers and sellers. This, of course, is at the heart of the recent literature on production networks. The recording cutoff also matters in this particular case, e.g. a $10,000 transaction is not the same as a collection of 100 $100 transactions. Not knowing the true distribution of transaction sizes can once again cause the model to overestimate the share of exporting firms.

These two examples highlight the importance of careful model design and the relevance of assumptions when taking a balls and bins model to the data.

6 Data Response to Shocks

The elementary model, in addition to matching cross-sectional stylized facts in production and trade networks, can guide our understanding of a diverse set of network responses to shocks. In other words, empirical regularities documented in the literature, in both levels and changes, may not be useful in improving our understanding of firm interactions beyond random matching. We present two examples, one examining the reshuffling importer-exporter links following the arrival of new suppliers, and the other looking at firm-level responses to exchange rate shocks.

6.1 Bipartite Outcomes

The elementary model is well-designed to understand the pattern of partner switching in response to shocks. Consider Sugita et al. (forthcoming), who combine several existing network-formation mechanisms to explain how US textile and apparel buyers match with Mexican exporters after the end of the Multi-Fibre Arrangement (MFA) and the ensuing arrival of new Chinese exporters.

Sugita et al. (forthcoming) cite five empirical findings that motivate and support their model of firm-to-firm matching. (1) US importers upgrade their Mexican partners more often when MFA quota are removed from their products. (2) Mexican exporters downgrade their US partners more often. (3) There are no systematic partner changes in other directions. (4) Among firms that switched their main partners, the new partners' ranks positively correlate with the ranks of the old partners. (5) The capability cutoff for Mexican exporters increases more if their products had quotas removed.

However, all of these five findings can be generated from the elementary model. Finding (4) is trivially true in a random matching world. As links are formed randomly, new partners should follow the same distribution as the old partners – in fact they should even have the same mean ranks in the absence of shocks. In the elementary model, we can view the end of the MFA as an increase of sellers to the economy. Then, the normalized size of Mexican firms decreases from s_i to

$\frac{s_i}{\delta}$, where $\delta > 1$.

Sugita et al. (forthcoming) ranked firms based on their product-level trade values prior to the liberalization; hence, their *firm rank* is directly comparable to our s and b. The *partner rank* in Sugita et al. (forthcoming) is computed as the rank of the firm's largest partner. Correspondingly, let $\bar{j}_i$ denote the highest-ranked buyer that seller i matches with in our model. Recall that when $s_i > s_{i'}$, the largest buyer that seller i matches with FSD the largest buyer with which i' matches. Let $s_{i'} = \frac{s_i}{\delta}$; we immediately have that $E_{preMPA}(\bar{j}_i) > E_{postMPA}(\bar{j}_i)$: exporters in Mexico experience partner downgrading.

Similarly, denote $\bar{i}_j$ as the size of the highest-ranked seller with which buyer j matches. Applying the inclusion-exclusion principle, the probability that $\bar{i}_j \leq z$ conditional on the buyer purchasing from Mexico is

$$\mathbb{P}[\bar{i}_j \leq z; \delta] = \frac{(\frac{\sum_{i \leq z} s_i}{\delta} + (1 - \frac{1}{\delta}))^{b_j} - (1 - \frac{1}{\delta})^{b_j}}{1 - (1 - \frac{1}{\delta})^{b_j}},$$

where the denominator on the RHS is the unconditional probability of at least one purchase from j landing in Mexico; the numerator is the unconditional probability that balls, if any, land in Mexican bins ranked lower than z netting out the probability that no purchases land in Mexico. Given $b_j > 1$ and $\delta \geq 1$, $\mathbb{P}[\bar{i}_j \leq z; \delta]$ is an increasing function in δ. Therefore, the distribution of $\bar{i}_j \mid \delta = 1$ FSD the distribution of $\bar{i}_j \mid \delta > 1$. Then, by the ranking theorem, $E_{preMPA}(\bar{i}_j) > E_{postMPA}(\bar{i}_j)$: importers in the US experience partner upgrading.

Given that we matched findings (1) and (2), finding (3) is immediately matched. Next, we consider finding (5). Denote $\underline{i}_j$ as the lowest-ranked seller that buyer j matches with; then,

$$\mathbb{P}[\underline{i}_j \leq z; \delta] = \frac{1 - (\frac{\sum_{i > z} s_i}{\delta} + (1 - \frac{1}{\delta}))^{b_j}}{1 - (1 - \frac{1}{\delta})^{b_j}}.$$

Given $b_j > 1$ and $\delta \geq 1$, $\mathbb{P}[\underline{i}_j \leq z; \delta]$ is a decreasing function in δ. Therefore, the distribution of $\underline{i}_j$ when $\delta > 1$ FSD is the distribution of $\underline{i}_j$ when $\delta = 1$. By the ranking theorem, this implies that $E_{preMPA}(\underline{i}_j) < E_{postMPA}(\underline{i}_j)$. Recall that by proposition 5, we can view the US as one buyer. Therefore, the rank cutoff for Mexican exporters increases after the end of MFA.

Because different studies look at different shocks and variables of interest, we cannot provide a general solution recipe as we did for the stylized facts. However, empirical regularities in changes are also about comparative statics, just like empirical regularities in levels. Therefore, the general results derived in Section 4 can also help pin down the elementary model's predictions on various firm-to-firm or aggregate responses to policy changes or economic shocks.

6.2 Aggregate Outcomes

The elementary model can also be used understand shock responses at a more aggregated level. For example, researchers are frequently interested in the heterogeneous response of firms to a common shock in terms of their sales, market entry/exit or pricing. A relatively standard empirical specification that examines the heterogeneous response of firms is given by:

$$ln(y_{ict}) = \alpha_p ln(\eta_{ct}) + \beta_p ln(a_{it-1}) + \gamma_p ln(a_{it-1}) * ln(\eta_{ct}) + d_t + d_{ic} + \epsilon_{ijt},$$

where a_{it-1} is a firm-specific characteristic at time $t-1$ and η_{ct} is a destination-specific shock at time t. We are interested in the estimated α_p and γ_p. Without loss of generality, the demeaned a_{it-1}, α_p, can be interpreted as the average impact of the shock on y_{ict}.

Consider, for example, firms' export-volume response to an exchange rate depreciation as in Berman et al. (2012). A fall in the exchange rate is equivalent to all buyers at destination market c increasing the number of purchases from b_j to δb_j, where $\delta > 1$. Let a_i be a variable of interest that is positively correlated with s_i.

From the elementary model, a firm i's export quantity to country c, x_{ic}, follows a binomial distribution $Bin(b_c, s_i)$ with $b_c = \sum_{j \in c} b_j$. As $Bin(\delta b_c, s_i) \succ Bin(b_c, s_i)$, it is immediate that

$$\alpha_p = E(ln(x_{ic}) \mid \delta b_c) - E(ln(x_{ic}) \mid b_c) > 0.$$

Next, we examine the heterogeneous response. Ignoring the untreated countries for now and using the Delta method, $E(g(x)) \approx g(E(x)) + \frac{1}{2} g''(E(x))V(x)$, one can show that

$$E(log(x_{ij})) \approx log(b_j s_i) - \frac{(1-s_i)}{2b_j s_i}.$$

Therefore,

$$\frac{\partial^2 E(log(x_{ij}))}{\partial b_j \partial s_i} \approx \frac{-2b_j^2 s_i - 2b_j^2(1-s_i)}{(2b_j^2 s_i)^2} = \frac{-1}{2b_j^2 s_i^2} < 0.$$

Hence, $\gamma_p < 0$, consistent with Berman et al. (2012)'s first finding that the elasticity of the firm's export volume to an exchange rate change decreases with the performance of the firm.

If suppliers tend to charge larger buyers lower prices, as documented by Morlacco (2020) and Rubens (2020), our model also predicts that the elasticity of the exporter price to the exchange rate change increases with the performance of the firm, consistent with the second finding of Berman et al. (2012).[17] In this example, the firm-level responses are not informative of any heterogeneous pricing-to-market behaviors. Those responses can simply be caused by buyer and seller heterogene-

[17] We treat export price as total export value divided by total export volume, just as how it was measured in Berman et al. (2012),

ity, coupled with a shock-induced random reshuffling of buyer-seller links.[18]

This example highlights the usefulness of the elementary model in understanding firm-level responses to shocks and policy changes. As with link reshuffling in the network, heterogeneity *in levels* at firm-to-firm level can lead to heterogeneous responses (*heterogeneity in changes*) for more aggregate outcomes. The elementary model can help identify which features of the data are useful in guiding more elaborate model development and which can be explained in a bins and balls framework.

7 "Instructive" Statistics and Theory Selection

Although our findings so far suggest that many existing empirical regularities may not be informative enough to guide the associated model construction, it simultaneously implies that the elementary model can be a powerful tool to select truly informative statistics and benchmark more elaborated theories. We illustrate both ideas in this section.

7.1 Use the Elementary Model to Discern Statistics

From the reduced-form perspective, we sketch two ideas on how to use our elementary model to develop "instructive" statistics and hence guide a more elaborated model construction. To be precise, "instructive" statistics are those that provide guidance on model structure and parameters values beyond those obtained in the elementary model.

First, we can design statistics that are by themselves "instructive" in that they do not match those from the elementary model. Then, these statistics can be used directly to inform the nonrandom matching process. Second, related to a suggestion of Armenter and Koren (2014), although the signs of some statistical relationships may not be informative, their magnitudes compared with the predictions from the elementary model can be useful in guiding model construction and model parameters.

Both approaches have pros and cons. The first approach is conceptually more demanding but it does not require us to compute any statistics using counterfactual data – a great advantage especially when transaction data is not available. The second idea is easy to implement, but can be sensitive to data features.[19]

[18]Precisely as the Berman et al. (2012) suggested, their findings are consistent with any model in which demand elasticity decreases with firm performance. The formal proof together with similar results on extensive margin adjustments are in Appendix A.5.

[19]See Armenter and Koren (2014)'s discussion on data skewness as one example.

7.1.1 *"Instructive Statistic" - Seller market shares*

One statistic where the data and the elementary model produce opposite signs is the relationship between the number of customers and the average market share of a seller in those customers. Bernard et al. (forthcoming) show that, in the Belgian domestic production network, large firms with many customers have smaller shares of their customers' purchases than do smaller firms with fewer customers, i.e. firm outdegree and average market share are negatively correlated.

To determine the relationship in the elementary model, we start by noting that the expected number of customers a seller matches with is given by N_i^s. By the law of iterated expectations, the expected average market share of seller i equals:

$$\bar{M}^b(i) = \frac{\sum_j M_{ij}^b p_{ij}}{\sum_j p_{ij}}, \tag{12}$$

where M_{ij}^b denotes the expected market share of seller i in buyer j's total purchases. Note that M_{ij}^b can be further simplified as

$$M_{ij}^b := E(\frac{x_{ij}}{b_j}) = \frac{E(x_{ij})}{b_j} = \frac{s_i b_j}{b_j} = s_i.$$

The intuition is simple. The probability that a purchase goes to seller i is s_i. Therefore, regardless of a buyers' size, its expected share of purchases from i is always s_i. Equation (12) can then simplified as:

$$\bar{M}^b(i) = \frac{\sum_j s_i p_{ij}}{\sum_j p_{ij}} = s_i.$$

We proved before that N_i^s is an increasing function of s_i; hence we establish the positive relationship between $\bar{M}^b(i)$ and N_i^s in the elementary model. This moment of the data can give guidance to the construction of a more elaborate model as in Bernard et al. (forthcoming). We provide a method for comparing and choosing among models in Section 7.2.

7.1.2 *"Instructive Statistic" - Correlated Matches*

A second instructive statistic comes from considering intertemporal versions of the elementary model. In our baseline design, the assignments are independent events at a point in time and over time; therefore, a seller matching with a buyer at time t will not increase their matching probability at time $t+1$. In the spirit of Chaney (2014), consider the following simple probit regression:

$$\mathbb{P}[x_{ij,t+1}|b_j, s_i] = \alpha 1[x_{ij,t} > 0] + \Psi(b_{j,t+1}, s_{i,t+1}) + \epsilon_{ij,t+1},$$

where $1[x_{ij,t} > 0]$ is the matching result of buyer j and seller i at time t. In the baseline random-matching model, the matching probability is a function of buyer and seller size, which should be fully captured by $\Psi(b_{j,t+1}, s_{i,t+1})$.

To test this, we use the transaction-level Norwegian customs data from 2005-2010. One unique feature of data is that both buyers and sellers are identified, see Bernard et al. (2018a) for details. Without loss of generality, we only consider sellers from Norway and buyers from China. The variables $b_{j,t+1}$ and $s_{i,t+1}$ are proxied by the yearly logged buyer purchases and logged seller sales, respectively.[20] We proxy for $\Psi(b_{j,t+1}, s_{i,t+1})$ with their third-order polynomials with full interactions. Note that this exercise has no interest in establishing causality – we are only interested in, conditional on b_{j+1} and s_{i+1}, whether the past matching results have additional explanatory power for future matches. The estimated $\hat{\alpha}$ is 2.844 and is statistically significant at the 1% level, suggesting that the formation of the production network is indeed path dependent.[21] Therefore, buyers are accustomed to buying from the same seller can be an important feature in explaining the network structure.

7.1.3 Informative Magnitudes - Bipartite Network Revisited

Even when the signs of some statistical relationships is not informative, their magnitudes compared with the predictions from the elementary model can be useful. We illustrate this point by examining the matching probability between Norwegian sellers and their international buyers in year 2005. In Figure 2, we group sellers (buyers) into 10 equal-sized bins by log sales (purchases) and present the fraction of buyers and sellers that are matched in real data across bins. With a slight abuse of language but no risk of confusion, we refer to the fraction of matched pairs as the matching probability p_{ij}. Panels (a) and (b) are the marginal distribution of p_{ij} for the top and bottom 10 percent of sellers and buyers, respectively. Panel (c) presents the surface plot of p_{ij} across 10×10 bins.

In Figure 3, we present the same statistic but use data from the random-matching simulation. Specifically, we take the transactions of each buyer and the total number of transactions of each seller as given from the data but randomly reshuffle the transactions. As shown in Figure 3, the elementary model captures the actual matching results well. In Figure 5-a, we present the prediction errors of the matching probability of the elementary model against the real data. The model slightly underpredicts the probability that the smallest buyers match with small sellers and overpredicts the probability that large buyers match with small sellers. If one wants to study how firms match, these prediction gaps are potentially useful in guiding model construction.[22]

[20] We obtain quantitatively very similar results when controlling for the number of transactions instead of sales.

[21] Note that in Section 4, we already proved that our model could generate a series of bipartite stylized facts even when the assignments are not independent. A bipartite relationship that can be explained by the elementary model or its extensions precisely suggests the underidentification of models that are designed to explain those stylized facts.

[22] Allowing for heterogeneity in ball size, matching the distribution of transaction sizes in the data, may help more

7.2 Use the Elementary Model to Select theory

Related to the idea presented above, we can use the elementary model as a benchmark to select applied theories. This is a slightly different thought experiment: instead of starting from data and the elementary model to ask whether any information left in the data demands a more elaborated theory to explain, we start from a theory and examine its explanatory power compared to the elementary model. This idea connects directly to the concept of Bayesian model selection.

7.2.1 Model Selection

Where there are several competing theoretical models, Bayesian model selection provides a formal way of evaluating their relative probabilities in light of the data and any prior information available. The basic idea is simple. The evaluation of a model M's performance in light of the data $\mathcal{D}$ is based on the likelihood $\mathbb{P}(\mathcal{D} \mid M)$.[23] Applying the Bayes' theorem to invert the order of conditioning, we then obtain the model posterior probability given the data:

$$\mathbb{P}(M \mid \mathcal{D}) \propto \mathbb{P}(M)\mathbb{P}(\mathcal{D} \mid M)$$

where $\mathbb{P}(M)$ is the prior probability assigned to the model itself.[24] Thus the posterior probability is the likelihood under the prior for a specific model choice.

In particular, when comparing two models, M_0 versus M_1, we can write down the ratio of the posterior probabilities and decompose it into two parts:

$$\frac{\mathbb{P}(M_1 \mid \mathcal{D})}{\mathbb{P}(M_0 \mid \mathcal{D})} = \underbrace{\frac{\mathbb{P}(D \mid M_1)}{\mathbb{P}(D \mid M_0)}}_{Bayes\ Factor} \times \underbrace{\frac{\mathbb{P}(M_1)}{\mathbb{P}(M_0)}}_{Priors}. \tag{13}$$

Bayes factor, which we denote by B_{01}, is the ratio of the likelihoods of the models. When $B_{01} >$ 1 (< 1), it represents an increase (decrease) of the support in favor of model 1 relative to model 0 given the observed data. *Priors* captures the belief about the relative plausibility of the models before the arrival of the data. Specifically, the extent to which we update our beliefs over the two models after observing the data is completely summarized by the Bayes factor.

In light of the idea of model selection, we can now think more formally about moving from data to a preferred theory. The extent to which a theory is preferred depends on (1) its relative ability in explaining the data compared to the the alternative theory, (2) the researcher's priors over the theory and the alternative.

Note that in contrast with the frequentist goodness-of-fit tests, Bayesian model selection high-

precisely matching moments from the data.

[23] In Bayesian analysis $\mathbb{P}(\mathcal{D} \mid M)$ is also called *Bayesian evidence*.

[24] We have dropped the constant $\mathbb{P}(\mathcal{D})$ that depends only on the data.

lights that it is pointless to reject a theory unless an alternative is available and fits the observed facts better (Loredo, 1990). In other words, the probability of a theory that makes a correct prediction can increase if the prediction is confirmed by observations, provided competitor theories do not make the same prediction. This lines up with our intuition that a verified prediction lends support to the theory that predicts it. In fact, this is very much in line with how theories in trade are evaluated in reality: for instance Krugman (1981) was the preferred theory for explaining intra-industry trade, because other contemporaneous theories fail to explain it.

In addition, the idea of model selection also contrast to the limited concept of falsifiability that scientific theories can only be tested by proving to be wrong (Popper, 1959). In particular, that a model can fail to pass certain falsification tests is not sufficient enough to lead to rejection. Unless the observations are totally impossible within a model, finding that the data have a small probability given a theory does not say anything about the probability of the theory itself unless we can compare it with an alternative (Trotta, 2008). In short, Bayesian model selection is not only a more transparent, but also a more complete concept of theory evaluation.

As an illustrative example, we compare the model featuring Beckerian matching and capacity constraints of Sugita et al. (forthcoming) with the elementary model in explaining firm-to-firm connections. Using the 2005 transaction-level Norwegian customs data with matched Norwegian exporters and international importers, we take each buyer and seller's transactions as given. For the elementary model, we match buyers and sellers by random reshuffling these transactions. For the Beckerian model, we simulate the links as follows: we take the total number of connections of each buyer and seller as given from the data. We starting from the largest buyer, we assign its largest connection (in terms of total sales) to the most-connected seller, its second-largest connection to the second-most-connected seller, and so forth. After completing assigning the links of the largest buyer, we move to the second-largest buyer and repeat the same process. In every round, we assign the buyer to the most connected seller that still has slots.

The matching probability generated from Beckerian matching is presented in Figure 4. It captures the fact that large sellers have a higher probability of matching with larger buyers. However, it predicts that for small sellers (buyers), their matching probability decreases in buyer (seller) size (4-a, b), in contrast to the data. Comparing the prediction errors of the matching probability of the elementary model vs. the Beckerian model (Figure 5), we find that the latter has a much higher prediction error for most of the buyer-seller bin. Through the lens of Bayes, this suggests that the data lends less credibility to the Beckerian model in explaining firm-to-firm connections compared to the elementary model. In other words, if we conclude that the Beckerian model is our preferred theory, the justification must comes from a strong prior we attached to it.

At first look, this may appear perplexing, as the Beckerian matching model is intended to explain matching, whereas the elementary model does not. However, despite the fact that the Beckerian matching model was designed to explain firm-firm linkages, whose data counterpart is

p_{ij}, in practice it only matches only *certain statistics* of p_{ij}, such as the sign of assortativity. To match these statistics, strong assumptions are imposed, which in turn generate greater prediction errors in matching the *distribution* of p_{ij}. Without comparing a model to alternatives, we may make incorrect conclusions about its plausibility, leading to mistaken conclusions on mechanisms, causal interpretations, and welfare or policy predictions.

7.2.2 Network Models: A Quantitative Comparison

In this subsection, we apply the idea of model selection to compare several candidate network models. From the Bayesian view, a model is pointless unless compared to an alternative. Therefore, the importance of an appropriately chosen alternative cannot be overstated. We show that our elementary model is so far the best alternative; hence a reasonable benchmark model for work on production networks.

We consider three class of network models besides our elementary model: (1) Erdös–Rényi model (ER), where every pair of buyer-sellers is independently linked with probability p; (2) Fixed-degree Random Graph model (FRG), where links are randomly allocated conditional on taking each buyer and seller' degree as given;[25] (3) the Beckerian extension of the hierarchy model from Sugita et al. (forthcoming). The first two models have been widely used in the broader network literature either as the foundation of more elaborate network models or as a baseline for falsification. Of course one can propose more complicated network models, but our focus is on selecting the benchmark.

Each of the aforementioned models can be specified by a set of parameters.[26] And for a fixed set of parameters, the model induces a probability distribution of the production network. If we hold a uniform prior, i.e., before the arrival of the data, we believe each model is equally probable, then selecting the best posterior model is equivalent to ask which model gives the largest likelihood to the observed data. In the interest of space, we discuss each model and its associated likelihood function in Appendix B.1.

Our estimated results are presented in Table 2. The first row of shows the maximized log-likelihood of a given model, which we normalize by the total number of links observed in data. Its value reflects the likelihood of observing our sample network if the associated model is the true

[25] For example, the Power-law Random Graph model proposed by Aiello et al. (2001) is one special case of the Fixed-degree Random Graph model, in which the degree distribution of firms (nodes in the network) is drawn from a Power-law distribution.

[26] To be precise, when we have parameter uncertainties, the posterior probability $\mathbb{P}(M \mid \mathcal{D})$ is proportional to $\mathbb{P}(\mathcal{D} \mid M, \theta^M_{ML})\mathbb{P}(\theta^M_{ML} \mid M)$ under uniform prior over model classes, where θ^M_{ML} is the vector of parameters that maximize the likelihood given the model class M. That is, the Bayesian evidence not only depends on the best-fit likelihood but also on our prior believes on how likely θ^M_{ML} appears given the model class M. Typically, a complicated model with many parameters each of which is free to vary will be penalized by having a smaller $\mathbb{P}(\theta^M_{ML} \mid M)$ than a simpler one, which naturally gives Occam's Razor in Bayesian inference. We intentionally avoid a more elaborate discussion as firstly $\mathbb{P}(\theta^M_{ML} \mid M)$ only depends on priors, and secondly our candidate models have maximum two free parameters, making over-fitting less a concern. Also, in the end our most preferred model in fact have no free parameters. Nevertheless whenever we report and discuss $\mathbb{P}(\mathcal{D} \mid M, \theta^M_{ML})$, we bear its difference between $\mathbb{P}(\mathcal{D} \mid M)$ in mind.

model. The rest rows report estimated parameter values if there are any. As we can see from Table 2, the elementary model yields the highest likelihood value, the baseline case of which is about 6 times as large as the likelihood value of the Erdös–Rényi model.[27] When we allow for optimal prediction errors in column (4), the likelihood value of the elementary model only improves marginally, suggesting that the baseline is a reasonable good description of the data. In column (3), we show that even when transaction-level data are unobservable, the elementary model with a simple assumption on ball size still gives a better fit compared to other models. Somewhat surprisingly, Erdös–Rényi model ranks the second in fitting data despite its simple structure. FRG model ranked third, followed by the Beckerian models. As expected, the naive prediction of the hierarchy model fits the data poorly, as suggested in columns (6) of Table 2. While introducing prediction errors in column (7) greatly improves the model fit, but it still has the least explanatory power among all the models. Overall, based on the log-likelihood values presented in Table 2, the elementary model is the logical choice as a baseline model for future comparisons.

8 Conclusion

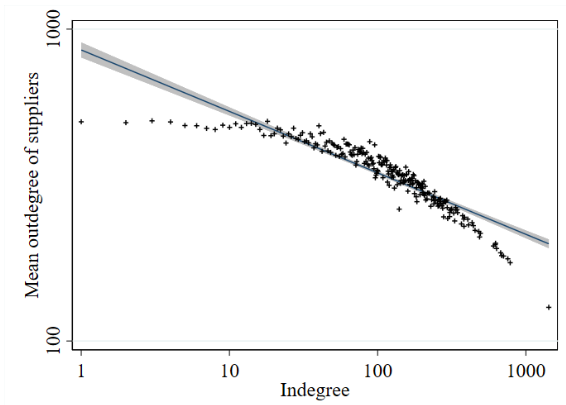
Firm heterogeneity and the sparse nature of the production network implicitly discipline the structure of the network, even when links are formed randomly. In this case, a large group of statistical relationships, in both levels and changes, may not be informative in improving our understanding of bipartite relationships beyond random matching.

We provide a benchmark elementary model to identify which statistics are sufficiently characterized by random matching and which require additional model assumptions. In addition to revisiting the set of stylized facts and empirical results commonly presented in the trade literature, we develop the general properties of the model, characterizing families of statistics and data generating processes that may, or may not, cause a model to suffer from underidentification concerns.

We propose ways to use our model to spot useful statistics and select applied models. We also provide an economic interpretation of the random-allocation process by showing that it could naturally emerge from a competitive environment, with Eaton and Kortum (2002) being a special case. This view not only disciplines our falsification designs but also helps us to select the appropriate data counterparts given the economic context of a researcher's studies. We hope that our work can be helpful for various research fields, in particular for the future applied work on firms using massive micro-level data sets. Although our focus is on production networks, we believe our work can be useful for many other research fields, as bipartite relationships are at the foundation of many economic problems and the prior-selection problem potentially exists whenever we link theory to data.

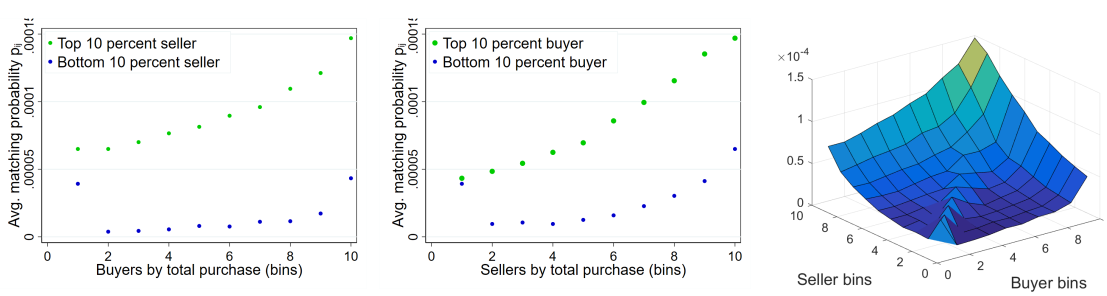
[27] $e^{-8.22}/e^{-9.99} \approx 6$.

Figures



Notes: This figure shows all possible values of the number of suppliers per firm on the x-axis and the average number of customer connections of those suppliers on the y-axis using simulated buyer-seller network based on 2005 Norwegian customs data. In particular take the total number of transactions of each buyer and seller as given, but randomly reallocated these transactions following the elementary model. Axes are in logs. The fitted regression line and 95% confidence intervals are denoted by the solid line and the gray area.

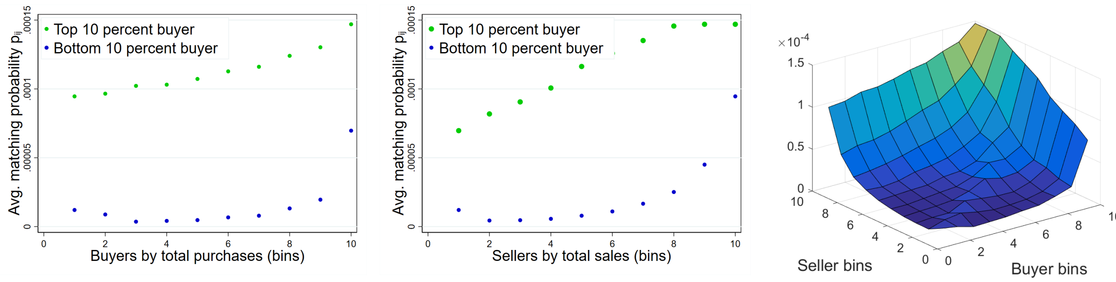
Figure 1: Degree assortativity with the correct falsification design: Norwegian data



(a) (b) (c)

Notes: This figure presents binned scatter plots where we group Norwegian exporters ad international importers into 10×10 equal-sized bins by log sales and purchases and compute the share of connected pairs (p_{ij}) in each bin pair. Panels (a) and (b) are the marginal distribution of p_{ij} for the top and bottom 10 percent of sellers and buyers, respectively. Panel (c) presents the surface plot of p_{ij} across 10×10 bins.

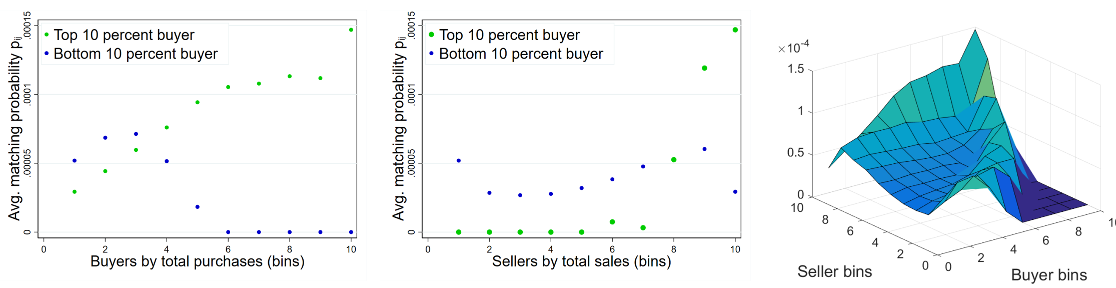
Figure 2: Matching probability: real data



(a) (b) (c)

Notes: This figure presents binned scatter plots exactly as in Figure 2 but using simulated buyer-seller links following the elementary model as in Figure 1.

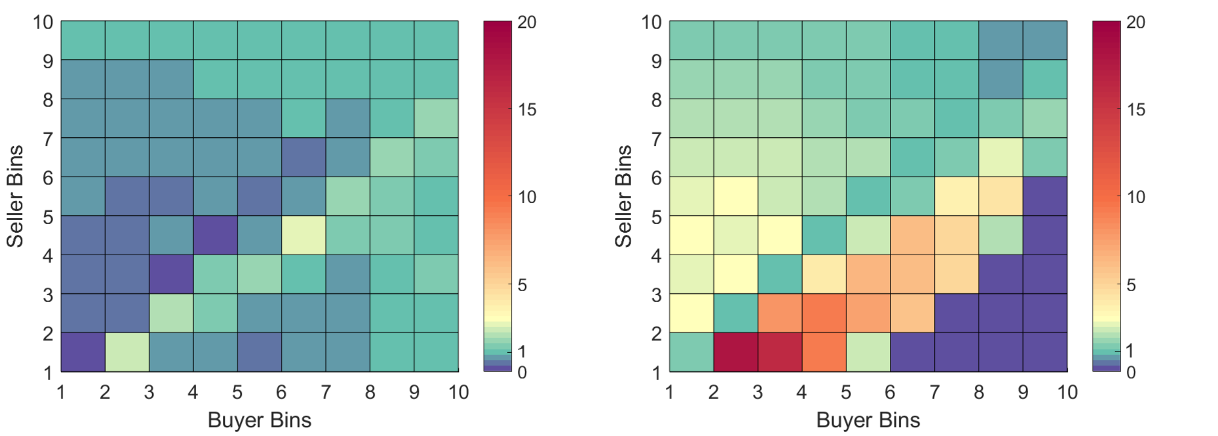
Figure 3: Matching probability: random-matching simulation



(a) (b) (c)

Notes: This figure presents binned scatter plots exactly as in Figure 2, but using simulated data that link buyers and sellers following a Beckerian matching rule. The data generating process is described in the text.

Figure 4: Matching probability: assortative matching simulation



(a) Balls-and-bins simulation (b) Assortative matching simulation

Notes: This figure presents predicted matching probability of the elementary model (panel a) and the Beckerian matching model (panel b) over the actual matching probability in data over the 10×10 buyer-seller bins, i.e. $\hat{p}_{ij}/p_{ij}$. If the value is greater (less) than 1 it implies the model over-predict (under-predict) the matching probability for the given buyer-seller group.

Figure 5: Prediction errors of matching probability

Tables

Table 1: Share of Manufacturing Exporters in the US: Balls and Bins Simulations

(a) US Firms as Ball Holders

	US Customs Reporting Cutoff					
	2000 USD (DATA)			36000 USD (AK)		
Bins are	Firms	Countries	RoW as one bin	Firms	Countries	RoW as one bin
	(1)	(2)	(3)	(4)	(5)	(6)
Ball Size = 100USD	**11.02 %**	79.22 %	90.72 %	0.41 %	38.51 %	59.84 %
Ball Size = Cutoff Size	92.34 %	92.05 %	92.85 %	65.38 %	64.32 %	66.95 % (AK)

(b) US Firms as Bins

	US Customs Reporting Cutoff					
	2000 USD (DATA)			36000 USD (AK)		
Bins are	Firms	Countries	RoW as one bin	Firms	Countries	RoW as one bin
	(1)	(2)	(3)	(4)	(5)	(6)
Ball Size = 100USD	11.18 %	79.49 %	90.87 %	0.43 %	38.88 %	60.18 %
Ball Size = Cutoff Size	92.66 %	92.63 %	92.63 %	66.53 %	66.61 %	66.61 %

Notes: This table provides the balls and bins simulation probing the share of direct manufacturing exporters in the US when linking random allocation model to economics (panel (a), first row of column (1)), compared to the exercise of Armenter and Koren (2014) (panel (a), second row of column (6)). For transparency we present how the choice of the economic counterpart of balls, ball holders, bins, and cutoff in turn affect the simulation results. Just as the theoretical results presented throughout the paper, the results of this simulation exercise are not quantitatively sensitive to assuming buyers or sellers throwing balls, as evident from comparing panel (a) and (b).

Table 2: Production Network Model Selection

Erdös-Rényi	Fixed-degree Random Graph	Elementary Model: Baseline	Elementary Model: $\hat{b} = \hat{\beta} b + \hat{\alpha}$	Elementary Model: Trans. uncertainty	Beckerian: Baseline	Beckerian: Optimal Error
(1)	(2)	(3)	(4)	(5)	(6)	(7)
-9.99	**-17.82**	**-8.22**	**-7.73**	**[-9.00, -8.18]**	**-73.47**	**-18.58**
$\hat{p} = 1.25e{-4}$			$\hat{\alpha} = 0.57$	$\hat{\alpha} = -0.21, 1.51$		$\hat{\epsilon} = 2.48e{-4}$
			$\hat{\beta} = 0.33$	$\hat{\beta} = 1.45, 1.35e{-05}$		

Notes: The maximized log-likelihood (per link) are presented in bold number for each given model. The rest are the parameter values that maximize the log-likelihood of a given model conditional on the observed Norwegian network.

References

Aiello, William, Fan Chung, and Linyuan Lu, "A random graph model for power law graphs," *Experimental Mathematics*, 2001, *10* (1), 53–66.

Albert, Réka and Albert-László Barabási, "Topology of evolving networks: local events and universality," *Physical review letters*, 2000, *85* (24), 5234.

_ **and** _ , "Statistical mechanics of complex networks," *Reviews of modern physics*, 2002, *74* (1), 47.

Alfaro-Urena, Alonso, Isabela Manelici, and Jose P Vasquez, "The effects of joining multinational supply chains: New evidence from firm-to-firm linkages," *Available at SSRN 3376129*, 2020.

Antras, Pol, Teresa C Fort, and Felix Tintelnot, "The margins of global sourcing: Theory and evidence from us firms," *American Economic Review*, 2017, *107* (9), 2514–64.

Armenter, Roc and Miklós Koren, "A balls-and-bins model of trade," *American Economic Review*, 2014, *104* (7), 2127–51.

Atalay, Enghin, Ali Hortacsu, James Roberts, and Chad Syverson, "Network structure of production," *Proceedings of the National Academy of Sciences*, 2011, *108* (13), 5199–5202.

Atkin, David, Amit K Khandelwal, and Adam Osman, "Exporting and firm performance: Evidence from a randomized experiment," *The Quarterly Journal of Economics*, 2017, *132* (2), 551–615.

Barabási, Albert-László and Réka Albert, "Emergence of scaling in random networks," *science*, 1999, *286* (5439), 509–512.

Berman, Nicolas, Philippe Martin, and Thierry Mayer, "How do different exporters react to exchange rate changes?," *The Quarterly Journal of Economics*, 2012, *127* (1), 437–492.

Bernard, Andrew B and Andreas Moxnes, "Networks and trade," *Annual Review of Economics*, 2018, *10*, 65–85.

_ **,** _ **, and Karen Helene Ulltveit-Moe**, "Two-sided heterogeneity and trade," *Review of Economics and Statistics*, 2018, *100* (3), 424–439.

_ **,** _ **, and Yukiko U Saito**, "Production networks, geography, and firm performance," *Journal of Political Economy*, 2019, *127* (2), 639–688.

_ , **Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes**, "The origins of firm heterogeneity: A production network approach," *Journal of Political Economy*, forthcoming.

_ , **Esther A Bøler, and Swati Dhingra**, "Firm-to-firm connections in Colombian imports," *in World Trade Evolution*, October 2018, *Lili Yan Ing and Miaojie Yu, Ed.s, Routlege.*

_ , **Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum**, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93* (4), 1268–1290.

Bezáková, Ivona, Adam Kalai, and Rahul Santhanam, "Graph model selection using maximum likelihood," in "Proceedings of the 23rd international conference on Machine learning" 2006, pp. 105–112.

Blum, Bernardo S, Sebastian Claro, and Ignatius J Horstmann, "Import intermediaries and trade costs: Theory and evidence," *Unpublished manuscript, Rotman School of Management, University of Toronto*, 2012.

Bollobás, B'ela, Oliver Riordan, Joel Spencer, and Gábor Tusnády, "The degree sequence of a scale-free random graph process," *Random Structures & Algorithms*, 2001, *18* (3), 279–290.

Borusyak, Kirill and Peter Hull, "Non-Random Exposure to Natural Experiments: Theory and Applications," *Working Paper, University College London*, 2020.

Cai, Jing and Adam Szeidl, "Interfirm relationships and business performance," *The Quarterly Journal of Economics*, 2018, *133* (3), 1229–1282.

Carballo, Jerónimo, Gianmarco IP Ottaviano, and Christian Volpe Martincus, "The buyer margins of firms' exports," *Journal of International Economics*, 2018, *112*, 33–49.

Chaney, Thomas, "The network structure of international trade," *American Economic Review*, 2014, *104* (11), 3600–3634.

_ , "Networks in international trade," in "The Oxford handbook of the economics of networks" 2016.

_ , "The gravity equation in international trade: An explanation," *Journal of Political Economy*, 2018, *126* (1), 150–177.

Costinot, Arnaud, "An elementary theory of comparative advantage," *Econometrica*, 2009, *77* (4), 1165–1192.

Dhyne, Emmanuel, Glenn Magerman, and Stela Rubínová, "The Belgian production network 2002-2012," Technical Report, NBB Working Paper 2015.

Ding, Jie, Vahid Tarokh, and Yuhong Yang, "Model selection techniques: An overview," *IEEE Signal Processing Magazine*, 2018, *35* (6), 16–34.

Dingel, Jonathan I and Felix Tintelnot, "Spatial Economics for Granular Settings," Technical Report, National Bureau of Economic Research 2020.

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

Eeckhoudt, Louis and Christian Gollier, "Demand for risky assets and the monotone probability ratio order," *Journal of Risk and Uncertainty*, 1995, *11* (2), 113–122.

Ellison, Glenn and Edward L Glaeser, "Geographic concentration in US manufacturing industries: a dartboard approach," *Journal of political economy*, 1997, *105* (5), 889–927.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin, *Bayesian data analysis*, CRC press, 2013.

Herkenhoff, Philipp, Sebastian Krautheim, and Philip U Sauré, "A Simple Model of Buyer-Seller Networks in International Trade," 2021.

Holst, Lars, "Asymptotic normality in a generalized occupancy problem," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 1972, *21* (2), 109–120.

Hooten, Mevin B and N Thompson Hobbs, "A guide to Bayesian model selection for ecologists," *Ecological Monographs*, 2015, *85* (1), 3–28.

Klenke, Achim and Lutz Mattner, "Stochastic ordering of classical discrete distributions," *Advances in Applied probability*, 2010, *42* (2), 392–410.

Kramarz, Francis, Julien Martin, and Isabelle Mejean, "Volatility in the small and in the large: The lack of diversification in international trade," *Journal of International Economics*, 2020, *122*, 103276.

Krugman, Paul, "Scale economies, product differentiation, and the pattern of trade," *The American Economic Review*, 1980, *70* (5), 950–959.

Krugman, Paul R, "Intraindustry specialization and the gains from trade," *Journal of political Economy*, 1981, *89* (5), 959–973.

Lim, Kevin, "Endogenous production networks and the business cycle," *Working Paper, The University of Toronto*, 2020.

Loredo, Thomas J, "From Laplace to supernova SN 1987A: Bayesian inference in astrophysics," in "Maximum entropy and Bayesian methods," Springer, 1990, pp. 81–142.

Madigan, David and Adrian E Raftery, "Model selection and accounting for model uncertainty in graphical models using Occam's window," *Journal of the American Statistical Association*, 1994, *89* (428), 1535–1546.

Melitz, Marc J, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *econometrica*, 2003, *71* (6), 1695–1725.

Mitzenmacher, Michael and Eli Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*, Cambridge university press, 2017.

Morlacco, Monica, "Market power in input markets: Theory and evidence from french manufacturing," *Working Paper, The University of Southern California*, 2020.

Onnela, J-P, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási, "Structure and tie strengths in mobile communication networks," *Proceedings of the national academy of sciences*, 2007, *104* (18), 7332–7336.

Popper, Karl, *The logic of scientific discovery*, Julius Springer, Hutchinson Co, 1959.

Raab, Martin and Angelika Steger, "“Balls into bins”—A simple and tight analysis," in "International Workshop on Randomization and Approximation Techniques in Computer Science" Springer 1998, pp. 159–170.

Roch, Sebastien, "Modern discrete probability: An essential toolkit," *Manuscript, The University of Wisconsin-Madison*, 2015.

Rubens, Michael, "Ownership consolidation, monopsony power and efficiency: Evidence from the Chinese tobacco industry," *Working Paper, The University of California, Los Angeles*, 2020.

Serrano, Ma Angeles and Marián Boguná, "Topology of the world trade web," *Physical Review E*, 2003, *68* (1), 015101.

Sugita, Yoichi, Kensuke Teshima, and Enrique Seira, "Assortative Matching of Exporters and Importers: Evidence From a Large Trade Liberalization Episode," *Review of Economics and Statistics*, forthcoming.

Trotta, Roberto, "Bayes in the sky: Bayesian inference and model selection in cosmology," *Contemporary Physics*, 2008, *49* (2), 71–104.

Weiss, Irving, "Limiting distributions in some occupancy problems," *The Annals of Mathematical Statistics*, 1958, *29* (3), 878–884.

Appendix A Theory Appendix

A.1 Coupling and Stochastic Dominance Theorem

Because this is the first time that the coupling technique has been employed in the network and trade literature, we provide a pedagogical introduction of the method and its relation with the stochastic dominance theorem with examples for interested readers.

Coupling is a quintessential probabilistic technique with a wide range of applications. The idea behind the coupling method is that to compare two random variables X and Y, it is sometimes useful to construct their joint probabilities. Formally, let X and Y take values in $(S, \mathcal{S})$, where S is a measurable space. A coupling of X and Y is a probability measure $Z := (X', Y')$ on the product space $(S \times S, \mathcal{S} \times \mathcal{S})$, such that the marginal distributions of Z equal to the distributions of X and Y.

In general, there could be multiple ways to couple random variables. Consider two Bernoulli random variables X_1 and Y_1 as an example. Let $\mathbb{P}[X_1 = 1] = q$ and $\mathbb{P}[Y_1 = 1] = p > q$. Here, $S = \{0, 1\}$. The following are two couplings of X_1 and Y_1:

- (Independent coupling) One coupling of X_1 and Y_1 is (X_1', Y_1'), where $X_1' = X_1$ and $Y_1' = Y_1$ and are independent. Then (X_1', Y_1') is a coupling of X_1 and Y_1 with law:

$$\mathbb{P}[(X_1', Y_1') = (i, j)]_{i,j \in \{0,1\}} = \begin{pmatrix} (1-q)(1-p) & (1-q)p \\ q(1-p) & qp \end{pmatrix}$$

- (Monotone coupling) Another possibility is to set $X_1'' = X_1$, Y_1'' equal to X_1'' whenever $X_1'' = 1$. When $X_1'' = 0$, there is a probability $(1-p)/(1-q)$ that $Y_1'' = 1$ and a probability $(p-q)/(1-q)$ that $Y_1'' = 1$. Then (X_1'', Y_1'') is a coupling of X_1 and Y_1 with law:

$$\mathbb{P}[(X_1'', Y_1'') = (i, j)]_{i,j \in \{0,1\}} = \begin{pmatrix} 1-q & p-q \\ 0 & q \end{pmatrix}$$

It is easy to verify that in both cases, the marginal distributions of the couplings equal the distributions of X_1 and Y_1.

Armed with the definition of coupling, we next briefly discuss its relationship with the stochastic dominance theorem. When ordering random variables, two sets of information are important for the ordering: (1) the values that the random variables can take and (2) the associated probability. If two random variables $X \succ Y$, we say that either X has a higher probability of obtaining larger values than Y, or X has a tendency to be greater than Y. The first view corresponds to the definition of first-order stochastic dominance. That is, conditional on values, we compares the two random variables' associated probability. Coupling takes another approach and corresponds to the second view: conditional on probability and comparing the values. This intuition is best seen from the construction of the following coupling: if two random variables X and Y with the cumulative distribution function (cdf) are F_X and F_Y, let $(X', Y') = (F_X^{-1}(U), F_Y^{-1}(U))$, where U is uniform in $[0, 1]$. By construction, (X', Y') is a coupling of X and Y. Giving this coupling, we transferred the problem from comparing the probability that X and Y take less than or equal to a certain value, to comparing the value of $F_X^{-1}(U)$ and $F_Y^{-1}(U)$ for a certain probability event (e.g., the probability of U equal to or smaller than a certain value). This is precisely the intuition behind stochastic dominance theorem. Specifically, if $X \succ Y$, the definition of FSD ensures that $\mathbb{P}[X' > Y'] = 1$. Therefore, (X', Y') is a monotone coupling of X and Y when $X \succ Y$.

We finish by providing an application of the stochastic dominance theorem. We believe this example will be helpful for interested readers to better understand the proof by construction of proposition 1 and 2. Consider two Poisson distributions $X_2 \sim Poi(a)$ and $Y_2 \sim Poi(b)$ with $a > b$. Note that a sum of independent Poissons is also Poisson. This fact leads to a natural construction of coupling: let $Y_2' = Y_2$, $Z_2 \sim Poi(a-b)$, and is independent of Y_2. Then, let $X_2' = Y_2' + Z_2'$. Then, (X_2', Y_2') is a coupling for X_2 and Y_2. Moreover, by construction, $\mathbb{P}(X_2' \geq Y_2') = 1$. Hence, we can apply the stochastic dominance theorem and conclude that $X_2 \succ Y_2$.

A.2 Robustness: Correlated Assignments

Consider a general case in which one purchase is allocated at a time. At time one, the probability of a seller i receiving a purchase from any buyer is given by s_i. The probability of receiving a purchase from buyer j at time $\zeta + 1$, for any $\zeta > 0$, is given by:

$$s_{ij,\zeta+1} = \frac{\Omega\left(s_{ij,\zeta}, 1(x_{ij,\zeta} > 0)\right)}{\sum_{i'} \Omega\left(s_{i'j,\zeta}, 1(x_{i'j,\zeta} > 0)\right)}, \tag{A1}$$

where Ω is an increasing function of both s and $1(x > 0)$. That is, the probability that a purchase assigned to i at time $\zeta + 1$ is positively associated with the past matching probability and the realized matching outcome in the previous period. The denominator reflects the fact that the unconditional probability needs to be normalized across all sellers.

Clearly, $s_{ij,\zeta+1}$ is a random variable. Therefore, we investigate the stochastic ordering of $s_{ij,\zeta+1}$ across all sellers and time periods. We first state the results: $s_{ij,\zeta}$ is FSD statistically ordered in ascending order in i, for all buyers j and periods $\zeta \geq 0$.

We prove the above statement using mathematical induction. Without loss of generality, consider two sellers with $s_i > s_{i'}$. When $\zeta = 1$, the statement is trivially true.

Suppose that when $\zeta = z$ and $s_{ij,\zeta} \succ s_{i'j,\zeta}$. Then, by the proof of proposition 1, $x_{ij,\zeta} \succ x_{i'j,\zeta}$, which in turn implies $1(x_{ij,\zeta} > 0) \succ 1(x_{i'j,\zeta} > 0)$. View $(s_{ij,\zeta}, 1(x_{i'j,\zeta} > 0))$ as the input vector; applying proposition 1, it is immediate that $s_{ij,\zeta+1} \succ s_{i'j,\zeta+1}$. Since both the base case and the inductive step have been shown, by mathematical induction, the statement $s_{ij,\zeta} \succ s_{i'j,\zeta}$ holds for every natural number ζ. $\square$

Given the above result, the normalized matching probability between i and j over any *period* also follows the FSD ordering. Note that whether s_i is deterministic or FSD ordered in i does not affect the stochastic ordering of x_{ij}. Therefore, all results derived in the paper remain unchanged when we allows for correlated assignment. Given our proof, it is also clear that the statement also holds for richer dynamics than (A1), as long as Ω and the stochastic ordering of the input vector are monotone.

A.3 Robustness: Unequal-sized Purchases

Two purchases of unequal size can be viewed as two purchases with the same price but different quantities (bundled purchase) or two purchases with the same quantity but different prices. Without the loss of generality, we choose the former view for our proofs.

We first consider the case in which the size of each purchase differs across buyers. Denote by q_j the purchases size for buyer j. All extensive-margin results remain unchanged, but the total sales from i to j now becomes $x_{ij}q_j$. When q_j is weakly increasing b_j, or FSD generated in ascending order in b_j, we will have $x_{ij}q_j \succ x_{ij'}q_{j'}$ for $b_j > b_{j'}$. In words, if bigger buyers are more likely to make more purchases when matching a seller, all results derived in the paper remain unchanged.

Proof: When q_j is weakly increasing b_j, then $q_j \geq q_{j'}$ for $b_j \geq b_{j'}$. By the proof of proposition 1, $(\hat{x}_{ij}, \hat{x}_{ij'})$ is the monotone coupling of x_{ij} and $x_{ij'}$, $\mathbb{P}(\hat{x}_{ij} > \hat{x}_{ij}) = 1$. Therefore, $\mathbb{P}(\hat{x}_{ij}q_j > \hat{x}_{ij}q_{j'}) = 1$ $\Rightarrow x_{ij}q_j \succ x_{ij'}q_{j'}$.

When q_j is FSD generated in ascending order in b_j, then $q_j \succ q_{j'}$ for $b_j \geq b_{j'}$. By the stochastic dominance theorem, monotone coupling also exists for q_j and $q_{j'}$, which we denote by $(\hat{q}_j,\, \hat{q}_{j'})$. Then, $(\hat{x}_{ij}\hat{q}_j,\, \hat{x}_{ij'}\hat{q}_{j'})$ form a coupling of $x_{ij}q_j$ and $x_{ij'}q_{j'}$. Clearly, $\mathbb{P}(\hat{x}_{ij}\hat{q}_j > \hat{x}_{ij}\hat{q}_{j'}) = 1$; therefore, $x_{ij}q_j \succ x_{ij'}q_{j'}$. □

Similar proofs can be constructed for the following two cases. When the size of each purchase differs across sellers, denote by q_i the purchases size from seller i. The extensive-margin results remain unchanged, but the total sales from i to j now becomes $x_{ij}q_i$. When q_i is weakly increasing s_i, or FSD generated in ascending order in s_i, we will have $x_{ij}q_i \succ x_{ij'}q_{i'}$ for $s_i > s_{i'}$. In words, if buyers are more likely to make more purchases when matching a bigger seller, all results derived in the paper also remain valid. Lastly, consider the case in which the size of each purchase is match specific. Denote it by $q(s, b)$. Again, the total sales from i to j now becomes $x_{ij}q_{ij}$. Again, it is obvious that when q is an increasing function of s and b or is FSD generated in ascending order in s_i and q_j, we will continue to have $x_{ij}q_{ij} \succ x_{ij'}q_{i'j'}$ for $b_i > b_{i'}$ and $s_i > s_{i'}$.

A.4 Robustness: Conditional on Realizations

In practise, empirical regularities are calculated based on statistics of realized assignment instead of fundamental parameters s or b. For instance, when examining the relationship between total sales and out-degree, we group sellers with the same realized number of connections, compute their average sales and compare it with the same statistic computed for another group of sellers with a higher number of realized connections. Will our previous results hold if we compute statistics based on realized assignment outcomes?

Formally, consider two statistics, $\hat{t}$ and $\tilde{t}$. Without loss of generality we consider the seller statistics, so both $\hat{t}_i$ and $\tilde{t}_i$ are statistics of realized assignment result and can be FSD ranked in ascending order of i. The question can be translated as: under what conditions can $\hat{t}$ be FSD ranked in ascending order of $\tilde{t}$? Our answer is that the sufficient condition is $\tilde{t}$ on the x-axis follows the monotonic probability ordering ($\succ_{MPR}$) in i.

The proof is the following. Let $u(\hat{t})$ be an increasing and piecewise differentiable function. Then, $E(u(\hat{t}) \mid \tilde{t})$ is given by

$$\begin{aligned} E(u(\hat{t}) \mid \tilde{t}) &:= \sum_{\hat{t}} u(\hat{t}) \mathbb{P}[\hat{t} \mid \tilde{t}] = E_i(E(u(\hat{t}) \mid \tilde{t}, i)) \\ &= \sum_i E(u_i(\hat{t}) \mid \tilde{t}) \mathbb{P}[i \mid \tilde{t}], \end{aligned}$$

where second equality is ensured by the law of iterated expectations. The variable $\mathbb{P}[i \mid \tilde{t}] = \frac{\mathbb{P}[\tilde{t}_i = \tilde{t}]}{\sum_i \mathbb{P}[\tilde{t}_i = \tilde{t}]}$ and is the probability that the realized value $\tilde{t}$ comes from seller i.

Consider two sellers with $s_i > s_{i'}$. We first prove that $E(u_i(\hat{t}) \mid \tilde{t})$ is an increasing function of i. For our purpose, it is sufficient to to consider the case when $\hat{t}$ is a monotone or order statistic. If $\hat{t}$ is a monotone statistic, based on the proof of proposition 1, we know that we can always construct coupling $(u_i'(\hat{t}), u_i'(\hat{t}))$ with $\mathbb{P}[u_i'(\hat{t}) > u_i'(\hat{t})] = 1$ for $u_i(\hat{t})$ and $u_i(\hat{t})$. If $\hat{t}$ is an order statistic, we proved its stochastic dominance in the proof of proposition 4. Hence by the stochastic dominance theorem we know the monotone coupling of $\hat{t}_i$ and $\hat{t}_{i'}$ also exist, hence so does that of $u(\hat{t}_i)$ and $u(\hat{t}_{i'})$. Then, by the monotonicity of conditional expectation, we have that $E(u_i(\hat{t}) \mid \tilde{t}) > E(u_{i'}(\hat{t}) \mid \tilde{t})$.

Now we proceed with the second step of our proof: if $\tilde{t}_i \succ_{MPR} \tilde{t}_{i'}$, then for any two values $\tilde{t}' > \tilde{t}''$. Note that if $\tilde{t}_i \succ_{MPR} \tilde{t}_{i'}$, we have

$$\frac{\mathbb{P}[\tilde{t}_i \leq \tilde{t}']}{\mathbb{P}[\tilde{t}_{i'} \leq \tilde{t}']} > \frac{\mathbb{P}[\tilde{t}_i \leq \tilde{t}'']}{\mathbb{P}[\tilde{t}_{i'} \leq \tilde{t}'']} \Rightarrow \frac{\mathbb{P}[\tilde{t}_i \leq \tilde{t}']}{\mathbb{P}[\tilde{t}_i \leq \tilde{t}'']} > \frac{\mathbb{P}[\tilde{t}_{i'} \leq \tilde{t}']}{\mathbb{P}[\tilde{t}_{i'} \leq \tilde{t}'']},$$

where the first inequality directly comes from the definition of MPR. The second inequality in turn implies that $\mathbb{P}[i \mid \tilde{t}]$ has the MPR property. Therefore, by the well-known result that $MLR \Rightarrow MPR \Rightarrow FSD$ (e.g., Eeckhoudt and Gollier, 1995), the distribution $\mathbb{P}[i \mid \tilde{t}']$ FSD the distribution $\mathbb{P}[i \mid \tilde{t}'']$. Using the ranking theorem, we then have

$$E[u(\hat{t}) \mid \tilde{t} = \tilde{t}'] > E[u(\hat{t}) \mid \tilde{t} = \tilde{t}''],$$

which holds and any increasing and piecewise differentiable function $u(\hat{t})$. Therefore, applying (the reverse of) the ranking theorem, we complete the proof that that $\hat{t}_{\tilde{t}}$ can be FSD ranked in ascending order of $\tilde{t}$.

In our model, the distribution of seller sales is binomial. The distribution of the number of buyers a seller matches with is Poisson binomial and is asymptotically Poisson when $\mathcal{N}$ is large and s is small. The total purchases of buyers are deterministic; the number of sellers that a buyer matches with is asymptotically normal (Holst, 1972; Weiss, 1958). To the best of our knowledge, these four statistics cover all x-axis variables researchers have used do date. In all cases, the distributions satisfy MPR stochastic orderings.[28]

A.5 Additional Proofs for Section 6.2

Proof of $\frac{\partial^2 E(log(x_{ij}))}{\partial b_j \partial s_i} < 0$.

Note that the partial derivative of a binomial distribution $Bin(b, s)$ is given by:

$$\frac{d\mathbb{P}(Bin(b,s)=x)}{dp} = \mathbb{P}(Bin(b,s)=x)(\frac{x}{s} - \frac{b-x}{1-s}),$$

$$\frac{d\mathbb{P}(Bin(b,s)=x)}{dn} = \mathbb{P}(Bin(b,s)=x)(log(1-s)).$$

Hence,

$$\begin{aligned}\frac{\partial^2\mathbb{P}(Bin(b,s)=x)}{\partial b\partial s} &= \mathbb{P}(Bin(b,s)=x)(\frac{x}{s} - \frac{b-x}{1-s})log(1-s) - \mathbb{P}(Bin(b,s)=x)\frac{1}{1-s} \\ &= \mathbb{P}(Bin(b,s)=x)\left((\frac{x}{s} - \frac{b-x}{1-s})log(1-s) - \frac{1}{1-s}\right),\end{aligned}$$

the first term on the RHS, $\mathbb{P}(Bin(b,s)=x)$, is always positive. The second term, $(\frac{x}{s} - \frac{b-x}{1-s})log(1-s) - \frac{1}{1-s}$, is increasing in x and approaches 0 when $x \to \infty$. Therefore, the whole term is negative, therefore we proved that $\frac{\partial^2\mathbb{P}(Bin(b,s)=x)}{\partial b\partial s} < 0$.

Then we can directly apply this result. Recall that a fall in the exchange rate is equivalent to all buyers at destination market c increasing the number of purchases from b_j to δb_j, where $\delta > 1$. And by proposition 3 we can view all buyers from country c as one buyer with $b_c = \sum_{j\in c} b_j$. Therefore, β_p have the same sign as

$$\begin{aligned}&\sum_x ln(x)(\mathbb{P}(Bin(\delta b_c, s_i) = x)) - (Bin(b_c, s_i) = x)) \\ &- \sum_x ln(x)(Bin(\delta b_c, s_{i'}) = x) - Bin(b_c, s_{i'}) = x))) < 0,\end{aligned}$$

[28] Poisson and normal distributions have the MLR property, which is sufficient to guarantee MPR ordering. The proof of the stochastic ordering of the discrete distributions can be found in, for example, Klenke and Mattner (2010).

where $s_i > s_{i'}$. Therefore $\beta_p < 0$. $\square$

Extensive margin adjustment

The probability that a seller matches which each buyer is an independent event with probability p_{ij}. The number of buyers that a seller matches with follows a Poisson binomial distribution:

$$\mathbb{P}(n_i^s = \kappa) = \sum_{A \in B_\kappa} \prod_{j \in A} p_{ij} \prod_{j' \in A^c} (1 - p_{ij'}),$$

where B_κ is the set of all subsets of κ integers that can be selected from $\{1, 2, ..., \mathcal{M}\}$. The set A^c is the complement set of A. Therefore

$$\mathbb{P}^{post}(n_i^s = \kappa) - \mathbb{P}^{post}(n_{i'}^s = \kappa) - \mathbb{P}(n_i^s = \kappa) + \mathbb{P}(n_{i'}^s = \kappa) = \sum_{A \in B_\kappa} \{\prod_{j \in A} p_{ij}^{post} \prod_{j' \in A^c} (1 - p_{ij'}^{post})$$
$$- \prod_{j \in A} p_{ij} \prod_{j' \in A^c} (1 - p_{ij'}) - \prod_{j \in A} p_{i'j}^{post} \prod_{j' \in A^c} (1 - p_{i'j'}^{post}) + \prod_{j \in A} p_{i'j} \prod_{j' \in A^c} (1 - p_{i'j'})\},$$

where $s_i > s_{i'}$, and the superscript *post* denote the post-shock variables. In our case, the latter refers to the case when the number of purchases increased from b_j to δb_j for all buyers in market c. As B_κ remains unchanged, to prove $\frac{\partial^2 \mathbb{P}(n_i^s = \kappa)}{\partial s_i \partial b_j} < 0$, it is sufficient to show that

$$\prod_{j \in A} p_{ij}^{post} \prod_{j' \in A^c} (1 - p_{ij'}^{post}) - \prod_{j \in A} p_{ij} \prod_{j' \in A^c} (1 - p_{ij'})$$
$$- \prod_{j \in A} p_{i'j}^{post} \prod_{j' \in A^c} (1 - p_{i'j'}^{post}) + \prod_{j \in A} p_{i'j} \prod_{j' \in A^c} (1 - p_{i'j'}) < 0 \quad \text{(A2)}$$

for all $A \in B_k$. Note that

$$\frac{\partial(\frac{p_{ij}^{post}}{p_{ij}})}{\partial s_i} \propto \frac{\partial p_{ij}^{post}}{\partial s_i} - \frac{\partial p_{ij}}{\partial s_i}$$
$$= \delta b_j (1 - s_i)^{\delta b_j - 1} - b_j (1 - s_i)^{b_j - 1} < 0,$$

and

$$\frac{\partial(\frac{1 - p_{ij}^{post}}{1 - p_{ij}})}{\partial s_i} \propto \frac{\partial p_{ij}}{\partial s_i} - \frac{\partial p_{ij}^{post}}{\partial s_i} + \frac{\partial p_{ij}^{post}}{\partial s_i} p_{ij} - \frac{\partial p_{ij}}{\partial s_i} p_{ij}^{post}$$
$$= b_j (1 - s_i)^{b_j + \delta b_j - 2} - \delta b_j (1 - s_i)^{b_j + \delta b_j - 2} < 0.$$

Therefore, $\frac{\prod_{j \in A} p_{ij}^{post} \prod_{j' \in A^c} (1 - p_{ij'}^{post})}{\prod_{j \in A} p_{ij} \prod_{j' \in A^c} (1 - p_{ij'})}$ decreases in s_i, which guarantees the equality in (A2). The remainder of the proof is simply identical to the proof of $\frac{\partial^2 E(log(x_{ij}))}{\partial b_j \partial s_i} < 0$. $\square$

Appendix B Empirical Appendix

B.1 Candidate Models and the Computation

ER model. Recall that $\mathcal{M}$ is the number of buyers and $\mathcal{N}$ is the number of sellers. Let $\mathcal{L}$ gives the total number of links observed in data. Therefore the probability that the ER model generates network $\mathcal{G}$ is $L_{ER}(\mathcal{G}) = p^{\mathcal{L}}(1-p)^{\mathcal{MN}-\mathcal{L}}$. In this model, the optimal probability p that maximize $L_{ER}(\mathcal{G})$ can be solved analytically and is given by $\hat{p} = \frac{\mathcal{L}}{\mathcal{MN}}$. The maximized likelihood is then given by $L_{ER}(\mathcal{G}) = \hat{p}^{\mathcal{L}}(1-\hat{p})^{\mathcal{MN}-\mathcal{L}}$.

FRG model. One difference of FRG model compared to our elementary model is that it is links rather than transactions that are randomized. Recall that the realized number of links of seller is denoted by n_i^s and buyer by n_j^b. As in Bezáková et al. (2006), for a network without multiple edges (i.e., more than one link between a pair), the probability of any particular matching is given by $L_{FRG}(\mathcal{G}) = \frac{\prod_i n_i^s! \prod_j n_j^b!}{\mathcal{L}!}$. Note that when we interested in the likelihood of the model conditional on the observed firm-degrees, there is no free parameters need to be estimated. The maximized likelihood is then directly given by $L_{FRG}(\mathcal{G})$.

Beckerian model. For the Beckerian model, the predicted matching is generated as in Section 7.1.3: we take the total number of connections of each buyer and seller as given from the data. We start from the largest buyer, assign its largest connection (in terms of total sales) to the most-connected seller, second-largest connection to the second-most-connected seller, and so forth. After completing assigning the links of the largest buyer, we move to the second-largest buyer and repeat the same process. In every round, we assign the buyer to the most connected seller that still has slots and with which the buyer has not yet been connected. By doing this, the model gives predicted probability either equal to 1 or 0. As $ln(p_{ij})$ is undefined when $p_{ij} = 0$, we set $p_{ij} = 1 - 0.1^{16}$ if the model predicts a match between i and j and $p_{ij} = 0.1^{16}$ otherwise. This treatment is equivalent to allowing for small prediction errors (see our later discussion on hierarchy model with optimal errors). The probability that the model generates the observed network structure is then given by $L_{Beckerian}(\mathcal{G}) = \prod_i \prod_j p_{ij}^{D_{ij}}(1-p_{ij})^{1-D_{ij}}$, where D_{ij} denote the matching outcome between i and j observed in data.

Beckerian model with prediction errors. Typically, when we conduct quantitative analysis, we always allow for some prediction errors as we do not expect a simple, stylized model to fit the data perfectly. Viewing through the lens of model selection, this practice, however, needs to be interpreted caution. Error term also has a structure and itself could also generate predictions. To illustrate, consider a general model $y = \mathcal{T}(x)$. The likelihood of $y = f(\mathcal{T}(x), \epsilon)$ is not an evaluation of the credibility of the model $\mathcal{T}(x)$, but a combination of $\mathcal{T}(x)$ and ϵ.[29] This is also the idea behind model averaging, i.e., $y = f(\mathcal{T}(x), \epsilon)$ should be viewed as an averaging of two models, $y = \mathcal{T}(x)$

[29] Think our often used form, $y = \mathcal{T}(x) + \epsilon$ is a special case of $y = f(\mathcal{T}(x), \epsilon)$.

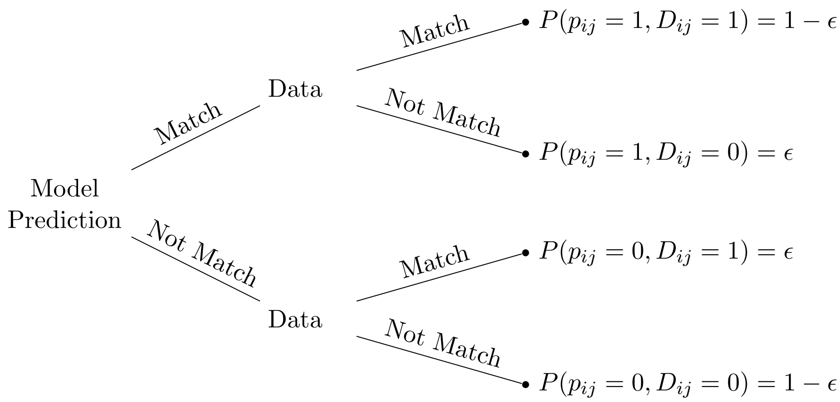


Figure A6: The error structure

and $y = \epsilon$.

Nevertheless, evaluating the likelihood of a hierarchy models with errors is still a valuable task. Firstly, one may believe that $y = f(\mathcal{T}(x), \epsilon)$ is the "real" description of the data. Correspondingly, it can also be viewed as a candidate model and be evaluated under model selection. For example, one can propose *Beckerian matching with (certain specified) prediction errors* being one candidate model describing production networks. Secondly, the estimated results also inform us to which extent the data can be explained by $\mathcal{T}(x)$. If the estimated likelihood improves a lot by introducing ϵ and the estimated ϵ is very large, it indicates the limited ability of $\mathcal{T}(x)$ in explaining data, even in the case when $\mathcal{T}(x)$ gives the lowest likelihood among all candidate models.

Following the above discussion, we consider a simple error structure: there is ϵ probability that the Beckerian matching model makes the wrong prediction for a buyer-seller pair. That is, we set $p_{ij} = 1 - \epsilon$ if the model predicts a match between i and j and $p_{ij} = \epsilon$ otherwise. We then optimally choose ϵ to maximize $L_{Beckerian}(\mathcal{G}) = \prod_i \prod_j p_{ij}^{D_{ij}} (1 - p_{ij})^{1-D_{ij}}$. We summarize the error structure in Figure A6.

Elementary model. We consider three cases of the elementary model, the baseline case as we presented throughout the paper, an extension with balls being a function of b_j, and the case when transaction data is not observable. We describe their calculation in turn. As before, the probability that a model generates the observed network structure is then given by $L_E(\mathcal{G}) = \prod_i \prod_j p_{ij}^{D_{ij}} (1 - p_{ij})^{1-D_{ij}}$, where $D_{ij} = 1$ if in data i and j are matched, zero otherwise. The value p_{ij} is the matching probability between i and j given by the model.

In the baseline case, we let $p_{ij} = 1 - (1 - s_i)^{b_j}$ and compute $L_E(\mathcal{G})$ directly. In the extension, we let $p_{ij} = 1 - (1 - s_i)^{\beta b_j + \alpha}$, and we optimally choose β and α to maximize $L_E(\mathcal{G})$. We expect β to be smaller than one, since we showed before that matches in reality tend to be path dependent. In the last case, we set s_i as the export share of firm i. We assume the number of balls, b_j, follows $b_j = \beta d_j + \alpha$. In case (1) we let d equal the total number of connections of the buyer, and in case (2) d equals the buyer's total purchase divided by the minimum order value \$100. Note that when

$\beta = 1$ and $\alpha = 0$, they corresponds to two extreme cases (1) one connection implies one transaction, and (2) each transaction's value equals the minimum order value, respectively. We compute s_i and b_j under the associated assumption, then let $p_{ij} = 1 - (1 - s_i)^{\beta b_j + \alpha}$ and optimally choose β and α to maximize $L_E(\mathcal{G})$.