University of Cambridge

# Three Essays in Firms, Trade and Development

John Albert Oteng Spray
Peterhouse

May 2020

This thesis is submitted for the degree of
*Doctor of Philosophy*

*"Dry dog taste sweet, ...*

*... but what do I eat while dog dry"*

**Liberian idiom on patience**

# Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed limit of 60,000 words.

John Spray
May 2020

# Acknowledgements

I am grateful to so many people who have helped me along the way towards this Ph.D. A few deserve special mention.

I would like to thank Vasco Carvalho, who pushed me further than I ever thought possible. Under his supervision, I have grown into an economist who cares about mechanisms and underlying forces, and who is equipped with a toolkit to answer challenging and insightful questions. His supervision propelled me to attain a Ph.D. I am proud of, and his guidance helped me cope with the emotional ups and downs.

I have been very lucky to have an amazing set of secondary supervisors who have helped at different points during my Ph.D. Matt Elliott has been a model advisor from our very first meeting where he encouraged me to take theory seriously. He is also a great person to be around and reminded me that economics can be fun as well as challenging. Dave Donaldson was extremely generous with his time and helped me develop an empirical and development-focused component to my thesis. Meredith Crowley gave me countless insights and pushed me to understand the underlying forces in international trade. Pramila Krishnan brought me to Cambridge and has remained a very helpful advisor, even after she moved to Oxford.

My Ph.D. has also benefited incalculably from the staff of the Government of Uganda, Government of Rwanda and the International Growth Centre in Uganda and Rwanda. Particular mention to Richard Newfarmer who kept my research grounded in the real world and pushed me to remember the policy-side of economics. I have also received amazing support and insights from Umulisa Adia, Astrid Hass, Nicole Ntungire, Jakob Rauschendorfer, Ritwika Sen, Victor Steenbergen, Anna Twum, and Sebastian Wolf.

Without doubt the people who deserve the most thanks are friends in Cambridge and beyond. I could not have completed this Ph.D. without the support of, among others, Anil Ari, Maarten De Ridder, Stephanie De Mel, Andrew Hannon, Zeina Hasna, Simone Hanebaum, Steph Mulhern, Nicolas Gustavo Paez, Alba Patozi, Monica Petrescu, Lida Smitkova, Nathan Smith, Anna Vitali, Dan Wales, and Alan Walsh. Thank you.

Finally, I would like to thank Granny, Mum, Dad, Clare and Alan who have kept me grounded throughout this long journey, and are a constant source of strength and encouragement.

# Preface

I present a collection of three essays exploring how firms in developing countries make supply-chain decisions and how those microeconomic decisions aggregate into macroeconomic outcomes.

Chapter 1 In the first chapter, which is co-authored with Vasco Carvalho and Matthew Elliott, we consider how a firm's position in a supply-chain can confer market power. We develop a tractable theory which introduces the notion of a bottleneck: a firm whose removal from the network leads to a sufficiently large fall in aggregate output such that supply can no longer meet demand. We develop a network algorithm to identify bottlenecks in an economy-wide production-network and apply these tools, at scale, in Uganda. We show that bottleneck firms have significantly larger profits, sales, wage bills, and higher mark-ups. They are also located in industries which have fewer new entrants.

Chapter 2 In the second chapter, I consider how firms form new supply-chain matches. I develop a model of firm-to-firm search and matching to show that the impact of falling trade costs on firm sourcing decisions and consumer welfare depends on the relative size of search externalities in domestic and international markets. These externalities can be positive if firms share information about potential matches, or negative if the market is congested. Using unique firm-to-firm transaction-level data from Uganda, I show empirical evidence consistent with positive externalities in international markets and negative externalities in domestic markets.

Chapter 3 In the third chapter, I build and estimate a dynamic quantitative version of the model presented in Chapter 2 to match transaction level tax data from Uganda. Structural estimates of the model's parameters provide evidence that the domestic market is more congested than the foreign market. I then show that a 25% reduction in trade costs will lead to a 5.2% increase in consumer welfare, 15% of which was due to search externalities. I also show that reducing search costs between firms could significantly increase welfare, but is best targeted on reducing international search costs when compared to domestic search costs.

# Contents

**1 Network Bottlenecks and Market Power** **1**
1.1 Introduction . . . . . . . . . . 2
1.2 A model of market power in a production network . . . . . . . . . . 6
1.2.1 Set up . . . . . . . . . . 6
1.2.2 Bottlenecks in uncompetitive equilibria . . . . . . . . . . 18
1.3 Scalable algorithm to find bottleneck firms . . . . . . . . . . 19
1.3.1 Identifying bottleneck firms in theory . . . . . . . . . . 20
1.3.2 Identifying bottleneck firms in practice . . . . . . . . . . 20
1.4 Proof of concept . . . . . . . . . . 25
1.4.1 Data . . . . . . . . . . 25
1.4.2 Characterization of bottlenecks . . . . . . . . . . 26
1.4.3 Consequences of bottlenecks . . . . . . . . . . 33
1.4.4 Robustness checks . . . . . . . . . . 35
1.5 Conclusion . . . . . . . . . . 40

**2 Search Externalities in Firm-to-Firm Trade** **41**
2.1 Introduction . . . . . . . . . . 42
2.2 A simple model of firm-to-firm search in two markets . . . . . . . . . . 45
2.2.1 Buyers, suppliers and consumers . . . . . . . . . . 46
2.2.2 Pricing and division of profits . . . . . . . . . . 47
2.2.3 Search and matching . . . . . . . . . . 47
2.2.4 Optimal search . . . . . . . . . . 48
2.2.5 Comparative statics . . . . . . . . . . 49
2.3 Data, context and descriptive statistics . . . . . . . . . . 51
2.3.1 Datasets . . . . . . . . . . 51
2.3.2 Context and trade cost reduction stylized facts . . . . . . . . . . 54
2.4 Reduced-form evidence of search externalities . . . . . . . . . . 60
2.4.1 Motivating evidence . . . . . . . . . . 60
2.4.2 Empirical strategy . . . . . . . . . . 60
2.4.3 Dataset . . . . . . . . . . 61

2.4.4 Main specification . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
2.4.5 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62
2.4.6 Mechanisms and alternative explanations . . . . . . . . . . . . . . 64
2.5 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65

**3 A Quantitative Model of Buyer-Supplier Search in Two Markets 67**
3.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 68
3.2 A quantitative model of buyer-supplier search in two markets . . . . . . . . 69
3.2.1 Buyers and suppliers . . . . . . . . . . . . . . . . . . . . . . . . . 70
3.2.2 Consumers . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71
3.2.3 Pricing and division of profits . . . . . . . . . . . . . . . . . . . . 71
3.2.4 Search and matching . . . . . . . . . . . . . . . . . . . . . . . . . 73
3.2.5 Search cost . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
3.2.6 Optimal search . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
3.3 Data . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
3.4 Estimation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 77
3.4.1 Estimating transfer equation . . . . . . . . . . . . . . . . . . . . . 78
3.4.2 Externally calibrated parameters . . . . . . . . . . . . . . . . . . . 79
3.4.3 Internally calibrated parameters . . . . . . . . . . . . . . . . . . . 79
3.4.4 Model fit . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
3.4.5 Heterogeneity . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 81
3.5 Counterfactual simulations . . . . . . . . . . . . . . . . . . . . . . . . . 82
3.5.1 Experiment 1: Transport cost reduction under structurally estimated parameters . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
3.5.2 Experiment 2: Transport cost reduction under constant returns to scale matching function . . . . . . . . . . . . . . . . . . . . . . . . 86
3.5.3 Experiment 3: Search cost reduction . . . . . . . . . . . . . . . . . 86
3.6 Concluding remarks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 90

**A Appendix to Chapter 1 93**
1 Empirical appendix . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93
A.1 Hierarchical firm clustering algorithm . . . . . . . . . . . . . . . . 93
A.2 Technology identification algorithm . . . . . . . . . . . . . . . . . 94
A.3 Identifying if there are no upstream bottlenecks . . . . . . . . . . . 96
2 Theoretical Appendix . . . . . . . . . . . . . . . . . . . . . . . . . . . 97
B.1 Examples . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 97
B.2 Omitted proofs . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 98

**B Appendix to Chapter 2** **105**
1 Context appendix . . . 105
A.1 Map of trade corridor . . . 105
A.2 Data comparison . . . 106
2 Mathematical appendix . . . 107
B.1 Comparative statics in the two-period model . . . 107
3 Empirical appendix . . . 109
C.1 Descriptive statistics . . . 109
C.2 Robustness tests . . . 111

**C Appendix to Chapter 3** **117**
1 Mathematical appendix . . . 117
A.1 Pricing, profits and transfer equation . . . 117

**References** **119**

# List of Figures


1.1 Example of technology DAG and Supply-Chain DAG . . . . . . . . . . . . . 8
1.2 Example of supply-chain with no bottleneck . . . . . . . . . . . . . . . . . . 16
1.3 Example of supply-chain with one bottleneck . . . . . . . . . . . . . . . . . . 16
1.4 Cement inputs example . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
1.5 Universe of firm-to-firm transactions in Uganda, 2010-2015 . . . . . . . . . 27
1.6 All firms . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29
1.7 Potential bottlenecks highlighted . . . . . . . . . . . . . . . . . . . . . . . . 29
1.8 Bottleneck in any period highlighted . . . . . . . . . . . . . . . . . . . . . . 29
1.9 Bottleneck in only one period highlighted . . . . . . . . . . . . . . . . . . . 29
1.10 Example of a bottleneck supply-chain . . . . . . . . . . . . . . . . . . . . . 31
1.11 Sales . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
1.12 Wage bill . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
1.13 Age . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
1.14 Gross profit . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
1.15 Markup . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
1.16 Betweenness centrality . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
1.17 Herfindahl-Hirschman Index . . . . . . . . . . . . . . . . . . . . . . . . . . 33
1.18 New entrants . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

2.1 Model environment . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
2.2 Locations of firms in Uganda . . . . . . . . . . . . . . . . . . . . . . . . . . 53
2.3 Transport costs for a 20’ container . . . . . . . . . . . . . . . . . . . . . . . 56
2.4 Transport costs and new importers . . . . . . . . . . . . . . . . . . . . . . . 57
2.5 $\beta$ coefficients from specification 2.12 . . . . . . . . . . . . . . . . . . . . . . 58
2.6 $\beta$ coefficients from specification 2.13 . . . . . . . . . . . . . . . . . . . . . . 59
2.7 Percentage of suppliers’ matches which have an existing buyer in location . 61

3.1 Model fit: buyer and supplier out-degree . . . . . . . . . . . . . . . . . . . . 82
3.2 Search by marginal cost . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
3.3 Search by marginal cost . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85

3.4 Mass of firms with $S_I$ international suppliers . . . . . 85
3.5 Mass of firms with $S_D$ domestic suppliers . . . . . 85
3.6 Average number of international and domestic suppliers for different reductions in trade costs when search externalities are shut down (CRS) compared to structurally estimated parameters (IRS) . . . . . 87
3.7 Consumer welfare gains from trade when search externalities are shut down (CRS) compared to structurally estimated parameters (IRS) . . . . . 88
3.8 Search by marginal cost if reduce domestic search costs by 25% . . . . . 89
3.9 Search by marginal cost if reduce international search costs by 25% . . . . 89

A.1 Hypothetical cement chain before and after algorithm . . . . . 96
A.2 Market clearing example . . . . . 98

B.1 Map of trade corridor . . . . . 105
B.2 Exports data comparison . . . . . 106
B.3 GDP and total output . . . . . 106
B.4 Transport costs and number of importers . . . . . 109
B.5 Transport costs and imports . . . . . 109

# List of Tables

1.1 Number of observations . . . 25
1.2 Bottleneck consequences . . . 34
1.3 Do distortions propagate? . . . 36
1.4 Centrality vs. MMF horserace . . . 37
1.5 Robustness table . . . 38

2.1 Descriptive statistics . . . 53
2.2 Import suppliers . . . 63
2.3 Domestic suppliers . . . 64

3.1 Estimating the transfer equation . . . 78
3.2 Model parameters . . . 80
3.3 Model fit . . . 81
3.4 Outcomes from 25% transport cost reduction . . . 84
3.5 Outcomes from 25% transport cost reduction under different matching functions . . . 87
3.6 Outcomes from 25% search cost reduction . . . 88

B.1 Newly added domestic suppliers among new importers . . . 110
B.2 Dropped suppliers' new matches . . . 110
B.3 Value from domestic suppliers among new importers . . . 110
B.4 Dropped suppliers' new matches - value . . . 111
B.5 Same and next-door balance table . . . 111
B.6 Imports suppliers from East African Community . . . 112
B.7 Domestic export suppliers . . . 113
B.8 Domestic suppliers . . . 114
B.9 Import suppliers . . . 115

# Chapter 1

# Network Bottlenecks and Market Power

Abstract We consider how a firm's position in a production network can confer market power. We develop a tractable theory of market power in production networks which introduces the notion of a bottleneck: a firm whose removal from the network leads to a sufficiently large fall in aggregate output such that supply can no longer meet demand. The location of these bottlenecks can depend both on a firm's immediate connections, and also on the entire structure of the network. We show that the existence of bottlenecks allows not only bottleneck firms to price above marginal cost, but that these distortions allow other non-bottleneck firms to also price above marginal cost in equilibrium. We develop a network algorithm to identify bottlenecks in an economy wide production-network and apply these tools, at scale, in Uganda. We show that bottleneck firms have significantly larger profits, sales, wage bills, and higher mark-ups. They are also located in industries which have fewer new entrants.

## 1.1. Introduction

Goods and services reach final consumers via often complex supply-chains. The routing of economic activity and efficiency of the system depends on prices set along these chains. In this context, an important question to consider is whether a firm's network position can confer market power, and if so, how these corresponding distortions then propagate through the supply network.

We seek to address these questions both theoretically and empirically. We build a tractable theoretical model of market power in a production network which introduces the notion of a bottleneck firm: a firm whose removal from the network leads to a sufficiently large fall in aggregate output, at current prices, such that supply can no longer meet demand. We then propose a scalable network algorithm to identify bottlenecks of market power in a supply chain. The algorithm takes, as an input, all transactions in an economy and gives, as an output, the set of bottleneck firms that preclude the economy from operating at first best. We then apply these tools, at scale, in a developing country using the near universe of Ugandan supply relationships, between 2010-2015. We identify bottlenecks across a wide array of industries and stages of production, and provide evidence that these firms have high market power based on their observable characteristics.

We develop a theory of market power in a production network in which firms are located on an exogenously given hierarchical supply network, such that goods flow downstream from raw materials to final goods producers via intermediary producers. Firms have Leontief production technology over the inputs they use from their suppliers, heterogeneous link capacities on the value of goods they can buy and sell from each connecting firm, and they cannot form new links. Firms simultaneously pick the price of their good which determines their demand which, in turn, determines the demand of firms up- and downstream in the network. This is non-trivial given supply chains are typically complex and interact with each other in important ways. For example, a given input can feature in the supply chains for several different types of goods. Therefore, identifying the allocation of demand through an economy-wide network is a complex problem. We resolve this by modelling the flow of demand as an auxiliary flow problem which is analogous to the flow of fluids through a system of pipes. Just as there are bottleneck conduits - pipes which, if blocked, will substantially decrease the total of flow of water - there are also bottleneck firms whose removal will substantially affect economy-wide production and leave final product demand for goods and services unmet. We show that these *bottleneck firms* will price above marginal cost in equilibrium.

These primary distortions will create secondary distortions. There will be horizontal spillovers if competitor firms price above marginal cost in response to the pricing of the bottleneck firm. There will be vertical spillovers if the absence of a bottleneck upstream allows firms to earn higher profits, as they can set the same price as a competitor who does have a bottleneck supplier, and hence has higher input costs. These secondary distortions highlight a key distinction between the policy implications arising from this paper, when compared to the more standard policy approach. Whereas a competition authority might decide to intervene with any firms operating with abnormal profits, in some instances these firms are *symptoms* and not the *cause* of the inefficiency. Indeed, to return the economy to first best, we show that it is sufficient to intervene only in the primary distortion i.e. markets with bottleneck firms.

Having introduced the notion of a bottleneck in theory, we then consider how to identify these bottlenecks in firm-to-firm transaction data. An advantage of modelling this problem as an auxiliary flow problem is that it gives us access to a rich toolkit of algorithms from the engineering and computer science literature. We utilize the Ford-Fulkerson method (Ford and Fulkerson, 1956) as implemented by the Edmonds-Karp (Edmonds and Karp, 1972) algorithm to identify the maximum flow of goods from raw materials to final goods when all supply-chains operate at full capacity. We then, one-by-one, drop firms from the network and measure their marginal contribution to the maximum flow. If supply can no longer meet demand when that firm is removed from the network, we classify this firm as a bottleneck.

We develop new tools in order to adapt the auxiliary flow problem to an economic setting. First, the theory requires that supply-chains can be represented by a Directed Acyclical Graph (DAG). We, therefore, prune the minimum set of edges to convert the dataset into a DAG using the Eades et al. (1993) algorithm.[1] Second, unlike the flow of fluid through a system of pipes, goods go through processing as they move from raw materials to final goods. We, therefore, develop a novel Hierarchical Clustering Algorithm (HAC) to transform the units of transactions between agents from a monetary value to a unit of the final production good.[2] Third, we infer the edge capacity and the production network structure utilizing the history of firm-to-firm trades.

We operationalize this approach on tax administration data from the Government of Uganda between 2010 and 2015. The most important dataset contains firm transaction receipts recorded for Value Added Tax (VAT) purposes. VAT-registered firms submit a monthly value added tax return form which includes the universe of their transactions over the previous month. Importantly, this dataset provides details on the transaction value and the tax identifier of the firm on the other side of the transaction. We combine this

[1] This also provides a partial ordering of nodes from those which are most likely to be raw material producers to final retailers.Intuitively, the algorithm classifies nodes which are "sink-like" or "source-like" in that their out-degree relative to their in-degree is low or high. Once pruned, we connect firms with no inputs to an artificial source. We connect firms to an artificial sink using their final goods sales.

[2] This ensures that the flow of goods is preserved at each node, a necessary condition to run the Ford-Fulkerson algorithm.

data with information from Corporate Income Tax and the business registration including information on the firm's 4-digit ISIC sector.

In our baseline model, we identify an average of 50 bottleneck firms, every period.[3] We find bottleneck status is persistent over time and that these firms are located in three characteristic sectors which are consistent with anecdotal evidence on market power in Uganda.[4] First, we observe bottlenecks in light manufacturing industries including agricultural, food and drinks supply-chains. In Uganda, these are sectors which have large fixed costs relative to the small market size, which only supports one or two large firms (Agarwal and Spray, 2016). Second, we observe bottlenecks in more traditional natural monopolies, including utilities sectors. This is what one might expect, given the importance of utilities to a large number of firms. Third, we identify several bottleneck firms which provide intermediary input services such as financial services. This demonstrates why focusing on the full economy is important, as in the utilities sectors, these firms are providing inputs into a large number of different sectors. Therefore, if we were to look only at firms with market power on a supply-chain by supply-chain basis we might miss bottlenecks which operate across multiple chains.

We find that bottleneck firms have observable characteristics consistent with market power. When looking across firms, we find that bottleneck status is correlated with age, higher sales, higher wage bills, higher profits, and higher mark-ups. They are also located in less competitive industries as calculated by the Herfindahl-Hirshman Index (HHI), and are located in industries with fewer new entrants, suggesting greater barriers to entry.

We find that when firms become bottlenecks they act in ways consistent with greater market power. For the same firm across time, we find that a 1% increase in a firm's marginal maximum flow[5] corresponds to a 1.7% increase in sales and 0.5% increase in profits, statistically significant at the 1% level. Once a firm becomes a bottleneck, this effect roughly doubles the increase in sales and profits. Becoming a bottleneck also results in a 1.2% increase in firm mark-ups with each 1% increase in maximum flow, statistically significant at the 1% level.

Finally, we find evidence consistent with bottleneck distortions propagating through vertical supply-chain spillovers. We find that having no bottleneck upstream corresponds to 7% higher sales and 11% higher profits, significant at the 1% and 10% levels, respectively. Additionally, having no bottleneck upstream corresponds to a 6% higher markup, although this is not statistically significant. This is consistent with predictions from the

[3]We use periods of six month intervals

[4]Given we are using confidential data, we are restricted in not revealing which firms are bottlenecks. However, we can provide some general information about which sectors they belong to.

[5]This is our measure of how much potential output would drop if this firm was removed from the network.

theory.

This paper relates to four main strands of the literature.

Several important theoretical works have shown that a firm's network position can confer market power (Manea, 2018; Goyal and Vega-Redondo, 2007; Choi et al., 2017; Condorelli et al., 2016; Kotowski and Leister, 2018; Farboodi, 2014; Siedlarek, 2015; Nava, 2015; Fainmesser, 2014; Di Maggio and Tahbaz-Salehi, 2015; Ostrovsky, 2008). A central theme of the theoretical literature, spanning several quite different modelling approaches, is that firms at bottleneck positions in the network have market power. The exact definition of bottleneck differs across models, but the idea is the same—where there is no viable, equally good, alternative path that can be taken through the network, such that trade must pass through a given firm, that firm has market power. In this paper, we take a direct approach to finding and identifying bottlenecks. We model the supply network and production process as a flow problem, in which raw materials transition through the supply network and end up (in a transformed state) at consumers. Bottleneck firms are then those firms that would constrain this flow were they removed from the network. The model captures the main features of the existing theoretical literature, whilst being tractable enough to fit the data.

We relate to the literature on market power and supply-chains in developing countries. Existing research has shown that intermediaries in agricultural supply-chains (Casaburi et al., 2013; Casaburi and Reed, 2017; Bergquist and Dinerstein, 2019) and among imported inputs (Atkin and Donaldson, 2015) are important sources of market power in developing countries. Fafchamps (2004) shows that, across many markets in Sub-Saharan Africa, firms organise into supply-chains influenced by their social network. This creates barriers to entry which can confer market power.[6] While informative, this literature is primarily focused on case-studies of particular sub-sectors, especially in agriculture. Moreover, the tools used to identify market power typically rely on pricing data. Our data and methodology allow us to look across all sectors of the Ugandan economy to assess holistically the major sources of market power using only firm transaction data, representing a significant departure from the existing literature in focus and methodology.

We also relate to a literature demonstrating the importance of modelling production networks for macroeconomic outcomes (Bigio and La'O, 2016; Grassi, 2017; Tintelnot et al., 2018; Grassi and Sauvagnat, 2019; Kikkawa et al., 2019; Acemoglu et al., 2019; Baqaee and Farhi, 2019; Liu, 2019). This literature has shown both theoretically and empirically the importance of network structure to the propagation of micro-economic shocks and has provided justification for policy intervention. We build on this literature by demonstrating that network structure is important for considerations of firm market

[6] There is also an important literature on information diffusion in networks in developing countries (Fafchamps and Quinn, 2018; Spray, 2020).

power which, in turn, impacts the efficiency and allocation on goods in the wider economy.

Finally, we contribute to the growing literature on measuring market power (De Loecker et al., 2019; Gutiérrez and Philippon, 2017; De Ridder, 2019). This literature primarily focuses on either the firm or the sector, by considering, for instance, mark-ups or market shares. We take a very different, but complementary, approach to considering market power. We focus on the full production network, allowing observation of competition across multiple supply-chains. This allows us to consider the local, and non-local, impact of network structure on a firm's market-power. We are also able to calculate our metric using only data on firm-to-firm transactions, obviating the need, for example, for mark-up estimation.

The remainder of this chapter is organized as follows: Section 1.2 presents a model of market power in a production network; Section 1.3 introduces an algorithm to find bottleneck firms in data; Section 1.4 applies the algorithm to data from Uganda and considers the empirical consequences of bottleneck firms; and Section 1.5 concludes.

## 1.2. A model of market power in a production network

We now present a model in which firms located on a production network transform raw materials into final goods through layers of intermediary goods using a given technology. We use the theory to demonstrate how a firm's network position can confer market power and how these primary distortions can create secondary distortions.

### 1.2.1. Set up

There is a finite set of raw materials $R$, a finite set of intermediate goods $I$ and a finite set of final goods $F$. Abusing notation we let $R$, $I$ and $F$ also denote the cardinality of these sets, and also use these terms as an index. A vector $x^R \in \mathbb{R}^R_+$ denotes quantities of the different raw materials, a vector $x^I \in \mathbb{R}^I$ denotes quantities of intermediate goods and $x^F \in \mathbb{R}^F$ denotes quantities of final goods. The vector $X := (x^R, x^I, x^F) \in \mathbb{R}^{R+I+F}_+$ combines these vectors.

Each final good and intermediate good is produced using a unique recipe which specifies the amount of the different intermediate goods and raw materials required to produce one unit. We represent the recipe for all goods in a matrix $A \in \mathbb{R}^{R+I+F}_+ \times \mathbb{R}^{R+I+F}_+$, where the entry $A_{\theta\theta'}$ represents the amount of good $\theta'$ that is required to make one unit of good $\theta \in R \cup I \cup F$. For each input quantity $x_{\theta'}$, the amount of good $\theta$ that can be made is $\min_{\theta':A_{\theta\theta'}>0} x_{\theta'}/A_{\theta\theta'}$. In other words, production is Leontief.[7]

We can represent this technological dependence of goods on each other in a directed

[7]Each smart phone has one screen, one speaker, one microphone, etc.

weighted network with $A$ as the adjacency matrix so that a link from a node $i$ to $j$ represents the amount of good $j$ required to produce a unit a good $i$. We assume that this technology network is a directed acyclic graph (DAG).[8]

### Supply network

We now introduce firms into a second network. Let $N$ denote the set of firms. These firms produce intermediate goods and final goods. We assume that raw materials are extracted at a given unit cost to be used by firms, and we assume there is no limit to the amount that can be extracted.[9] Each firm is able to produce a single good. We let this be the type of the firm and index product types by $\theta \in R \cup I \cup F$. Let $Z(\theta)$ denote the group (set) of firms of type $\theta$. If $\theta$ is an intermediate good or raw material, it may appear in the supply chain for many final goods and may be used as an input for multiple goods within the same supply chain. Further, there may be many firms producing the same good, or just one firm.

These firms are embedded in a weighted *supply network* $G$. This differs from, but is closely related to, the technology network already introduced. Each node in the supply-network represents a firm, and an edge from firm $i$ to firm $j$ with weight $w_{ij}$ represents that firm $i \in Z(\theta)$ can supply at most $w_{ij}$ units of good $\theta$ to firm $j$. We let $w_{ij} = 0$ if firm $i$ produces a product that is *not* an input for firm $j$ or these firms do not have a supply-relationship. The supply network $G$ is a directed acyclical graph, which is a consequence of the structure of the technology network.

Firms also face capacity constraints on their overall output. Let $\phi_i$ denote the capacity constraint of firm $i$.

Finally, we assume there is a representative consumer who can purchase from all producers which we index $c$.

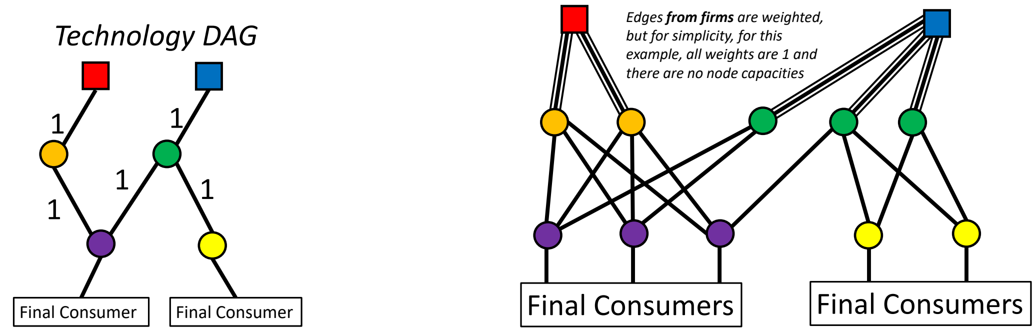
To focus on the role of the network structure, and abstract from heterogeneities in production technology, we assume a given transformation of raw materials and/or intermediate goods costs the same amount per unit, for all firms performing this transformation. Thus, there is a unique marginal cost of production (transformation) associated with each intermediate and final good.

To demonstrate the set up of the model, we now introduce a simple example of a production network. We will also use this example, later, to demonstrate key concepts in the theory.

**Example 1.** Example of technology DAG and Supply-Chain DAG

[8]For example trees can be used in the production of pulp which can be used in the production of

Figure 1.1. Example of technology DAG and Supply-Chain DAG




*Notes:* The left hand figure shows the underlying technology DAG e.g., one units of the orange intermediate and one unit of the green intermediate are required to produce one unit of the purple final good. The right hand figure shows a hypothetical supply chain DAG. Red and blue nodes are raw materials which provide large finite capacity to orange and green intermediary producers. There are multiple firms producing each good. The capacity on each edge is 1.

## Partially ordering the goods

We associate each intermediate good and each firm in the supply chain for $j$ with a demand-level.

**Lemma 1.** *There exists a non-empty set of goods that are sold only to final customers.*

The proof of Lemma 1 is relegated to Appendix B.2. Denote the non-empty set of goods that sell only to final producers as demand-level 1 goods. We can then define demand-level 2 goods as those goods that are not demand-level 1 goods and sell only to demand-level 1 producers and/or final consumers. By the same logic as that used to prove Lemma 1, the set of demand-level 2 goods producers must be non-empty. We can then proceed iteratively until all producers have been assigned a demand-level. We let $L^D$ denote the last demand-level and index these demand-levels by $l^D \in \{1, \ldots, L^D\}$.

We also associate each intermediate good in the supply chain for $j$ with a supply-level. The same argument used to prove Lemma 1 can be applied to show the following corollary.

**Corollary 1.** *There exists a non-empty set of goods that are produced only with raw materials.*

The proof of Corollary 1 is in Appendix B.2. Denote the non-empty set of goods that are produced only with raw materials by supply-level 1 goods. We can then define supply-level 2 goods as those goods that are not supply-level 1 goods and use only supply-level 1 inputs and/or raw materials. The set of supply-level 2 goods producers must be

paper. However, paper cannot be used in the production of either trees or pulp

[9]Constraints limiting extraction could be easily incorporated into the model—we omit them for simplicity.

non-empty (by the same logic as Corollary 1). We can then proceed iteratively until all producers have been assigned a supply-level. We let $L^S$ denote the last supply-level and index these supply-levels by $l^S \in \{1, \dots, L^S\}$.

Assigning the goods demand and supply-levels provides two partial orders of the goods based on their position in the technology network. This will be helpful later.

### Consumer problem

A representative consumer has utility $u(x^F)$, where $u(\cdot)$ is a continuous, twice differentiable, strictly increasing and strictly quasi-concave function. Let $x$ be a vector which records quantities of goods the consumer purchases from each final goods producer. Given that producers set prices $p$, the consumer's problem is to maximise utility by choosing the quantity of each final good to purchase from each producer,

$$\max_x u(x^F) \text{ subject to } x \cdot p \leq \omega,$$

where $\omega$ is the consumer wealth and each entry of the vector $x^F$ is $x^F_\theta = \sum_{i \in Z(\theta)} x_i$ for $\theta \in F$ (recall that $x^F$ records only a quantity for each type of good, and not which firms that type of good was purchased from, while in comparison the vector $x$ has an entry for each final good producer).

The solution to this problem generates a demand correspondence. However, as we have assumed that consumer utility is strictly quasi-concave and continuous, the demand correspondence is single-valued and continuous (see, for example, Barten and Böhm (1982)). We let $D_{c\theta}(p)$ denote the representative consumer's demand function for good $\theta$ given prices $p$. We also assume that $D_{c\theta}(p)$ is decreasing in the price of good $\theta$ for all $\theta$—i.e., none of the goods is a Giffen good.[10]

### Timing

Formally the timing of the game is as follows.

(i) All firms simultaneously set prices.

(ii) Pairwise demands and supplies are determined.

(iii) Consumption occurs.

[10] A sufficient but not necessary condition for this is that all goods are normal, and a sufficient condition for all goods to be normal is that are all Auspitz-Lieben-Edgeworth-Pareto complements (Chipman, 1977).

### Marginal cost pricing

Let $\kappa_i \in \mathbb{R}_+$ be $i$'s constant marginal cost of transformation. This is the same for all firms producing the same good $\theta$. The overall marginal costs of a firm will depend on their cost of sourcing their required inputs as well as their processing costs. However, if all firms of the same type set the same price, the input costs faced by each producer of type $\theta$ will be the same. In this case, the overall marginal cost of production of a firm $i$ of type $\theta$ will be

$$\gamma_\theta := \kappa_\theta + \sum_{\theta'} A_{\theta\theta'} p_{\theta'},$$

where $p_{\theta'}$ is the price charged by producers of goods of type $\theta'$. We will be particularly interested in equilibrium in which all firms set prices equal to their overall marginal cost. In this case,

$$\gamma_\theta = \kappa_\theta + \sum_{\theta'} A_{\theta\theta'} \gamma_{\theta'}.$$

Because of the DAG structure of the network, this formula recursively pins down the price of all goods in terms of the marginal costs of transformation only.[11] We let $\gamma^*$ denote the unique vector of prices that solves this system of equations.

### Market clearing

We fix the price of raw material inputs equal to the cost of extracting them and impose no capacity constraints on the amount of raw materials firms can extract at this price.

[11]Recall that, by assumption, the price of raw material goods is fixed at their extraction cost. Thus, for $\theta \in R$,

$$\gamma_\theta = \kappa_\theta.$$

For a good $\theta$ in supply-level 1 (using only raw materials in production),

$$\gamma_\theta = \kappa_\theta + \sum_{\theta'=1}^{R} A_{\theta\theta'} \kappa_{\theta'}.$$

For those goods in supply-level 1 (using only these intermediate goods and raw materials in production),

$$\begin{aligned}
\gamma_\theta &= \kappa_\theta + \sum_{\theta'=1}^{R} A_{\theta\theta'}\kappa_{\theta'} + \sum_{\theta'=R+1}^{R+I} A_{\theta\theta'}\gamma_{\theta'} \\
&= \kappa_\theta + \sum_{\theta'=1}^{R} A_{\theta\theta'}\kappa_{\theta'} + \sum_{\theta'=R+1}^{R+I} A_{\theta\theta'} \left(\kappa_{\theta'} + \sum_{\theta''=1}^{R} A_{\theta'\theta''}\kappa''_{\theta}\right).
\end{aligned}$$

And so on.

Firms, however, choose their prices. After these prices have been selected, the market attempts to clear (taking these prices as given).

Modelling market-clearing in this context is a non-trivial exercise for four main reasons. First, many of the firms produce intermediate goods, and demand for these goods is, at least partially, derived from the demand other firms have for this input. Second, the constraints on who can trade how much and with whom limit the demands that firms can receive. Third, if one firm offers a lower price than another for the same input, then demand should be preferentially allocated to the firm with the lower price.[12] Fourth, when firms offer the same prices, demand must be rationed in some way across them. While this problem is complicated, the assignment of firms to levels, which we derived from the DAG structure, allows demand and supply to be defined in a consistent way satisfying the above constraints.

For a supply profile, $S = S_{ij}$ for all $ij$, we can inductively define a set of *supply-constrained* relationships. We say firm $j$'s supply to $i$ is supply-constrained if either (i) $w_{ji} = S_{ji}$; or (ii) $\phi_j = \sum_k S_{jk}$; or (iii) there exists an input type $\theta$ used by $j$ such that $k$'s supply to $j$ is supply-constrained for all suppliers $k \in Z(\theta)$.[13]

If a firm $i$ has an upstream supply relationship with $j$ that is constrained, then there is no scope for $i$ to source more from $j$. Even if supply in the relationship is not directly constrained ($w_{ji} > S_{ji}$ and $\phi_j > \sum_k S_{jk}$), were $j$ to ask its suppliers to increase their supply to it, $j$ would find itself unable to source one of the inputs it requires to increase production.

It is also helpful to define what we'll call the total transacted cost associated with each firm's output. In general, given supplies $S$ and demands $D$, the costs associated with $i$'s costs of production are given by $c_i + \sum_j p_j \min\{S_{ji}, D_{ij}\}$. Because of the DAG structure of the supply network, it is easy to also find the cost associated with all these inputs, the costs associated with all their inputs, and so on. We define the total transacted cost associated with $i$'s output, denoted by $\Psi_i$, recursively

$$\Psi_i := \kappa_i + \frac{\sum_j (p_j + \Psi_j) \min\{S_{ji}, D_{ij}\}}{\sum_k S_{ik}},$$

and set $\Psi_i = 0$ for raw materials. The DAG structure of the supply network implies that $\Psi_i$ is well defined and unique for all $i$.[14] As there is a unique technology for producing

[12] as permitted by the aforementioned constraints on trade.

[13] Although this definition is recursive, it is well defined because of the DAG structure inherited by the supply network from the technology network. supply-level 1 producers use only raw material as inputs and so have no suppliers that are supply-constrained. Each such producer therefore has a supply-constrained relationship with a downstream firm $i$ if and only if (i)$w_{ji} = S_{ji}$ or (ii) $\phi_j = \sum_k S_{jk}$. This allows all relationships with supply-level 1 producers to be assigned to being supply-constrained or not. But then supply-level 2 producers can be assigned, and so on.

[14] To see this, suppose we have supply and demand $(S, D)$ and we start with supply-level 1 producers,

each good, and all firms of the same type have the same constant marginal cost of transformation, a difference in total transacted cost between suppliers of the same product must be due to a difference in prices set in their supply chains.[15] As our solution concept has firms simultaneously setting prices, letting markets clear in a way that resolves supply indifferences by preferentially allocating supply to lower total transacted cost suppliers, allows market clearing to respond in a realistic way to relatively high upstream prices.

**Definition 1.** Given prices $\mathbf{p}$, demand $D$ and supply $S$ clear the market if and only if,

(i) demand for firm $i \in Z(\theta)$'s output induces $i$'s input demand (if input $\theta'$ is required by firm $i$, then $\sum_{j \in Z(\theta')} D_{ji} = \frac{\sum_{k \in Z(\theta')} D_{ik}}{A_{\theta\theta'}}$ for all $i \in N \cup \{c\}$); and

(ii) network supply constraints are satisfied ($S_{ij} \leq w_{ij}$ for all $i \in N \cup \{c\}$ and all $j \in \{N \cup R\}$ and $\sum_i S_{ij} \leq \phi_j$ for all $j \in N$).

(iii) pairwise demands are met ($S_{ij} = D_{ji}$ for all $i \in N \cup \{c\}$ and all $j \in \{N \cup R\}$), with $\sum_{i \in Z(\theta)} D_{ci} = D_{c\theta}(p)$).

(iv) no firm can source any input cheaper via a supply relationship that is not supply-constrained.

(v) indifference about whom to buy from is resolved in favor of lower total transacted cost suppliers for all $i \in N \cup \{c\}$.

We can now specify how pairwise demands and supplies are determined given prices $\mathbf{p}$. When markets can clear, we select demands and supplies that clear them, but pinning down agents' incentives to deviate requires also specifying what happens when markets can't clear.

Given prices $\mathbf{p}$, pairwise demands $D$ and supplies $S$ are selected as follows:

1. If there exist demands $D$ and supplies $S$ such that $(\mathbf{p}, D, S)$ clears the market (satisfy market clearing conditions (i)-(v)), such a demand and supply profile pair is selected. We are agnostic about which such pair is selected when there are many.

---

who use only raw material as their input. In this case $\Psi_i = \kappa_i + \sum_j p_j \min\{S_{ji}, D_{ij}\}$. supply-level 2 firms then have well defined total transacted cost associated with supply and demand $(S, D)$ as they only use level 1 inputs and raw materials, and so on. Note, the total transacted cost is a somewhat artificial quantity—input prices represent transfers rather than true costs and so the total transacted cost will tend to exceed the value of the final good produced. Nevertheless, it will be useful for keeping track of upstream price increases that are not passed on.

[15] If one of firm $i$'s indirect upstream suppliers increases its price, but $i$ still faces the same prices from its direct suppliers, the total transacted cost of production will increase while $i$'s input costs remain fixed.

2. Otherwise we consider whether there exists demands $D$ and supplies $S$ such that $(\mathbf{p}, D, S)$ satisfies market clearing conditions (i)-(iii). If so, such a demand and supply profile are selected. Again, we are agnostic about which one.

3. Finally, if there do not exist demands $D$ and supplies $S$ $(\mathbf{p}, D, S)$ such that conditions (i)-(iii) in the market clearing definition are satisfied, we look for the minimum reduction in the value of final consumer demands such that, with these demands imposed, there exist demands and supplies $(\mathbf{p}, D, S)$ such that conditions (i)-(iii) in the market clearing definition are satisfied.

Conditions (i)-(iii) in the market clearing condition concern just the matching of demand to supply and the feasibility of supply. Condition (iv) just requires firms to source cheaper inputs when they are available. Condition (v) is more technical. To see why it is required and how it captures realistic market forces, we present an example in Appendix B.1.

A mapping from prices to pairwise demands $D$ and supplies $S$ is feasible if given prices $p$, pairwise demand and supplies are selected in a way consistent with the above criteria. There will typically be many consistent mappings.

### Equilibria

An equilibrium is prices, demand and supplies $(\mathbf{p}, D, S)$ such that

(i) Firms choosing prices $\mathbf{p}$ is a Nash equilibrium of the pricing game for any feasible mapping from prices to demands and supplies.

(ii) Prices, demand and supplies $(\mathbf{p}, D, S)$ clear the market.

The conditions for an equilibrium are quite demanding. It requires firms' price choices to be best responses to other firms' pricing decisions for any determination of pairwise demands and supplies consistent with the process outline in the previous section. We define equilibrium in this way just to emphasize that we are not using an intricate selection to create artificial incentives at odds with basic market forces.

**Definition 2.** An economy is *competitive* if there is an equilibrium in which all goods are priced at marginal cost.

### Planner's Problem

Consider a social planner choosing supplies $S$ to maximize consumer surplus subject to technology and resource constraints. Specifically, the planner's problem is to

$$\max_S U(x^F)$$

subject to

(i) Resource constraint $\sum_i \sum_k S_{ik} x_i \leq \omega$

(ii) Leontief production constraints and capacity constraints are satisfied (for each $x_\theta$ and all $i$ $\min_{\theta': A_{\theta\theta'}>0} x_{\theta'}/A_{\theta\theta'}$ and $\sum_{j\in N\cup\{c\}} D_{ji} \leq \phi_i$ for all $i$)

(iii) consumption of good $\theta$ is equal to amount of good $\theta$ supplied to the consumer ($x_\theta = \sum_{i\in Z(\theta)} S_{ic}$)

### Bottleneck firms

We now introduce an auxiliary flow problem which will allow us to identify the key notion in the theory, bottleneck firms. For each final good $\theta$ we add a consumer node we label $\theta$. We also add a source node and a sink node. The raw material producers are linked with large but finite capacities to the source node $s$. The final goods producers are linked with large but finite capacities to their respective consumer nodes.[16] The sink node $t$ is linked to by all consumer nodes $\theta$, and the capacities of these links is set equal to final good demand given marginal cost pricing ($w_{\theta t} = D_{c\theta}(\gamma^*)$). Combining the sinks in this way means that the flow can at most be equal to demand for the final good. It also means that the aggregate demand constraint is the only individual demand constraint that final goods producers face. We refer to this supply network including sinks, sources and consumer nodes as the augmented supply network $\bar{G}(G)$.

We let $f_{ki}$ denote the flow from $k$ to $i$ and $f_k = \sum_i f_{ki}$ denote the total flow out of $k$. The objective is to maximize the flow, which can be measured as the flow into the sink.[17]

$$\sum_k f_{kt}. \tag{1.1}$$

Given the Leontief structure of production, the conservation of flow constraints are min functions. For example, if producer $j \in Z(\hat{\theta})$ uses inputs of types $\theta$ and $\theta'$ only, then the flow out of $j$ is given by:

[16] These capacity constraints just need to be large enough to not matter. For example, it is sufficient to set each of these capacities equal to the sum of all other capacities in the graph prior to these links being added.

[17] We care about when production is supply-constrained. The key observation is that when production is not supply-constrained, it is demand constrained and the minimum cut of the network will be across the consumer to sink links (which reflect demands given marginal cost pricing).

$$f_j = \min\left\{\frac{1}{A_{\theta\hat{\theta}}}\sum_{i\in Z(\theta)} f_{ij}, \frac{1}{A_{\theta'\hat{\theta}}}\sum_{i\in Z(\theta')} f_{ij}\right\}.$$

These flow constraints can be represented by a series of simple inequalities. For example:

$$f_j \leq \frac{1}{A_{\theta\hat{\theta}}}\sum_{i\in Z(\theta)} f_{ij},$$

and

$$f_j \leq \frac{1}{A_{\theta'\hat{\theta}}}\sum_{i\in Z(\theta')} f_{ij}.$$

This helps us map our flow problem into a more standard set up, with linear constraints. So, our flow constraints for firms $i \in Z(\hat{\theta})$ are:

$$f_i \leq \frac{1}{A_{\hat{\theta}\theta}}\sum_{k\in Z(\theta)} f_{ki}, \quad \text{for all } \theta \text{ required to produce good } i. \tag{1.2}$$

We also have node capacity constraints.

$$f_i \leq \phi_i \tag{1.3}$$

Next, edge capacity and non-negativity constraints are also present, implying that

$$f_{ki} \leq w_{ki} \quad \text{for all } k, i. \tag{1.4}$$

$$f_{ki} \geq 0 \quad \text{for all } k, i. \tag{1.5}$$

Given an augmented supply network $\bar{G}$, the linear program is to maximize 1.1 subject to 1.2, 1.3, 1.4, and 1.5. We let $f(\bar{G})$ denote the maximum flow in the graph $\bar{G}$, and let $f(\bar{G} - i)$ denote the maximum flow in the graph after node $i$ is removed.

**Definition 3.** A firm $i$ is a *bottleneck* if and only if $f(\bar{G}) > f(\bar{G} - i)$.

Finally, it will be helpful to define the concept of a demand-constrained system.

**Definition 4.** A system is demand constrained if the maximum flow is equal to the capacities on the links from the consumer nodes to the sink (i.e., if $f(\bar{G}) = \sum_{\theta\in F} w_{\theta t}$).

We present a simple example of a supply-chain with and without a bottleneck firm which may help to clarify the definitions above.

Figure 1.2. Example of supply-chain with no bottleneck

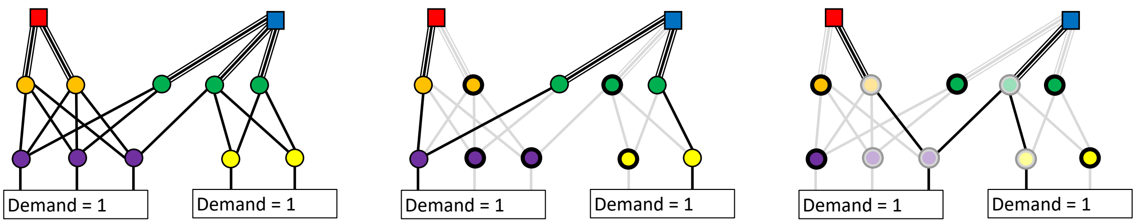
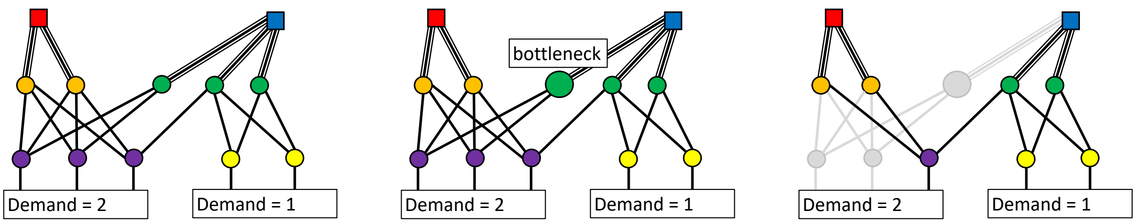



Figure 1.3. Example of supply-chain with one bottleneck




*Notes:* The figures show a hypothetical supply chain. Red and blue nodes are raw materials which provide large finite capacity to orange and green intermediary producers. Purple uses 1 unit of orange and one unit of green to make one unit of purple. Yellow uses one unit of green to make one unit of yellow. The capacity on each edge is 1.

**Example 2.** In Figure 1.2, we present a supply-chain with no bottleneck firms. Assume that the underlying technology is the same as presented in Example 1. Assume, for simplicity, that the capacity on each edge is one and that final demand for purple and yellow is both one. Recall that a firm is a bottleneck, if and only if, supply can meet demand with the firm removed from the network at current prices. In panel 2 of Figure 1.2, we observe that five firms can be dropped with supply still meeting demand. These firms can therefore not be bottlenecks, and so in panel 3 we fade them to grey. In panel 3, we then observe that all of the remaining firms can also be routed around. Therefore, if any firm was to deviate by pricing above marginal cost, their demand would fall to zero. Therefore, there is no bottleneck.

Now consider Figure 1.3, where an increase in demand for purple from 1 to 2 generates a bottleneck firm. If this bottleneck firm were removed (panel 3), two of the purple firms could no longer source green. This leaves a maximum purple production of 1 (less than consumer demand of 2). Therefore, the green firm is a bottleneck.[18] We will show that this firm can make a profitable deviation by pricing above marginal cost.

[18]A similar process can be used to show that none of the remaining firms are bottlenecks.

Market power and competitive outcomes

Recall that $D_{c\theta}(p)$ denotes the representative consumer's demand for good $\theta$ at price $p$ (by the assumptions we have made on the representative consumer's preferences, the demand correspondence is single valued and can be represented by such a function). We say an outcome $S$ is competitive if (i) all supplies are feasible and non-wasteful (i.e., for all $\theta'$ we have $\sum_{k\in Z(\theta')} S_{ki}/A_{\theta\theta'} = \sum_j S_{ij}$) and (ii) final good production equals the representative consumer's demand at marginal cost pricing (i.e., for all $\theta \in F$ we have $\sum_{j\in Z(\theta)} S_{jc} = D_{c\theta}(\gamma^*)$.

Competitive economies admit competitive outcomes. By definition there exists an equilibrium in which all firms price at marginal cost if an economy is competitive. In such an equilibrium all final goods are priced at marginal cost, and, for the market to clear, final goods production equals consumer demand. On the other hand, there need not exist a competitive outcome. Because of the capacity constraints on production, it may be infeasible to produce sufficient quantities of the final goods to satisfy consumer demand when goods are priced at marginal cost. We will tie the competitiveness of economies (and thus competitive outcomes) to the presence of the bottleneck firms later in this subsection, but first we show that (when they exist) competitive outcomes are an appropriate benchmark.

**Proposition 1.** *If a competitive outcome exists it solves the planner's problem.*

We relegate the proof to Appendix B.2

As competitive economies admit a competitive outcome, Proposition 1 shows that it is socially desirable for an economy to be competitive, raising the question of the conditions that are required on the supply network for an economy to be competitive. This leads us to our main theoretical result.

**Theorem 1.** *An economy is competitive if and only if no firm is a bottleneck.*

We relegate the proof of Theorem 1 to Appendix B.2, but do provide some intuition here. When no firm is a bottleneck, it can be shown that if all firms price at marginal cost the system will be demand constrained and it is possible to satisfy consumer demand at these prices. It is then possible to construct demands and supplies that clear the market given marginal cost pricing from a maximum flow. Further, if at marginal cost pricing a firm deviates by pricing above marginal cost, the market clearing will select demand and supplies that give the deviating firm zero demand, rendering the deviation unprofitable. This establishes that if no firm is a bottleneck then marginal cost pricing is an equilibrium. To see that if a firm is a bottleneck there does not exist a competitive equilibrium, suppose all firms priced at marginal cost but there was a bottleneck firm. If this firm increased its price it would still be possible to satisfy all induced demands and supplies, but only if this deviating firm continued to receive positive demand. Market clearing would select such demands and supplies, and hence the deviation would be profitable.

The following corollary provides an argument for tackling the sources of market power (i.e., bottleneck firms) rather than the symptoms (i.e., all firms pricing above marginal cost in equilibrium).

**Corollary 2.** *To restore the existence of an efficient equilibrium it is sufficient to intervene only in the markets containing bottleneck firms.*

### 1.2.2. Bottlenecks in uncompetitive equilibria

While Theorem 1 provides a nice benchmark, it is silent on where market power resides when there are bottleneck firms. To explore this further, it is helpful to introduce a new notion of a bottleneck firm. Suppose there is an equilibrium in which firms are setting prices $\mathbf{p}$ (a $\mathbf{p}$-*equilibrium*). These prices include the prices being charged to all consumers for all goods. Given this pricing schedule for the different final goods, consumer demand for each of the goods can be calculated. Let $\hat{D}_{c\theta}(\mathbf{p})$ denote the overall consumer demand for good $\theta$, given prices $\mathbf{p}$. We can then construct an *adjusted production network* by taking our production network and replacing each link from a producer $i \in Z(\theta)$ to a final consumer with a link of weight $\hat{D}_{c\theta}(\mathbf{p})$, and each link from a consumer of product $\theta$ to the sink node $t$, by a link of weight $\hat{D}_{c\theta}(\mathbf{p})$. Denote this adjusted production network by $\hat{G}(\mathbf{p})$.

**Definition 5.** A firm $i$ is a $\mathbf{p}$-*equilibrium bottleneck* if and only if $\mathbf{p}$ is an equilibrium price vector and $f(\hat{G}(\mathbf{p})) > f(\hat{G}(\mathbf{p}) - i)$.

Note that on this new network the connection between bottlenecks and equilibrium pricing is more subtle. While before all firms of the same product charged the same price, now firms of the same product price differently. By construction, the maximum flow calculation pays no heed to these price differentials. It seeks only to maximize the flow through the network given the capacity and technological constraints. Nevertheless, in any $\mathbf{p}$-equilibrium, a maximum flow must be achieved. We argue this more carefully in the proof of Proposition 2 below, but the intuition is simple. The maximum flow can never exceed consumer demand given prices $\mathbf{p}$ and were the maximum flow strictly below consumer demand at prices $\mathbf{p}$, there would be a firm facing excess demand from consumers and it would have a profitable deviation to increase its price.

Thus, in a $\mathbf{p}$-equilibrium we must still obtain a maximum flow through the (new) network. This connection is enough for us to show that bottleneck firms must always be pricing above marginal cost in equilibrium. However, non-bottleneck firms can sometimes also price above marginal cost in equilibrium. While we cannot rule out the possibility that many more firms than there are bottlenecks will end up pricing above marginal cost, there is a concrete sense in which these firms pricing above marginal cost is a *symptom* of the problem rather than the *cause*.

**Proposition 2.** *If* $\mathbf{p}$ *is an equilibrium price vector, and firm* $i \in Z(\theta)$ *is a* $\mathbf{p}$-equilibrium bottleneck*, then firm* $i$ *makes positive profits* ($p_i > \kappa_i + \sum_j p_j \frac{S_{ji}}{\sum_k S_{ik}}$)*.*

Proof in Appendix B.2.

**Proposition 3.** *If consumer demand for each product is independent of the prices of other goods (i.e.,* $D_{c\theta}(p)$ *depends only on* $(p_i)_{i\in Z(\theta)}$*),* $\mathbf{p}$ *is an equilibrium price vector, and firm* $i$ *is a* $\mathbf{p}$-equilibrium bottleneck*, then firm* $i$ *is also a bottleneck firm.*

Proof in Appendix B.2.

An immediate corollary of Propositions 2 and 3 is that if consumer demand for each product is independent of the prices of other goods, if there is $\mathbf{p}$-equilibrium bottleneck firm, then there does not exist a competitive equilibrium.

### Secondary distortions

Distortions created by $\mathbf{p}$-*equilibrium bottleneck* firms can propagate. For example, suppose firms $i$ and $j$ both produce product $\theta$, both firms have access to the same suppliers (i.e., $w_{kj} = w_{ki}$ for all firms $k$), and for simplicity, let firm $i$ be a $\mathbf{p}$-equilibrium bottleneck firm. We know by Proposition 2 that firm $i$ charges a price above its marginal cost. As firm $j$ has the same access to suppliers as firm $i$, it has the same marginal cost as firm $i$. Suppose that firm $j$ would be a $\mathbf{p}$-equilibrium bottleneck firm on the network with $i$ removed, then the same argument used to prove that firm $i$ can't be pricing at marginal cost in equilibrium also applies to firm $j$. Further, downstream from the $\mathbf{p}$-equilibrium bottleneck firm, there may be a firm that has to source from $i$, while a competitor firm is able to source from a cheaper supplier of the same input. Now that these firms face different marginal costs of production (including their sourcing costs), we should not in general expect them to both price at their (different) marginal costs either.

## 1.3. Scalable algorithm to find bottleneck firms

In this section, we develop an algorithm for identifying firm bottlenecks in data. We start by considering how to identify bottlenecks in the theoretical setting outlined in section 1.2. We then consider how to address the challenges of identifying bottlenecks in a world in which we observe real data on the universe of firm-to-firm transactions and final sales.

### 1.3.1. Identifying bottleneck firms in theory

Assume we observe an economy in equilibrium described by some price vector, $\mathbf{p}$. Assume we also observe the full production network, the capacities on each edge, the underlying technology DAG, a source and a sink, and final demand, $D$.

Under these conditions we can exploit the auxiliary maximum flow problem as discussed in Section 1.2.1 to obtain a measure of the maximum network flow from raw materials to final consumers given three constraints: (i) firms must use inputs as defined in the technology DAG, (ii) firms must use the existing production network and cannot form new links, (iii) firms cannot exceed the link capacities.

In order to identify the maximum flow of goods through the production network, we utilize the Ford-Fulkerson method (Ford and Fulkerson, 1956) as implemented by the Edmonds-Karp (Edmonds and Karp, 1972) algorithm. Intuitively, the algorithm begins by assuming no flow. It then adds more flow on any path for which there is available capacity on all edges in the path. These paths are called augmenting paths. When there are no more augmenting paths available, the algorithm terminates and reports the maximal flow, $f(G)$.

Recall from Definition 4, that if prices $\mathbf{p}$ are an equilibrium, a firm $i$ is a bottleneck if and only if supply cannot meet final demand at these prices with $i$ removed from the network.

To identify firm bottlenecks we, therefore, calculate the maximum flow dropping each firm $i$, $f(G-i)$. A firm is defined as a bottleneck if $f(G-i) < D$ i.e. the supply cannot meet demand, at current prices, when removing firm $i$ from the network.

### 1.3.2. Identifying bottleneck firms in practice

In order to implement this algorithm on data we must overcome several hurdles.

A necessary starting point is to observe the universe of firm-to-firm transactions and final sales. This type of data is increasingly becoming available to researchers through the availability of government tax datasets and from financial data through peer-to-peer lending platforms and credit-card companies (Elliott et al., 2019).

Assuming this data is available, there are four main challenges to implementing the algorithm defined above. First, supply-chains are not typically represented by a DAG structure. Second, the dataset does not have a source and a sink node. Third, we don't observe the underlying technology used by each firm. Fourth, we do not observe the edge capacities. In this section, we address each of these problems in turn.

### Supply-chain is not a DAG

The model assumes that the supply-chain can be represented by a DAG. For instance, wood can be used in the production of pulp which can be used in the production of paper. However, paper cannot be used to produce wood.

In reality, this assumption will sometimes be violated, and so we must prune violating edges in order to eliminate directed cycles. However, we wish to do this by dropping the weighted minimum set of edges causing cycles. This is known as a feedback arc set problem. We adopt the algorithm developed by Eades et al. (1993) which is both fast and prunes fewer edges than other existing algorithms (Simpson et al., 2016). The algorithm has already been used in economics, by Tintelnot et al. (2018), who apply it to Belgium VAT data.[19].

Intuitively, the algorithm begins by calculating the in-degree minus the out-degree which is called $\delta$. Nodes are then classified into three buckets: sink ($out - degree = 0$), source ($in - degree = 0$) or $\delta$ (the remainder). Iteratively, the algorithm takes the vertex from the $\delta$ group with the most "source-like node" ($\delta$ is low) and adds it to the source group, pruning any violating edges. After each iteration, each bucket is updated with the new set and the process is repeated. The algorithm stops when the $\delta$ group is empty. As a by-product, the algorithm will provide a partial ordering of nodes from those which are most likely to be raw material producers to final retailers.

### What is a source and what is a sink?

The model assumes that final goods producers are connected to a sink node which represents final demand. We represent this in the data by connecting firms which have positive final goods sales to an artificial sink node using their maximum final goods sales.[20]

The model also assumes that raw materials producers are connected to a source node with a large but finite edge capacity. As previously discussed, the Eades et al. (1993) algorithm provides a partial ordering of firms which classifies firms into most "source-like" to most "sink-like". Indeed, it will always produce a non-empty set of firms with no upstream linkages.[21] We, therefore, connect this set to a new artificial source with large but finite edge capacity.[22] This is equivalent to assuming firms pay a constant marginal

[19]They find that at most 18% of edges in the Belgium network data violate acyclicity

[20]This makes an implicit assumption that firms cannot sell more than the previous maximum they have ever sold, even if supply was to increase. You can think of this assumption as a short-run friction meaning firms cannot quickly scale up demand.

[21]As if a network has no firm with zero input suppliers then there must be a cycle and so it cannot be a DAG

[22]An alternative strategy would be to connect only the bottom $x$% of firms in the partial ordering to the source node, including firms which have input suppliers. This would be problematic, as any input supplier to a firm with a source node connection could not be a bottleneck as their input could be

cost to obtain more raw materials. What constrains the size of these firms is downstream demand.

### No information on technology

The maximum flow algorithm is more commonly used to measure flow through a system of pipes. An important distinction between the flow of goods through a production network and the flow of fluid through a system of pipes is that goods are transformed at each node in the network. It is, therefore, necessary to transform the edge-weights from a unit of currency into a unit which is comparable throughout the production network. In particular, the flow should be conserved at each node across all observations.

We build an algorithm which converts each transaction into a unit of the final consumption good. The algorithm is described in detail in Appendix A.2, but below we summarise the main features.

Starting with the firm most upstream,[23] the algorithm consists of three main steps which are conducted at the firm level: (i) identify and group all inputs of the same type, (ii) estimate a firm-level production function, (iii) attribute a value to the edges in terms of units of the final good.

### Step i: Identify and group all inputs of the same type

In order to estimate a firm-level production function, we first need to know which inputs for each firm are the same type. This is to maintain a tight connection with the model, where we assume that all firms producing the same good use the same production technology. Detailed product codes are not available in most transaction-level datasets. Instead, we develop a new Hierarchical Firm Clustering Algorithm (HAC) which exploits the Leontief technology assumption to find firm-specific sets of suppliers which provide inputs in the same constant proportion. The HAC algorithm is described in detail in Appendix A.1, but the basic intuition can be described with an example as shown in Figure 1.4. In this example, to produce 1 unit of cement (red) you require 1 unit of limestone (green) and 1.2 units of gypsum (yellow). The Leontief assumption ensures that limestone can not be substituted for gypsum.

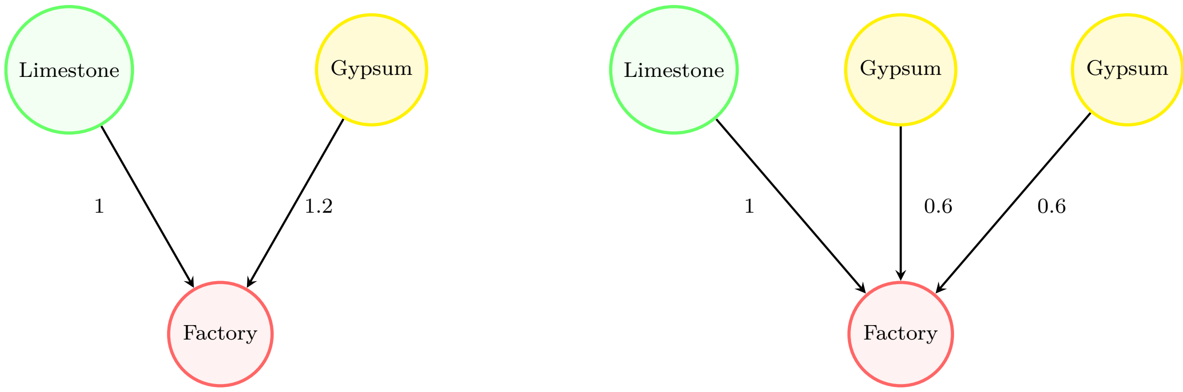
Suppose, in period 1, we observe a cement manufacturer using inputs from two firms

---

substituted for through greater flow from the source node.

[23] It is important to begin with the firm at the top of the partial ordering (i.e. the firm which is considered by the Eades et al. (1993) algorithm to be the furthest downstream). This is because all edges must be converted into a unit of the final consumption good, so it is necessary that the most downstream firm has its inputs in units of the final good, before a firm higher up the production chain has its inputs transformed. A worked example is given in appendix A.2.

Figure 1.4. Cement inputs example




*Notes:* The left figure shows a cement factories inputs in period 1, the right figure shows input in period 2. Each node represents a firm and colours represent different inputs. The amount sourced is given on the edge.

in the required ratio, 1 and 1.2. Then, in period 2, we observe the firm using inputs from three firms with ratio 1, 0.6 and 0.6. Observe, that if inputs from firms 2 and 3 are combined then we return to the same input ratio, 1 and 1.2. Therefore, we can infer from the firms repeated sourcing patterns that firms 2 and 3 are likely to be producing the same input.

We can use this simple example to build a generalisable algorithm to identify which input goods are likely to be the same.

The algorithm begins by assuming that all inputs are the same (and therefore should be used in the same proportion). We define a quadratic loss function which has a higher value for inputs which violate the Leontief constant proportions assumption (over time). We then consider all possible single partitions of the inputs into two groups and calculate a new loss function and the proportionate loss from making the new partition. If one of these partitions exceeds a threshold level, we then separate the goods into two groups. We continue making partitions until either all goods are considered different, or making an additional partition will lead to a sufficiently small change in the loss function that the goods are effectively used in constant proportions over time.

As a robustness check, we later supplement the HAC algorithm by using information from the firm's ISIC 4-digit industry as an alternative measure of whether the goods are the same.

### Step ii: Estimate a firm-level production function

Having grouped inputs into goods of the same type, we can now use the panel nature of the dataset to estimate a firm-specific production function. This allows us to infer how each input good is transformed into the firm's output good.

We regress output for firm $i$ on the sum of inputs in each partition $p$, which yields a set of transformation coefficients $\tau_{ip}$.

$$TotalSales_{it} = \sum_{p} \tau_{ip} Input_{ipt} + u_{it} \tag{1.6}$$

Within the regression, we force the transformation coefficients $\tau_{ip}$ to be greater than or equal to 1. This condition is necessary as otherwise firms would be losing value on inputs and so the assumption of conservation of flow would be violated.

#### Step iii: Attribute a value to the edges in terms of units of the final good

The final step is to multiply each of the input edges by the corresponding $\tau_{ip}$. This transforms the units of the input from a currency value to a unit of the final consumption good. We then begin the algorithm again with the next firm in the partial ordering, until all firms have their inputs transformed, such that, we have a dataset of firm-to-firm transactions in real units of the final consumption good.

### No information on edge/node capacities

In the model, we assume that each edge is endowed with a maximum capacity reflecting a notion that in the short-run there are substantial adjustment costs which restrict firms from increasing bi-lateral sales beyond a fixed value. In the data, we observe the realization of a series of transactions and not the underlying capacity. We, therefore, adopt two strategies to approximate these capacities.

In the first strategy, we set each edge capacity to the maximum observed bi-lateral transaction over a given time period. This is equivalent to saying that a firm could not increase capacity on a given edge beyond the maximum previously observed. They can, however, use all of their edges at their previously observed maximum, in any period. We call this strategy *edge-capacity.*

In the second strategy, we set each out-degree edge to have the observed output of a firm in that particular time period. By contrast to the first strategy, this is equivalent to saying that, in each period, a firm could direct all sales through one of its customer firms. We call this strategy *node-capacity.*

Both strategies introduce some slack into the system but in different ways. The edge-capacity strategy introduces slack by using the time dimension, i.e. allowing firms to utilize each edge at the previous maximum, in any period, but not allowing firms to substitute across customers. For instance, a cement manufacturer could sell cement to different retailers at the previously observed maximum, but could not substitute cement sales

Table 1.1. Number of observations

| **Variable** | **Value** |
|---|---|
| Number of Firms | 37,000 |
| Number of Edges | 89,000 |
| Average Degree | 2.9 |
| Transactions | 12m |

*Notes:* Numbers are calculated for all periods before edges have been trimmed by the min arc feedback set algorithm.

intended for one retailer to another beyond the maximum edge-capacity. This captures the idea that retailers in different parts of the country could not absorb the additional supply intended for a different region, at least in the short term.

The node-capacity strategy introduces slack by allowing firms to substitute their sales through different customers but keeps output fixed to what is observed in each period. This captures a notion that firms may have short-run capacity constraints which limit their production in a given period, but that they have flexibility to sell their goods through their existing customers.

We will run the model using both strategies to observe any differences in the set of firms identified as bottlenecks.

Having addressed these four empirical issues, the dataset is now ready to run the algorithm defined in Section 1.3.1

## 1.4. Proof of concept

In this section, we deploy the algorithm outlined in Section 1.3 on supply-chain transaction data from Ugandan VAT declarations. We first outline the data, then provide a characterization of the firms we identify to be bottlenecks before considering the consequences of these bottlenecks on economic outcomes.

### 1.4.1. Data

The data we use comes from the Ugandan Revenue Authority Tax Administration datasets covering the period 2010-2015. We use four main data series which all contain unique firm tax identifiers which allow records to be linked across firms and time.

The first and most important dataset contains firm transaction records recorded for VAT purposes. VAT-registered firms submit a monthly value added tax return form which includes the universe of their transactions over the previous month. Importantly, this dataset provides details on the transaction value and the tax identifier of the firm on the other side of the transaction. This allows us to observe a dynamic production network for the universe of tax paying firms.

The second dataset contains information from Corporate Income Tax which provides details on income tax for all non-individuals. We use variables on firm profits, employment, and balance sheet information.

The third dataset provides details from the business's registration. From this dataset, we obtain information on the firm's 4-digit ISIC sector.[24]

Finally, we obtained monthly CPI data from the Ugandan Central Bank.

For the baseline scenario, we collapse the data into 6 month periods, deflated using the CPI index. We choose a 6-month window as it is long enough to capture most frequent input trades, but short enough to allow us to observe firms over multiple periods.

A summary of the number of firms and observations is given in Table 1.1. In total we observe 37,000 firms with 89,000 firm-to-firm connections. A visualisation of the network is given in Figure 1.5.

### 1.4.2. Characterization of bottlenecks

#### Deploying the algorithm

Equipped with the data, we now implement the algorithm defined in Section 1.3.

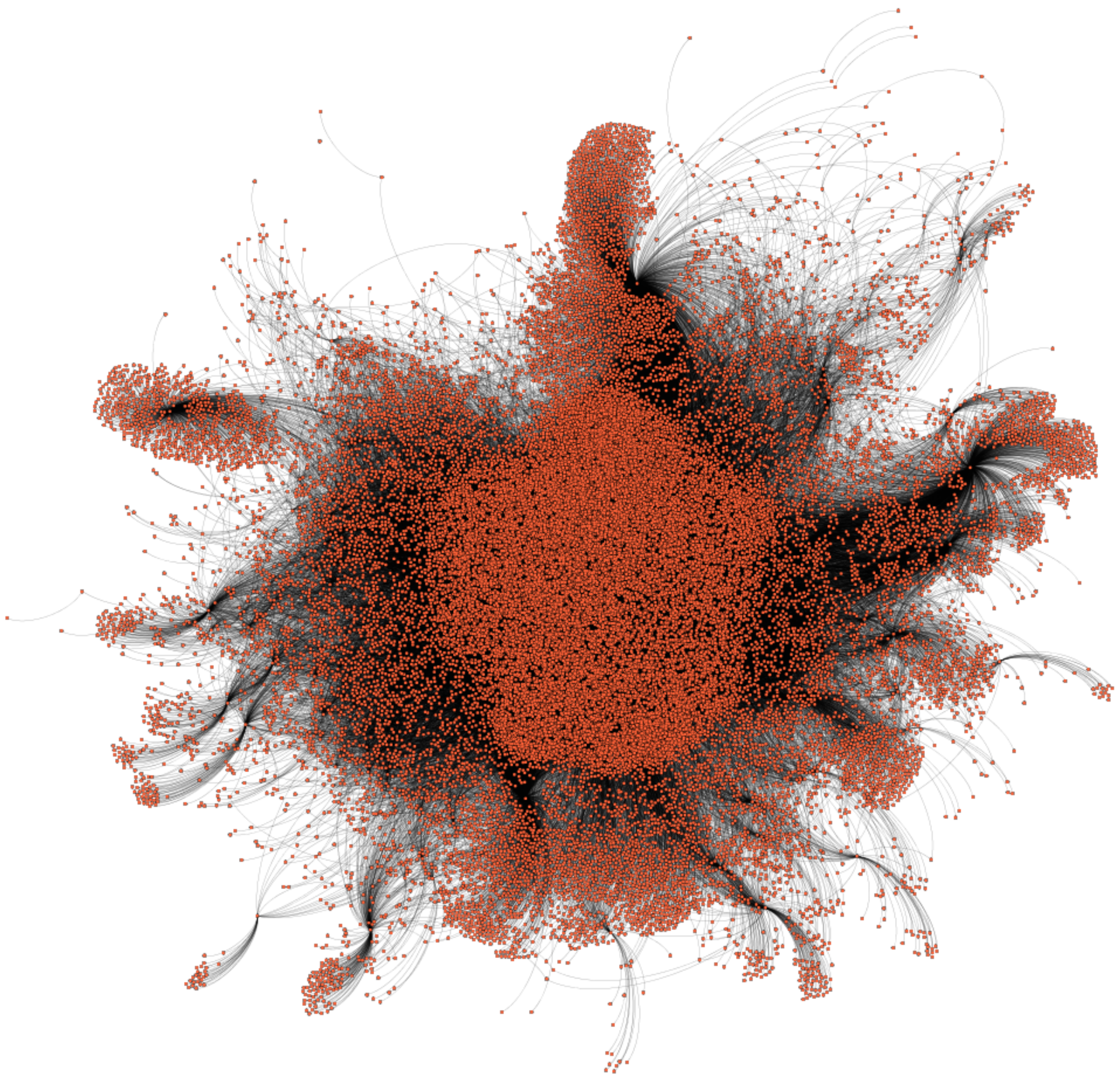
As shown in Figure 1.5, the production network is clearly not structured into a DAG. Instead, we can see multiple cycles, no layers from raw materials to final goods, and no sink nor source nodes.

We, therefore, implement the Eades et al. (1993) algorithm to prune edges. The algorithm cuts a relatively large number of links (14.7% of total links), however a small proportion of the total value (3.7%). This is consistent with the possibility that firms do use some inputs in their production from downstream firms, but that this is a small value relative to the total value of trades which flow downstream. It is also consistent with results found when the algorithm was run on Belgium VAT data. Tintelnot et al. (2018) find that at most 18% of edges in the Belgium network data violate acyclicity.

We then identify the Leontief production technology in order to convert transactions in the network into units of the final consumption good. We initially undertake this activity using the HAC algorithm described in Appendix A.1. The implementation of the HAC algorithm corresponds relatively closely to the ISIC sectoral categories: we find 28% of HAC clusters are in the same ISIC 4-digit sectors and 43% of HAC clusters are in the same ISIC 2-digit clusters. Additionally, as a robustness test we use ISIC sector

[24]Standard industrial classification of economic activities (ISIC) is a classification system for industry categories. The URA classifies firms at a 4 digit level. An example of the level of disaggregation would be "Mining of metal ores".

Figure 1.5. Universe of firm-to-firm transactions in Uganda, 2010-2015



*Notes:* Each node on the graph represents a firm, an edge represents that a trade took place between the two firms at some point between 2010 and 2015. The location of nodes is determined by forceatlas2 which is run on software Gephi.

classifications instead of the HAC algorithm in the identification of bottlenecks. Results from this exercise are presented in Section 1.4.4.

We next assign each edge a capacity using the maximum observed trade over all six month periods. This, on average, yields an edge capacity equal to 3.7 times the average trade.

We then connect firms to the sink using their final goods sales. This means connecting 37% of firms to the artificial sink node. We connect all firms with no inputs to the artificial source. This constitutes 15% of firms in the dataset.

Finally, we run the Max-Flow Ford-Fulkerson algorithm for the full DAG to obtain max-flow $f(G)$, and then 37,000 more times (one for each dropped firm $i$) to obtain each firm's contribution to max-flow $f(G-i)$.

### What do Ugandan Bottlenecks look like?

Figure 1.6 shows each of the 37,000 firms once edges have been trimmed to remove cycles and an artificial sink and source node has been added. Nodes towards the top of the graph are the most upstream firms such as raw material producers or primary processors. Nodes in the middle of the graph are secondary and tertiary producers. Nodes at the bottom of the graph are final goods producers and retailers, selling primarily directly to the consumer.

In Figure 1.7, we highlight firms for which their removal from the network, leads to a drop in the maximum flow, $f(G-i) < f(G)$. We refer to these as potential bottlenecks given the theory predicts that this is a necessary but not sufficient condition for firms to have hold up power.[25] Notice that these firms are spread at all layers of the supply-chain indicating that just looking for bottlenecks in particular subsectors will not be sufficient to identify potential bottlenecks.

In Figure 1.8, we highlight the set of firms which are identified at least once at some point over the whole time period as being a bottleneck firm (i.e. $f(G-i) < D_t$ at some $t$). In Figure 1.9, we show the set of bottleneck firms which are identified in one particular period.[26]

We identify a small number of persistent bottlenecks in Uganda. On average, we find 50 bottlenecks every semester and the probability of firm $i$ being a bottleneck at time $t$, given $i$ was a bottleneck at time $t-1$, is 0.77. This suggests there is a small set of entrenched firms which have substantially greater market power.

Given we are using confidential data, we are restricted in not revealing which firms are bottlenecks. However, we can provide some general information about which sectors they belong to.

Bottleneck firms are organised into three main sub-groups. Firstly, we observe bottlenecks in light manufacturing industries especially in the agricultural, food and drinks supply-chains. Interestingly, we identify bottlenecks at all stages of the supply-chain from primary production, to manufacturing to wholesale. These are sectors which would not normally be associated with substantial market power. However, given the context of Uganda this makes intuitive sense. Many of these sectors are structured around one or

[25] If $f(G-i) > D$ then these firms can still be routed around as supply can still meet demand at current pricies, even if $f(G-i) < f(G)$.

[26] We use the the first half of 2014, but it could be any period.

Figure 1.6. All firms

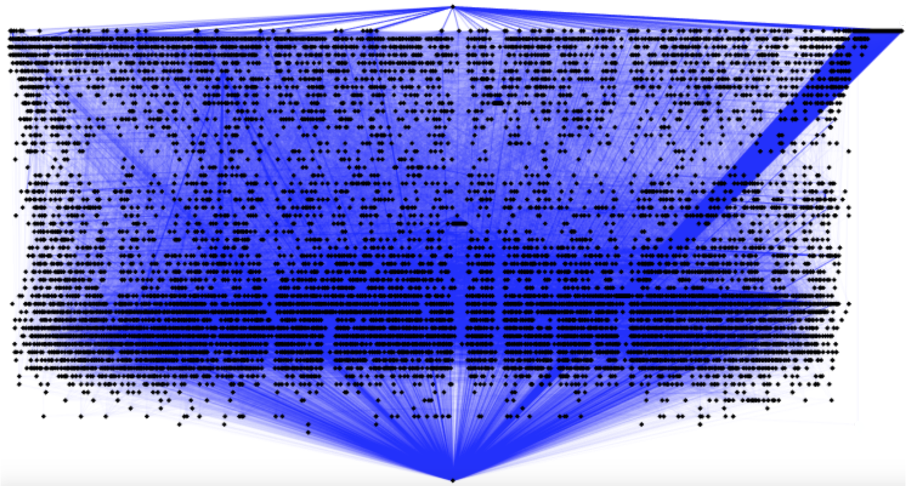


Figure 1.7. Potential bottlenecks highlighted

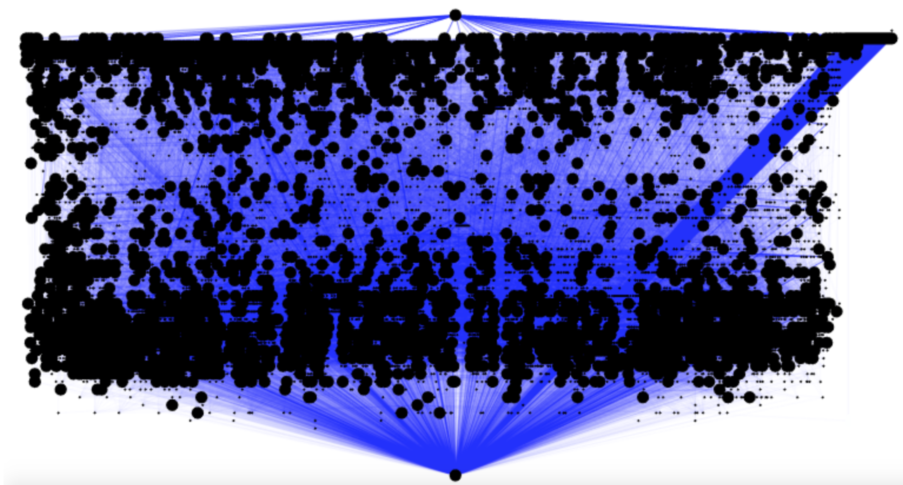


Figure 1.8. Bottleneck in any period highlighted

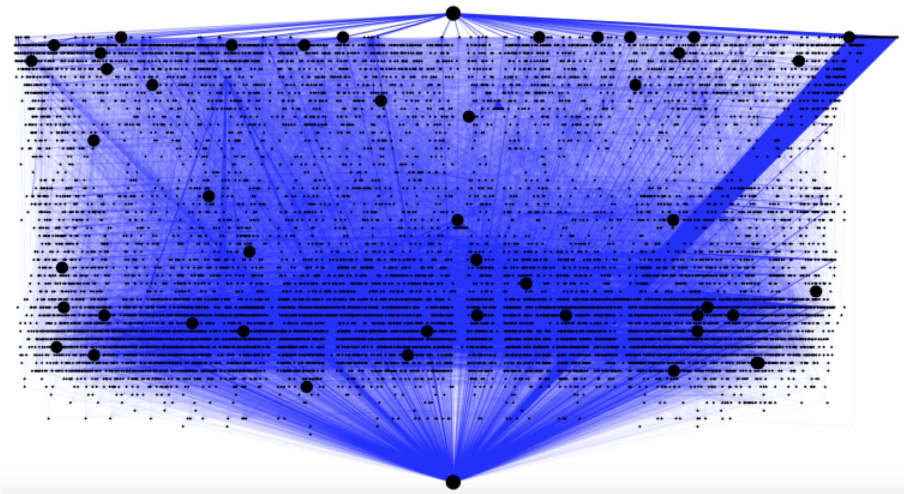
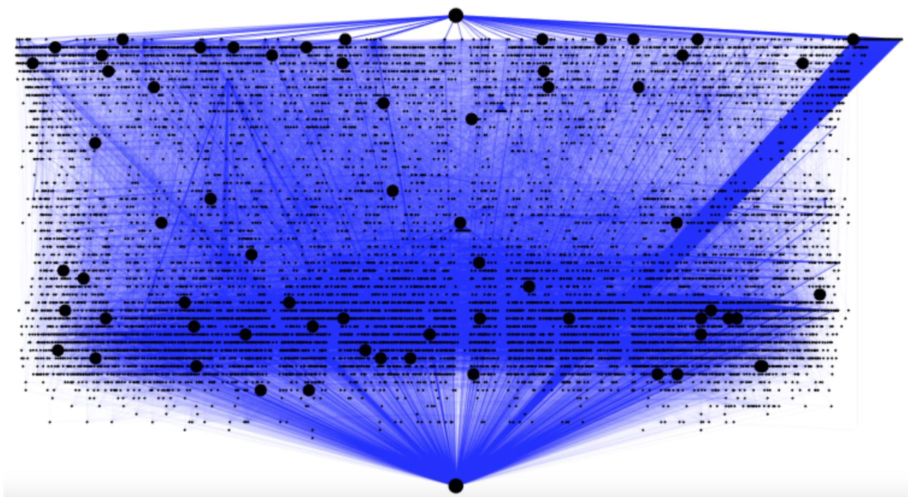


Figure 1.9. Bottleneck in only one period highlighted



*Notes:* Each node on the graph represents a firm, an edge represents that a trade took place between the two firms at some point between 2010 and 2015 once edges have been trimmed according to the Eades et al. (1993) algorithm. An artificial source node is added to any node which has no inputs. An artificial source is attached at the bottom to any firm by its final goods sales. The location of nodes is determined by their layer which is run on software Matlab. A potential bottleneck is any firm for which $f(G-i) < f(G)$, a bottleneck in any period is a firm for which $f(G-i) < D$ in at least one period, a bottleneck in only one period is a firm for which $f(G-i) < D$ in the first half of 2014.

two very large firms. This is because the market size is not big enough to support multiple firms with a large fixed cost of production (Agarwal and Spray, 2016).

The second major subgroup of bottlenecks is more traditional natural monopolies in the utilities sectors. This is what one might expect, given the algorithm identifies the set of firms for which their removal from the network would lead to the biggest fall in output. Firms in the utilities sector are essential inputs into a large number of sectors and so their removal would have a significant negative impact.

The final subgroup of bottlenecks is services inputs. In particular, we identify firms in the financial services sectors. As in the utilities group, these firms provide inputs into multiple sectors. This highlights the value of looking for market power across the entire economy and not on a supply-chain by supply-chain basis. As the former approach might miss firms which operate across many interconnected supply-chains.

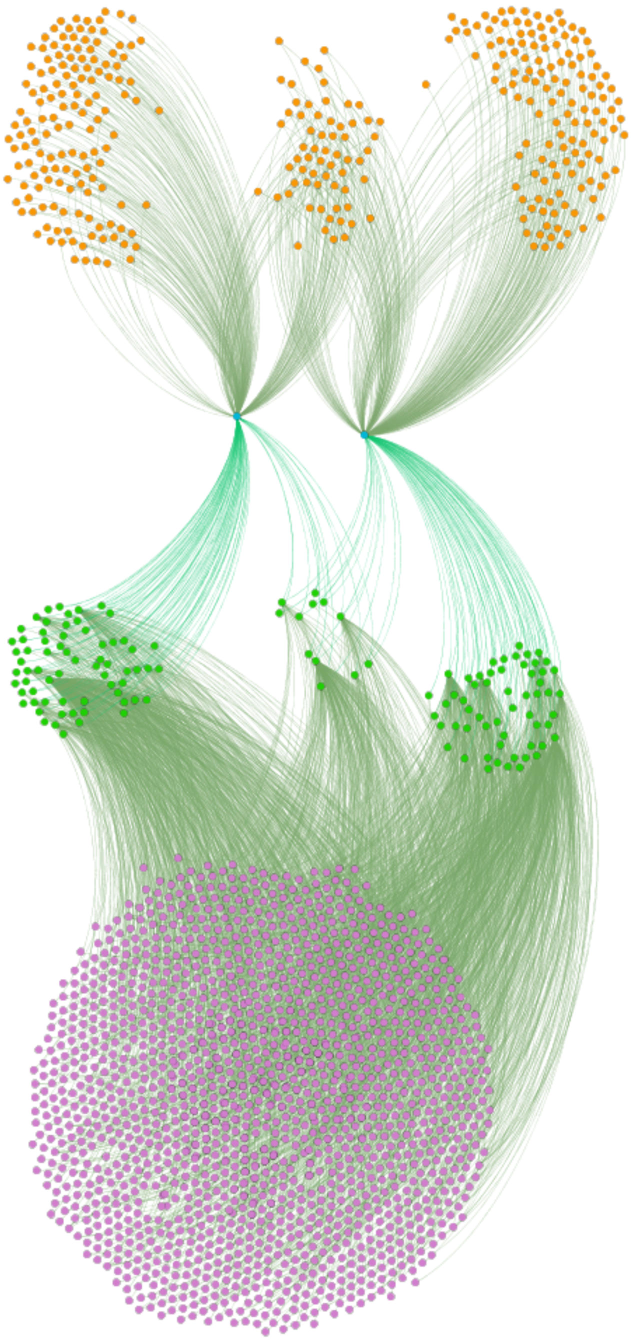
In Figure 1.10, we investigate one particular supply-chain. We cannot reveal the sector for data confidentiality reasons, but can reveal it is in light manufacturing. Both bottlenecks in this sector are highlighted in blue and both happen to be factories. These two firms purchase inputs from 340 different suppliers (orange) of which they share 60 of these suppliers. The two bottleneck firms sell onto 135 different customer firms (green) who in turn sell onto over 1500 additional firms (purple). The important take away is that these two factories anchor an industry of over 2000 firms. If either of the manufacturers can't produce, the other manufacturer cannot take up the slack in the short-run; the network maximum flow is necessarily diminished and output drops. Therefore, the manufacturers (blue nodes) are both bottlenecks.

Whilst this supply-chain is relatively simple, analyzed in isolation, supply chains are typically much more complex and interact with each other in important ways. Even when firms multisource, overlap in suppliers can mean there is less spare capacity than there might seem. This highlights why our methodology which is scalable to an economy is so important.

In Figures 1.11 to 1.16, we show how bottleneck firms differ in terms of observable characteristics which relate to market power. We find that relative to non-bottlenecks, these firms are substantially larger in terms of their sales, wage bills and profit. We also find that they are older and are more central to the production network. This is consistent with bottlenecks being established large firms which may have existed for a long time with potentially entrenched market power.

Figures 1.15 to 1.18 show more direct evidence of market power. Figures 1.15 shows bottleneck firms can be seen to have higher mark-ups as calculated based on sales over costs. As shown in Figure 1.17, bottleneck firms are located in less competitive industries as calculated by the Herfindahl-Hirshman Index. However, this is not significant at the 5% level. This is an initial indication that we are observing more information than simply

Figure 1.10. Example of a bottleneck supply-chain



*Notes:* Each node on the graph represents a firm, an edge represents that a trade took place between the two firms at some point between 2010 and 2015 once edges have been trimmed according to the Eades et al. (1993) algorithm. The supply-chain has been trimmed to remove connections to other sectors. The two blue firms are identified as bottlenecks. 135 suppliers (orange), 135 customers (green), 1548 customers of customers (purple).

Figure 1.11. Sales

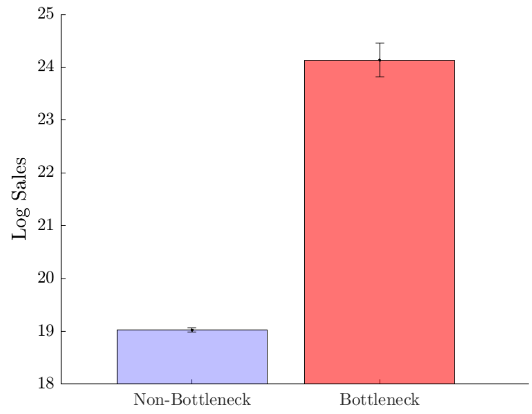



Figure 1.12. Wage bill

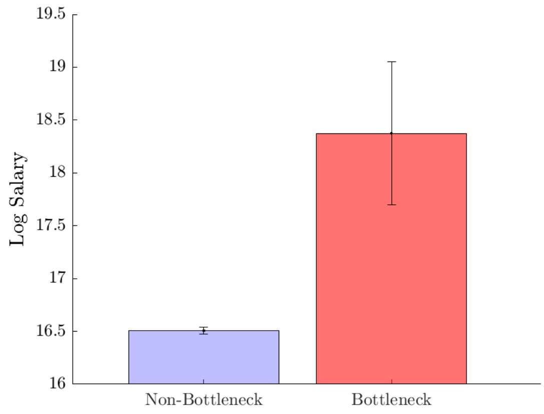



Figure 1.13. Age

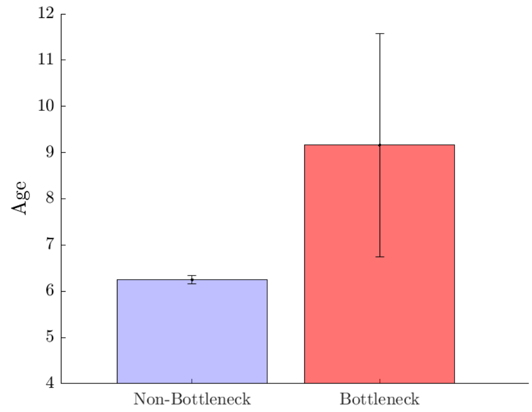



Figure 1.14. Gross profit

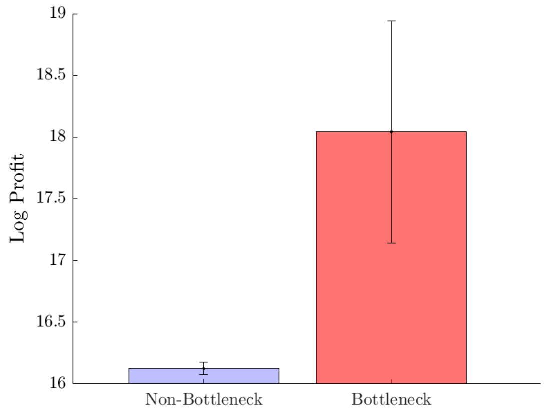



Figure 1.15. Markup

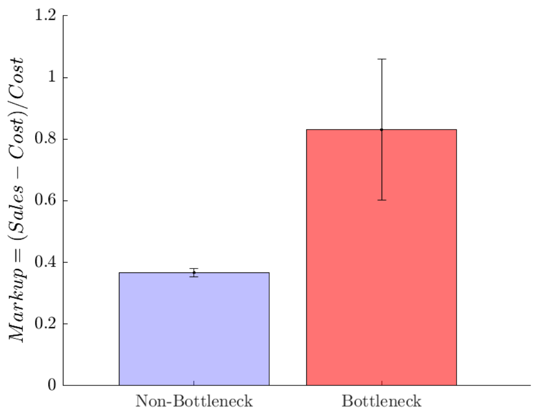
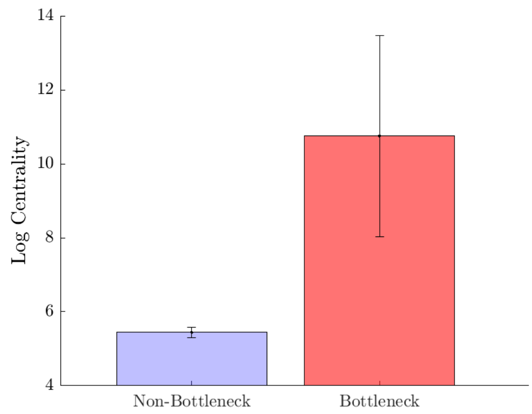



Figure 1.16. Betweenness centrality




*Notes:* Mean and 95% confidence interval of variable for bottlenecks and non-bottlenecks. Sales is the log of annual revenue as reported into CIT. Wage bill is the log of the total annual wage bill. Age is the number of years since the firm first registered for a tax identification number. Gross profit is the log of profit before tax. Markup costs are reported in CIT receipts. Centrality is the Bonacich Centrality.

Figure 1.17. Herfindahl-Hirschman Index

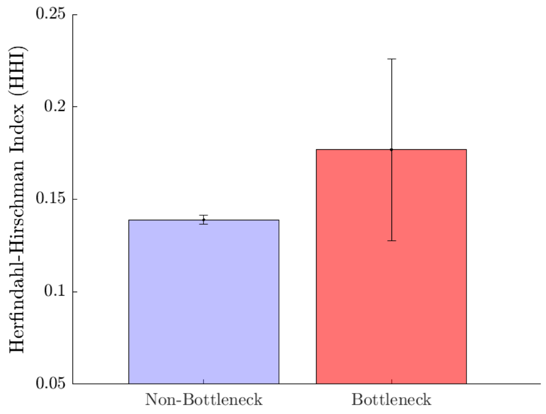
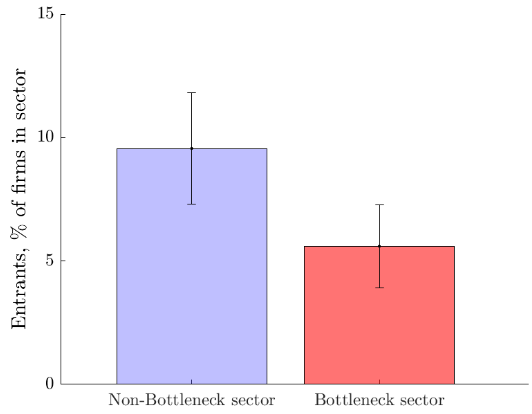



Figure 1.18. New entrants




*Notes:* Mean and 95% confidence interval of variable by ISIC 4-digit sector for sectors containing a bottleneck firm and sectors not containing a bottleneck firm.

correlating with existing measures of market power. Finally, Figure 1.18 shows that bottleneck firms are located in sectors with fewer new entrants, significant at a 5% level. This is consistent with bottleneck firms having market power due to higher barriers to entry.

### 1.4.3. Consequences of bottlenecks

Having characterized firm bottlenecks in Uganda, we now turn to the question of what the presence of these firms means for the economy, through first considering whether bottleneck status over time correlates with firm-level variables, and then by considering if there is evidence of propagation of distortions to other non-bottleneck firms.

#### Does bottleneck status influence firm-level variables?

In order to test whether a firm's status as a bottleneck is correlated with firm-level variables, we must transform our measure of market power to be calculated on a rolling basis. We calculate $MarginalMaxFlow(MMF)_t = f(G)_t - f(G-i)_t$ using the *edge-capacity* approach described before, which calculates maximum observed trade, between two firms, over a rolling three periods. We define a firm as $Bottleneck_{it}$ if, at time $t$, supply can no longer meet demand if the firm was to be removed from the network, $f(G-i) < D$.

We then estimate the following specification via OLS,

$$Y_{it} = \beta_1 MMF_{it} + \beta_2 MMF_{it} \times Bottleneck_{it} + \alpha_t + \delta_i + u_{it} \tag{1.7}$$

where we consider a vector of outcome variables $Y = \{markup, sales, profits\}$, $\alpha_t$ is a set of time dummies and $\delta_i$ is a firm fixed effect.

Table 1.2. Bottleneck consequences

| | (1) | (2) | (3) |
|---|---|---|---|
| | log sales | log profit | markup = (sales-cost)/cost |
| log marginal max flow | 0.0167*** | 0.00561*** | -0.000752 |
| | (0.00134) | (0.00205) | (0.00180) |
| log marginal max flow * bottleneck | 0.0279*** | 0.0143** | 0.0123** |
| | (0.00660) | (0.00659) | (0.00557) |
| Observations | 25517 | 14604 | 20411 |

*Notes:* Unit of observation is buyer $i$ and year $t$. Log Marginal Max Flow = $f(G) - f(G-i)$, Bottleneck = 1 if $f(G-i) < D$. Standard errors in parentheses clustered at the buyer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The regression allows us to test whether, within a firm, an increment on our metric of market power corresponds with a change in relevant outcome variables. Strictly speaking, the theory predicts that a firm will only have hold-up power once it breaches the bottleneck threshold. We, therefore, expect that $\beta_1 = 0$ and $\beta_2 > 0$. In practice, moving from the theory to the data, we might expect $\beta_1 > 0$, given there could be measurement error in the threshold for bottleneck, or there may be a friction outside of the model which causes rerouting supply-chains to be costly.

Results are given in Table 1.2. Row 1 of columns 1 and 2 show that an increase in marginal max flow by 1% leads to a 1.7% and 0.5% increase in sales and profits, respectively. This is significant at the 1% level. By contrast, it has no significant impact on firms' markups. This is consistent with the model's predictions, as firms cannot exploit their network position to make higher mark-ups if they are below the bottleneck threshold, but do receive higher demand.

Row 2 of Table 1.2 demonstrates the impact of breaching the bottleneck threshold. We can see that this roughly doubles the impact of an increase in marginal max flow on firm sales and profits, significant at the 1% and 5% significance levels. It also shows a statistically significant impact on firms' mark-ups.[27] In effect, once a firm breaches the bottleneck threshold, it is able to exploit its network position to earn higher profits through higher mark-ups.

[27]For the purposes of considering magnitude: a one standard deviation increase in MMF leads to 13% increase in sales for non-bottleneck firms, 36% for bottleneck firms; a one standard deviation increase in MMF leads to 13% increase in profits, 26% for bottleneck firms; a one standard deviation increase in MMF leads to -0.6% increase in markup, 9.2% for bottleneck firms.

### Do distortions propagate?

We now consider whether there is empirical evidence that the existence of bottleneck firms leads to other distortions elsewhere in the network, as predicted in Section 1.2.2. We, therefore, begin by defining a variable called $NoBottlenecksUpstream_{it}$ which is a dummy variable equal to 1 if at time $t$ there are no bottlenecks upstream of firm $i$, conditioning on $i$ not being a bottleneck itself. A description of how we identify whether there are any bottleneck firms upstream is given in Appendix A.3.

In order to identify if there are indirect effects along the supply chain from having bottleneck firms, we estimate the following specification by OLS for firm $i$ in 4-digit ISIC industry $c$ at time $t$,

$$Y_{ict} = \delta_c + \alpha_t + \beta NoBottlenecksUpstream_{it} + u_{it} \tag{1.8}$$

where $Y_{it} = \{markup, sales, profits\}$ is a vector of outcome variables, $\delta_c$ is a industry fixed effect and $\alpha_t$ is a time fixed effect.

We expect that $\beta_1 > 0$, because firms which do not have a bottleneck upstream are able to either: (i) price at the same level as their competitors (who are upstream distorted) but have lower input costs and therefore are more profitable or (ii) price below their competitors (who are upstream distorted) and sell more. We include an industry fixed effect because we expect the mechanism to take place within the industry level, but wish to condition for industry characteristics.[28]

Results, given in Table 1.3, are consistent with bottleneck distortions propagating vertically through the network. Columns 1 and 2 show that having no bottleneck upstream corresponds to 7% higher sales and 11% higher profits, significant at the 1% and 10% significance levels, respectively. Having no bottleneck upstream also corresponds to having a 6% higher markup, although this is not statistically significant.

### 1.4.4. Robustness checks

There are three topics which deserve further consideration.

First, we consider whether our measure of market power has additional explanatory power, when compared to a more straight forward measure of a firm's network position. Second, we consider whether using the firms ISIC industry classification instead of the HAC algorithm yields qualitatively different results. Finally, we consider whether altering the assumption about edge-capacity influences results.

[28]We also run the regression with firm fixed-effects finding qualitatively similar effects. Results available upon request.

Table 1.3. Do distortions propagate?

| | (1) | (2) | (3) |
|---|---|---|---|
| | log sales | log profit | markup = (sales-cost)/cost |
| No Bottlenecks Upstream | 0.0700*** | 0.112* | 0.0594 |
| | (0.0258) | (0.0601) | (0.0370) |
| Observations | 69769 | 44655 | 54368 |

*Notes:* Unit of observation is buyer $i$, from sector $c$ and year $t$. No Bottlenecks Upstream is a dummy equal to 1 if at time t there are no bottlenecks upstream of firm. Standard errors in parentheses clustered at the buyer level. $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

### Centrality or Maximum Flow?

Extant research has shown that a firm's centrality in a network may confer market power and the possibility to extract higher rents (Goyal and Vega-Redondo, 2007), although to the best of our knowledge this has not previously been tested empirically. We, therefore, first test whether a firm's Bonacich Centrality is correlated over time with firm-level outcome variables. We then test whether our measure of network market power has explanatory power, over and above, the simple centrality measure. We estimate the following specification via OLS,

$$Y_{it} = \beta_1 MMF_{it} + \beta_2 MMF_{it} \times bottleneck_{it} + \gamma_1 Centrality_{it} + \alpha_t + \delta_i + u_{it} \tag{1.9}$$

where $Y_{it} = \{markup, sales, profits\}$ is a vector of outcome variables for firm $i$ at time $t$, and $Centrality_{it}$ is a firm-level Bonacich centrality.

Columns 1 and 2 from Table 1.4 shows that firm network centrality is positively and significantly correlated with sales and profits, as predicted by the theory. More difficult to explain, is that there is a negative correllation between firm centrality and markups. A possible explanation for this result, is that large firms adopt a revenue maximising strategy lowering their markup to gain revenue share and a more central position.

Columns 4-6 from Table 1.4 show the comparison between centrality and our new measure of market power. Including both measures demonstrates that both metrics have relevant explanatory power in predicting outcome variables.

Indeed, adding the bottleneck variable to the regression with centrality does not greatly impact either the magnitude nor the significance of the coefficients on centrality. Likewise, the coefficients on the bottleneck variables are very similar to those in Table 1.2 which do not include centrality.

Table 1.4. Centrality vs. MMF horserace

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | log sales | log profit | markup | log sales | log profit | markup |
| log centrality | 0.0185*** | 0.00852** | -0.0147*** | 0.0174*** | 0.00836** | -0.0146*** |
| | (0.00229) | (0.00385) | (0.00318) | (0.00228) | (0.00385) | (0.00318) |
| log marginal max flow | | | | 0.0163*** | 0.00550*** | -0.000404 |
| | | | | (0.00133) | (0.00205) | (0.00179) |
| log marginal max flow * bottleneck | | | | 0.0280*** | 0.0145** | 0.0121** |
| | | | | (0.00661) | (0.00658) | (0.00555) |
| Observations | 25517 | 14604 | 20411 | 25517 | 14604 | 20411 |

*Notes:* Unit of observation is buyer $i$, from sector $c$ and year $t$. No Bottlenecks Upstream is a dummy equal to 1 if at time t there are no bottlenecks upstream of firm. Standard errors in parentheses clustered at the buyer level. $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table 1.5. Robustness table

| | **Edge Capacity** | | **Node Capacity** | | **ISIC** | |
|---|---|---|---|---|---|---|
| | **Non-Bottleneck** | **Bottleneck** | **Non-Bottleneck** | **Bottleneck** | **Non-Bottleneck** | **Bottleneck** |
| **Profit** | 16.12 | 18.04 | 16.11 | 19.27 | 16.13 | 21.32 |
| **Salary** | 16.51 | 18.37 | 16.50 | 19.50 | 16.51 | 20.89 |
| **Age** | 6.25 | 9.16 | 6.23 | 11.15 | 6.23 | 13.89 |
| **HHI** | 0.14 | 0.18 | 0.14 | 0.16 | 0.14 | 0.27 |
| **Sales** | 19.03 | 24.14 | 19.03 | 24.44 | 19.03 | 25.38 |
| **Centrality** | 5.44 | 10.75 | 5.44 | 11.15 | 5.45 | 12.41 |
| **Markup = Profit/Sales** | 0.09 | 0.19 | 0.09 | 0.20 | 0.09 | 0.13 |
| **Markup = (Sales - Cost)/Cost** | 0.37 | 0.83 | 0.37 | 1.20 | 0.37 | 1.25 |
| **Entrants** | 9.56 | 5.60 | 9.48 | 2.21 | 9.58 | 2.19 |

*Notes: Table shows the average value of different descriptive statistics by bottleneck status for three different calculation strategies: edge capacity, node capacity and edge-capacity using ISIC classification instead of HAC classification.*

### Edge Capacity vs. Node Capacity

In our baseline model, we assign each edge a capacity equal to the maximum observed transaction along that edge. This allows slack over time, but does not allow reallocation beyond the maximum previously observed trade, between edges sharing the same source-node. We, therefore, re-run the model choosing to give each edge a capacity equal to the total one period production of the firm. The implicit assumption is that production could be routed through any of the customers. We call this *Node capacity*. This generates slack in a different way, in that firms can substitute across their customers, but cannot increase production beyond what is observed.

Changing to *node-capacity* does alter results. Unsurprisingly, there are fewer firms identified as bottlenecks (maximum of 8). However, the specific the specific sets of firms identified as bottlenecks using the two methods are similar. The node capacity bottlenecks are almost a subset of the edge capacity bottlenecks (7 of 8). Similarly, the correlation in $f(G-i)$ remains high at 0.56.

Table 1.5 shows descriptive statistics on bottlenecks and non-bottlenecks from the robustness exercises next to the baseline model. In all instances the node capacity results are are in line with the edge capacity results. The gap between bottlenecks and non-bottlenecks is larger in the node capacity model. This is because we identify a smaller set of firms which are bottlenecks given the more testing criteria.

### HAC vs. ISIC Sector classification

In this final section, we rerun the algorithm using each firm's self reported ISIC 4-digit sector classification instead of the HAC algorithm for clustering the same inputs. We do this through two strategies. First, we assume all suppliers with the same ISIC sector are considered in the same industry. This is likely to be over confident on grouping suppliers as an industry classification at 4-digits is still fairly coarse. For instance, one classification is "Mining of metal ores" which will include a wide array of different types of mining.

Changing the sectoral classification has some impact on measurement of $f(G-i)$, although we still find a strong correlation between the two methods (equal to 0.55). We identify fewer bottlenecks (19 compared to 50), of which 11 are bottlenecks in both classification methods.

The last two columns of Table 1.5 show results on observable outcomes between bottlenecks and non-bottlenecks in different robustness exercises. Regardless of methodology, we find that the difference between bottlenecks and non-bottleneck firms is consistent.

The final strategy is to define two firms to be in the same sector if the two buyers have suppliers from exactly the same set of industries. Again, we find a high correlation in $f(G-i)$ (equal to 0.54). Results for bottleneck descriptive statistics are qualitatively similar to those presented in Table 1.5 and are available upon request.

## 1.5. Conclusion

This paper has built a new framework for considering market power in a supply-chain, demonstrating that firms located in bottlenecks in a production network are able to price above marginal cost which leads to inefficiencies. Modelling firms' production as a flow problem allows us to take the model to the data, operationalizing tools developed in computer science to a production network. This represents a significant departure from existing techniques for identifying market power which focus either on the firm or sector. By contrast, we are able to observe, firm-by-firm, the economy-wide impact of each firms removal from the production network, and hence its hold up power.

The immediate application for the technology developed in this paper is as a diagnostic tool for competition authorities identifying market power using transaction data. We see this method as a first-cut, before further investigative measures can take place.

An additional application for this toolkit is the investigation of weak points in supply-chains. Our measure can be considered as a metric of systemic importance, in that we pick up the set of firms whose removal from the network would lead to the largest fall in output. This can be useful to policy makers in considering responses to major shocks such as recessions or pandemics where it is vital that certain supply-chains keep functioning.[29]

[29] See Carvalho et al. (2020) for a discussion of how this can be done in practice.

# Chapter 2

# Search Externalities in Firm-to-Firm Trade

Abstract I develop a model of firm-to-firm search and matching to show that the impact of falling trade costs on firm sourcing decisions and consumer welfare depends on the relative size of search externalities in domestic and international markets. These externalities can be positive if firms share information about potential matches, or negative if the market is congested. Using unique firm-to-firm transaction-level data from Uganda, I show empirical evidence consistent with positive externalities in international markets and negative externalities in domestic markets.

## 2.1. Introduction

In a developing country, finding and maintaining an efficient and reliable supplier can be a costly and a time consuming process (Allen, 2014; Macchiavello and Morjaria, 2015; Startz, 2016). One factor which can make this process more difficult, is if many other firms are simultaneously searching for a supplier (Arnosti et al., 2018). This *congestion externality* will occur when trading frictions mean supply cannot instantaneously meet demand from multiple buyers. This is plausibly a large concern in developing countries where contracting frictions cause high adjustment costs (Macchiavello and Morjaria, 2019), and a lack of access to credit can cause firm supply-constraints (Manova, 2012). One policy response is to open to international trade, giving firms access to a large pool of suppliers which are less inhibited by these trading frictions.[1]

In this chapter, I show that reducing international trade costs will lead to a greater number of matches in the international market, alongside an important and novel secondary benefit - the alleviation of the consequences of congestion in the domestic supplier market. I formalise this new mechanism for a domestic market consumer welfare gain from trade and consider its effects in Uganda. I show empirical evidence consistent with the Ugandan supplier market suffering from greater congestion than the international supplier market.

This analysis requires a unique combination of data on firm-to-firm domestic and international transactions. I use Ugandan administrative Value Added Tax (VAT) data that includes information on every transaction between domestic tax-paying firms. I also use the government's import customs dataset which includes details on both the buyer in Uganda and the foreign seller. The combination of these datasets amounts to a dynamic transaction-level firm-to-firm input-output matrix. Using the firm's unique ID, I link this dataset with other tax administration datasets: firm balance-sheet data, firm employment information and detailed firm geographic location. Together, this constitutes a dynamic picture of the entire Ugandan formal economy from 2010 to 2016. To the best of my knowledge, this is the first paper to combine this breadth of administrative firm-level transaction-data in a developing country.

I start by developing a simple model of optimal search in two markets with different search externalities. The model serves to highlight the key mechanism proposed in this chapter - after a trade cost reduction, firms increase search in international markets as these goods become relatively lower-cost to source. This is mitigated by two forces. First, as firms move into the import market, this increases import market-tightness, thus

[1] These firms are likely to be less inhibited given international exporters tend to be larger (Bernard and Jensen, 2004) and with better access to credit (Manova, 2012). Indeed this channel should exist in any two markets, where one is more congested due to firms being supply-constrained.

decreasing the probability of an import match. Second, as firms move out of the domestic-market, domestic market-tightness decreases, therefore increasing the probability of a domestic match. The scale of these congestion effects depends on the relative size of search externalities in each market. These parameters also determine the welfare consequences of a reduction in international trade costs. If there is a greater positive externality to search in international markets compared to domestic markets, then a reallocation of search towards international markets not only leads to more matches in the international market, but also alleviates congestion in the domestic market. This will lead to a greater number of overall matches which benefits consumers with taste-for-variety.[2]

Motivated by the simple model, I undertake two empirical exercises. In the first empirical exercise, I study the impact of a 25% reduction in international transport costs that Uganda implemented in 2010-2011.[3] I test the model's predictions on number and type of matches and show that: (i) there was a 80% increase in the number of new importing firms; (ii) the firms that began importing in 2011 simultaneously adjusted their supply-chain by dropping domestic suppliers; (iii) the suppliers that were dropped as a consequence of this readjustment re-matched primarily with firms which were not importers.

In the second empirical exercise, I look for evidence of search externalities in a reduced-form setting. In the case of Uganda and consistent with previous literature,[4] I show that firms located in the same building adopt sequentially the same import suppliers. I then show that this effect is substantially larger for firms located in the same building compared to firms located in a next-door building. This is consistent with information diffusing among firms about suppliers at a very local level. When looking at domestic suppliers, however, this effect is not significantly different from zero. By contrast, in the domestic market, a buyer adding a specific new supplier actually reduces the probability of buyers in a different region of the country matching with that supplier. This is consistent with geographically distant firms not benefiting from the information externality, but still subject to the congestion externality, and has not been tested in the literature to date.

The results are in line with qualitative evidence that I collected through structured interviews with firms in East Africa.[5] I interviewed 25 managers from firms in a variety

[2]An alternative way of thinking about the model is through a lens of trading frictions. In this sense, buyers may be aware of the existence of suppliers, however, there is no centralized market where buyers and suppliers meet and trade at a single price (Rogerson et al., 2005). In order to form a partnership they must undertake costly investments which involve externalities.

[3]The reforms are discussed in detail in Section 2.3.2. Given the reforms were exclusively conducted outside Uganda or on the border crossing, I assume they had no impact on domestic trade costs.

[4]See for instance Bisztray et al. (2018) and Kamal and Sundaram (2016)

[5]Interviews were conducted with firms in Kampala in Uganda and in Kigali in Rwanda which has a very similar structure of firm market.

of sectors[6] who reported that: (i) it is common for buyers to share information about international suppliers; (ii) international suppliers have the size or the ability to scale up to service multiple buyers; (iii) domestic suppliers are limited in their ability to service multiple buyers and this means that sometimes there is wasted search effort.[7]

This chapter relates to three main strands of the literature. This chapter contributes to the literature on firm-to-firm search. The literature has shown that the competitive equilibrium does not necessary result in the socially optimal level of search (Krolikowski and McCallum, 2017), that search frictions explain firm's export market decisions (Chaney, 2014), and that search influences predictions on gains from trade (Antras and Costinot, 2011). I build on work by Eaton, Jinkins, Tybout and Xu (2016), who write a dynamic quantitative model of optimal two-sided buyer-supplier search which is rich enough to take to the data.[8] This chapter's contribution is to separately model search in domestic and international markets and incorporating different matching technologies in either market, providing new predictions on a search channel for consumer welfare gains following a reduction in international trade costs and using novel data.

In addition, this chapter relates to the literature on the firm supply-chain impacts following a trade liberalization in the absence of search frictions. Arkolakis et al. (2012) show that gains from trade are higher in models with intermediate goods. Tintelnot et al. (2018) and Fieler et al. (2018) build quantitative models to show that the gains from trade depend on domestic firm-to-firm linkages and how firms are directly or indirectly connected to the international market. Antras et al. (2017) build a quantitative model of global sourcing.[9] I build on this literature by incorporating intermediate goods into a model of domestic and international sourcing whilst also including a search channel. Moreover, I consider not only firms' international sourcing decisions but also the interdependencies between this and domestic sourcing decisions.

Finally, the chapter contributes to the empirical literature on firm-to-firm search externalities. The closest paper to the reduced-form work is Bisztray et al. (2018), which has extremely detailed geographic data on firms in Hungary. The authors show that firms

---

[6]I interviewed firms in logistics, retail and wholesale, coffee and tea, hotels and tourism, agribusiness, service input sectors. A full list of firms is available on request.

[7]For instance, in one interview I undertook with a firm they stated that they had asked another firm in the same business park about their input supplier for imported packaging. In another interview a firm had stated that they had tried to find a domestic transport firm but that another similar firm had already taken the contract.

[8]Other important contributions on sourcing include Rauch (2001), Rauch and Trindade (2002), Rauch and Watson (2003). A parallel literature also exists on exporter search for buyer markets (See for instance Eaton et al. (2017), Allen (2014), Albornoz et al. (2012)).

[9]A connected literature considers the role of production networks in firm performance and the propagation of shocks (Lim (2017), Carvalho (2014), Carvalho (2014), Bernard and Moxnes (2018), Bernard et al. (2018)).

in the same building sequentially add imports from the same country and in the same product category. The chapter also relates to Kamal and Sundaram (2016), who show a similar effect for matched importer-exporter data but without detailed geographic data. Cai and Szeidl (2017) show that when firms are randomly allocated into different business groups they refer each other leading to a 9% increase in the number of suppliers.[10] I build on this literature in three ways. First, the Ugandan dataset contains details on both the geographic location of firms and the matched supplier which gives more detail than the existing literature. Second, I compare firms searching domestically to firms searching internationally, providing evidence of the comparative size of domestic and international externalities for the first time. Third, in addition to looking for a positive search externality, I also show results consistent with a negative search externality.

The remainder of this chapter is organized as follows: Section 2.2 sets out a simple two-period model of firm-to-firm search and shows comparative statics; Section 2.3 describes the dataset and the context of the trade cost reduction, it also provides descriptive statistics on how firms responded to the trade cost fall; Section 2.4 provides empirical empirical evidence of search externalities in Uganda; and Section 2.5 concludes.

## 2.2. A simple model of firm-to-firm search in two markets

To illustrate the key mechanisms in this chapter, I build a simple model of buyers purchasing intermediary goods from suppliers in international and domestic markets.

The simple model is shown graphically in Figure 2.1. Buyers sell a single differentiated product to consumers in a frictionless retail market. Buyers purchase these products from suppliers, who are either domestic or international,[11] and each produces one differentiated product. International suppliers produce a higher quality product, but must pay a higher transportation cost. Buyers and suppliers cannot costlessly match, but must instead undertake search to find a match. In both markets, a match between a buyer and a supplier depends on the intensity of search effort and the equilibrium market tightness. In order to incorporate differences in search externalities between markets, I allow the matching technology to differ when looking for domestic or international suppliers.

I demonstrate the main mechanisms of the model by showing comparative statics of a reduction in trade costs leading to a reallocation of search between markets, but with some mitigation due to congestion.[12]

[10] A number of related empirical papers highlight additional aspects of the search frictions among firms (Bernard et al. (2015), Startz (2016), Steinwender (2018), Fafchamps and Quinn (2016)).

[11] International here implies a foreign exporter

[12] The simple model, however, misses some salient features observed in the data. In order to make the model match key moments from the Ugandan data, I extend the framework in Chapter 3 to include a number of these features and estimate the quantitative model.

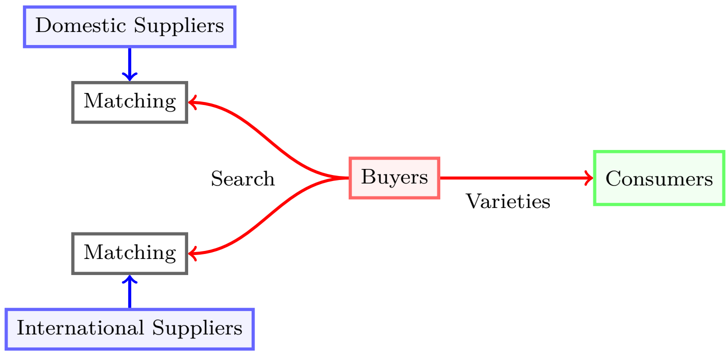



Figure 2.1. Model environment

### 2.2.1. Buyers, suppliers and consumers

There is a measure $B$ continuum of buyers, measure $S_D$ continuum of domestic suppliers and measure $S_I$ continuum of international suppliers. For simplicity, I assume for the simple model that $S_D = S_I = S$.

Suppliers produce differentiated products which they sell to buyers once they match. Let $B(s_I)$ denote the set of buyers who match with international suppliers. Similarly, let $B(s_D)$ denote the set of buyers who match with domestic suppliers. For simplicity I assume all suppliers have the same marginal cost.

Buyers pay an iceberg trade cost $\tau_I$ on each unit of international goods and iceberg trade cost $\tau_D$ on each unit of domestic goods, where I normalize $\tau_D = 1$.

Buyers begin with marginal cost $c$ and no matches. Buyers have a fixed search intensity $\sigma$ but choose the proportion of search they exert domestically, $a$ such that $a \in [0, 1]$, and internationally, $1 - a$.

Consumers demand differentiated products from buyers $b$ with a CES utility function, which shows their taste-for-variety over products sold by buyers

$$C = \left[ \int_{b \in B(s_I)} \psi_I C_b^{\frac{\eta-1}{\eta}} + \int_{b \in B(s_D)} C_b^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}, \tag{2.1}$$

where I assume all international products have the same demand shifter, $\psi_I$, and all domestic products have the same demand shifter, $\psi_D$, which I normalize to 1. If imports are higher quality products, we might expect $\psi_I > 1$ for imported goods, although I do not impose this. $\eta > 1$ is the elasticity of substitution between goods which does not vary between imports and domestic products.

### 2.2.2. Pricing and division of profits

In period one, buyers search and matches materialize. In period two, buyers compete using Bertrand competition in the retail market. This leads to the standard CES constant mark-up rule

$$\frac{p_b - c_b}{p_b} = \frac{1}{\eta}, \tag{2.2}$$

where $p_b$ is the price charged by buyer $b$.

Substituting the mark-up into the profit function yields the instantaneous profit flow for a buyer and a matched supplier which depends on whether the supplier is domestic or international

$$\pi(s_L) = \frac{E}{\eta P^{1-\eta}} \left[ \left( \frac{\eta}{\eta - 1} \right) \frac{\tau_L c}{\psi_L} \right]^{1-\eta} \quad \text{for } L \in \{D, I\}, \tag{2.3}$$

where $P$ is the standard CES aggregate price index and $E$ is household expenditure. Once I make the standard CES assumption that the elasticity of substitution $\eta > 1$, the profit function behaves as one would expect - increasing in the aggregate price index, decreasing in marginal cost. If there is a domestic good then $\tau_D = \psi_D = 1$. For higher international trade costs ($\tau_I$) or smaller international demand shifter ($\psi_I$) profits from matching with an international supplier are smaller.

I assume profits are split via Nash bargaining where $\Lambda \in [0, 1]$ is the bargaining coefficient for the seller and $1 - \Lambda$ is the bargaining coefficient for the buyer. This assumption means I do not need to consider inefficiencies lost due to double marginalization.[13]

### 2.2.3. Search and matching

I assume two aggregate matching functions which are homogeneous of degree one in the search of buyers and sellers, respectively. In the simple model, all sellers search such that their aggregate search is simply given by their mass $S$. The aggregate buyers' search in each market is given by the mass of buyers multiplied by the amount they search in each market, such that

$$\begin{aligned} B_D &= a\sigma B \\ B_I &= (1 - a)\sigma B. \end{aligned} \tag{2.4}$$

Following the labor literature, I assume that the aggregate measure of matches per unit time $(X^D, X^I)$ is homogeneous of degree one and increasing in the aggregate search

[13] In practice, Bernard and Dhingra (2015) show this assumption may not hold, but it is a necessary simplification for the purposes of this chapter as firm pricing is not a main feature of the chapter's focus.

of buyers and suppliers

$$
\begin{aligned}
X^D &= S^{\gamma_S} B_D^{\gamma_B} \\
X^I &= S^{\beta_S} B_I^{\beta_B}.
\end{aligned}
\tag{2.5}
$$

The matching function exponents are key objects in the model. A positive externality to search would be indicated by high $\gamma_S, \gamma_B$ and $\beta_S, \beta_B$. This is because, at the margin, an increase in buyers or sellers will lead to a large increase in the number of matches. There are increasing returns to scale in domestic matching if $\gamma_S + \gamma_B > 1$, in which case an increase in the mass of firms by 10% would have a greater than 10% increase in the number of matches.[14] By contrast, a congestion externality to search would be indicated by low $\gamma_S, \gamma_B$ and $\beta_S, \beta_B$, as more firms entering leads to very few new matches. There are decreasing returns to scale in domestic matching if $\gamma_S + \gamma_B < 1$. A low $\gamma_S$ would indicate that congestion is largely on the domestic supplier-side. Whereas, a low $\gamma_B$ would indicate that there is high congestion among domestic buyers. It is common in the labor literature to assume a constant returns to scale matching function, as this guarantees a single equilibrium and has some empirical support (Petrongolo and Pissarides, 2001). However, this has not been as extensively tested in firm-to-firm search. In Section 2.4, I show reduced-form evidence on the relative size of search externalities between markets. In Chapter 3, I structurally estimate the exponents in a richer version of the simple model to verify reduced-form results and to demonstrate further mechanisms within the model.

The match flow per unit of buyer search $\theta$ is a measure of market tightness and is defined separately in the domestic and international markets, given by

$$
\theta_D = \frac{S^{\gamma_S} B_D^{\gamma_B}}{B_D} \qquad \theta_I = \frac{S^{\beta_S} B_I^{\beta_B}}{B_I}. \tag{2.6}
$$

A higher value of $\theta$ simply indicates that the hazard-rate of finding a match is higher.

### 2.2.4. Optimal search

Buyers solve a maximization problem by picking an optimal search intensity in the domestic market $a$ to maximize profits

$$
\max_a \left\{ a\sigma\theta_D\pi(s_D) + (1-a)\sigma\theta_I\pi(s_I) - k(a) \right\}, \tag{2.7}
$$

[14] Allowing for the matching function to not be constant returns to scale generates a possibility for multiple equilibria (Petrongolo and Pissarides, 2001). For simplicity, I assume that firms obtain an equilibrium with the highest level of search.

where $a\sigma\theta_D$ and $(1-a)\sigma\theta_I$ are the endogenous hazard rates of making a domestic and international match, respectively. $k$ is a convex search cost on the amount that buyers search in each market such that $\frac{\partial^2 k}{\partial a^2} > 0$ and $k$ is minimized at $a = \frac{1}{2}$.[15] The rationale for this assumption is that it is relatively easy to undertake a light search in either market by, for instance, browsing the internet. However, undertaking a comprehensive search might involve travel or hiring a consultant, which would increase costs rapidly.

Taking the first order condition of Equation 2.7 yields a policy function which determines the optimal level of domestic search depending on the relative market tightness, the difference in profit from a domestic and an international supplier, and the change in search costs.

$$\sigma\theta_D\pi(s_D) - \sigma\theta_I\pi(s_I) - \frac{\partial k}{\partial a} = 0 \tag{2.8}$$

The intuition behind Equation 2.8 is that the firm wishes to choose their proportion of domestic search to equate the profit from matching with a domestic supplier multiplied by the probability of a domestic match with the profit from matching with a international supplier multiplied by the probability of a international match.

### 2.2.5. Comparative statics

To demonstrate the main search channel in the model, I present comparative statics of how firms respond to a reduction in transportation costs.

#### Buyer search decisions

The first comparative static shows how the proportion of search intensity in the domestic market changes when international trade costs change. In order to obtain this comparative static, I totally differentiate equation 2.8 as shown in Appendix 2, which yields equation 2.9.

$$\begin{aligned}\frac{\partial a}{\partial \tau_I} &= \frac{-\sigma\theta_I\frac{\partial \pi(s_I)}{\partial \tau_I}}{\sigma\frac{\partial \theta_I}{\partial a}\pi(s_I) - \sigma\frac{\partial \theta_D}{\partial a}\pi(s_D) + \frac{\partial^2 k}{\partial a^2}} \\ &= \frac{-\sigma\theta_I\frac{\partial \pi(s_I)}{\partial \tau_I}}{\sigma^2(1-\beta^B)\theta_I B_I\pi(s_I) + \sigma^2(1-\gamma^B)\theta_D B_D\pi(s_D) + \frac{\partial^2 k}{\partial a^2}}\end{aligned} \tag{2.9}$$

For the purposes of exposition, I discuss the case of a fall in transport costs to match the case study of Uganda. The numerator of equation 2.9 shows the direct effect of a change in trade costs; when trade costs decrease, the proportion of domestic search ($a$) falls as returns to importing increases.

[15] Picking the minimum point at $\frac{1}{2}$ is based on the assumption that searching equally in both markets is the minimum cost. Changing this to an alternative minimum would not alter results.

This is mitigated by two main forces. First, as firms increase import search, the international market becomes tighter driven by international congestion $\frac{\partial \theta_I}{\partial a}$. Second, as firms move out of the domestic market, domestic market-tightness decreases $\frac{\partial \theta_D}{\partial a}$. Together, these forces reduce the amount of reallocation towards imports following the international trade cost reduction.[16]

To reinforce the idea, consider a positive search externality in the international market ($\beta_B$ is large). Assuming that $\beta_B < 1$, then each additional buyer entering the international market reduces the probability of other firms matching, but only by a small amount. Therefore, a substantial volume of buyers can be absorbed by the international market before market-tightness increases sufficiently to stop this flow.

If $\beta_B > 1$, then each additional buyer joining the international market actually increases the chance of existing buyers matching. Even in this case, the model predicts that not all firms will search internationally, as buyers have convex search costs and there would be a reduction in market-tightness in the domestic market, as discussed below.

If there is a negative externality in the domestic market then $\gamma$ would be small. When buyers leave the domestic market, this causes a large reduction in market tightness in the domestic market. Consequently, it becomes easier for firms to match domestically, causing a smaller reallocation towards imports following the trade cost reduction.

### Consumer welfare and matching efficiency

The second comparative static concerns consumer welfare. Given all buyers are ex-ante identical, I can rewrite consumer welfare as the consumption from each buyer ($C$) multiplied by the matching probability of each type ($A$). $A$ is made up of the probability of a domestic match ($a\sigma\theta_D$) plus the probability of an international match ($(1-a)\sigma\theta_I$) multiplied by the international match demand shifter $\psi_I$.

$$\begin{aligned} W(a) &= \left[ \int_{b \in B(s_I)} \psi_I C_b^{\frac{\eta-1}{\eta}} + \int_{b \in B(s_D)} C_b^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}} \\ &= \underbrace{\Big[ a\sigma\theta_D + \psi_I(1-a)\sigma\theta_I \Big]}_{A} \underbrace{\left[ \int_{b \in B} C_b^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}}}_{C} \end{aligned} \tag{2.10}$$

The impact on welfare is therefore split into two parts. The first part is due to a reduction in trade costs leading to higher consumption acting through lower marginal costs $\frac{\partial C}{\partial \tau_I} < 0$. Meanwhile, the second part considers how the matching probability changes as trade costs change $\frac{\partial A}{\partial \tau_I}$. As shown in Appendix B.1, the change in welfare from matching

[16]In addition to these two forces there is a third force coming from the convexity of the search costs

following a fall in trade costs will be greater than zero if and only if the inequality in equation 2.11 holds.

$$\frac{\partial A}{\partial \tau_I} < 0 \iff \gamma_B a^{\gamma_B - 1} < \beta_B \psi_I (1-a)^{\beta_B - 1} S^{\beta_S - \gamma_S} B^{\beta_B - \gamma_B} \tag{2.11}$$

Equation 2.11 shows that for sufficiently large $a$ and $\psi_I \geq 1$, the change in welfare due to matching depends on the relative size of the matching exponents. If $\gamma_B < \beta_B$ and $\gamma_S < \beta_S$ then the returns to search are higher in the international market. Consequently, a fall in trade cost will increase welfare, given firms will move from matching in the decreasing returns to scale domestic market to the increasing returns to scale international market. The intuition for this result is that a reallocation of search leads to more matches for the same search intensity. Given consumers have a taste-for-variety, this generates an increase in consumer welfare.[17]

In summary, following a fall in trade costs, both the level of reallocation between markets and the degree to which consumer welfare increases depend on the relative size of search externalities in domestic and international markets.

## 2.3. Data, context and descriptive statistics

Having demonstrated the main mechanism in the simple model, I now look for empirical evidence of reallocation in firm supply-chains following a reduction in international trade costs. In this section, I first describe the datasets I use in this study, present descriptive statistics on firms and their connections in Uganda, and discuss the context and consequences of a reduction in trade costs.

### 2.3.1. Datasets

The data used in this chapter comes from four linked datasets collected by the Ugandan Revenue Authority (URA) which are administered for taxation purposes and cover the period 2010-2016. This data is confidential and is made available for the purposes of this research. Each tax dataset contains a unique tax identification number which allows the datasets to be linked across firms and time. The datasets contain the universe of firms

[17] An alternative consideration is to compare welfare in the decentralized market economy to the level of welfare should a social planner pick the optimal level of search in the presence of search frictions. This is similar to the Hosios (1990) condition, which shows in a wide array of search models that the socially optimal level of search occurs when buyers' share of the joint match surplus equals the elasticity of the matching function with respect to buyers (Mangin and Julien, 2018). However, the model does not fall into this class of models given the matching function is not constant returns to scale and there are two search markets.

paying tax in Uganda; consequently they are representative of the entire formal sector. It also contains the universe of importing firms in Uganda, as all firms choosing to import must go through a customs office and must be registered to pay tax. Inference on the informal sector is outside the scope of this study.[18]

The first dataset contains details on domestic firm transactions. Ugandan firms are required to record every transaction with any other tax-paying firm alongside the transacting firm's unique tax ID for Value-Added-Tax (VAT) purposes. This gives a line-by-line account of the good transacted, the value of the transaction, the date it took place, and the tax identification number of the linked firm. This dataset, therefore, constitutes a dynamic input-output matrix for the entire Ugandan formal economy.[19]

The second dataset contains transaction-level international trade data. The dataset includes variables of import origin, value, product and the matched foreign exporter on the other side of the transaction.[20]

The third dataset is monthly balance-sheet data from VAT records from 2010-2016. Ugandan firms are required to report on their total sales and total inputs each month. I winsorize these variables at the 5% level and collapse to annual frequency.

The fourth dataset is a firm registration dataset and contains descriptive details on the firm itself. This includes the ISIC industrial sector classification[21] and a more general description of its main operations. It also includes firms' addresses which I show on a map of Uganda in Figure 2.2.[22]

Descriptive statistics are presented in Table 2.1. The consolidated dataset contains 7,000 import buyers 13,000 domestic buyers, 24,000 import suppliers and 86,000 domestic suppliers. There are in total over 12 million transactions and over 490,000 firm-to-firm connections.

To the best of my knowledge, this is the first paper to link VAT transaction level data with firm employee and importer-exporter matched customs data. This allows observa-

[18] While I do not observe non-tax paying firms, this is not a major concern given tax paying firms in Uganda are much larger and more technically adept (Kathage, 2018) and represent the sample of firms I am most interested in. Between 2009-2011, 58% of Uganda's workforce was working in the informal sector, 13% of informal-sector workers were paid employees, 23% were unpaid helpers and 63% were working proprietors (mainly subsistence farmers) (Overseas Development Institute, 2015). There is a possibility that there is greater missed data domestically to internationally given import customs checks are likely to be more thorough.

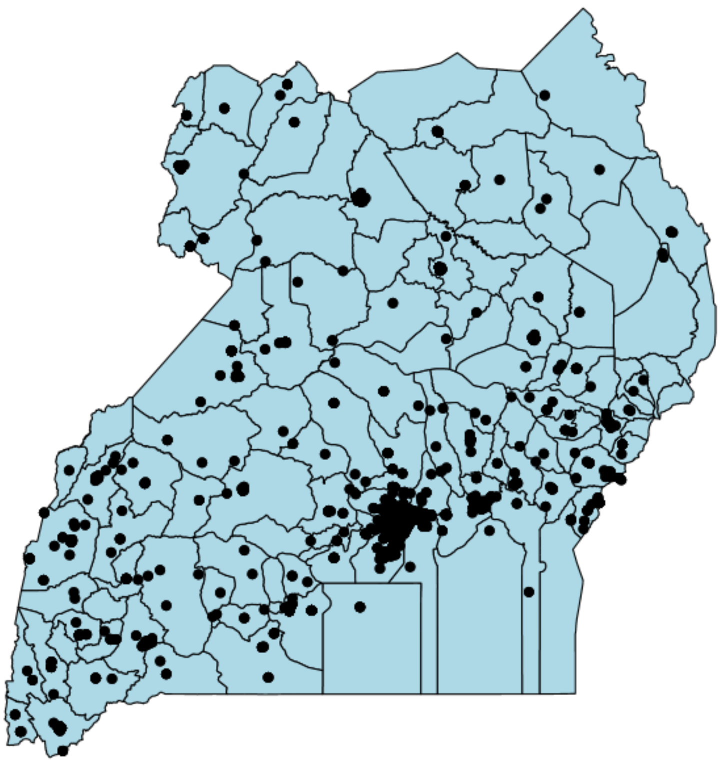
[19] It also allows a product-specific calculation of inputs, although this is not done for the time being given the complexity of the data management process since records are manually entered without product codes.

[20] There is also data on firm exports, although I do not use this for the purpose of this project given I am primarily interested in firm sourcing behaviour.

[21] Standard industrial classification of economic activities (ISIC) is a classification system for industry categories. The URA classifies firms at a 4 digit level.

[22] Address geo locations were mapped using google maps API

Figure 2.2. Locations of firms in Uganda



*Notes:* Each point on the graph represents a unique location, although there are likely to be multiple firms in each location. There are a small number of firms located on islands in lake Victoria located in the bottom right.

Table 2.1. Descriptive statistics

| **Variable** | **Import Sample** | **Domestic Sample** |
|---|---|---|
| Number of buyers | 6788 | 12984 |
| Number of suppliers | 24133 | 86689 |
| Number of buyers (> 3 matches) | 3373 | 7294 |
| Number of suppliers (> 3 matches) | 3451 | 17293 |
| Firm-to-firm connections | 71,000 | 420,000 |
| Transactions | 1.3m | 11m |
| Mean Age | 8.7 | 8.5 |
| Median Wage Bill (USH) | 100900 | 40100 |
| Median Sales (USH) | 1468800 | 972800 |

*Notes:* Data combined from Uganda administrative tax datasets from 2010-2016. The import sample comes from import trade data and the domestic sample comes from the VAT transaction dataset. Mean age comes from the firm registration dataset. Mean wage and sales comes from the firm balance sheet dataset.

tions on the complete and dynamic picture of the formal economy of Uganda. As research using tax data remains rare, one potential concern might be that the data is inconsistent with other datasets. In Appendix A.2, I address this concern by comparing the tax data used in this study to other freely-available data sources on firms in Uganda.

### 2.3.2. Context and trade cost reduction stylized facts

#### Ugandan Economy

Uganda is a landlocked country in East Africa which has experienced high and sustained growth driven by high investment levels and strong international trade performance. The economy is made up of a large services sector (56.6%); agriculture, forestry and fishing (24.2%); and industry (19.2%) (World Bank, 2019).

Uganda is open to the external sector with imports reaching 25.9% as a share of GDP in 2016/17 (International Monetary Fund, 2019). The largest components of imports are consumables and capital goods for investment (World Bank, 2019).

As shown in Figure 2.4, only a small proportion of Ugandan firms import. As shown in Table 2.1, importers are on average larger than firms who only source domestically, with median sales and wage bill 1.5 and 2.5 times higher, respectively. This is consistent with previous research on this topic (e.g. Bernard and Jensen (1999)).

On average, each Ugandan firm has 2.7 domestic suppliers. The sectors with the largest number of connections are in service and manufacturing industries including construction services, telecommunication services, accounting services, and the manufacturing of plastic products, metals, and paper products.[23]

#### Buyer-supplier search and search externalities in Uganda

Finding a buyer or supplier in Uganda is a costly process. Sen (2018) argues that a lack of information about Ugandan suppliers is one of the main reasons behind a lack of oil and gas sector supplier development. Steenbergen and Sutton (2017), in neighboring Rwanda, suggest that “international firms often do not have extensive local networks, and so are unfamiliar with all the inputs that domestic suppliers may be able to provide.” Buyers also have limited information about international suppliers given that the cultural, language and/or knowledge barriers are difficult for Ugandan firms to navigate. Moreover, as few firms in Uganda import, there are a limited number of firms to approach for importing advice.[24] This has led the Ugandan Government to target reducing search costs by 25% in 2019 (Government of Uganda, 2019).[25]

[23] This topic is covered in detail in Spray and Wolf (2016)

[24] Indeed, making new connections internationally has been shown in other countries to be easier if other firms in the same location are already importing (Bisztray et al., 2018).

[25] A similar goal is being targeted by the government of Rwanda through the Made in Rwanda policy via establishing a publicly available supplier database to make information about firms operating in Rwanda easier to find Spray and Steenbergen (2017).

If information about potential new suppliers diffuses among firms, either deliberately due to firms sharing knowledge or through buyers and suppliers meeting for instance in the same business location, then this would imply a positive search externality.[26] Qualitative interviews I undertook with firms in East Africa suggest that in some instances knowledge about new import suppliers is, indeed, passed among businesses.[27]

By contrast, it may be difficult for firms to make matches if there is a congestion externality. Congestion occurs when one firm's search reduces another firm's chance of matching. For instance, a buyer may spend resources looking for a supplier only to match with a firm who is unable to meet the demand because they have recently matched with another buyer (Arnosti et al., 2018). This effect has been shown to occur in multiple contexts where there is a search friction. For instance, Fradkin (2015) shows congestion in online platform AirBnB and Horton (2010) shows congestion in online labor markets.[28]

Given Ugandan suppliers are characterized by being small and with limited access to credit (Spray and Wolf, 2016), one might expect that congestion effects are larger among these firms compared to foreign importers. In an interview with a hotel in Uganda, the CEO stated that they had tried to find a domestic fruit and vegetable supplier, but that another similar hotel had recently signed up the supplier to an exclusive contract. In Section 2.4, I look for evidence for both of these effects empirically.

### Trade cost reduction and descriptive statistics

Despite having a high import volume, Uganda has some of the highest transportation costs in the world. In 2017, Uganda ranked 136 out of 190 countries on World Bank's Trading Across Border Index (World Bank, 2016). The majority of goods entering Uganda must first transit through the port of Mombasa in Kenya. In 2010, 68% of Ugandan imports arrived from the Kenyan border.[29] In 2010, the Mombasa port was described as having "persistent congestion", being "behind international standards" and facing issues of "corruption and incompetence" (Bulzomi et al., 2014). Once goods are cleared from the port, they are required to be transported over 1000km by road through Kenya, before crossing the border into Uganda. A map of the main trade corridor, and location of the

[26] If this information is priced then it would no longer represent an externality, however, this was never mentioned by firms.

[27] For instance, one tea processor explained that to find a foreign supplier of packaging products they would speak to multiple other business owners to obtain advice before purchasing. This was recounted to me in an interview with a tea factory CEO

[28] Fradkin (2015) shows a congestion effect for matches made on the online platform AirBnB, where 49% of inquiries are rejected or ignored by the host, and only 15% of inquiries lead to a transaction. An initial rejection decreases the probability that the guest eventually books any listing by 50%.

[29] Based on customs dataset. 25% of imports arrived through the airport, and the remainder came through the Tanzanian, Rwandan, Congolese borders or through the lake port in Jinja.

Figure 2.3. Transport costs for a 20' container


*Notes:* Data comes from the World Bank Trading Across Borders Index. The y-axis shows the import cost in US dollars per 20-foot container. The reform took place between 2010 and 2011.

six weighbridge truck stops is shown in Figure B.1 in Appendix A.1.

High transport costs have been shown in other research in Africa to severely constrict international trade (see Donaldson et al. (2017) for summary). Given Uganda's high trade costs, the effects of reducing transportation costs may be substantial.

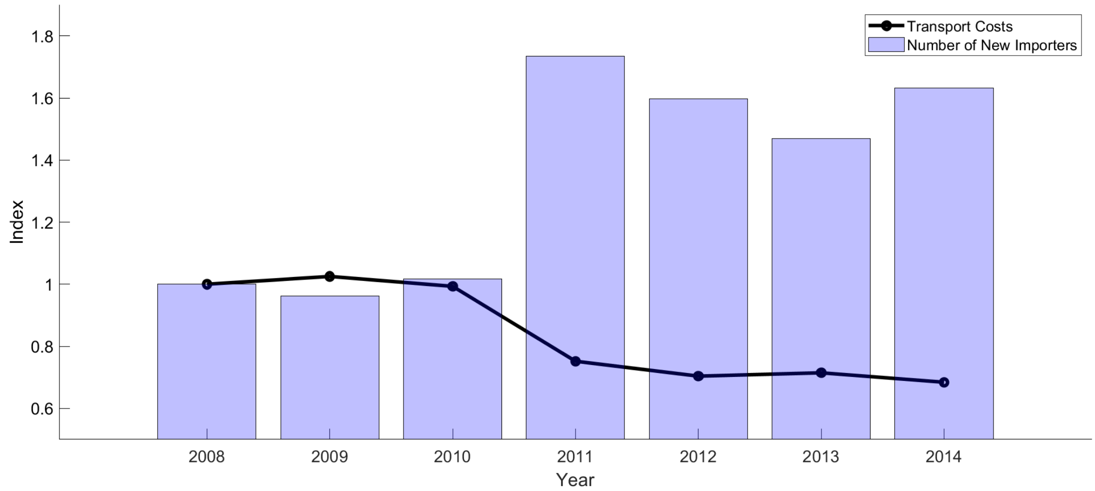
In 2011, Uganda implemented reforms to reduce the cost of importing. The main reforms were longer border opening hours and improved port infrastructure at the main port in Mombasa (World Bank, 2011). In addition, Uganda rehabilitated roads thanks to a large grant from the European Union and removed several weighbridges along the route (Bulzomi et al., 2014). These reforms were negotiated at the East African Community (EAC) level and so can be thought to be outside the direct control of the Ugandan government, thus making them quasi-exogenous. The combination of these reforms led to a 25% fall in transport costs in 2011, which then reduced the cost of importing a 20-foot container from USD5807 to USD4396 (-24.3%). As shown in Figure 2.3, this effect happened rapidly over one year and was later stable.

I present three descriptive statistics on how firms responded to the reduction in trade costs.

*(i) Falling transport costs corresponded with an increase in importers*

As shown in Figure 2.4, the fall in transport costs corresponded with an increase in the number of new importers. The fall in transport costs was very rapid between 2010 and 2011, and was then followed by a period of flat costs. Similarly, the increase in

Figure 2.4. Transport costs and new importers




*Notes:* The black line shows transport cost in USD per 20-foot container from the World Bank's Trading Across Border Index between 2007-2014, the bars show the number of new importers. The data for the bars comes from customs dataset. Reforms took place between 2010 and 2011.

importing also happened very rapidly followed by a corresponding period of zero growth. I show in Appendix Figures B.4 and B.5 that the total number of importers, the average number of suppliers and the proportion of firms which import also increase in line with the falling transport costs.[30] Although I do not observe a counterfactual of what would have happened in the absence of falling trade costs, this fits with what one would expect based on the previous literature in Africa (Donaldson et al., 2017).

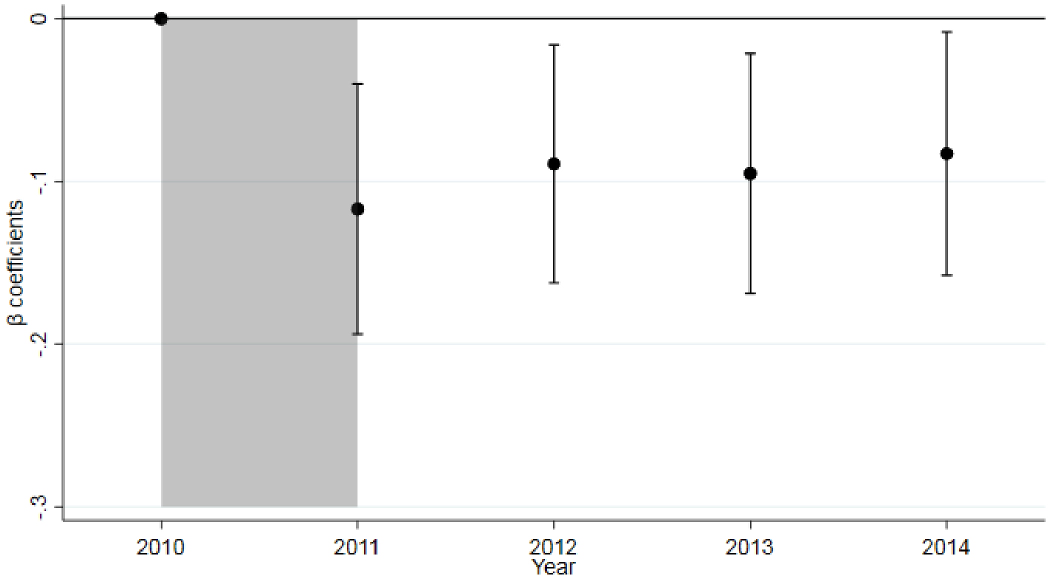
*(ii) Falling transport costs corresponded with new importers reducing domestic suppliers relative to other firms*

To demonstrate that the change in transportation costs also corresponded with firms making readjustments to their domestic supply-chains, I compare the number of domestic suppliers used by firms who first imported in 2011 to all other buyers in a difference-in-difference specification as shown in equation 2.12. Note, that I do not have domestic transaction data prior to 2010, so I can not look for pre-trend differences in treatment and control groups.

$$DomesticSuppliers_{it} = \sum_{t} \beta_t(\delta_t \times FirstImportIn2011_i) + \alpha_i + \delta_t + u_{it} \qquad (2.12)$$

[30]There is also an increase in exporting, although this happens slightly later, this is discussed in detail in Spray (2017)

Figure 2.5. $\beta$ coefficients from specification 2.12




*Notes:* The figure plots the $\beta$ point estimates from specification 2.12 and the 95% confidence interval. The red vertical line shows the period of reduced trade costs. The outcome variable is the log of the number of domestic suppliers.

where $DomesticSuppliers_{it}$ is the log of the number of domestic suppliers supplying firm $i$ at time $t$, $FirstImportIn2011_i$ is a dummy variable indicating whether firm $i$ first imported in 2011, $\alpha_i$ is a set of buyer fixed effects, and $\delta_t$ is a set of year dummies.

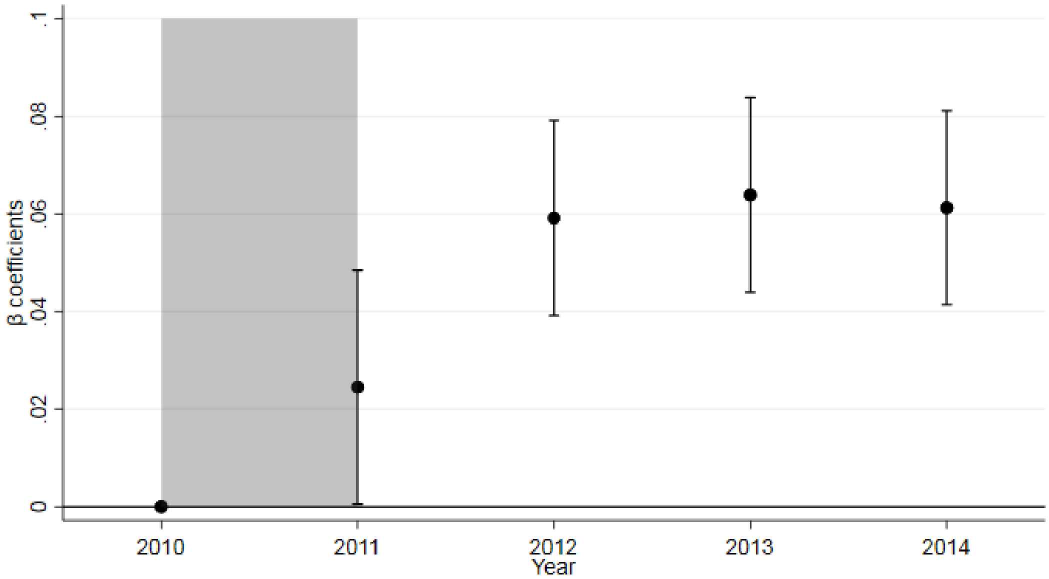
I plot the $\beta$ coefficients in Figure 2.5, and present the results in regression format in Table B.1 in Appendix. I also show in Appendix Table B.3 that this descriptive statistic is robust to using the value of domestic inputs as opposed to the number of suppliers. Relative to the control group,[31] new importers reduced their number of domestic suppliers by 10% in the year of the international trade cost reduction. The effect declines over time, but is still significant at the 5% level two years later. This result is non-trivial, as we might expect new-importers to be generally expanding and hence adding both domestic and international suppliers. The fact that this is not the case suggests that firms choose either domestic or international sourcing strategies.

*(iii) Suppliers which were dropped by new-importers rematched with non-importing firms*

As I observe which specific suppliers were dropped by first-time importers in 2011, I now consider whether these dropped suppliers managed to replace their lost buyers with

[31]In this case the control group is all other firms.

Figure 2.6. $\beta$ coefficients from specification 2.13




*Notes:* The figure plots the $\beta$ point estimates from specification 2.13 and the 95% confidence interval. The red vertical line shows the period of reduced trade costs. The outcome variable is the log of the number of domestic suppliers.

buyers who were importers, or buyers who only sourced goods domestically in Uganda. In order to show this, I estimate equation 2.13.

$$PropNonImportingBuyers_{ft} = \sum_t \beta_t(\delta_t \times Dropped_f) + \delta_t + \alpha_f + u_{ft} \qquad (2.13)$$

where $Dropped_f$ is a dummy variable for whether supplier $f$ was dropped by a buyer who first imported in 2011 and $PropNonImportingBuyers_{ft}$ is the proportion of buyers for supplier $f$ at time $t$ which do not import, excluding any buyers which were 2011 first-time importers to avoid a spurious correlation.

As shown in Figure 2.6 and Table B.2, suppliers which lost a buyer to a 2011 first-time importer rematched with buyers who were not importers. This effect is significantly different to zero even four years after the event. I also show in Table B.4 that it is robust to using the value of inputs as opposed to the number of suppliers.

We must treat these three descriptive statistics with caution as they show correlations as opposed to causal relationships. However, together, the results are consistent with the mechanism laid out in the simple model. When trade costs fell, importing became more attractive which led to a rebalancing of search in favour of international markets. This movement out of domestic search created space in the domestic market, allowing non-importing firms to match with the dropped suppliers.

## 2.4. Reduced-form evidence of search externalities

In this section, I look for evidence consistent with search externalities in both markets in reduced-form, and also present evidence on the relative size of these externalities between markets.

### 2.4.1. Motivating evidence

Figure 2.7 shows the percentage of supplier matches which have at least one buyer in the same neighborhood. The first bar shows that 21% of suppliers' new matches with domestic or import suppliers are in the same building as an existing customer.

This tight proximity between suppliers' customers is consistent with the fact that it is easier to sell to customers in similar locations. One explanation for this is that information about potential suppliers may diffuse more easily among closely located buyers. This could be because closely located buyers have stronger relationships or because suppliers may bump into potential buyers operating close to their existing customers. This narrative is supported by comparing the percentage of matches with a buyer in the same building (21%) to the percentage of matches in the same or next-door buildings (25%), an increase of just 4% from adding next-door buildings. Firms in the same building are unlikely to be substantially different to their next-door neighbors, except in the ease with which information can diffuse. However, even when moving from one building to the next, the diffusion of knowledge appears to reduce substantially.

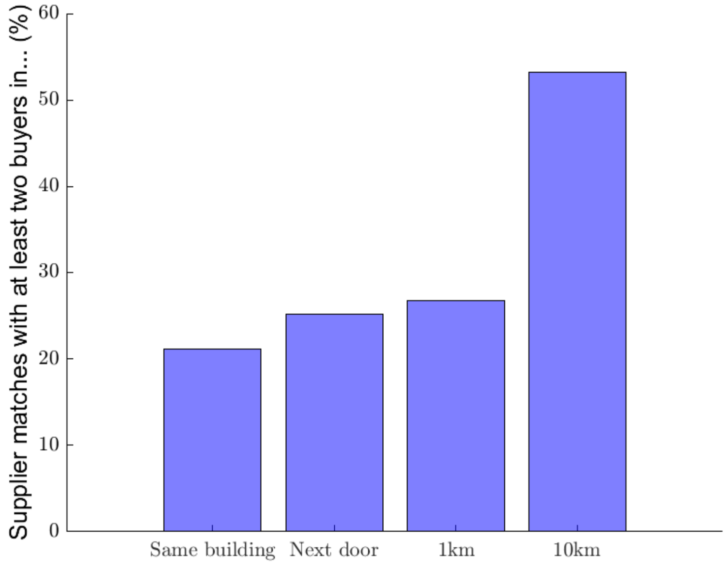
While these results are consistent with a positive information spillover, they do not exploit the richness of the data, and have nothing to say on the possible negative externality. In the next section, I move into a more formal characterization of this effect.

### 2.4.2. Empirical strategy

In order to explain the empirical strategy, consider the following example. Two firms in Kampala, {A,B}, are looking for a new supplier. Each firm can look for this supplier either locally or abroad. As discussed in Section 2.3.2, there are two ways $A$'s search might influence $B's$ probability of matching; either $B$ may pass information to $A$ (a positive externality) or $B$ may crowd-out $A$'s chance of matching (a negative externality).

If information is easier to diffuse among firms located close to one another, then the spatial diffusion of firms can be used to identify different externalities. In order to test for a positive search externality, I consider whether one firm making a match increases the probability of geographically close firms making the same supplier match. To test for a negative externality, I consider whether one firm making a match decreases the probability of geographically distant firms making the same supplier match.

Figure 2.7. Percentage of suppliers' matches which have an existing buyer in location




*Notes:* On the y-axis is the percentage of supplier matches with at least two buyer in the same location. On the x-axis, the location progressively gets wider away, such that *Next door* refers to the proportion of supplier matches with an existing buyer either in the same building or in the next-door building.

### 2.4.3. Dataset

I begin by generating a dataset of every buyer-supplier-year triplet separately for domestic and international suppliers. Given that I observe over 13,000 domestic buyers and 86,000 domestic suppliers over 6 years, this generates a dataset with 6.8 billion observations.

However, many matches are unlikely to ever be formed. For instance, you would not expect an iron ore mine to supply a tea factory. Instead, I trim this dataset to obtain a sample of likely matches. First, I drop suppliers which have never sold to the buyer's ISIC 4-digit industry. Second, I drop any buyer or supplier which does not make at least three matches over the entire sample period. Third, I drop any observations from the sample following the first observed match. This restricts the sample to only consider the first-time matches between firms which are active and which are in sectors which are likely to trade.

### 2.4.4. Main specification

The main specification is given by the linear probability model shown in equation 2.14

$$Y_{ift} = \mu X_{if,t-1}^{neighborhood} + \gamma X_{if,t-1}^{other-city} + \alpha_i + \delta_t + u_{ift} \tag{2.14}$$

where $Y_{ift}$ is a dummy $= 1$ if buyer $i$ adds supplier $f$ for the first-time in period $t$. $X^{neighborhood}_{if,t-1}$ is a count of number of firms who matched with supplier $f$ in $i$'s neighborhood in period $t-1$.[32] $X^{other-city}_{i,t-1}$ is a count of number of firms who added supplier $f$ in $t-1$ but are not in $i$'s city.

If information diffuses among firms about suppliers, we would expect these effects to occur more strongly among geographically closer firms. Therefore, $\mu > 0$ would be consistent with a positive externality.

If suppliers have a limited capacity to add multiple buyers at once, then firms making matches elsewhere in the country should decrease the probability of buyers in other locations making a match. Therefore, $\gamma < 0$ would be consistent with a negative congestion externality.

I consider four different definitions of neighborhood. The first two definitions of neighborhood consider any firm located in 10 km and 1 km radii, respectively. While these measures include a wide array of firms which could cause an information spillover, however, they suffer from the possibility that location-specific shocks hit geographically close firms. This motivates the use of two additional measures of neighborhood that consider firms located in the same building and firms located in next-door buildings. The second specification, shown in equation 2.15, compares the latter two definitions of neighborhood simultaneously, given that one might expect firms in the same building to be structurally very similar to those located in next-door buildings in all respects except that information is harder to diffuse across buildings than within buildings. Results would be consistent with a positive spillover if $\mu_1 > \mu_2 > 0$.

$$Y_{ift} = \mu_1 X^{same}_{if,t-1} + \mu_2 X^{nextdoor}_{if,t-1} + \gamma X^{other-city}_{if,t-1} + \alpha_i + \delta_t + u_{ift} \tag{2.15}$$

In both specifications, I include buyer and time fixed effects ($\alpha_i$ and $\delta_t$) which control for unobserved buyer characteristics and time trends.

I consider domestic and international suppliers in separate regressions, and test whether the respective coefficients are different.

### 2.4.5. Results

As can be seen in column 1 of Table 2.2, each additional importer of supplier $f$ within a 10 km radius increases the probability of buyer $i$ matching with supplier $f$ by 0.086%. This is a significant magnitude given that the baseline probability of a match is very low: 0.00393 for imports and 0.00398 for domestic samples. Column 3 demonstrate that this

[32] I run a robustness on this specification in Tables B.8 and B.9 where I test alternative functional forms showing results are robust to including a continuous measure of the number of new buyers in a neighborhood.

Table 2.2. Import suppliers

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $Y_{ift}$ | $Y_{ift}$ | $Y_{ift}$ | $Y_{ift}$ |
| $X^{10km}_{if,t-1}$ | 0.0864*** | | | |
| | (0.00693) | | | |
| $X^{1km}_{if,t-1}$ | | 0.0819*** | | |
| | | (0.00658) | | |
| $X^{same}_{if,t-1}$ | | | 0.0910*** | 0.0907*** |
| | | | (0.00624) | (0.00651) |
| $X^{nextdoor}_{if,t-1}$ | | | | 0.00128 |
| | | | | (0.00994) |
| $X^{other-city}_{if,t-1}$ | -0.00347* | -0.00242 | -0.00240 | -0.00234 |
| | (0.00179) | (0.00172) | (0.00171) | (0.00177) |
| Observations | 4834635 | 4834635 | 4834635 | 4834635 |
| Year and Firm FE | YES | YES | YES | YES |

*Notes:* Unit of observation is buyer $i$, supplier $f$ and year $t$. Dependent variable $Y_{ift}$ indicates a first match took place between buyer and supplier. $X^k_{ift}$ is a count of buyers in region $k$ which added supplier $f$ in $t-1$. Coefficients are multiplied by 100 to read as percentage point marginal effects. Standard errors in parentheses are clustered at the buyer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

effect is larger when just looking at firms in the same building, which is consistent with information diffusion having a larger effect at shorter distances. Column 4 shows that a firm in the same building adding a new supplier has a much larger marginal effect, when compared to a firm in a next-door building adding a new supplier (0.09% vs. 0.001%, respectively). This is consistent with a local information spillover among firms in the same building, but that this becomes more difficult across buildings. Taken together, these results are consistent with qualitative evidence that firms share information on import suppliers presented in Section 2.3.2.

Evidence on negative spillovers is also consistent across specifications. Where an additional buyer being added in a different city to buyer $i$ in the previous year reduces the probability of $i$ matching by between 0.0023% and 0.0035%. This effect is small and not statistically significant.

In Table 2.3, I show results for the same specification run on the sample of domestic suppliers. As in the import case, having an additional buyer in the same neighborhood increases the probability of buyer $i$ matching with supplier $f$ for all definitions of neighborhood. Unlike the import case, this effect is not significantly different from zero. Additionally, the magnitude of this positive coefficient is in all cases smaller than in the

Table 2.3. Domestic suppliers

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $Y_{ift}$ | $Y_{ift}$ | $Y_{ift}$ | $Y_{ift}$ |
| $X^{10km}_{if,t-1}$ | 0.00513 | | | |
| | (0.00606) | | | |
| $X^{1km}_{if,t-1}$ | | 0.00502 | | |
| | | (0.00612) | | |
| $X^{same}_{if,t-1}$ | | | 0.00509 | 0.00465 |
| | | | (0.00613) | (0.00631) |
| $X^{nextdoor}_{if,t-1}$ | | | | 0.000616 |
| | | | | (0.000322) |
| $X^{other-city}_{if,t-1}$ | -0.00515*** | -0.00515*** | -0.00515*** | -0.00516*** |
| | (0.000962) | (0.000967) | (0.000972) | (0.000938) |
| Observations | 27975967 | 27975967 | 27975967 | 27975967 |
| Year and Firm FE | YES | YES | YES | YES |

*Notes:* Unit of observation is buyer $i$, supplier $f$ and year $t$. $Y_{ift}$ indicates a first match took place between buyer and supplier. $X^k_{ift}$ is a count of buyers in region $k$ which added supplier $f$ in $t-1$. Coefficients are multiplied by 100 to read as percentage point marginal effects. Standard errors in parentheses are clustered at the buyer level. $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

import case.

Unlike on the import side, evidence in Table 2.3 is consistent with congestion effects among domestic suppliers. In all specifications, an additional buyer in a different city in the previous year decreases the probability of the firm matching by 0.0052%. This is statistically significant at the 1% level.

Taking the results from Tables 2.2 and 2.3 together provides evidence consistent with a positive externality to search in international markets and a negative externality to search in the domestic market. As shown in the simple model, this should lead to higher welfare gains following a reduction in international trade costs.

### 2.4.6. Mechanisms and alternative explanations

I now consider two main possible alternative explanations for these results; either that very local shocks are driving results or that spillovers do exist, but that they are not search related. The full detail is provided in Appendix C.2.[33]

[33] A key point to keep in mind is that the main role of this section is to demonstrate a difference in imports and domestic suppliers externalities. As long as these concerns do not differ systematically across domestic and international suppliers, then we should be less concerned.

If firms in the same building were systematically different to firms in next-door buildings, then this might raise a concern that local shocks to specific industries drive results. To address this concern, in Appendix Table B.5 I compare the proportion of firms in the same ISIC 4-digit sector in the same building to those in the next-door building. While there is a small difference, it is not statistically significant. However, when I look at firms further away, I do see this difference increasing. I therefore conclude that there is some firm agglomeration, but that it is happening at a block level and not at a building level. Moreover, the fact that the agglomeration decreases over space, but that the impact of an additional buyer in the neighborhood does not dramatically decrease between columns 1 and 3 of Table 2.2 suggests this is not a major concern.

A second alternative explanation is that a spillover is taking place, but that it is not search related. To allay these concerns, I test if the marginal effect is smaller among firms where one would expect search frictions to be less prevalent. In Appendix Table B.6, I interact the independent variables with whether the import supplier exported from the East African Community (EAC). This is because one would expect search frictions to be smaller in local neighbors such as Kenya when compared to more distant locations.

Another prediction consistent with search frictions, is that suppliers which are not supply-constrained will be able to match with multiple buyers, and so we should not observe a negative congestion effect. As discussed in Section 2.3.2, this is the reason why we did not expect to find a strong congestion externality on foreign imports, given international suppliers are characterized by being large firms with cheap access to credit and multiple customers. Results in Appendix Table B.7 show that domestic suppliers which are exporters, and hence less supply constrained, have a smaller negative effect from making a match elsewhere in the country. This is again consistent with the search narrative.

## 2.5. Conclusion

Using novel data on both domestic and international firm-to-firm transactions from Uganda, I show that the presence of search frictions between buyers and suppliers, in a low-income country, can have a significant impact on how firms respond to a trade liberalization.

I show in a model of firm-to-firm search and matching in two markets that the relative size of search externalities determines the extent of sourcing reallocation, as well as changes to consumer welfare. I then show that a change in trade costs led to firms reallocating their supply-chain in line with the predictions of the model. Given the importance of the search externality parameters, I then show empirical evidence consistent with stronger positive externalities in international markets compared to domestic markets. Together, this predicts that a change in trade costs will lead to an increase in welfare through not only lower prices, but also through increased matching efficiency.

In order to investigate the magnitude of these effects further it is necessary to build and estimate a quantitative model. This exercise is undertaken in Chapter 3.

# Chapter 3

# A Quantitative Model of Buyer-Supplier Search in Two Markets

Abstract I build and estimate a dynamic quantitative model of firm-to-firm search and matching in two markets. I then estimate the model to match VAT data from Uganda. Structural estimates of the model's parameters provide evidence that the domestic market is more congested than the foreign market. I then show that a 25% reduction in trade costs will lead to a 5.2% increase in consumer welfare, 15% of which was due to search externalities. I also show that reducing search costs between firms could significantly increase welfare but is best targeted on reducing international search costs when compared to domestic search costs.

## 3.1. Introduction

The previous chapter identified a new channel via which a reduction in trade costs could impact consumer welfare. Namely, a reduction in trade costs will lead firms to reallocate search towards the import market and away from the domestic market. This will free up space in the domestic market allowing other firms to find suppliers. If the domestic market is less congested than the foreign market, then this process will lead to an improvement in matching efficiency, which will increase welfare. In Chapter 2, I presented empirical evidence consistent with this mechanism.

In order to investigate the magnitude of this channel and to consider counterfactual policy analysis requires a rich quantitative model. In this chapter, I build and estimate a dynamic model of buyer-supplier search and matching in two markets. This extends the simple model presented in the previous chapter substantially. In a dynamic setting, both buyers and suppliers choose optimal search intensity and the proportion of search in each market. The model builds on existing work by Eaton, Jinkins, Tybout and Xu (2016) (hereafter EJTX (2016)), adding both a domestic and an international search decision and market-specific matching functions, as well as adding firm heterogeneity and additional structure to search costs.

The most important structural parameters are those which govern the returns to scale in the matching function. The structurally estimated parameters substantiate the reduced-form findings using a different yet complementary methodology.[1] I find that there are decreasing returns to scale to searching in domestic markets and increasing returns to scale to searching in international markets, as is consistent with the reduced-form results. I then test the external validity of the model by simulating the effect of a reduction in international trade costs and comparing the results to what is observed in the data. The proportion of firms that import increases from 20% to 23%, the average number of import suppliers increases by 20% and the average number of domestic suppliers decreases by 6.5%. The change observed in the data is the same direction and of a similar magnitude to that seen in the simulation.

Using the model, I run two counterfactual experiments. In the first experiment, I consider how much the increase in consumer welfare is due to differences in search externalities between markets. I again simulate the reduction in trade costs, but assume both markets have the same constant returns to scale matching function. The average

[1] The reduced-form methodology has the advantage of being clearer where the estimated coefficients come from. However, in this paper the reduced-form structure is restrictive and one might expect that there are multiple channels for search externalities to pass which are not picked up by the reduced-form. I therefore turn to a structural model which allows a more clearly model-driven pass through of externalities and has the large advantage of allowing the consideration of policy counterfactuals.

number of import suppliers increases by a smaller amount (11.1% vs. 20.1%), as there is a larger increase in import market tightness. There is also a larger decrease in the average number of domestic suppliers (-9.8% vs. -6.5%), this is because the reduction in search domestically does not have the mitigating effect of reducing congestion in the domestic market. This results in an increase in consumer welfare which is 15% smaller than when I allow there to be differences in externalities between markets, demonstrating that allowing for search externalities has a quantitatively important impact on welfare.

Second, I simulate the government of Uganda's goal of a "25% reduction in search costs for suppliers" as one of its four goals in trade (Government of Uganda, 2019).[2] I show that this leads to a 3-5% increase in consumer welfare, depending on where the reduction is targeted. If the government reduces international search costs, then this will significantly increase the number of matches in the same manner as the trade cost reduction. If, however, the government reduces domestic search costs then the impact, albeit still positive, is dampened by the increase in domestic congestion caused by a greater number of searching firms.

In addition to the literature discussed in the previous chapter, this chapter relates to a literature on estimation of trade models with search. Eaton et al. (2016) build and estimate a dynamic model of search and matching between importers and exporters using Columbian matched data. Tintelnot et al. (2018) estimates a model using Belgium matched domestic trade data. I build on this by providing structural evidence of search externalities which differ between markets, which I use to show welfare consequences of different counterfactual experiments. I also utilize a natural experiment to obtain external validity for the models predictions.

The remainder of this chapter is organized as follows: Section 3.2 presents the quantitative model; Section 3.4 structurally estimates the model; Section 3.5 provides counterfactual simulations; and Section 3.6 concludes.

## 3.2. A quantitative model of buyer-supplier search in two markets

The simple model presented in the previous chapter (Section 2.2) highlights the key mechanism of how a change in trade costs can impact welfare, but misses a number of salient features in the real world. The full model builds on the dynamic empirical model developed by Eaton, Jinkins, Tybout and Xu (2016) (EJTX (2016)) to address

[2]The specific sub targets are i) establishing a internet platform support programme (e.g. organize quarterly trainings on the use of Ali Baba), ii) encourage firms peer-to-peer learning (e.g. organize quarterly peer groups with Uganda business groups), iii) target key firms in supplier development programmes (e.g. establish anchor firm support unit and annual public-supplier meetings).

these limiations. The main departure is that I add international and domestic suppliers, different search costs and matching functions, and a greater degree of firm heterogeneity.

The most important extension from the simple model is to incorporate firm heterogeneity. As shown in Table 2.1 in the previous chapter, only a subset of firms in Uganda import and these firms are on average significantly larger. In order to incorporate this feature, I allow firms to draw a marginal cost and then pay a fixed cost for searching internationally, therefore in equilibrium this means that only the lowest marginal cost firms import.

A second source of buyer heterogeneity comes in the number of matches made by firms. I observe in the data that a large mass of firms have a small number of suppliers, however, I also observe many firms with over 30 suppliers. I therefore allow buyers and suppliers to make multiple matches by making the model dynamic, adding an additional search intensity decision, and exogenous link death probability. In addition to matching buyer size distributions, I also match supplier size distributions by allowing suppliers to make an optimal search decision.

### 3.2.1. Buyers and suppliers

There is a measure $B$ continuum of buyers, measure $S_D$ continuum of domestic suppliers and measure $S_I$ continuum of international suppliers.

Suppliers produce differentiated products $(x)$ which they sell to buyers $(b)$ once they match. Let $B(s_I)$ denote the set of buyers who match with international suppliers. Similarly, let $B(s_D)$ denote the set of buyers who match with domestic suppliers. Suppliers choose search intensity $\sigma_j^S(n)$. There is an exogenously given probability $\delta$ of an existing match being severed.

There are $\Gamma$ buyer types indexed $i \in \{1, 2, ..., \Gamma\}$ with marginal cost $c_i$ drawn from a known distribution, and match with $\mathbf{s} = \{s_I, s_D\}$ suppliers. This now warrants a change of subscripts from buyer $b$ to buyer type $i$. Buyers choose their search intensity $\sigma_i^B(\mathbf{s})$ and choose the proportion of search they exert domestically, $a$ such that $a \in [0, 1]$, and internationally, $1 - a$.

Buyers pay an iceberg trade cost $\tau_I$ on each unit of international goods and iceberg trade cost $\tau_D$ on each unit of domestic goods, where I normalize $\tau_D = 1$.

### 3.2.2. Consumers

Consumers have a nested CES utility function which shows their taste-for-variety over buyers ($b$) and products ($x$), such that

$$C = \left[ \int_{b \in B} C_b^{\frac{\eta-1}{\eta}} \right]^{\frac{\eta}{\eta-1}} \tag{3.1}$$

$$C_b = \left[ \sum_{x \in J(s_I)_b} (\psi_I C_b^x)^{\frac{\alpha-1}{\alpha}} + \sum_{x \in J(s_D)_b} (C_b^x)^{\frac{\alpha-1}{\alpha}} \right]^{\frac{\alpha}{\alpha-1}}, \tag{3.2}$$

where $J(s_I)_b$ is the set of international products $x$ offered by buyer $b$ and $J(s_D)_b$ is the set of domestic products $x$ offered by buyer $b$, $C_b^x$ is consumption of product $x$ from buyer $b$, and $C_b$ is consumption of the set of products offered by $b$. $\eta$ and $\alpha$ are the elasticities of substitution among products and buyers, respectively. I assume all international products have the same demand shifter, $\psi_I$, and all domestic products have the same demand shifter, $\psi_D$, which I normalize to 1. If imports are higher quality products, we might expect $\psi_I > 1$ for imported goods, although I do not impose this.

### 3.2.3. Pricing and division of profits

As buyers now match with multiple suppliers, they sell multiple goods. They, therefore, internalize the price set on one good on the demand of their other goods. This yields a first order condition on prices given by

$$q_{xb} + \sum_{x' \in J_b} \frac{\partial q_{x'b}}{\partial p_{xb}} (p_{x'b} - c_{x'b}) = 0 \quad \forall x \in J_b, \tag{3.3}$$

where $c_{x'b}$ is the marginal cost of supplying product $x'$ to consumers through buyer $b$. The intuition behind Equation 3.3 is that buyers internalize that their pricing on one good alters demand on other goods.

The instantaneous profit flow created by buyer $b$ and its set of suppliers is now given by a summation over the profit provided by each product $x$ in buyer $b$'s bundle ($J_b$), such that

$$\pi_b(\mathbf{s}) = \frac{E}{\eta P^{1-\eta}} \left[ \sum_{x \in J_b} \left( \frac{\eta}{\eta - 1} \right)^{1-\alpha} \tau_L \tilde{c}_b^{1-\alpha} \right]^{\frac{1-\eta}{1-\alpha}}, \tag{3.4}$$

for $L \in \{D, I\}$ and where $\tilde{c}_b = c_b/\psi_L$ is the quality-adjusted marginal cost, $\mathbf{s} = \{s_I, s_D\}$ is a vector of the number of international and domestic suppliers, $P$ is the standard CES aggregate price index and $E$ is household expenditure. As long as $\alpha > \eta > 1$, then the profit function is increasing in the aggregate price index and decreasing in marginal cost. This condition also ensures that there are diminishing returns to the number of suppliers, given that adding a new supplier appears in the summation $x \in J_b$ which leads to an increase in profit but at a decreasing rate, as long as the exponent $\frac{1-\eta}{1-\alpha} < 1$.[3] If the buyer matches with a domestic supplier then $\tau_D = \psi_D = 1$. For higher international trade costs ($\tau_I$) or smaller international demand shifter ($\psi_I$) profits from matching with an international supplier are smaller.

As buyers now have multiple suppliers, division of profits becomes more complex. I assume Stole and Zwiebel (1996) bargaining which gives each seller a profit flow $z_{ji}$ equal to their bargaining share multiplied by their marginal contribution to profit which depends on whether the good is domestic or international $L \in \{D, L\}$.[4]

$$\begin{aligned} z_{ji}(\mathbf{s}) &= \Lambda \frac{\partial \pi_i^T(\mathbf{s})}{\partial s_L} \\ &= \frac{\Lambda}{\alpha - 1}\left(\frac{\eta}{\eta - 1}\right)^{-\eta} \frac{E}{P^{1-\eta}} \left[\sum_{j \in J_b} \tau_L \tilde{c}_i^{1-\alpha}\right]^{\frac{\alpha-\eta}{1-\alpha}} \tau_L \tilde{c}_i^{1-\alpha} \end{aligned} \tag{3.5}$$

Equation 3.5 is very close to being a structural equation which would be estimatable in the data, therefore allowing the recovery of key parameters. However, the seller's profit $z_{ji}$ is not observable in the data. Instead, the data shows a firm-to-firm transaction which includes both profit and compensation for marginal costs in production of each good. If a constant fraction $\lambda$ of the variable costs is attributable to the seller,[5] then the revenue transfer can be expressed between firms $r_{ji}$ in terms of fixed effects and observables

$$r_{ji}(\mathbf{s}) = (h_{j|i})^{\frac{\alpha-\eta}{\alpha-1}} \frac{E}{P^{1-\eta}} \left(\frac{\eta}{\eta-1}\right)^{-\eta} \left(\tau_L \tilde{c}_{ji}\right)^{1-\eta} \left[\frac{\Lambda}{\alpha-1} + \lambda\right], \tag{3.6}$$

where $r_{ji}$ is the revenue for seller $j$ from buyer $i$, $h_{j|i} = \frac{\tau_L \tilde{c}_j^{1-\alpha}}{\sum_{l=1}^{J} s_l \tau_L \tilde{c}_i^{1-\alpha}}$ is the within buyer-$i$ revenue share of a type-$j$ seller, $\lambda$ is the seller's fraction of marginal cost. Equation 1.4 is a structural equation which I follow EJTX (2016) in estimating from the data in order to obtain elasticity of substitution parameters $\eta$.

[3] In this way profit depends on the number of suppliers, however, this is not to be confused with diminishing returns to scale in the matching function discussed in Section 3.2.4.

[4] Stole and Zwiebel (1996) is a generalization of Rubinstein bargaining to multiple firms based on Shapley value which gives firms a constant fraction of revenue

[5] This assumption only influences the estimation of the structural equation for the purpose of extracting the elasticity of substitution parameters. In all other aspects I consider the buyer and supplier to be jointly maximising profits.

### 3.2.4. Search and matching

Relative to the simple model, modelling search-and-matching is made more complex by the addition of a search intensity choice for buyers and suppliers ($\sigma^B, \sigma^S$ respectively) and given that buyers have a choice on the proportion of search done domestically ($a$).[6]

Following EJTX (2016), I define a new variable, visibility ($H$) of a type-$i$ buyer in domestic and international markets, respectively, as

$$\begin{aligned} H^B_{i,D}(\mathbf{s}) &= a_i(\mathbf{s})\sigma^B_i(\mathbf{s})M^B_i(\mathbf{s}) \\ H^B_{i,I}(\mathbf{s}) &= (1-a_i(\mathbf{s}))\sigma^B_i(\mathbf{s})M^B_i(\mathbf{s}), \end{aligned} \tag{3.7}$$

where $M^B_i(s_D, s_I)$ is a measure of type-$i$ buyers with $\mathbf{s}$ sellers. Intuitively, buyers of type-$i$ are more visible if they are searching more ($a_i\sigma_i, (1-a_i)\sigma_i$) and if there is a larger mass of them ($M^B_i$).

The overall visibility of buyers in the domestic and international market is a summation over all buyer types and for any number of existing matches.

$$H^B_L = \sum_{i=1}^{I}\sum_{s_L=0}^{s_{L_{max}}} H^B_{i,L}(\mathbf{s}) \text{ for } L \in \{D, I\} \tag{3.8}$$

Domestic and international sellers' visibility ($H^S_D, H^S_I$) are defined symmetrically to buyers

$$\begin{aligned} H^S_D(n) &= \sigma^S_D(n)M^S_D(n) \\ H^S_I(n) &= \sigma^S_I(n)M^S_I(n). \end{aligned} \tag{3.9}$$

The matching function is similar to the simple model, but is now increasing in buyer and seller visibility

$$X_D(H^S_D, H^B_D) = \left(H^B_D\right)^{\gamma_B}\left(H^S_D\right)^{\gamma_S} \tag{3.10}$$

$$X_I(H^S_I, H^B_I) = \left(H^B_I\right)^{\beta_B}\left(H^S_I\right)^{\beta_S}. \tag{3.11}$$

As in the simple model discussed in Section 2.2.3, the matching function exponents are key objects in the model. A positive externality to search would be indicated by high $\gamma_S, \gamma_B$ and $\beta_S, \beta_B$. This is because, at the margin, an increase in buyers or sellers visibility will lead to a large increase in the number of matches. There are increasing returns to scale in domestic matching if $\gamma_S + \gamma_B > 1$. By contrast, a congestion externality to search would be indicated by low $\gamma_S, \gamma_B$ and $\beta_S, \beta_B$, as more firms entering leads to very few new matches. There are decreasing returns to scale in domestic matching if $\gamma_S + \gamma_B < 1$. A

[6] Where as in the simple model $a \in [0, 1]$ and the amount of search internationally is $1 - a$.

low $\gamma_S$ would indicate that congestion is largely on the domestic supplier-side. Whereas, a low $\gamma_B$ would indicate that there is high congestion among domestic buyers. In Section 3.4, I structurally estimate the exponents using simulated method of moments.

The match flow per unit of buyer visibility $\theta$ is a measure of market tightness and is defined separately in the domestic and international markets, given by

$$\theta_D = \frac{X_D(H_D^S, H_D^B)}{H_D^B} \qquad \theta_I = \frac{X_I(H_I^S, H_I^B)}{H_D^B}. \tag{3.12}$$

A higher value of $\theta$ simply indicates that the hazard-rate of finding a match is higher.[7]

### 3.2.5. Search cost

In order to make sure that buyers do not enter a sorting equilibrium of only searching domestically or internationally, I assume positive and convex search costs[8] with a fixed cost of search $F_S$ and an additional fixed cost of international search $F_I$ only paid if the firm chooses to search internationally.

$$k^B = \left(\left(a\sigma^B\right)^v + \left((1-a)\sigma^B\right)^v\right)^v + F_S + F_I, \quad v > 1 \tag{3.13}$$

Fixed costs are common in the trade literature following Melitz (2003) as they ensure that high marginal cost firms only sourcing domestically. They represent the up-front costs firms pay in entering international trade (see for instance Antras et al. (2017). I structurally estimate $F_S, F_I$ in Section 3.4.

Sellers have a parallel set of search costs which are convex in the seller search intensity

$$k_L^S = \left(\sigma^S\right)^v, \qquad \text{for } L \in \{D, I\} \text{ and } v > 1, \tag{3.14}$$

which for simplicity are assumed to be the same for domestic and international suppliers.

### 3.2.6. Optimal search

Buyers solve the following maximization problem by picking their optimal search intensity $\sigma$ and the proportion of that search intensity in the domestic market $a$

[7] $\theta_{S_L}$ is defined symmetrically for $L \in \{D, I\}$ type suppliers.
[8] See Section 2.2.4 for further justification of this assumption.

$$V_i^B(\mathbf{s}) = \max_{a,\sigma^B}\left\{\frac{1}{A}\left(\pi_i^B(\mathbf{s}) - k^B(a,\sigma) + s_D\delta V_i^B(s_D-1,s_I) + a\sigma^B\theta_D^B V_i^B(s_D+1,s_I)\right.\right.$$
$$\left.\left. + s_I\delta V_i^B(s_I-1,s_D) + (1-a)\sigma^B\theta_I^B V_i^B(s_I+1,s_D)\right)\right\} \quad (3.15)$$

where $A = \rho + s_D\delta + s_I\delta + a\sigma^B\theta_D^B + (1-a)\sigma^B\theta_I^B$, $V_i^B(\mathbf{s})$ is the present value of a type$-i$ buyer that is matches with vector $\mathbf{s} \in \{s_I, s_D\}$ sellers, $\rho$ time preferences, $\delta$ is an exogenously given link death parameter.

Buyers receive profit equal to gross profit minus search costs, $(\pi_i^B(\mathbf{s}) - k^B(a_i, \sigma_i^B))$, until one of four events occurs with an endogenously given hazard: either (i) a buyer drops a domestic supplier $\left(V_i^B(s_D-1)\right)$, (ii) adds a domestic supplier $\left(V_i^B(s_D+1)\right)$, (iii) drops an international supplier $\left(V_i^B(s_I-1)\right)$, or (iv) adds an international supplier $\left(V_i^B(s_D+1)\right)$.

This yields policy functions for optimal search and the proportion of search in the domestic market where the change in cost of search is equal to the change in the value function from adding an additional domestic or international supplier multiplied by the hazard of these events occurring

$$\frac{\partial k^B(\sigma^B,a)}{\partial \sigma^B} \leq a\theta_D^B\Delta_{s_D}V_i^B + (1-a)\theta_I^B\Delta_{s_I}V_i^B \quad (3.16)$$

$$\frac{\partial k^B(\sigma^B,a)}{\partial a} \leq \sigma^B\theta_D^B\Delta_{s_D}V_i^B - \sigma^B\theta_I^B\Delta_{s_I}V_i^B \quad (3.17)$$

where $\Delta_{s_L}V_i^B = V_i^B(s_L+1) - V_i^B(s_L)$ for $L \in \{D, I\}$. Equation 3.16 and 3.17 hold with equality when a firm searches both internationally and domestically ($a < 1$).

Suppliers solve a parallel problem, where the value $V$ to any seller matching with a type-$i$ buyer who has $\mathbf{s}$ suppliers depends on their type $L$ and is given by

$$\begin{aligned} V_{D,i,s}^S &= \frac{r_i(\mathbf{s}) + (s_D-1)\delta V_{D,i,s_D-1}^S(s_D-1) + a_i\sigma_i^B\theta_D^B V_{D,i,s_D+1}^S}{\rho + s_D\delta + a_i\sigma_i^B(\mathbf{s})\theta_D^B} \\ V_{I,i,s}^S &= \frac{r_i(\mathbf{s}) + (s_I-1)\delta V_{I,i,s_I-1}^S(s_I-1) + (1-a_i)\sigma_i^B\theta_I^B V_{I,i,s_I+1}^S}{\rho + s_I\delta + (1-a_i)\sigma_i^B(\mathbf{s})\theta_I^B} \end{aligned} \quad (3.18)$$

Intuitively, the supplier gets revenue $r_i(\mathbf{s})$ as defined in equation 1.4, until they either lose a match with probability $(s_L-1)\delta$ or gain a match with probability depending on whether the supplier is domestic or international $a_i\sigma^B\theta_D^B, (1-a_i)\sigma_i^B\theta_I^B$. Taking expected value of a match is a summation over buyer types:

$$V_L^S = \sum_i \sum_{s=0}^{\infty} V_{L,i,s+1}^S P_i^B(\mathbf{s}), \qquad \text{for Ł} \in \{D, I\} \tag{3.19}$$

where $P_i^B(\mathbf{s}) = H_i^B(\mathbf{s})/H^B$ is the share of matches involving buyers of type$-i$ with $\mathbf{s}$ sellers.

Optimal seller search is then given by a parallel set of policy functions

$$\frac{\partial k_D^S(\sigma_D^S, s_D)}{\partial \sigma_D^S} = \theta_D^S V_D^S \tag{3.20}$$

$$\frac{\partial k_I^S(\sigma_I^S, s_I)}{\partial \sigma_I^S} = \theta_I^S V_I^S. \tag{3.21}$$

The optimal level of seller search is, therefore, the expected value of a new relationship multiplied by the probability of a match.

## EQUILIBRIUM

The model is completed via an equation of motion, where the change in the mass of buyers with $\mathbf{s}$ sellers is given by,

$$\begin{aligned}\dot{M}_i^B(s) = \Big[&\underbrace{a_i\sigma_i^B\theta_D^B M_i^B(s_D-1, s_I)}_{i} + \underbrace{\delta(s_D+1)M_i^B(s_D+1, s_I)}_{ii} + \underbrace{(1-a_i)\sigma_i^B\theta_I^B M_i^B(s_D, s_I-1)}_{iii} \\ &+ \underbrace{\delta(s_I+1)M_i^B(s_I+1, s_D)}_{iv}\Big] - \Big[\underbrace{a_i\sigma_i^B\theta_D^B}_{v} + \underbrace{\delta s_D}_{vi} + \underbrace{(1-a_i)\sigma_i^B\theta_I^B}_{vii} + \underbrace{\delta s_I}_{viii}\Big] M_i^B(s_D, s_I).\end{aligned} \tag{3.22}$$

Equation 3.22 shows the change in mass of type$-i$ buyers with $\mathbf{s}$ sellers is equal to flows in $(i{+}ii{+}iii{+}iv)$ minus flows out $(v{+}vi{+}vii{+}viii)$. Flows in is made up of the mass of type$-i$ buyers who have: $(i)$ $s_D - 1$ suppliers multiplied by the probability of adding a domestic supplier; $(ii)$ $s_D + 1$ suppliers multiplied by the probability of losing a domestic supplier; $(iii)$ $s_I - 1$ suppliers multiplied by the probability of adding a international supplier; $(iv)$ $s_I + 1$ suppliers multiplied by the probability of losing a international supplier. Flows out is made up of the mass of type$-i$ buyers who have $\mathbf{s}$ suppliers multiplied by the probability of: $(v)$ adding a domestic supplier; $(vi)$ losing a domestic supplier; $(vii)$ adding a international supplier; $(viii)$ losing a international supplier. Finally, the measure of buyers of type$-i$ with $s_L = 0$ is given by

$$\dot{M}_i^B(0,s_I) = \Big[\delta M_i^B(1,s_I) + (1-a_i)\sigma_i^B\theta_I^B M_i^B(0,s_I-1) + \delta(s_I+1)M_i^B(0,s_I)\Big] - \Big[a_i\sigma_i^B\theta_D^B + (1-a_i)\sigma_i^B\theta_I^B + \delta s_I\Big] M_i^B(0,s_I). \tag{3.23}$$

$$\dot{M}_i^B(s_D,0) = \Big[\delta M_i^B(s_D,0) + a_i\sigma_i^B\theta_D^B M_i^B(s_D-1,0) + \delta(s_D+1)M_i^B(s_D,0)\Big] - \Big[(1-a_i)\sigma_i^B\theta_I^B + a_i\sigma_i^B\theta_D^B + \delta s_D\Big] M_i^B(s_D,0). \tag{3.24}$$

A symmetric set of equations exists for suppliers.

As in EJTX (2016), I look for a stationary equilibrium at the steady state, I set $\dot{M}_i^B(s) = \dot{M}_j^S(n) = 0$ and solve the system of equations for all buyer types and suppliers given in equations 3.22, 3.23 and 3.24. I treat each buyer type as exogenously given.

## 3.3. Data

The data used in this chapter comes from the same source as presented in the previous two chapters. To see more detail refer to sections 3.3 and 3.3. There are a couple of ways in which I clean the data differently in order to estimate the model.

In the following section, I will estimate a transfer equation directly from the data. For this, I collapse the data into annual periods in order to be confident that I observe the firms full cycle of supplier relationships over an extended period. I also trim outliers, winsorizing at the 5% level.

For the estimation of the model, I collate a number of moments from the data. Each of these moments is collected prior to the policy change of the fall in transport costs taking place in 2011. I then collect another iteration of the moments taking an average from the periods after the policy change has taken place.

## 3.4. Estimation

Model estimation takes place in three steps: 1) Estimating the transfer equation to obtain elasticity of substitution parameters; 2) Externally calibrating parameters using the literature, and; 3) Structurally estimating the model using simulated method of moments.

Table 3.1. Estimating the transfer equation

| | (1) OLS-FE | (2) IV-FE | (3) OLS-FE |
|---|---|---|---|
| $\ln h_{j\|i,t}$ | 0.869*** | 0.957*** | |
| | (0.00373) | (0.00391) | |
| $\ln n_{it}$ | | | -0.300*** |
| | | | (0.0130) |
| Match FE | yes | yes | no |
| Buyer FE | no | no | yes |
| Importer FE | no | no | yes |
| Year FE | yes | yes | yes |
| N | 686170 | 686170 | 686170 |

*Notes:* Unit of observation is buyer $i$ supplier $j$ and year $t$. Standard errors in parentheses clustered at the buyer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

### 3.4.1. Estimating transfer equation

I follow EJTX (2016)'s methodology in estimating a transfer equation between buyers and suppliers in order to identify the elasticities of substitution between buyers. I estimate the structural equation 1.4 via Ordinary Least Squares (OLS). Equation 1.4 relates the revenue passed between buyers and suppliers ($r_{ji}$) to the within buyer$-i$ revenue share of seller $j$. When taking logs and adding time dummies ($d_t$) and a stochastic noise parameter ($\epsilon$), I can recover the coefficient on $\ln h_{j|i}$ which incorporates the elasticity of substitution between products ($\alpha$) and elasticity of substitution across buyers ($\eta$)

$$\ln r_{ji}(\mathbf{s}) = \frac{\alpha - \eta}{\alpha - 1} \ln h_{j|i} + 1 - \eta \ln \tilde{c}_{ji} + d_t + \epsilon_{jit} \tag{3.25}$$

where $r_{ji}$ is the revenue passed from buyer $i$ to supplier $j$ and $h_{j|i}$ is the within buyer-$i$ revenue share of seller $j$.

In order to address the term $\ln \tilde{c}_{ji}$, I include different fixed effects options. As in EJTX (2016), I address the concern that there is comovement in $\ln h_{j|i}$ and $\ln r_{ji}$, not driven by the components of the model, by using an instrument for $\ln h_{j|i}$ which is equal to a share-weighted average of the number of buyers of the *other* sellers at buyer $j$. The instrument should be correlated with $h$ through common shocks for similar products but should not influence revenue through any other channel.

I also run a separate model where I assume that all suppliers are identical except in allowing fixed effects to differ between import and domestic suppliers. In this case, I include just the log of the number of suppliers as the explanatory variable.

The first result from Table 3.1 is that the coefficient $\frac{\alpha-\eta}{\alpha-1} < 1$. Therefore, I conclude, as in EJTX (2016), that the elasticity of substitution across varieties ($\alpha$) exceeds the elasticity of substitution across buyers ($\eta$). Therefore, as shown in equation 1.3, there are decreasing returns to adding new suppliers.[9] Note that this is not to be confused with returns to scale in the matching function, which I estimate within the model. In column 2 of Table 3.1, I adopt the IV strategy and observe that the estimate increases but remains below 1.

Finally in column 3, I estimate the transfer equation where I assume all suppliers have the same marginal costs. Intuitively, for a given buyer adding another supplier lowers the revenue transferred to all other suppliers. As shown in Appendix equation 1.5, the coefficient on $\ln n$ is equal to $-\frac{\alpha-\eta}{\alpha-1}$. Intuitively, for a given buyer adding another supplier lowers the revenue transferred to other suppliers.

This gives a smaller coefficient than that in columns 1 and 2, but the result is still below 1 $\left(\frac{\alpha-\eta}{\alpha-1} = 0.3\right)$. Given the model's assumption that all suppliers have the same marginal cost, I use column 3 as my preferred specification.

### 3.4.2. Externally calibrated parameters

There are 8 parameters that are externally calibrated. The elasticity of substitution with respect to products $\alpha$ is set to 4.35 as in EJTX (2016). Using $\alpha = 4.35$, I can infer from column 3 of Table 3.1 that $\eta = 3.35$. This is coincidentally identical to the value estimated in EJTX (2016).[10] Firms' productivities are assumed to be Pareto distributed with shape parameter $\kappa = 4.25$ following Melitz and Redding (2015). The remaining parameters are adopted from the literature and are displayed in Table 3.2.

### 3.4.3. Internally calibrated parameters

I structurally estimate 7 key parameters of the model ($\xi = \{F_D, F_I, \psi_I, \gamma^S, \gamma^B, \beta^S, \beta^B\}$) using Simulated Method of Moments (SMM). This method selects the model parameters to minimize the difference between the simulated model generated moments and the moments in the data, by minimizing the following objective function

$$\hat{\zeta} = \text{argmin}_{\zeta}\mathscr{L}(\zeta) = \text{argmin}_{\zeta}\frac{1}{N}[M_m(\zeta) - M_d]'W_N\frac{1}{N}[M_m(\zeta) - M_d] \tag{3.26}$$

where $\zeta$ is a vector of moments to be targeted internally, $\mathscr{L}(\zeta)$ quadratic loss function to be minimized, $M_m(\zeta)$ vector of model moments, $M_d$ vector of corresponding data

[9] As discussed after equation 1.3, this condition ensures that there are diminishing returns to the number of suppliers, given that adding a new supplier appears in the summation $x \in J_b$ which leads to an increase in profit but at a decreasing rate, as long as the exponent $\frac{1-\eta}{1-\alpha} < 1$.

[10] EJTX (2016) use Colombian data finding a coefficient of -0.382 for rubber products and -0.289 for textiles. They take a middle point of these estimates to obtain -0.3 which works out as an $eta = 3.35$

Table 3.2. Model parameters

| | Externally Calibrated Parameter | Value | | Data source |
|---|---|---|---|---|
| $\alpha$ | Elasticity of sub. products | 4.35 | | Eaton et al. (2016) |
| $\eta$ | Elasticity of sub. buyers | 3.35 | | Estimated in transfer equation |
| $\Lambda$ | Bargaining coefficient | 0.5 | | Eaton et al. (2016) |
| $v$ | Convexity of search cost | 2 | | Eaton et al. (2016) |
| $\delta$ | Death parameter | 0.4 | | Calculated in data |
| $\tau$ | Iceberg trade cost | 1.45 | | Atkeson and Burstein (2008) |
| $\kappa$ | Pareto shape parameter | 1.45 | | Melitz and Redding (2015) |
| $\rho$ | Time preference | 0.05 | | Eaton et al. (2016) |
| | **Internally Calibrated Parameter** | **Value** | | **Most important moment** |
| $\psi_I$ | Import premium | 1.92 | (0.0211) | Ratio of imports to domestic among importers |
| $F_D$ | International fixed cost | 0.24 | (0.0061) | Prop of firms import |
| $F_I$ | Domestic fixed costs | 0.001 | (0.0001) | Number of active firms |
| $\gamma_B$ | D buyer matching CD share | 0.45 | (0.0093) | Prob. of a new match for dom. buyer |
| $\gamma_S$ | D supplier matching CD share | 0.50 | (0.0087) | Prob. of a new match for dom. supplier |
| $\beta_B$ | I buyer matching CD share | 0.60 | (0.0112) | Prob. of a new match for imp. buyer |
| $\beta_S$ | I supplier matching CD share | 0.66 | (0.0106) | Prob. of a new match for imp. supplier |

*Notes:* Standard errors in parentheses based on 25 bootstrapped samples drawn with replacement.

counterparts of the moments of interest, $M_m(\zeta) - M_d$ is the orthogonality condition and $W_N$ is a positive semi-definite weighting matrix which for simplicity is the identity matrix.

As shown in Table 3.3, I obtain 10 moments from the data using periods prior to the trade cost reduction. Intuitively, the proportion of buyers which are importers and the ratio of imports to domestic inputs among importers ties down the import premium and the import fixed cost. The mass of active firms ties down the domestic fixed cost. Each of the matching parameters are tied down by the combination of the probability of a new match for their type (domestic, international, buyer, supplier) and also the mass of active buyers and suppliers of their type in the population.

The results are given in Table 3.2. Importantly, I find that imports have a 1.92 times quality premium over domestic goods which is consistent with imported goods being of a higher standard. However, fixed costs of searching for imports are 240 times higher than the fixed cost of searching for domestic goods.

The most important parameters are the matching coefficients $\gamma$ and $\beta$. Consistent with the reduced form evidence, I find that there are decreasing returns to search in the domestic market ($\gamma^S + \gamma^B < 1$). By contrast, there are increasing returns to search in the international market ($\beta^S + \beta^B > 1$). In Section 3.5, I show numerically that this results in higher consumer welfare following a fall in transport costs.

### 3.4.4. Model fit

Table 3.3 compares the simulated model moments with their data counterparts, highlighting a close fit. The model also does well in matching untargeted moments. For example, as shown in the top two charts of Figure 3.1, the model's generated mass distribution of

Table 3.3. Model fit

| Moment | Model Value | Data Value |
|---|---|---|
| Ratio of imports to domestic among importers | 0.58 | 0.59 |
| Proportion of firms which import | 0.24 | 0.20 |
| Prob. of a new match for international suppliers | 0.20 | 0.28 |
| Prob. of a new match for domestic suppliers | 0.30 | 0.31 |
| Prob. of a new match for international buyer | 0.32 | 0.35 |
| Prob. of a new match for domestic buyer | 0.18 | 0.24 |
| Number of active international suppliers | 11,100 | 8,400 |
| Number of active domestic suppliers | 14,400 | 13,600 |
| Number of active international buyer | 5,700 | 4,800 |
| Number of active domestic buyer | 18,300 | 19,200 |

*Notes:* Table shows model generated moments and corresponding data moments. The ratio of imports to domestic among importers is calculated by dividing the total import value by the total value of inputs (imports + domestic goods). The proportion of firms which import is simply the proportion of buyers which imported in 2010 divided by the total number of buyers. The probability of a new match for an each type of buyer and supplier is calculated by seeing the proportion of firms which add a new match. The number of active firms is calculated as the number of firms in the dataset with positive sales in 2010.

buyers with different numbers of domestic and international suppliers closely matches its data counterpart.

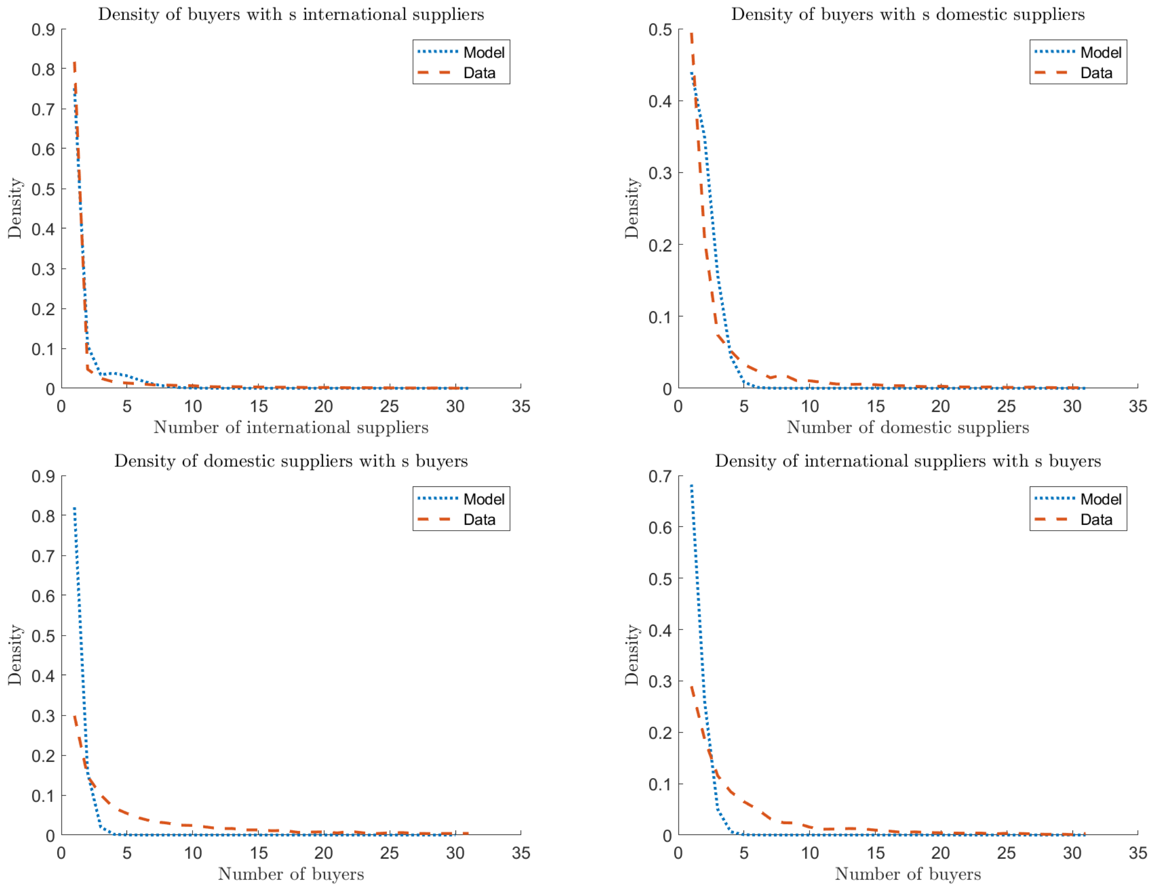
However, as shown in the bottom two charts of Figure 3.1, the model does less well in matching the distribution of supplier with different numbers of buyers. Although the shape of the distribution is similar, the model overestimates the density of suppliers with a small number of buyers. This is because the model has less flexibility on the supplier side relative to the buyer side given I assume all buyers have the same marginal costs. It is also consistent with fit of the quantitative model in Lim (2017) which also underpredicts the extent of connections of the most connected firms.

### 3.4.5. Heterogeneity

In addition to the model's aggregate predictions, it also demonstrates that firms behave differently depending on their marginal cost. In Figure 3.2, I group firms into marginal cost bins from 1 to 10 on the x-axis, and show the average level of search for each firm in each bin in international (red) and domestic markets (blue) on the y-axis. Due to the large fixed cost of importing, only the lowest marginal cost firms choose to search internationally. These firms also search domestically due to the convex costs to searching in each market.

Firms just below the threshold of paying the import fixed cost end up spending more on searching in the domestic market than the lower marginal cost firm, causing the peak

Figure 3.1. Model fit: buyer and supplier out-degree



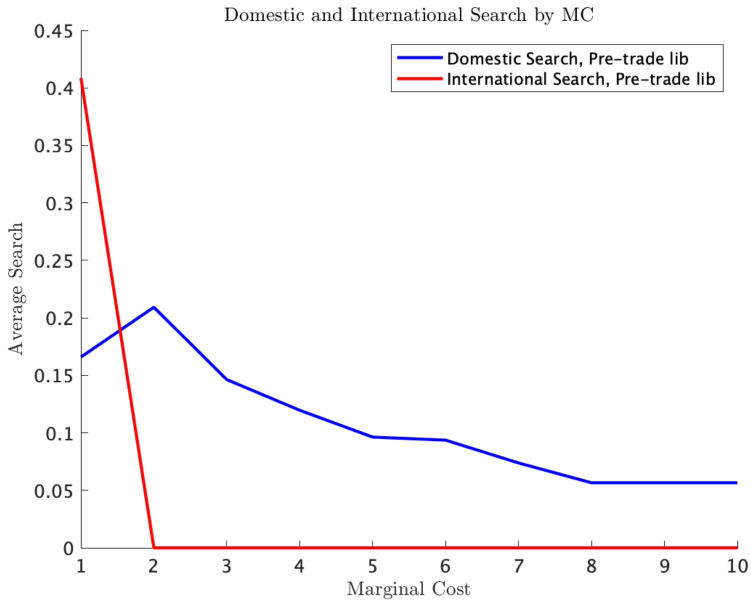

*Notes:* The top two figures shows the density of buyers with different numbers of international and domestic suppliers, respectively. the bottom two figures shows density of international and domestic suppliers with different numbers buyers. The blue lines show the model predicted density and the orange lines show the true value observed in the data.

in domestic search for firms in the second marginal cost bin. This is because, the lower marginal cost firms (in marginal cost group 1) have higher convex search costs given that they search both domestically and internationally. Following this peak, as marginal costs increase, firms spend progressively less on search given the diminishing marginal returns to adding new suppliers is more binding to firms with higher marginal cost.

## 3.5. Counterfactual simulations

I now test the external validity of the model by simulating a reduction in transport costs to match the observed reduction in East African trade costs shown in Section 2.3.2 in the previous Chapter. I then demonstrate the role of search externalities through two counterfactual experiments.

Figure 3.2. Search by marginal cost




*Notes:* The x-axis breaks buyers into 10 different marginal cost bins, where 1 indicates the lowest marginal costs and 10 equals the highest marginal costs. The y-axis shows the average search undertaken by buyers in each of these groups. The solid red and blue lines show the amount of domestic and international search, respectively, before the trade cost reduction.

### 3.5.1. Experiment 1: Transport cost reduction under structurally estimated parameters

As discussed in detail in Section 2.3.2, between 2010 and 2011, the cost to import a shipping container into Uganda fell rapidly by 25% driven by policy at the East African Community level.

Results from simulating this reduction in the model are shown in Table 3.4. The proportion of firms that import increases from 20% to 23%, as it becomes profitable for more firms to pay the fixed cost of importing. The average import search intensity increases by 21% and domestic search intensity decreases by 3%. The large increase in import search translates into a 20% increase in the average number of import suppliers.

The aggregate figures hide important heterogeneity which demonstrates the influence of search externalities. It also maps to the descriptive statistics shown in Section 2.3.2 and the comparative statics shown in 2.2.5. As shown in Figure 3.3, firms in the second marginal cost group become importers and existing importers increase their search leading to the average number of import suppliers increasing from 2.05 to 2.47. This directly maps to descriptive statistic (i): *as transport costs fall, imports increase.* As they do this, they are pushed up their convex search cost constraint and so reduce the amount they search

Table 3.4. Outcomes from 25% transport cost reduction

| Outcome | High $\tau$ | Low $\tau$ | Change | Data |
|---|---|---|---|---|
| Percentage of Importers | 20.01 | 23.05 | 15.2% | 16% |
| Av. Import Suppliers | 2.05 | 2.47 | 20.1% | 19% |
| Av. Domestic Suppliers | 2.70 | 2.52 | -6.5% | -1.6% |
| Domestic Search ($a\sigma$) | 0.119 | 0.115 | -3.14% | |
| Import Search ($(1-a)\sigma$) | 0.704 | 0.851 | 20.88% | |
| Consumer Welfare | | | 5.2% | |

*Notes:* Table shows the model generated outcome variables under the high and low trade cost equilibriums and the percentage change. This is compared to the observed percentage change in the real data. Average refers to the average number of suppliers over all firms.

domestically (domestic search for marginal cost bin 2 firms decreases from 0.21 to 0.14). This maps to descriptive statistic (ii): *new importers drop domestic suppliers.* This then increases market tightness in the international market and reduces market tightness in the domestic market. Consequently, higher marginal cost firms, which do not import, increase their domestic search as the probability of finding a domestic match increases (average search for firms in marginal cost bin 3 increases from 0.15 to 0.18). This maps to descriptive statistic (iii) *dropped domestic suppliers re-match with non-importing buyers.*

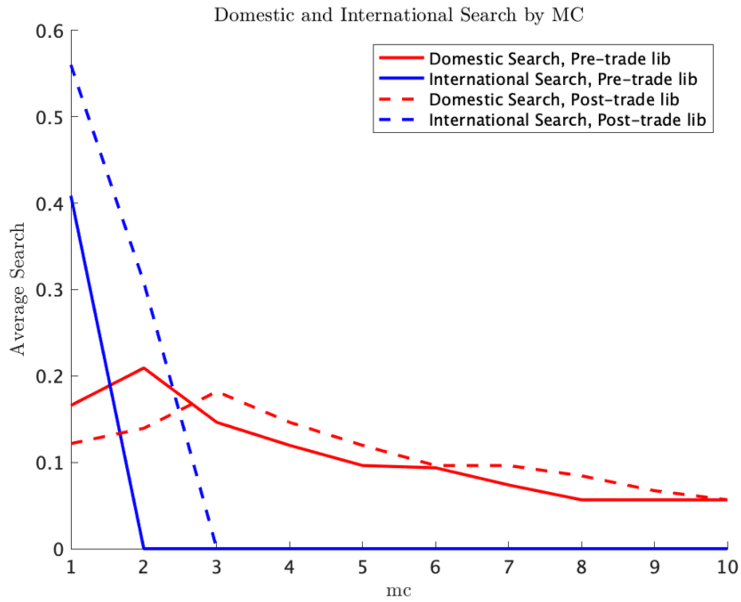
Table 3.4 also reports the observed changes in firm outcomes as seen in the Government of Uganda tax data. The observed change is the same direction and of a similar magnitude to that seen in the simulation. The main disparity is in domestic suppliers, where the reduction is overestimated by the model. This is because there was growth in the domestic economy outside of the influence of the trade cost reduction. As the results from the trade cost reduction were not used in the parametrization of the model, the fit to the observed shift provides external validity to the model.

Figures 3.4 and 3.5 provide more detail on the change in the distribution of firm size. The trade cost reduction lead to an increase in the number of international suppliers for firms of all sizes. The biggest shift, however, comes at the tails of the distribution where the number of firms with greater than 15 suppliers increases by 1.7%. There is also a shift in the number of medium-sized importers as the proportion of firms which import increases by 16%.

Finally, the model shows that a 25% transport cost reduction led to a 5.2% increase in consumer welfare. As shown in Section 2.2.5, this is due to: i) the lower marginal cost of importing having an income effect, and ii) the increase in matching efficiency from moving to the increasing returns to scale international market.[11]

[11] An extension would consider the short and long-run effects from the intervention. In the short-run, the model predicts that the reallocation of search towards the international market frees up space in the domestic market given domestic suppliers can now re-match. However, in the long-run these firms may no longer be profitable causing firm exit and reversing the gains from a reduction in domestic market

Figure 3.3. Search by marginal cost




*Notes:* The x-axis breaks buyers into 10 different marginal cost bins, where 1 indicates the lowest marginal costs and 10 equals the highest marginal costs. The y-axis shows the average search undertaken by buyers in each of these groups. The solid red and blue lines show the amount of domestic and international search, respectively, before the trade cost reduction. The red and blue dashed lines show the amount of domestic and international search, respectively, after the trade cost reduction.

Figure 3.4. Mass of firms with $S_I$ international suppliers

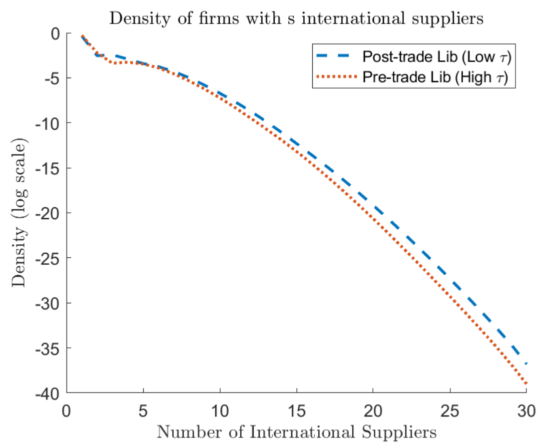
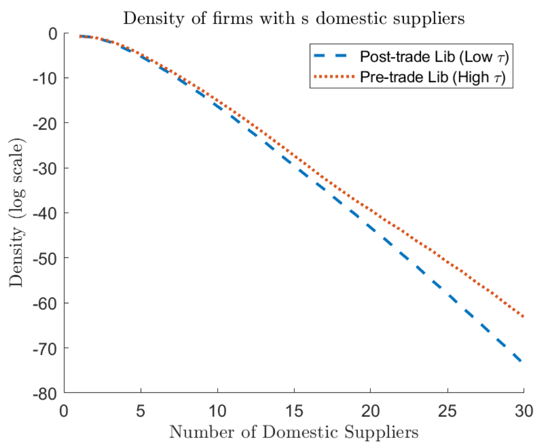



Figure 3.5. Mass of firms with $S_D$ domestic suppliers




*Notes:* Figures show model predictions on the density of buyers with different number of suppliers before and after the trade cost fall. the left hand panel shows the density of buyers with $s_I$ international suppliers and the right hand panel shows the density of buyers with $s_D$ domestic suppliers. The orange line shows the density prior to the trade cost fall and the blue line shows the density after the trade cost fall.

tightness. This could be incorporated into the model with a fixed-cost on suppliers.

### 3.5.2. Experiment 2: Transport cost reduction under constant returns to scale matching function

The second counterfactual experiment tests how much search externalities influence consumer welfare. I shut down the difference in search externalities between markets by assuming that both markets have the same constant returns to scale matching function.

Table 3.5 compares the results of the second experiment to those with structurally estimated matching parameters. When both matching functions are constant returns to scale, the most obvious difference between the two experiments is the smaller magnitude by which the average number of import suppliers increases (11.1% vs. 20.1%). This is due to the import market becoming tighter, making it relatively harder for firms to match for each unit of search.

Domestic search also decreases in the CRS experiment. This leads to a larger reduction in the average number of domestic suppliers suppliers (-9.8% vs. -6.5%). This is because the reduction in search domestically does not have the mitigating effect of reducing congestion in the domestic search market.

Figure 3.6 shows the average number of suppliers for buyers on the y-axis, and different trade cost reductions on the x-axis. This is plotted for both the case of different search externalities (IRS) and where both matching functions are constant returns to scale (CRS). Figure 3.6 shows that for larger trade cost reductions, the difference in the predicted number of suppliers diverges. For a 10% reduction in search costs the average number of international suppliers increases by 2.4% in the increasing returns to scale simulation and 1.7% in the constant returns to scale model. Whereas for a 25% reduction in search costs the average number of international suppliers increases by 20% in the increasing returns to scale simulation and 11% in the constant returns to scale model, a larger difference. This non-linearity in the model is due to the non-linearity in the two matching function - as more firms switch into the increasing returns to scale sector from the decreasing returns to scale sector there is an increasingly large impact on matching efficiency.

This non-linearity is also shown in Figure 3.7, where consumer welfare is increasing as trade costs fall, and is increasing more rapidly in the simulation which allows for different externalities. A 25% reduction in trade costs results in a 15% larger increase in consumer welfare in the simulation with different search externalities, compared to the simulation with the same externalities in both markets.

### 3.5.3. Experiment 3: Search cost reduction

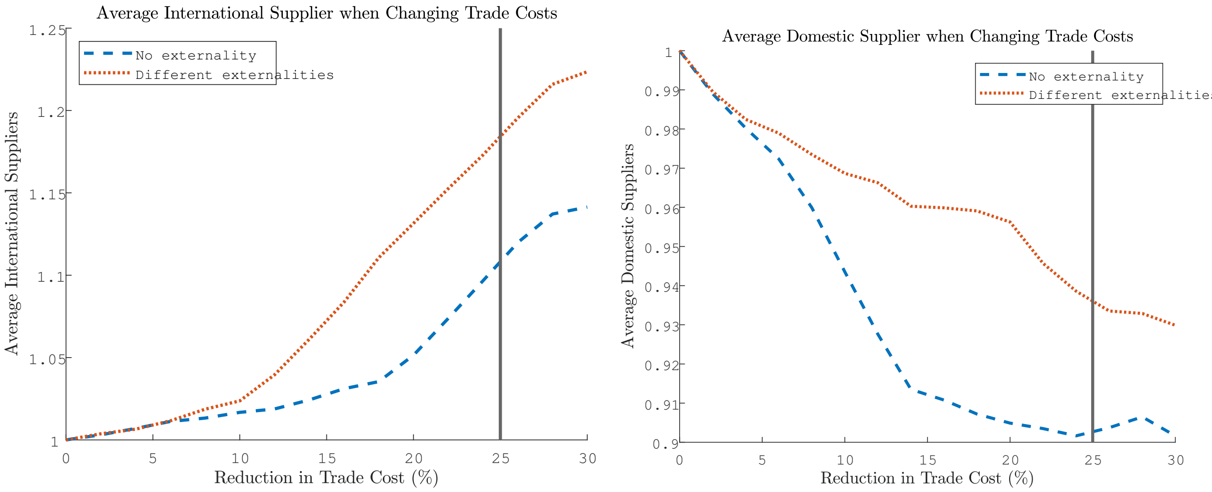
In experiment 3, I simulate the Ugandan government's stated target for 2019 to reduce search costs for suppliers by 25% (Government of Uganda, 2019). The specific sub targets are i) establishing a internet platform support programme (e.g. organize quarterly

Table 3.5. Outcomes from 25% transport cost reduction under different matching functions

| Outcome | Change IRS | Change CRS | Real Change |
|---|---|---|---|
| Percentage of Importers | 15.20% | 12.77% | 16% |
| Av. Import Suppliers | 20.1% | 11.10% | 19% |
| Av. Domestic Suppliers | -6.5% | -9.77% | -1.6% |
| Domestic Search ($a\sigma$) | -3.14% | -5.82% | |
| Import Search ($(1-a)\sigma$) | 20.88% | 17.65% | |
| Consumer Welfare | 5.2% | 4.4% | |

*Notes:* Table shows the change in the model generated outcome variables under the model estimated parameters on the matching function which allow different externalities between both markets (IRS), under the case where the matching function is assumed to be constant returns to scale for both markets (CRS), and the observed change in the data. Average refers to the average number of suppliers over all firms.

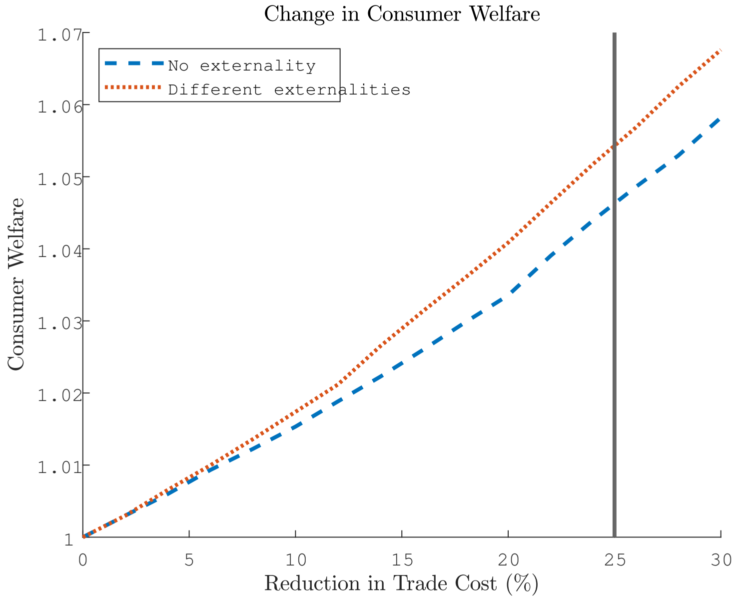
Figure 3.6. Average number of international and domestic suppliers for different reductions in trade costs when search externalities are shut down (CRS) compared to structurally estimated parameters (IRS)




*Notes:* The y-axis shows the change in the average number of suppliers where the baseline is normalized to 1. The x-axis shows the reduction in trade costs from 0 to 30%. The orange line (IRS) shows the change in the average number of suppliers when the model is estimated using the structurally estimated parameters which allows for increasing returns to scale in matching internationally and decreasing returns to scale in matching domestically. The blue line (CRS) shows the change in the average number of suppliers when the model is estimated shutting down differences in the returns to scale in matching between domestic and international markets.

trainings on the use of Ali Baba), ii) encourage firms peer-to-peer learning (e.g. organize quarterly peer groups with Uganda business groups), iii) target key firms in supplier development programmes (e.g. establish anchor firm support unit and annual public-supplier meetings). Intervention (ii) mimics the work done by the Chinese government and documented by Cai and Szeidl (2017), where firms which meet regularly for business meetings have been shown to increase the number of clients by 12% and the number of

Figure 3.7. Consumer welfare gains from trade when search externalities are shut down (CRS) compared to structurally estimated parameters (IRS)




*Notes:* The y-axis shows the change in consumer welfare where the baseline is normalized to 1. The x-axis shows the reduction in trade costs from 0 to 30%. The orange line (IRS) shows the change in the average number of suppliers when the model is estimated using the structurally estimated parameters which allows for increasing returns to scale in matching internationally and decreasing returns to scale in matching domestically. The blue line (CRS) shows the change in the average number of suppliers when the model is estimated shutting down differences in the returns to scale in matching between domestic and international markets.

Table 3.6. Outcomes from 25% search cost reduction

| Outcome | Change following 25% decrease in domestic search costs | Change following 25% decrease in import search costs |
|---|---|---|
| Percentage of Importers | -0.48% | 10.16% |
| Av. Import Suppliers | -0.74% | 35.1% |
| Av. Domestic Suppliers | 10.02% | -4.54% |
| Domestic Search ($a\sigma$) | 9.93% | -1.62% |
| Import Search ($(1-a)\sigma$) | -0.97% | 40.57% |
| Consumer Welfare | 3.4% | 4.3% |

*Notes:* Table shows the change in the model generated outcome variables under a 25% decrease in domestic search costs and a 25% decrease in international search costs. Average refers to the average number of suppliers over all firms.

suppliers by 9%.

The idea behind this experiment is to consider whether the government's stated target would improve firm outcomes and where the search cost reduction would be best targeted. In order to consider this question, I run two separate counterfactual experiments - first lowering the domestic search costs and then the import search costs.

Figure 3.8. Search by marginal cost if reduce domestic search costs by 25%

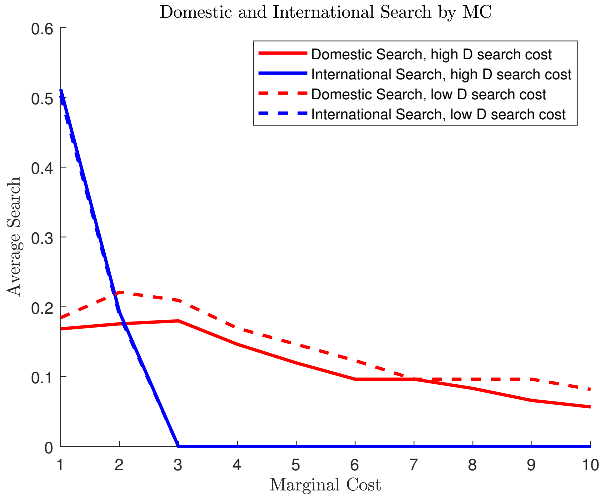
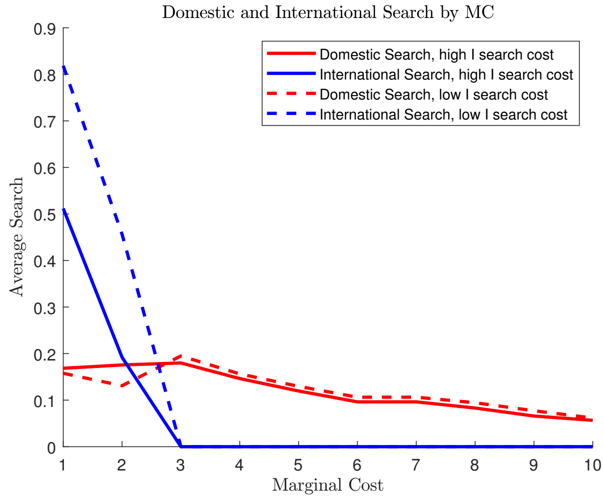



Figure 3.9. Search by marginal cost if reduce international search costs by 25%




*Notes:* The x-axis breaks buyers into 10 different marginal cost bins, where 1 indicates the lowest marginal costs and 10 equals the highest marginal costs. The y-axis shows the average search undertaken by buyers in each of these groups. The solid blue and red lines show the amount of domestic and international search, respectively, before the reduction in search costs. The blue and red dashed lines show the amount of domestic and international search, respectively, after the search cost reduction. The left graph shows the impact for reducing domestic search costs. The right graph shows the impact from reducing international search costs.

The outcomes from the experiment are given in Table 3.6 and Figure 3.8. When reducing domestic search costs, there is a sharp increase in buyers' domestic search and consequently the average number of domestic suppliers increases by 10%. This is of a similar magnitude to the 9% increase in suppliers found in Cai and Szeidl (2017) following the business-meeting intervention. As can be observed in Figure 3.8, this increase in domestic search is observed across all levels of buyer marginal cost. However, the increase in the number of domestic matches is relatively modest (10%), as the increase in domestic search leads to an increase in domestic market congestion. There is also a small decline in international search (-1%), as firms make a substitution decision away from international markets.

As shown in Table 3.6 and Figure 3.9, when reducing international search, there is a large increase in import search (40.6%) leading to a 35% increase in import suppliers. As can be observed in Figure 3.9, this is concentrated among the low marginal cost firms, as for all other firms they still do not choose to pay the import fixed cost. These firms, reduce the amount they search domestically, given they are still subject to a convex cost of searching in both markets. This then frees up space in the domestic market, captured by higher marginal cost firms. Therefore, the second experiment acts in a similar way to the trade cost reduction in leading to welfare gains through both the lower marginal costs and the benefit of moving from the decreasing returns to scale market to the increasing returns to scale market. As a consequence, reducing international search costs increases consumer welfare by 4.3%.

By contrast, when domestic search costs fall, firms increase domestic search, however, this leads to a large increase in domestic market tightness due to the domestic congestion. Therefore, the impact of the reform is muted.

These results provide support for the government of Uganda's policy of lowering search costs as the impact on welfare is of a similar magnitude to lowering international trade costs by 25%. The results show that the impact of the reforms will be greater if the government focusses on lowering international search costs. Therefore, the government may focus on their planned interventions to train firms on using platforms such as Ali Baba and Amazon and by having firms meet with firms who have experience of importing in a similar vein to Cai and Szeidl (2017).

## 3.6. Concluding remarks

In this chapter I have provided evidence consistent with the domestic Ugandan market being more congested than the foreign import market. I then show that this has important welfare benefits when considering opening to international trade. I then demonstrate that lowering search costs would have a substantial impact on welfare, especially if targeted towards reducing international search costs.

While, the estimates in this chapter are specific to the Ugandan context, however, the mechanisms are general to any setting which has search frictions between buyers and suppliers. There is reason to believe that the relative size of the effects maybe larger in a low-income country setting where search frictions are substantial, although, this is speculative without obtaining similar data in a different setting. This does suggest a channel for future work.

The results in this chapter provide support for policy intervention to address search frictions. As is the case with all search frictions, the first-best outcome would be to remove the search friction entirely. In the context of the model presented in this chapter, this would mean all firms finding and matching with suppliers costlessly. In practice this is not feasible, instead, governments can focus on reducing search costs. The Ugandan government's goal of providing training on platforms such as Ali Baba and Amazon to Ugandan businesses will have a large impact as these channels directly target lowering international search costs. Similarly, encouraging firms to learn from each other has been shown in other contexts to improve firm-to-firm matching (Cai and Szeidl, 2017). Results from this chapter suggest the Ugandan government should focus on interventions that target reducing the cost to international search as opposed to domestic search. This is because lowering the cost to domestic search may simply increase congestion leading to a small increase in matches. However, lowering international search costs will increase both international matches and reduce domestic congestion.

# Appendix A

# Appendix to Chapter 1

## 1. Empirical appendix

### A.1. Hierarchical firm clustering algorithm

We first present the algorithm and then discuss it's implementation in Uganda.

#### Description of technology algorithm

The Hierarchical Firm Clustering Algorithm aims to identify firms providing the same inputs. It achieves this by using the Leontief production and panel dataset to infer when two inputs are likely to be the same.

This is best explained with a simple example as shown in Figure 1.4. In this example, to produce 1 unit of cement (red) you require 1 unit of limestone (green) and 1.2 units of gypsum (yellow). The Leontief assumption insures that limestone can not be substituted for gypsum.

Suppose, in period 1, we observe a cement manufacturer using inputs from two firms in the required ratio, 1 and 1.2. Then, in period 2, we observe the firm using inputs from three firms with ratio 1, 0.6 and 0.6. Observe, that if inputs from firms 2 and 3 are combined then we return to the same input ratio, 1 and 1.2. Therefore, we can infer from the firms repeated sourcing patterns that firms 2 and 3 are likely to be producing the same input.

We can use this simple example to build a generalisable algorithm to identify which input goods are likely to be the same. The algorithm has six steps.

1. Take a firm. Initially assume all its inputs are in one cluster.

2. Define a loss function for firm $i$ of making a single partition of an input cluster, where $P_{min} \in P$ is the partition that minimises the loss function from all possible single partitions. Define $e$ as an element in partition $e \in P_{min}$. Then loss function

$L$ given by,

$$L_i = \frac{1}{|e|}\frac{1}{|t|}\sum_{t,e\in P_{min}}\left(\frac{(X_{et}-\bar{X}_e)-\frac{\bar{X}}{\bar{Y}}(Y_{it}-\bar{Y})}{\bar{Y}}\right)^2 \tag{1.1}$$

where $X$ denotes input value and $Y$ denotes firm sales. Note the loss function is quadratic, causing partitions that violate Leontieff constant proportions (over time) to generate greater losses.

3. Calculate a proportionate loss from making the new partition

$$\epsilon_{il} = \epsilon_{i,l-1}[1 - L_i] \tag{1.2}$$

4. Define cut-off $c > \epsilon_{il}$ for whether to stop the algorithm at layer $l$. We choose $c = 0.3$.

5. If proportionate loss greater than cut-off $\epsilon_{il} > c$ or number of inputs = number of partitions, then go to next firm, if not then repeat algorithm for all new possible partitions.

### Results of algorithm in Uganda

We implement the HAC algorithm on data from Uganda. To give an indication of the algorithm's performance we compare results of inputs the algorithm groups together to the ISIC sector the firm reports to the revenue authority. In total, we find 28% of HAC clusters are in the same ISIC 4 digit. However, we also compared many more which did not fall into these categories but may well be providing the same input. For instance,

- 5224 - Cargo handling
  5320 - Courier activities

- 4220 - Construction of utility projects
  4100 - Construction of buildings

- 4663 - Wholesale of construction materials, hardware, plumbing and heating equipment and supplies
  4100 - Construction of buildings

When we compare a broader grouping, we find that 43% of HAC clusters are in the same ISIC 2 digit industry.

## A.2. Technology identification algorithm

In this section, we first lay out the algorithm in detail and then present an example of its implementation on a simple production network. The algorithm is organised into five steps.

### Description of technology algorithm

Step 1. Begin with firm 1 at the top of partial ordering $i \in \{1, ..., N\}$ and consider all $j \in \{1, ..., K\}$ inputs which are in units of the real value of sales. We must start at the top of the partial ordering as we need to transform each layer in order, so that a firm which is two layers below the final product can still have its transactions classified in units of the final consumption good.

Step 2. Identify which inputs are of the same good. This can be done using the HAC algorithm discussed in Appendix A.1, or by using the firm's ISIC classification. This step is necessary because we make the assumption that all firms producing the same good use the same production technology. Therefore, the same input should be transformed using a consistent technology.

Step 3. Regress output for firm $i$ on sum of inputs in each partition $p$, generate $\tau$ under the condition that $\tau_p > 1$ for all partitions $p$. This condition is necessary as otherwise firms would be losing value on inputs and so the assumption of conservation of flow would be violated.

$$TotalSales_{it} = \sum_{p} \tau_p Input_{pt} + u_{it} \tag{1.3}$$

Step 4. Multiply edges by $\tau_p$.

Step 5. Repeat from step ii for next firm in partial ordering until $i = N$

### Stylized example of algorithm

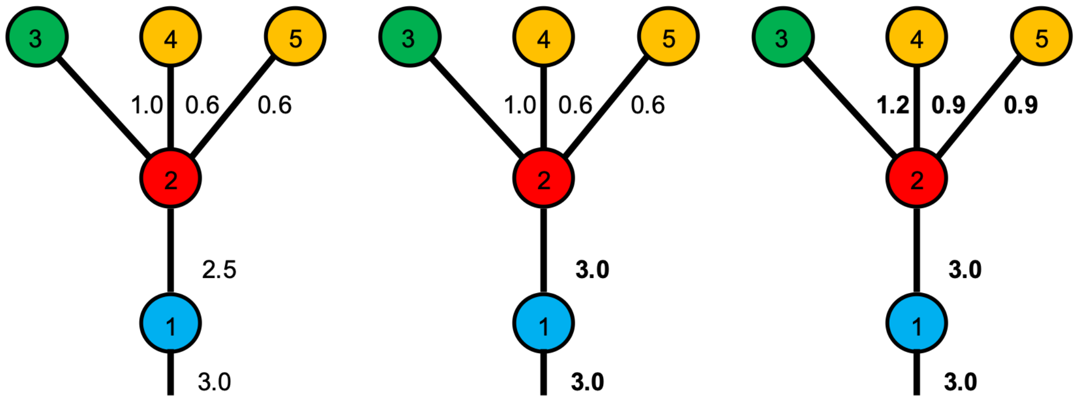
To demonstrate the algorithm, we now present how it would be performed on a very simple supply network as shown in Figure A.1. As in the HAC example, I will again assume we observe a cement chain. As shown in Figure A.1, there is one cement retailer (blue), one cement manufacturer (red), one limestone mine (green), and two gypsum input providers (yellow), although, we assume each firms product types are not observable to the researcher. The edges of the graph show the value of trades in one period, although assume the researcher in fact observes multiple periods. The numbers inside the nodes show the firm's position in the partial ordering as given by the Eades et al. (1993) FAS algorithm.

In Step 1, we begin with the firm at the top of the partial ordering - the cement retailer (blue) and observe it only has one input supplier. Given it has one supplier, we know it only has one input type, completing Step 2. In Step 3, we must regress output on inputs to obtain $\tau_{12}$.

$$TotalSales_{1t} = \tau_{12} Input_{1t} + u_{1t} \tag{1.4}$$

where we find $\tau_{12} = 1.2$.

Figure A.1. Hypothetical cement chain before and after algorithm




*Notes:* The Figure shows a cement supply chain. Each node represents a firm and colours represent different firm types: retailer (blue), manufacturer (red), input 1 (green) and input 2 (yellow). The amount transacted is given on the edge. The firms position in the partial ordering is given inside the node. Panel 1 shows the initial raw transaction values in one period. Panel 2 shows the values on the edges once the value between firm 1 and 2 has been transformed into units of the final consumption good. Panel 3 shows the edge values in units of the final consumption good, once the technology algorithm has been completed.

In Step 4, we then transform the input by $2.5 * \tau_{12} = 3.0$. This is shown in the second panel of Figure A.1. You can observe that flow is now conserved along the chain between firms 1 and 2.

Step 5 tells us to repeat the exercise with the next firm in the partial ordering - the cement manufacturer (red). This firm has three input suppliers. Using the HAC algorithm, we identify firms 4 and 5 to be of the same input type. We now run the following regression to obtain technology parameters over the two inputs,

$$TotalSales_{2t} = \tau_{23} Input_{1t} + \tau_{24} Input_{2t} + u_{2t} \tag{1.5}$$

where we find $\tau_{23} = 1.2$ and $\tau_{24} = 1.8$. We then transform the inputs by the technology parameters to transform the edge value into units of the final consumption good. This is shown in the third panel of Figure A.1. Now observe that flow is conserved along each layer of the graph.

## A.3. Identifying if there are no upstream bottlenecks

Calculate No Upstream Bottlenecks (NBU) variable using following algorithm.

1. Take list of all nodes and select node $i$ (initialize counter at $i = 1$).
2. Check if current node is in NBU. If so update counter at $i = i + 1$ and go to 1. If not go to 3.

3. Take DAG-R and do a depth-first search of DAG-R, starting at current node. We obtain the upstream subgraph of node $i$.

4. Check if there are any bottleneck nodes in the current upstream subgraph. If yes, label current node $i$ as belonging to the set Uncompetitive Upstream (UU) (update UU vector with $i$ id). If not, label current node $j$ as belonging to set NBU (same)

5. Update counter to $i+1$ and go to step 1.

## 2. Theoretical Appendix

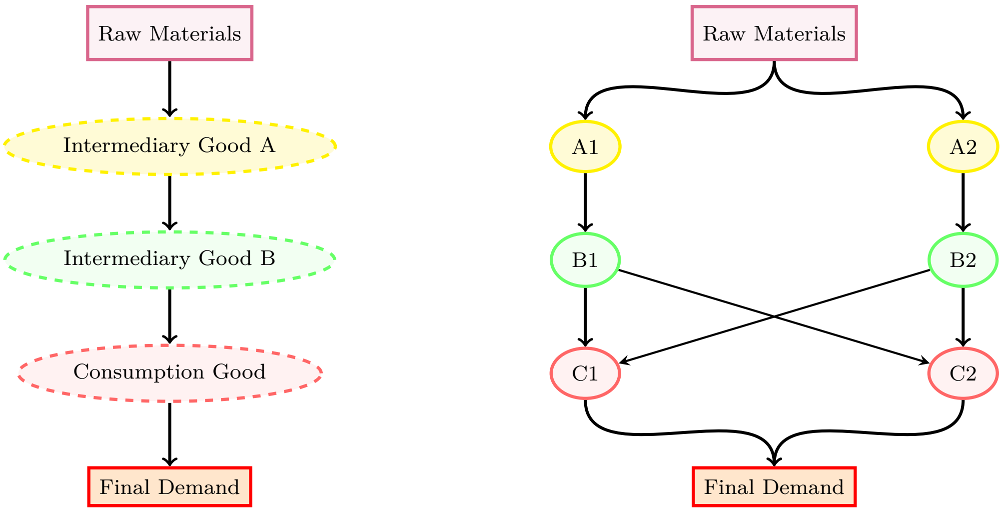
### B.1. Examples

**Example 3.** Consider the supply-network shown in Figure A.2. There are two intermediary goods and one consumption good and two producers of each. Suppose that each producer has capacity 1, and production technologies all require one unit of input to produce one unit of output. Finally, let the cost of extracting one unit of the raw material be 1, and there are no processing costs. Consider then the possibility of an equilibrium in which all firm set prices of 1, and suppose that at a price of 1 demand for good $C$ is 1. There are many ways in which demand at these prices might be met, but for simplicity suppose that each firm produces 1/2 a unit. Now consider whether each firm has a profitable deviation to set a different price. At a lower price each firm would make a loss, so we can restrict attention to higher prices.

Suppose then that $B1$ deviated and set its price above 1. At this higher price, condition (iv) guarantees that the market cannot clear with $B1$ receiving positive demand, so the deviation is not profitable for $B1$. The same is true for $B2$. However, the argument does not work for $A1$ or $A2$. If, for example, $A1$ increased its price above 1, $B1$ has no alternative supplier and so cannot switch suppliers. So, without condition (v), there is nothing to stop the market clearing with $A1$ making positive sales. However, with condition (v), $C1$ and $C2$ will choose to route all their demand through $A2$. Although $B1$ and $B2$ charge the same price, $B1$'s costs are higher. This means that $A1$ will not make any sales and the deviation will not be profitable. Of course, in practice, prices throughout the supply chain would adjust dynamically. Condition (v) provides a simple means for capturing this force while retaining the simplicity of firms simultaneously setting prices for their output.[1]

[1] An alternative that would work for this example is that firms could simultaneously choose markups instead of prices and condition (v) could be dropped from the market clearing condition. Unfortunately this approach gets complicated in general because of the possibility of firms sourcing the same type of input at different prices from different suppliers.

Figure A.2. Market clearing example




*Notes:* The left figure shows the underlying texhnology DAG—product A uses 1 unit of raw materials to make 1 unit of B, which makes one unit of C. The right figure shows the full supply-network. For instance, B1 can obtain one unit from A1, C1 can source from either B1 or B2 with a maximum capacity of 1 from either node, but can only sell to final demand a maximum capacity of 1. On each edge the maximum capacity is 1.

## B.2. Omitted proofs

### Proof of lemma 1

*Proof.* Towards a contradiction suppose that all final goods are also used as an intermediate good. Consider now the subgraph induced by the producer nodes (so ignoring raw materials and end consumers). In this subgraph all nodes have a strictly positive out-degree, and thus there exists a cycle. This contradicts the assumption that the supply network is a DAG. We therefore conclude that at least some final goods are not also intermediate goods. □

### Proof of corollary 1

*Proof.* Towards a contradiction suppose that all goods use intermediate goods as inputs. Consider now the subgraph induced by the producer nodes (so ignoring raw materials and end consumers). In this subgraph all nodes have a strictly positive in-degree, and thus there exists a cycle. This contradicts the assumption that the supply network is a DAG. We therefore conclude that at least some goods are produced using only raw materials. □

## Proof of theorem 1

*Proof.* Throughout this proof we map back and forth between the flow problem and market clearing. The mapping we use is the following. We map flows into demands and supplies by setting $D_{ji} = S_{ij} = f_{ij}$ for all $i, j \in N$ and $S_{ic} = \sum_\theta f_{i\theta} = f_{i\theta'}$, for $i \in G(\theta')$. And we map supplies $S$ into flows by setting flows between raw materials, producers and consumers as follows $f_{ij} = S_{ij}$ for all $i, j \in N \cup C$, setting flows into the sink as follows $f_{\theta t} = \sum_{i \in N} f_{i\theta} = \sum_{i \in N} S_{i\theta}$ for all $\theta \in F$ and setting flows from the source into raw materials by $f_{sr} = \sum_{i \in N} f_{ri} = \sum_{i \in N} S_{ri}$ for all $r \in R$. We refer to these mappings as, respectively, the demands and supplies induced by given flows, and the flows induced by given supplies.

We now provide a first formal link between the maximum flow problem and market clearing with a Lemma.

**Lemma 2.**

*(i) If, given prices $p$, the system is demand-constrained then supplies $S$ and demands $D$ induced by any maximum flow satisfy market clearing conditions (i)-(iii).*

*(ii) If prices, demands and supplies $(p, D, S)$ clear the market, then the flow induced by these demands and supplies achieves a maximum flow.*

*Proof.* **Part (i):** Given a weighted directed flow graph $G$ with finite weights (i.e., a directed weighted graph with a source, a sink and a path from the source to the sink), let $\mathcal{F}$ denote the set of maximum flows. This set is always non-empty and compact and so there exists a flow $\hat{f} \in \arg\min_{f \in \mathcal{F}} \sum_j f_{ij}$. For our supply network $W$ we do this exercise to find such a flow $\hat{f}$, and then use this to induce demands $D$ and supplies $S$. We will show that if, given prices $p$, $W$ is demand constrained, then these demands and supplies with prices $p$ satisfy conditions (i)-(iii) in the market clearing algorithm.

First, as $\hat{f}$ is a maximum flow in a demand constrained system, $f_{\theta t} = D_{c\theta}(p)$ for all final goods $\theta \in F$. Thus, consumer demand at prices $p$ is satisfied. Moreover, as $\hat{f} \in \arg\min_{f \in \mathcal{F}} \sum_j f_{ij}$ the conservation of flow constraints through each node must bind—were the flow to diminish through a node we'd have $\hat{f} \notin \arg\min_{f \in \mathcal{F}} \sum_j f_{ij}$, a contradiction. The binding conservation of flow constraints immediately imply that condition (i) holds. Condition (ii) holds by the capacity constraints in the flow problem, and condition (iii) holds by construction of demands and supplies.

**Part (ii):** By construction there is a cut of the flow network severing just the artificial links denoted $\theta$ for each good from the sink (i.e., all links into the sink). The maximum flow through the network must be weakly less than the sum of these capacities by the max flow min cut theorem.[2]

As prices, demands and supplies $(p, D, S)$ clear the market, consumer demand at prices $p$ is satisfied. Setting $(f_{ij})_{ij} = (S_{ij})_{ij}$ for all $i, j$, this implies that $f_{\theta t} = D_{c\theta}(p)$. As the flow is equal to the capacities, the overall flow must achieve the upper bound we established, and hence $f$ is a maximum flow. □

We now prove the theorem:

*If:* We need to show that if there is no bottleneck firm, there is an equilibrium with marginal cost pricing. Suppose all firms set prices equal to their marginal costs. We need the market to clear, and no firm to have a profitable deviation.

As there is no bottleneck firm, no firm can be part of a minimum cut. This implies that the minimum cut must be the links $(w_{\theta t})_\theta$. But this means that $f_{\theta t} = w_{\theta t}$ and by construction $w_{\theta t} = D_{c\theta}(p)$. Hence the system is demand constrained. Thus, by Lemma 2(i), there exist demands and supplies that satisfy conditions (i)-(iii). Further, as all firms are pricing at marginal cost and there is a unique technology for making each good, demands and supplies satisfying conditions (i)-(iii) also trivially satisfy conditions (iv) and (v). Thus, the market clears.

Consider now the possibility that a firm $k$ does not best respond to others pricing at marginal cost by also pricing at marginal cost. As pricing below marginal cost cannot be more profitable, consider firm $k$ setting a price above its marginal cost. As firm $k$ is not a bottleneck, a maximum flow such that $f_{ki} = 0$ for all $i$ exists. Take any such flows $(\hat{f}_{ij})_{i,j}$ and consider demands $D$ and supplies $S$ induced by this flow. By the earlier argument, these flows clear the market (satisfy conditions (i)-(v)). As it is possible to clear the market, demands and supplies will be selected that do clear the market. We, instead, show now that demands and supplies derived from flows with $f_{ki} > 0$ for some $i$ cannot clear the market.

There are two possibilities. Suppose first that the representative consumer purchases directly from $k$ (so $f_{k\theta} > 0$). However, as $k$ is not a bottleneck and supply through $k$ is positive, the consumer must have another supplier of this product that can supply more (i.e., is not supply-constrained). Moreover, as only firm $k$ deviated, this alternative supplier is pricing at marginal cost and hence market clearing condition (iv) is violated.

The second possibility is that the representative consumer does not purchase directly from $k$ (so $f_{k\theta} = 0$), but $f_{ki} > 0$ for some $i$. Nevertheless, the representative consumer must be purchasing indirectly from $k$ as demand equals supply throughout—$k$'s supply is positive and all the demand $k$ faces is induced by the representative consumer's demand. Moreover, as firm $k$ is not a bottleneck, there exists an alternative flow that excludes

[2] This is a well known theorem in computer science which states the maximum flow from a source to a sink equals the minimum cut to the network to completely stop flow (Ford and Fulkerson, 2015).

$k$. Setting demands and supplies equal to this alternative flow, the total transacted cost of the representative consumer's purchases must decline (as all other firms set prices at marginal cost). Hence, condition (v) is violated.

Thus, we have shown that after the deviation, $k$ will make no sales and the deviation was unprofitable.

*Only if:* This is a corollary of Proposition 2 below.

□

## Proof of Proposition 1

*Proof.* We can substitute out the last constraint in the planner's problem (i.e., the constraint requiring that consumption equals supply) simplifying the planner's problem to

$$\max_S U(S)$$

subject to

(i) Resource constraint $\sum_i \sum_k S_{ik}\kappa_i \leq \omega$

(ii) Leontief production constraints and capacity constraints are satisfied

Suppose supplies $\hat{S}$ satisfy the conditions that make them a competitive outcome. Thus, $\sum_{j\in G(\theta)} \hat{S}_{jc} = D_{c\theta}(\gamma^*)$ and supplies $\hat{S}$ are feasible and non-wasteful. We begin by showing that supplies $\hat{S}$ satisfy all the planner's constraints. First, as these supplies are feasible, the Leontief production constraints and capacity constraints are satisfied.

Second, by the DAG structure of the supply network, each demand-level 1 good (i.e., final goods) can be thought of as a bundle of demand-level 2 goods, each demand-level 2 good can be thought of as a bundle of demand-level 3 goods, and so on. Tracing back these demand-levels, the total costs incurred to produce one unit of a final good $\theta \in F$ are given by $\gamma^*$. Thus, as supplies $\hat{S}$ are non-wasteful and each firm $j \in G(\theta)$ sources only as much of a good $\theta'$ as it requires (i.e., $\sum_{i\in G(\theta')} \hat{S}_{ij} = A_{\theta\theta'} \sum_k \hat{S}_{jk},$) the total production costs associated with supplies $\hat{S}$ are $\sum_i \sum_k \hat{S}_{ik}\kappa_i = \sum_{\theta\in F} \gamma_\theta \sum_{i\in G(\theta)} \hat{S}_{ic}$. As the outcome is competitive, $\sum_{i\in G(\theta)} \hat{S}_{ic} = D_{c\theta}(\gamma^*)$, and as these demands solve the consumer's problem given prices $\gamma^*$, $\sum_{\theta\in F} D_{c\theta}(\gamma^*)\gamma_\theta = \omega$. Combining these equations we have $\sum_i \sum_k \hat{S}_{ik}\kappa_i = \omega$, and so the resource constraint holds with equality.

We have shown that supplies $\hat{S}$ satisfy the constraints of the planner's problem. We now show that they solve it. Towards a contradiction suppose there are supplies $\tilde{S}$ satisfying conditions (i)-(ii) above such that $U(\tilde{S}) > U(\hat{S})$. At prices $\gamma^*$ the solution to the representative consumer's problem is given by demands $D_{c\theta}(\gamma^*) = \sum_{i\in G(\theta)} \hat{S}_{ic}$. Thus, as supplies $\tilde{S}$ are feasible (satisfy condition (ii)), these supplies must have been unaffordable—$\sum_\theta \gamma^*_\theta \sum_{i\in G(\theta)} \tilde{S}_{ic} > \omega$. But this implies that they violate the resource constraint (as $\gamma^*_\theta$ is the total production of producing a unit of good $\theta$). Hence supplies $\tilde{S}$ violate condition (i)—a contradiction. □

## Proof of Proposition 2

*Proof.* As $(\mathbf{p}, D, S)$ is an equilibrium, the market clears. Thus, final consumer demand is satisfied and the system is demand-constrained and this determines the maximum flow. Towards a contradiction, suppose $i$ is a bottleneck and $p_i = \kappa_i + \sum_j p_j \frac{S_{ji}}{\sum_k S_{ik}}$. This implies that $i$ makes 0 profits. Suppose that instead $i$ were to charge a price $p'_i = \kappa_i + \sum_j p_j \frac{S_{ji}}{\sum_k S_{ik}} + \epsilon$. We will show that there exists a $\bar{\epsilon} > 0$ such that for all $\epsilon < \bar{\epsilon}$ $i$ makes strictly positive profits. Let $\mathbf{p}' := (\mathbf{p}_{-i}, p'_i)$ be the vector of prices $\mathbf{p}$ but with $i$'s price changed to $p'$.

First suppose that for all $\epsilon > 0$, at prices $\mathbf{p}'(\epsilon)$ we have demand $D'(\epsilon)$ and supplies $S'(\epsilon)$. If $S'_{ic}(\epsilon) > 0$, then $i$ makes positive sales to final consumers and the deviation is profitable.

We can therefore restrict attention to the case in which $S'_{ic}(\epsilon) = 0$ for all $\epsilon > 0$. We'll show that this implies aggregate demand is unaffected (i.e., $\sum_k D_{ck} = \sum_k D'_{ck}$). As aggregate consumer demand ($\sum_\theta D_{c\theta}$) is decreasing in the marginal price the representative consumer faces, aggregate demand can only fall if the marginal price the representative consumer faces increases. Thus, for demand to be affected, the marginal price must be $p'_i(\epsilon)$ (as all other prices are the same) and, further, we must have $S'_{ic}(0) > 0$. As aggregate consumer demand ($\sum_\theta D_{c\theta}$) is continuous in the marginal price the representative consumer faces, there then exists a $\bar{\epsilon}$ such that for all $\epsilon < \bar{\epsilon}$, $S'_{ic}(\epsilon) > 0$—but we have already dealt with this case. Thus, if $S'_{ic}(\epsilon) = 0$, then $\sum_k D_{ck} = \sum_k D'_{ck}$.

Then, as supply constraints are unaffected and final consumer demand is the same, the system must remain demand constrained at prices $\mathbf{p}'$. There are two cases to consider.

Case 1: Suppose there exist demands $\hat{D}$ and supplies $\hat{S}$ that clear the market at prices $\mathbf{p}'$ (satisfy market clearing condition (i)-(v)), then the selected demands $D'$ and supplies $S'$ must also clear the market. As at these demands and supplies consumer demand would be satisfied, so setting $f_{ij} = S'_{ij} = D'_{ji}$ for all $i, j$, the flows $(f_{ij})_{ij}$ must constitute a maximum flow. But as firm $i$ is a bottleneck, we must then have $S'_{ij} > 0$ for some $j$, and the deviation is again profitable.

Case 2: Suppose there do not exist demands $\hat{D}$ and supplies $\hat{S}$ that clear the market at prices $\mathbf{p}'$. However, as $(\mathbf{p}, D, S)$ is an equilibrium, we know that $D$ and $S$ clear the market. Thus, demands $D$ and $S$ can also be selected at prices $\mathbf{p}'$, and the tuple $(\mathbf{p}', D, S)$ will satisfy market clearing conditions (i)-(iii). This implies that demands $D'$ and supplies $S'$ must also satisfy clearing conditions (i)-(iii). This implies that setting $f_{ij} = S'_{ij} = D'_{ji}$

for all $i, j$, the flows $(f_{ij})_{ij}$ must constitute a maximum flow. But as firm $i$ is a bottleneck, we must then have $S'_{ij} > 0$ for some $j$, and the deviation is again profitable. $\square$

## Proof of Proposition 3

*Proof.* Let $\bar{\mathbf{p}}$ be the competitive equilibrium prices. Then, $\mathbf{p} \geq \bar{\mathbf{p}}$ element wise. Thus, as final demand for each final good is independent of final demand for other goods, demand is weakly higher for all products when prices are competitive. Thus, the only change in the supply network is that the capacity of the edges that terminate at the sink (the demand edges) weakly increase. If a firm $i$ was a bottleneck before these changes, then $f(\hat{G}(\mathbf{p})) > f(\hat{G}(\mathbf{p}) - i)$. Thus, there is at least one good, call it $k$, where demand at prices $\mathbf{p}$ is greater than can be supplied once $i$ is removed from the network. Suppose prices are now competitive. As demand for this product is weakly higher at prices $\bar{\mathbf{p}}$, demand for this product still cannot be satisfied once $i$ is removed from the network, and more can be supplied with $i$ than without, and so $i$ is still a bottleneck firm. $\square$

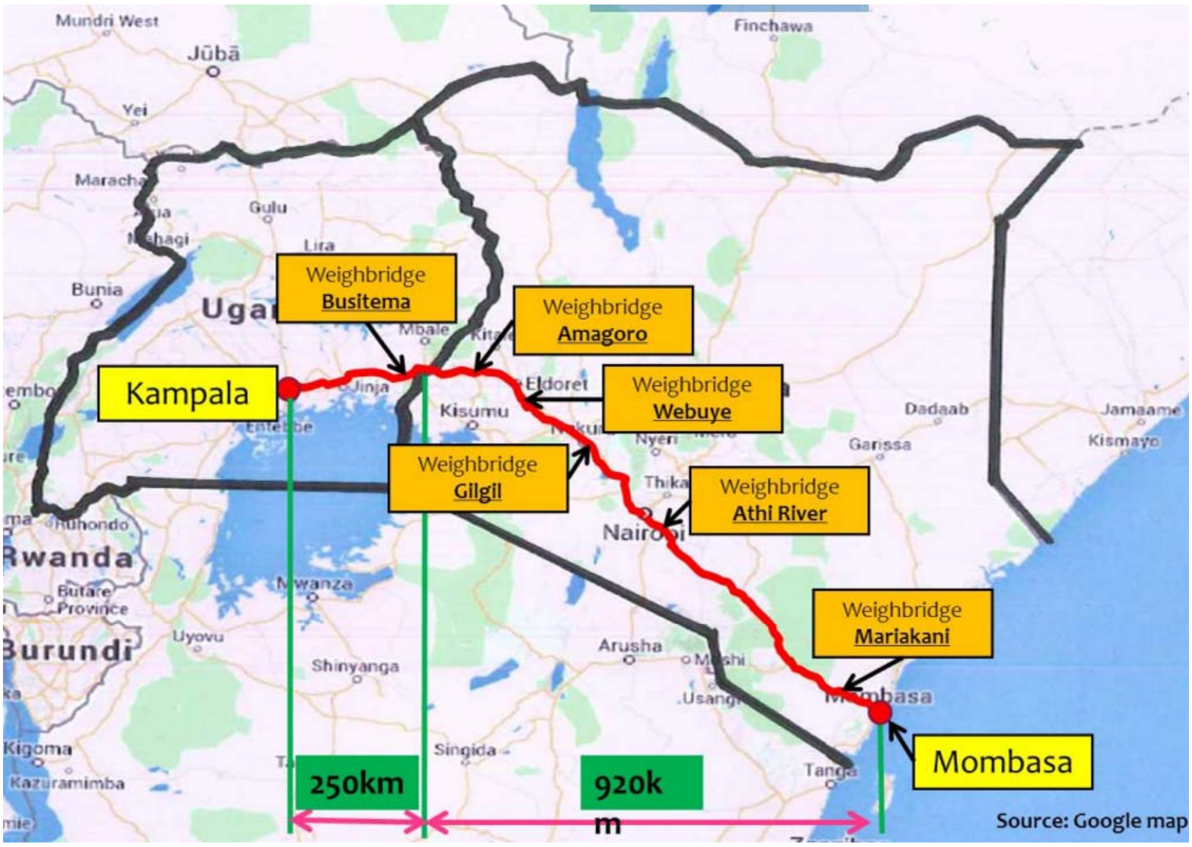
# Appendix B

# Appendix to Chapter 2

## 1. Context appendix

### A.1. Map of trade corridor

Figure B.1. Map of trade corridor




Osawa, WCO


Figure B.2. Exports data comparison

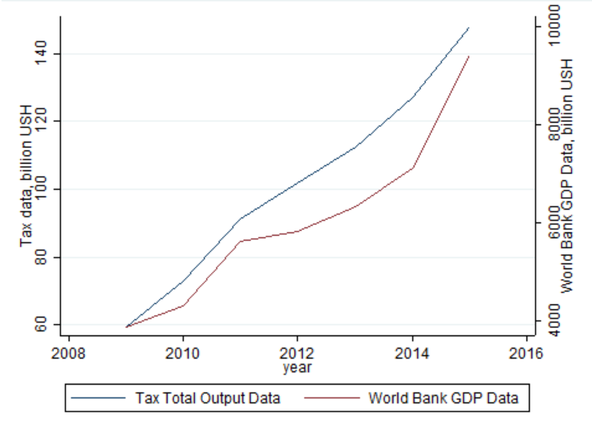



Figure B.3. GDP and total output

*Notes:* The left-hand figure compares the Uganda Revenue Authority (URA) export data with data obtained from the World Trade Organization. The right-hand figure compares total output data from the URA's tax data with GDP data from the World Bank.

## A.2. Data comparison

Given research using tax data remains rare, one potential concern might be that the data is of low quality. This section addresses this concern by comparing the tax data used in this study to other freely available data sources.

Figure B.2 shows a comparison between the raw export trade data used in this study and trade data from the WTO. From the graph it appears as if the WTO data is understating the actual export volumes. However, for the purposes of this study, the important fact is how closely the two lines track one another showing that the data is strongly correlated with the external source.

Figure B.3 shows a comparison between the total output variable used in the tax data and GDP data from the World Bank. Unsurprisingly, the tax data is smaller than the GDP data given the tax data only observes formal sector firms. Importantly, like in B.2, the correlation between the two lines is very strong again supporting the reliability of the tax data.

Finally, Spray and Wolf (2016) show the distribution of firms in each sector is consistent with those in the Uganda Business Census.

# 2. Mathematical appendix

## B.1. Comparative statics in the two-period model

The buyer picks their optimal $a$ in order to solve the following maximization problem

### Reallocation

$$\pi_b = \frac{E}{\eta P^{1-\eta}}\left(\frac{\eta}{\eta-1}\right)^{1-\alpha}\tau_I \tilde{c}_{xb}^{1-\alpha} \tag{2.1}$$

$$\max_a \left\{a\sigma\theta_D\pi(s_D) + (1-a)\sigma\theta_I\pi(s_I) - k(a)\right\}. \tag{2.2}$$

This yields a first order condition

$$\sigma\theta_D\pi_{s_D} - \sigma\theta_I\pi_{s_I} - \frac{\partial k}{\partial a} = f(a, \tau_I) = 0. \tag{2.3}$$

Totally differentiating 2.3 and rearranging yields the comparative static of how $a$ changes as $\tau$ changes

$$\frac{\partial f}{\partial a}\frac{\partial a}{\partial \tau_I} + \frac{\partial f}{\partial \tau_I} \implies \frac{\partial a}{\partial \tau_I} = -\frac{\frac{\partial f}{\partial \tau_I}}{\frac{\partial f}{\partial a}}. \tag{2.4}$$

Solving for each of these terms separately gives an explicit solution,

$$\frac{\partial a}{\partial \tau_I} = \frac{-\sigma\theta_I\frac{\partial \pi_i^B(s_I)}{\partial \tau_I}}{\frac{\partial^2 k}{\partial a^2} - \sigma\frac{\partial \theta_D}{\partial a}\pi_i^B(s_D) + \sigma\frac{\partial \theta_I}{\partial a}\pi_i^B(s_I)} \tag{2.5}$$

$$\frac{\partial a}{\partial \tau_I} = \frac{-\sigma\theta_I\frac{\partial \pi_i^B(s_I)}{\partial \tau_I}}{\frac{\partial^2 k}{\partial a^2} - \sigma^2(\gamma^B - 1)\theta_D B_D\pi(s_D) - \sigma^2(\beta^B - 1)\theta_I B_I\pi(s_I)} \tag{2.6}$$

### Matching efficiency

Consumer Welfare is broken into matching efficiency $A$ and consumption $C$.

$$\begin{aligned} W(a) &= \left[\int_{b\in B(s_I)}\psi_I C_b^{\frac{\eta-1}{\eta}} + \int_{b\in B(s_D)} C_b^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}} \\ &= \underbrace{\Big[a\sigma\theta_D + \psi_I(1-a)\sigma\theta_I\Big]}_{A}\underbrace{\left[\int_{b\in B} C_b^{\frac{\eta-1}{\eta}}\right]^{\frac{\eta}{\eta-1}}}_{C}. \end{aligned} \tag{2.7}$$

Rewriting the matching efficiency $A$ by expanding the market tightness yields the following equation,

$$A = a^{\gamma_B} S^{\gamma_S} B^{\gamma_B - 1} + \psi_I (1-a)^{\beta_B} S^{\beta_S} B^{\beta_B - 1}. \tag{2.8}$$

taking a partial derivative of $A$

$$\begin{aligned} \frac{\partial A}{\partial \tau_I} =& \gamma_B a^{\gamma_B - 1} S^{\gamma_S} B^{\gamma_B - 1} \frac{\partial a}{\partial \tau_I} - \beta_B \psi_I (1-a)^{\beta_B - 1} S^{\beta_S} B^{\beta_B - 1} \frac{\partial a}{\partial \tau_I} \\ & \frac{\partial a}{\partial \tau_I} \left[ \gamma_B a^{\gamma_B - 1} S^{\gamma_S} B^{\gamma_B - 1} - \beta_B \psi_I (1-a)^{\beta_B - 1} S^{\beta_S} B^{\beta_B - 1} \right] \end{aligned} \tag{2.9}$$

The first term $> 0$ as shown in equation 2.6, the second term determines the direction of the effect

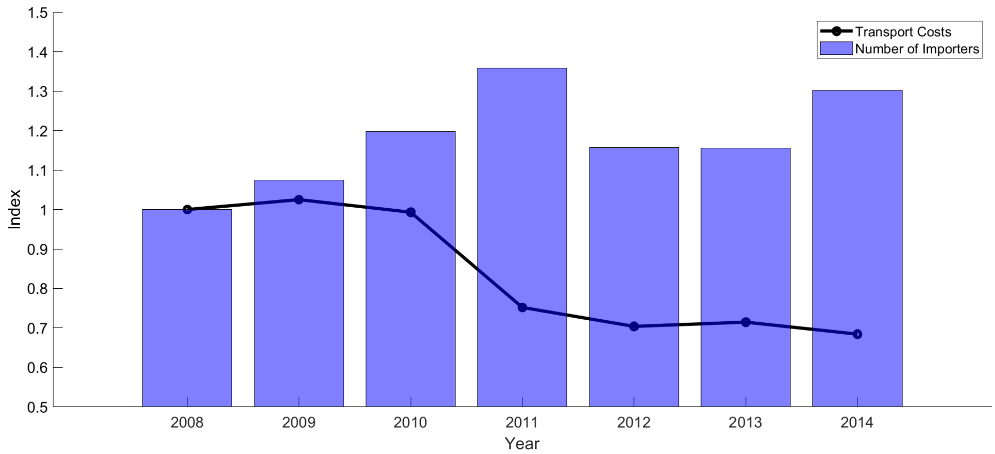
$$\begin{aligned} \frac{\partial A}{\partial \tau_I} < 0 \iff & \gamma_B a^{\gamma_B - 1} S^{\gamma_S} B^{\gamma_B - 1} < \beta_B \psi_I (1-a)^{\beta_B - 1} S^{\beta_S} B^{\beta_B - 1} \\ & \gamma_B a^{\gamma_B - 1} < \beta_B \psi_I (1-a)^{\beta_B - 1} S^{\beta_S - \gamma_S} B^{\beta_B - \gamma_B} \end{aligned} \tag{2.10}$$

Therefore, the change in welfare due to matching efficiency following a fall in trade costs depends on $a, \psi_I$ and the matching exponents $\gamma_B, \gamma_S, \beta_B, \beta_S$. The main takeaway from equation 2.10 is that for $a$ sufficiently large and $\psi \geq 1$, the change in welfare due to matching depends on the relative size of the matching exponents. If $\gamma_B < \beta_B$ and $\gamma_S < \beta_S$ i.e. returns to search are higher in the international market, then an increase in trade cost will lower welfare given firms move from matching in the increasing returns to scale international market to the decreasing returns to scale domestic market.

# 3. Empirical appendix

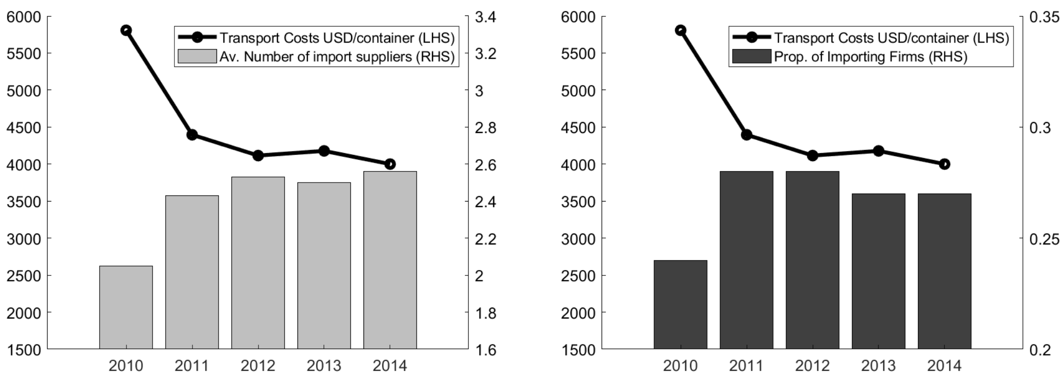
## C.1. Descriptive statistics

Figure B.4. Transport costs and number of importers




*Notes:* The black line shows transport cost in USD per 20-foot container from the World Bank's Trading Across Border Index between 2007-2014, the bars contains data on the total number of importers. The data for comes from customs dataset. Reforms took place between 2010 and 2011.

Figure B.5. Transport costs and imports




*Notes:* The black line shows transport cost in USD per 20-foot container from the World Bank's Trading Across Border Index, light grey bars on the left-hand graph show the average number of import suppliers for importers, and dark grey bars on the right-hand graph show the proportion of firms which import. The reason for the shorter time series is that I do not know the identity of import suppliers prior to 2010.

Table B.1. Newly added domestic suppliers among new importers

| | (1) Number of domestic suppliers | (2) Number of domestic suppliers |
|---|---|---|
| First Time Import in $2011_i \times 2011_t$ | -0.167*** | -0.104*** |
| | (0.0243) | (0.0244) |
| First Time Import in $2011_i$ | 0.712*** | |
| | (0.00981) | |
| Observations | 162190 | 162190 |
| Year FE | YES | YES |
| Buyer FE | NO | YES |

*Notes:* Unit of observation is buyer $i$ and year $t$. Standard errors in parentheses clustered at the buyer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.2. Dropped suppliers' new matches

| | (1) Proportion of buyers don't import | (2) Proportion of buyers don't import |
|---|---|---|
| Dropped by 2011 first time $\text{importer}_f \times 2011_t$ | 0.0370*** | 0.0404*** |
| | (0.00821) | (0.00608) |
| Dropped by 2011 first time $\text{importer}_f$ | 0.0267*** | |
| | (0.00329) | |
| Observations | 96470 | 96470 |
| Year FE | YES | YES |
| Buyer FE | NO | YES |

*Notes:* Unit of observation is supplier $f$ and year $t$. Standard errors in parentheses clustered at the buyer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.3. Value from domestic suppliers among new importers

| | (1) Value of domestic suppliers | (2) Value of domestic suppliers |
|---|---|---|
| First Time Import in $2011_i \times 2011_t$ | -0.354*** | -0.304*** |
| | (0.0572) | (0.0381) |
| First Time Import in $2011_i$ | 1.658*** | |
| | (0.0230) | |
| Observations | 160138 | 108380 |
| Year FE | YES | YES |
| Buyer FE | NO | YES |

*Notes:* Unit of observation is buyer $i$ and year $t$. Standard errors in parentheses clustered at the buyer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.4. Dropped suppliers' new matches - value

| | (1) Proportion of value from buyers which don't import | (2) Proportion of value from buyers which don't import |
|---|---|---|
| Dropped by 2011 first time importer$_f$ $\times$ 2011$_t$ | 0.0292*** | 0.00631 |
| | (0.00697) | (0.00486) |
| Dropped by 2011 first time importer$_f$ | 0.00292 | |
| | (0.00360) | |
| Observations | 96103 | 84908 |
| Year FE | YES | YES |
| Buyer FE | NO | YES |

*Notes:* Unit of observation is supplier $f$ and year $t$. Standard errors in parentheses clustered at the buyer level. $^*$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Table B.5. Same and next-door balance table

| **Distance** | **Proportion of firms in same sector** | **Difference with same building** |
|---|---|---|
| Same building | 0.097 | |
| | (0.296) | |
| Next-door building | 0.088 | -0.009 |
| | (0.284) | (0.014) |
| Next-door building $< distance < 0.1km$ | 0.060 | -0.037*** |
| | (0.237) | (0.012) |
| $0.1km < distance < 0.15km$ | 0.051 | -0.046*** |
| | (0.219) | (0.017) |
| $0.15km < distance < 0.2km$ | 0.044 | -0.053** |
| | (0.204) | (0.021) |
| $0.2km < distance < 0.25km$ | 0.040 | -0.057** |
| | (0.196) | (0.026) |

## C.2. Robustness tests

### Very local shocks drive results

To address the concern that shocks drive reduced form results, I look at the proportion of firms in the same building which are in the same ISIC 4-digit sector and compare that to the proportion of firms in the next-door building. Results are shown in Table B.5. While there is a small difference, it is not statistically significant. However, when I look at firms further away, I do see this difference increasing. I therefore conclude that there is some firm agglomeration, but that it is happening at a block level and not at a building level.

### Spillover exists but is not search related

A second alternative explanation is that a spillover is taking place, but that it is not search related. For instance, we might expect that transport costs could be driving the results. To allay these concerns, I test if the marginal effect is smaller among firms where one would expect search frictions to be less prevalent. To test for this, I interact the independent variables with whether the import supplier exported from the East African Community (EAC). This is because one would expect search frictions to be smaller in

Table B.6. Imports suppliers from East African Community

| | (1) |
|---|---|
| | $Y_{ift}$ |
| $X_{t-1}^{same}$ | 0.0931*** |
| | (0.00665) |
| $X_{t-1}^{same} \times EAC_f$ | -0.0346** |
| | (0.0151) |
| $X_{t-1}^{other-city}$ | -0.00223 |
| | (0.00176) |
| $X_{t-1}^{other-city} \times EAC_f$ | -0.00486 |
| | (0.00552) |
| Observations | 4834635 |

*Notes:* Unit of observation is buyer $i$, supplier $f$ and year $t$. $Y_{ift}$ indicates a first match took place between buyer and supplier. $X_{ift}^{k}$ is a count of buyers in region $k$ which added supplier $f$ in $t-1$. $EAC_f$ indicates the supplier operates in the East African Community. Coefficients are multiplied by 100 to read as percentage point marginal effects. Standard errors in parentheses clustered at the buyer level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

local neighbors like Kenya or Tanzania when compared to more distant locations. We would therefore expect when estimating equation 3.1 that the positive search externality for EAC suppliers is weaker ($\mu_2 < 0$).

$$Y_{ift} = \mu_1 X_{if,t-1}^{same} + \mu_2 X_{if,t-1}^{same} \times EAC_f + \gamma_1 X_{if,t-1}^{other-city} + \gamma_2 X_{if,t-1}^{other-city} \times EAC_f + \alpha_f + \alpha_i + \alpha_t + u_i \tag{3.1}$$

Results shown in Table B.6 confirm that suppliers in the EAC have a smaller positive spillover. This is again consistent with a narrative in which search is driving results.

Another prediction consistent with search frictions, is that suppliers which are not supply-constrained will be able to match with multiple buyers, and so we should not observe a negative congestion effect.

As discussed in Section 2.3.2, this is the reason why we did not expect to find a strong congestion externality on foreign imports, given international suppliers are characterized by being large firms with cheap access to credit and multiple customers. By contrast, domestic Ugandan firms are characterized by being small with limited access to credit. You might therefore expect that Ugandan firms cannot make multiple matches in a given period, thus making the domestic market more congested.

If this is indeed the case, I would expect domestic Ugandan suppliers which are also exporters to act in a similar way to foreign exporters, as they are less likely to be supply

Table B.7. Domestic export suppliers

| | (1) |
|---|---|
| | $Y_{ift}$ |
| $X_{t-1}^{same}$ | 0.00236 |
| | (0.00358) |
| $X_{t-1}^{same} \times exporter_f$ | 0.00358 |
| | (0.00802) |
| $X_{t-1}^{other-city}$ | -0.00574*** |
| | (0.000680) |
| $X_{t-1}^{other-city} \times exporter_f$ | 0.00268** |
| | (0.000609) |
| Observations | 27975967 |

*Notes:* Unit of observation is buyer $i$, supplier $f$ and year $t$. $Y_{ift}$ indicates a first match took place between buyer and supplier. $X_{ift}^{k}$ is a count of buyers in region $k$ which added supplier $f$ in $t-1$. $exporter_f$ indicates supplier $f$ is an exporter. Coefficients are multiplied by 100 to read as percentage point marginal effects. Standard errors in parentheses clustered at the buyer level. $^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

constrained. This is tested in equation 3.2.

$$Y_{ift} = \mu_1 X_{if,t-1}^{same} + \mu_2 X_{if,t-1}^{same} \times Exporter_f + \gamma_1 X_{if,t-1}^{other-city} + \gamma_2 X_{if,t-1}^{other-city} \times Exporter_f + \alpha_i + \alpha_t + u_{ift} \tag{3.2}$$

Results in Table B.7 show that domestic suppliers which are exporters, and hence less supply constrained, have a smaller negative effect from making a match elsewhere in the country. This is again consistent with the search narrative.

Table B.8. Domestic suppliers

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $Y_{ift}$ | $Y_{ift}$ | $Y_{ift}$ | $Y_{ift}$ |
| $Z_{t-1}^{10km^2}$ | 0.00453 | | | |
| | (0.00524) | | | |
| $Z_{t-1}^{1km^2}$ | | 0.00743 | | |
| | | (0.00538) | | |
| $Z_{t-1}^{same}$ | | | 0.00920* | |
| | | | (0.00540) | |
| $Z_{t-1}^{same}$ | | | | 0.00919* |
| | | | | (0.00557) |
| $Z_{t-1}^{nextdoor}$ | | | | -0.0235 |
| | | | | (0.0169) |
| $X_{t-1}^{other}$ | -0.00778*** | -0.00771*** | -0.00768*** | -0.00795*** |
| | (0.00188) | (0.00187) | (0.00187) | (0.00188) |
| Year and Buyer FE | Yes | Yes | Yes | Yes |
| Observations | 27975967 | 27975967 | 27975967 | 27975967 |

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.9. Import suppliers

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $Y_{ift}$ | $Y_{ift}$ | $Y_{ift}$ | $Y_{ift}$ |
| $Z_{t-1}^{10km^2}$ | 0.115*** | | | |
| | (0.00140) | | | |
| $Z_{t-1}^{1km^2}$ | | 0.136*** | | |
| | | (0.00158) | | |
| $Z_{t-1}^{same}$ | | | 0.133*** | |
| | | | (0.00158) | |
| $Z_{t-1}^{same}$ | | | | 0.134*** |
| | | | | (0.00158) |
| $Z_{t-1}^{nextdoor}$ | | | | 0.0185*** |
| | | | | (0.000692) |
| $X_{t-1}^{other}$ | -0.00178*** | -0.00130*** | -0.00173*** | -0.00155*** |
| | (0.000357) | (0.000357) | (0.000357) | (0.000357) |
| Year and Buyer FE | Yes | Yes | Yes | Yes |
| Observations | 4834635 | 4834635 | 4834635 | 4834635 |

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

# Appendix C

# Appendix to Chapter 3

## 1. Mathematical appendix

### A.1. Pricing, profits and transfer equation

The first order condition for pricing yields the following equation:

$$q_{xb} + \sum_{x' \in J_b} \frac{\partial q_{x'b}}{\partial p_{xb}} (p_{x'b} - c_{x'b}) = 0 \quad \forall x \in J_b, \tag{1.1}$$

where $c_{x'b}$ is the marginal cost of supplying product $x'$ to consumers through buyer $b$. The intuition behind Equation 1.1 is that buyers internalize that their pricing on one good alters demand on other goods.

This in turn yields a condition for the mark-up which is a constant over marginal cost and equal to one over the elasticity of substitution over products

$$\frac{p_{xb} - c_{xb}}{p_{xb}} = \frac{1}{\eta} \tag{1.2}$$

Substituting this mark-up into the profit function, yields an instantaneous profit flow created by buyer $b$ and its set of suppliers given by a summation over the profit provided by each product $x$ in buyer $b$'s bundle ($J_b$), such that

$$\pi_b(\mathbf{s}) = \frac{E}{\eta P^{1-\eta}} \left[ \sum_{x \in J_b} \left( \frac{\eta}{\eta - 1} \right)^{1-\alpha} \tau_L \tilde{c}_b^{1-\alpha} \right]^{\frac{1-\eta}{1-\alpha}}, \tag{1.3}$$

for $L \in \{D, I\}$ and where $\tilde{c}_b = c_b/\psi_L$ is the quality-adjusted marginal cost, $\mathbf{s} = \{s_I, s_D\}$ is a vector of the number of international and domestic suppliers, $P$ is the standard CES aggregate price index and $E$ is household expenditure.

Using the assumption of Stole and Zweibel bargaining, I obtain an equation for the revenue transfer between buyer $i$ and supplier $j$,

$$r_{ji}(\mathbf{s}) = (h_{j|i})^{\frac{\alpha-\eta}{\alpha-1}} \frac{E}{P^{1-\eta}} \left(\frac{\eta}{\eta-1}\right)^{-\eta} \left(\tau_L \tilde{c}_{ji}\right)^{1-\eta} \left[\frac{\Lambda}{\alpha-1} + \lambda\right], \tag{1.4}$$

where $r_{ji}$ is the revenue for seller $j$ from buyer $i$, $h_{j|i} = \frac{\tau_L \tilde{c}_j^{1-\alpha}}{\sum_{l=1}^{J} s_l \tau_L \tilde{c}_i^{1-\alpha}}$ is the within buyer-$i$ revenue share of a type-$j$ seller, $\lambda$ is the seller's fraction of marginal cost.

If cost per unit quality $\tilde{c}$ is fixed across products within buyers then the transfer equation collapses to the following as $h_{j|i}$ becomes constant within buyers,

$$r_{ji}(\mathbf{s}) = \frac{E}{P^{1-\eta}} \left(\frac{\eta}{\eta-1}\right)^{-\eta} s^{\frac{\alpha-\eta}{1-\alpha}} \left(\tau_L \tilde{c}_{ji}\right)^{1-\eta} \left[\frac{\Lambda}{\alpha-1} + \lambda\right]. \tag{1.5}$$

# References


**Acemoglu, Daron, Alireza Tahbaz-Salehi et al.**, "Firms, Failures, and Fluctuations," in "2019 Meeting Papers" number 1148 Society for Economic Dynamics 2019.

**Agarwal, Vivek and John Spray**, "Enterprise Mapping in Rwanda Dairy, Maize and Tourism," Technical Report 2016.

**Albornoz, Facundo, Héctor F Calvo Pardo, Gregory Corcos, and Emanuel Ornelas**, "Sequential exporting," *Journal of International Economics*, 2012, *88* (1), 17–31.

**Allen, Treb**, "Information frictions in trade," *Econometrica*, 2014, *82* (6), 2041–2083.

**Antras, Pol and Arnaud Costinot**, "Intermediated trade," *The Quarterly Journal of Economics*, 2011, *126* (3), 1319–1374.

**_ , 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.

**Arkolakis, Costas, Arnaud Costinot, and Andres Rodriguez-Clare**, "New trade models, same old gains?," *American Economic Review*, 2012, *102* (1), 94–130.

**Arnosti, Nick, Ramesh Johari, and Yash Kanoria**, "Managing congestion in matching markets," *Available at SSRN 2427960*, 2018.

**Atkeson, Andrew and Ariel Burstein**, "Pricing-to-market, trade costs, and international relative prices," *American Economic Review*, 2008, *98* (5), 1998–2031.

**Atkin, David and Dave Donaldson**, "Who's getting globalized? The size and implications of intra-national trade costs," 2015.

**Baqaee, David Rezza and Emmanuel Farhi**, "The macroeconomic impact of microeconomic shocks: beyond Hulten's Theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

**Barten, Anton P and Volker Böhm**, "Consumer theory," *Handbook of mathematical economics*, 1982, *2*, 381–429.

**Bergquist, Lauren Falcao and Michael Dinerstein**, "Competition and entry in agricultural markets: Experimental evidence from Kenya," 2019.

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

**_ and J Bradford Jensen**, "Exceptional exporter performance: cause, effect, or both?," *Journal of international economics*, 1999, *47* (1), 1–25.

**_ and _** , "Why some firms export," *Review of Economics and Statistics*, 2004, *86* (2), 561–569.

**_ and Swati Dhingra**, "Contracting and the Division of the Gains from Trade," 2015.

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

**_ , _ , and Yukiko Saito**, "Geography and firm performance in the Japanese production network," *NBER Working Paper No. 21082*, 2015.

**Bigio, Saki and Jennifer La'O**, "Distortions in production networks," Technical Report, National Bureau of Economic Research 2016.

**Bisztray, Márta, Miklós Koren, and Adam Szeidl**, "Learning to import from your peers," *Journal of International Economics*, 2018, *115*, 242–258.

**Bulzomi, Anna, Peter Danssaert, Sergio Finardi, and Ken Matthysen**, "Supply Chains and Transport Corridors in East Africa," *International Peace Information Service and TransArms-Research*, 2014.

**Cai, Jing and Adam Szeidl**, "Interfirm relationships and business performance," *The Quarterly journal of economics*, 2017, *133* (3), 1229–1282.

**Carvalho, Vasco M**, "From micro to macro via production networks," *The Journal of Economic Perspectives*, 2014, *28* (4), 23–47.

**_ , Matthew Elliott, and John Spray**, "Supply Chain Bottlenecks in a Pandemic," *Mimeo*, 2020.

**Casaburi, Lorenzo and Tristan F Reed**, "Competition in agricultural markets: An experimental approach," 2017.

**_ , Rachel Glennerster, and Tavneet Suri**, "Rural roads and intermediated trade: Regression discontinuity evidence from Sierra Leone," *Available at SSRN 2161643*, 2013.

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

**Choi, Syngjoo, Andrea Galeotti, and Sanjeev Goyal**, "Trading in networks: theory and experiments," *Journal of the European Economic Association*, 2017, *15* (4), 784–817.

**Condorelli, Daniele, Andrea Galeotti, and Ludovic Renou**, "Bilateral trading in networks," *The Review of Economic Studies*, 2016, *84* (1), 82–105.

**Donaldson, Dave, Amanda Jinhage, and Eric Verhoogen**, "Beyond borders: Making transport work for African trade," *IGC Growth Brief Series 009*, 2017.

**Eades, Peter, Xuemin Lin, and William F Smyth**, "A fast and effective heuristic for the feedback arc set problem," *Information Processing Letters*, 1993, *47* (6), 319–323.

**Eaton, Jonathan, David Jinkins, James R Tybout, and Daniel Yi Xu**, "Two-sided Search in International Markets," *Mimeo*, 2016.

**_ , Samuel Kortum, Francis Kramarz et al.**, "Firm-to-Firm Trade: Imports, exports, and the labor market," *Brown University, unpublished manuscript*, 2017.

**Edmonds, Jack and Richard M Karp**, "Theoretical improvements in algorithmic efficiency for network flow problems," *Journal of the ACM*, 1972.

**Elliott, Matthew L, Sanjeev Goyal, and Alexander Teytelboym**, "Networks and economic policy," *Oxford Review of Economic Policy*, 2019, *35* (4), 565–585.

**Fafchamps, Marcel**, *Market institutions in Sub-Saharan Africa: theory and evidence*, Cambridge, MA: MIT Press, 2004.

**_ and Simon Quinn**, "Networks and manufacturing firms in africa: Results from a randomized field experiment," *The World Bank Economic Review*, 2016, *32* (3), 656–675.

**_ and _** , "Networks and manufacturing firms in Africa: Results from a randomized field experiment," *The World Bank Economic Review*, 2018, *32* (3), 656–675.

**Fainmesser, Itay P**, "Exclusive intermediation," *Available at SSRN 2410408*, 2014.

**Farboodi, Maryam**, "Intermediation and voluntary exposure to counterparty risk," *Available at SSRN 2535900*, 2014.

**Fieler, Ana Cecília, Marcela Eslava, and Daniel Yi Xu**, "Trade, quality upgrading, and input linkages: Theory and evidence from colombia," *American Economic Review*, 2018, *108* (1), 109–46.

**Ford, Lester Randolph and Delbert Ray Fulkerson**, "Flows in networks," 2015, *54*.

**Ford, L.R and D.R Fulkerson**, "Maximal flow through a network," *Canadian Journal of Mathematics*, 1956.

**Fradkin, Andrey**, "Search frictions and the design of online marketplaces," 2015.

**Government of Uganda**, "Uganda Economic Growth Forum: Import Substitution and Export Promotion for Growth," 2019.

**Goyal, Sanjeev and Fernando Vega-Redondo**, "Structural holes in social networks," *Journal of Economic Theory*, 2007, *137* (1), 460–492.

**Grassi, Basile**, "IO in IO: Competition and volatility in input-output networks," *Unpublished Manuscript, Bocconi University*, 2017.

_ **and Julien Sauvagnat**, "Production networks and economic policy," *Oxford Review of Economic Policy*, 2019, *35* (4), 638–677.

**Gutiérrez, Germán and Thomas Philippon**, "Declining Competition and Investment in the US," Technical Report, National Bureau of Economic Research 2017.

**Horton, John J**, "Online labor markets," in "International workshop on internet and network economics" Springer 2010, pp. 515–522.

**Hosios, Arthur J**, "On the efficiency of matching and related models of search and unemployment," *The Review of Economic Studies*, 1990, *57* (2), 279–298.

**International Monetary Fund**, "Article IV Consultation," *International Monetary Fund*, 2019.

**Kamal, Fariha and Asha Sundaram**, "Buyer–seller relationships in international trade: Do your neighbors matter?," *Journal of International Economics*, 2016, *102*, 128–140.

**Kathage, Angus Morgan**, "Understanding the informal economy in African cities: Recent evidence from Greater Kampala," *World Bank Blogs*, 2018.

**Kikkawa, Ayumu Ken, Glenn Magerman, and Emmanuel Dhyne**, "Imperfect competition in firm-to-firm trade," *National Bank of Belgium, Working Paper No. 363*, 2019.

**Kotowski, Maciej H and C Matthew Leister**, "Trading networks and equilibrium intermediation," 2018.

**Krolikowski, Pawel Michal and Andrew H McCallum**, "Goods-market frictions and international trade," 2017.

**Lim, Kevin**, "Firm-to-firm Trade in Sticky Production Networks," Technical Report, Society for Economic Dynamics 2017.

**Liu, Ernest**, "Industrial policies in production networks," *The Quarterly Journal of Economics*, 2019, *134* (4), 1883–1948.

**Loecker, Jan De, Jan Eeckhout, and Gabriel Unger**, "The Rise of Market Power," Technical Report 2019.

**Macchiavello, Rocco and Ameet Morjaria**, "The value of relationships: evidence from a supply shock to Kenyan rose exports," *American Economic Review*, 2015, *105* (9), 2911–45.

_ **and** _ , "Competition and Relational Contracts in the Rwanda Coffee Chain," 2019.

**Maggio, Marco Di and Alireza Tahbaz-Salehi**, "Collateral shortages and intermediation networks," *Columbia Business School Research Paper*, 2015, (14-40).

**Manea, Mihai**, "Intermediation and resale in networks," *Journal of Political Economy*, 2018, *126* (3), 1250–1301.

**Mangin, Sephorah and Benoît Julien**, "Efficiency in search and matching models: A generalized hosios condition," *Becker Friedman Institute for Research in Economics Working Paper*, 2018, (2018-10).

**Manova, Kalina**, "Credit constraints, heterogeneous firms, and international trade," *Review of Economic Studies*, 2012, *80* (2), 711–744.

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

_ **and Stephen J Redding**, "New trade models, new welfare implications," *American Economic Review*, 2015, *105* (3), 1105–46.

**Nava, Francesco**, "Efficiency in decentralized oligopolistic markets," *Journal of Economic Theory*, 2015, *157*, 315–348.

**Ostrovsky, Michael**, "Stability in supply chain networks," *American Economic Review*, 2008, *98* (3), 897–923.

**Overseas Development Institute**, "Productive employment and transformation in Uganda," Technical Report, ODI 2015.

**Petrongolo, Barbara and Christopher A Pissarides**, "Looking into the black box: A survey of the matching function," *Journal of Economic literature*, 2001, *39* (2), 390–431.

**Rauch, James E**, "Business and social networks in international trade," *Journal of economic literature*, 2001, *39* (4), 1177–1203.

_ **and Joel Watson**, "Starting small in an unfamiliar environment," *International Journal of industrial organization*, 2003, *21* (7), 1021–1042.

_ **and Vitor Trindade**, "Ethnic Chinese networks in international trade," *Review of Economics and Statistics*, 2002, *84* (1), 116–130.

**Ridder, Maarten De**, "Market power and innovation in the intangible economy," 2019.

**Rogerson, Richard, Robert Shimer, and Randall Wright**, "Search-theoretic models of the labor market: A survey," *Journal of economic literature*, 2005, *43* (4), 959–988.

**Sen, Ritwika**, "Enhancing local content in Uganda's oil and gas industry," Technical Report, WIDER Working Paper 2018.

**Siedlarek, Jan-Peter**, "Intermediation in networks," 2015.

**Simpson, Michael, Venkatesh Srinivasan, and Alex Thomo**, "Efficient computation of feedback arc set at web-scale," *Proceedings of the VLDB Endowment*, 2016, *10* (3), 133–144.

**Spray, John**, "Reorganise, Replace or Expand? The role of the supply-chain in first-time exporting," *Cambridge INET Working Paper 1741*, 2017.

_ , "Search Externalities in Firm-to-Firm Trade," *Mimeo*, 2020.

_ **and Sebastian Wolf**, "Industries without smokestacks in Uganda and Rwanda," *Mimeo*, 2016.

_ **and Victor Steenbergen**, "Made in Rwanda: Establishing a publicly available supplier database for Rwanda," *Policy Note. Technical report, International Growth Centre*, 2017.

**Startz, Meredith**, "The value of face-to-face: Search and contracting problems in Nigerian trade," *Available at SSRN 3096685*, 2016.

**Steenbergen, Victor and John Sutton**, "Establishing a Local Content Unit for Rwanda," *Policy Note. Technical report, International Growth Centre*, 2017.

**Steinwender, Claudia**, "Real effects of information frictions: When the states and the kingdom became united," *American Economic Review*, 2018, *108* (3), 657–96.

**Stole, Lars A and Jeffrey Zwiebel**, "Intra-firm bargaining under non-binding contracts," *The Review of Economic Studies*, 1996, *63* (3), 375–410.

**Tintelnot, Felix, Ayumu Ken Kikkawa, Magne Mogstad, and Emmanuel Dhyne**, "Trade and domestic production networks," Technical Report, National Bureau of Economic Research 2018.

**World Bank**, "Doing Business 2011," *World Bank Group*, 2011.

_ , "Doing Business 2016 : Trading Across borders," *World Bank Group*, 2016.

_ , "Uganda Economic Update, 13th Edition, May 2019 : Economic Development and Human Capital in Uganda - A Case for Investing More in Education," *World Bank Group*, 2019.