Essays on Economic Linkages and Resilience



Verena Wiedemann

St. Antony's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Trinity Term 2023

Essays on Economic Linkages and Resilience

Verena Wiedemann

St. Antony's College
University of Oxford

Trinity Term 2023

This thesis consists of five chapters that explore the theme of economic linkages and resilience. My research investigates the structure of Kenya's formal firm network and its connections to both international markets and the domestic informal sector.
Chapter 1 discusses how tax returns can be used to map Kenya's domestic firm network and describes the network's fundamental properties. In Chapter 2, my co-authors and I address the informal sector's significance in shaping spatial patterns of firm-to-firm trade observed in formal sector data. We document stylized facts about trade among formal firms, revealing a high degree of spatial concentration, beyond the concentration of aggregate economic activity. Using census data and a network formation model we show that solely relying on formal sector network data might underestimate both the connectivity but also vulnerability of smaller regional markets to shocks, especially those passing through hubs like Nairobi. Chapter 3 draws on qualitative interviews to study the role of Internet infrastructure in shaping supply chains. I find that a firm's choice to adopt Internet-based technologies is influenced by its capacity to make complementary investments, which in turn correlate with to firm size. Turning to linkages of Kenya's supply chains with international markets in Chapter 4, my co-authors and I show that firms with strong direct or indirect connections to international markets tend to be older, larger, and account for a substantial share of Kenya's formal economy. We find that both importers and exporters adjust their domestic supply chains in response to international trade shocks – before and during the COVID-19 crisis alike. Sourcing from international markets does not crowd out domestic purchases, while sales abroad and at home can act as substitutes. In Chapter 5, my co-authors and I study the impact of changes in mobility patterns during COVID-19 pandemic on food price dynamics in Kenya and Uganda. Using geo-referenced price data collected through a crowdsourcing approach, we find that a 10 percentage point reduction in mobility leads to a 0.3% and 1.5% increase in food prices in Kenya and Uganda, respectively.

Supervisors This thesis was supervised by Christopher Woodruff and Alexander Teytelboym.

Statement of authorship The work in this thesis is based exclusively on research carried out by me as a DPhil student in the Department of Economics at the University of Oxford. No part of this thesis has been submitted for a degree at the University of Oxford or any other university. Chapter 1 is joint work with Peter Wankuru Chacha and Benard Kipyegon Kirui. Chapter 2 is joint work of the same team with the addition of Vatsal Khandelwal. Chapter 4 is again joint work with Peter Wankuru Chacha and Benard Kipyegon Kirui. Chapter 5 is joint work with George Kariuki Kinyanjui and Doreen K. Rubatsimbira.

Acknowledgements

I gratefully acknowledge funding and support from the German National Scholarship Foundation, the Department of Economics, and the Dartmouth College International Economics Ph.D. Fellowship. My research was funded by the Foreign, Commonwealth & Development Office (FCDO) and the Centre for Economic Policy Research (CEPR) through the Private Enterprise Development in Low-Income Countries (PEDL) and the Structural Transformation and Economic Growth (STEG) initiatives, the International Growth Centre (IGC), the Centre for the Study of African Economies (CSAE), and the Oxford Economics Papers (OEP) Fund.

To my supervisors, Chris Woodruff and Alex Teytelboym: thank you for your unwavering support and mentorship throughout this journey. I consider myself incredibly fortunate to have had two supervisors who are deeply passionate about research and have so generously shared their time. My research is the sum of their parts: firms and networks. To Chris: thank you for your guidance, kindness, and countless inspiring conversations. I am truly grateful for the many opportunities you have created for me to discuss, explore, and develop my ideas. Alex, thank you for your open-mindedness, for sharing your excitement and knowledge on all things networks.
To Doug Gollin, who has been a mentor since my MSc days: thank you for your continued support and your encouragement to pursue the questions that fueled my enthusiasm and curiosity.

This DPhil thesis would not have been possible without my co-authors Benard Kirui and Peter Chacha. They took an unbelievable leap of faith to embark with me on this journey in the midst of a pandemic. To Benard and Peter: thank you for your trust, persistence, and patience throughout. We have been extremely fortunate to work with the Kenya Revenue Authority on our research agenda. I would like to thank the research team at the Strategy, Innovation & Risk Management Department for their support. I am particularly indebted to Joseph Sirengo. My gratitude further extends to the dedicated team of data scientists at the data warehouse, specifically Benard Sang, Romeo Ekirapa, and Simon Mwangi. To my co-authors Vatsal Khandelwal, George Kinyanjui, and Doreen Rubatsimbira: thank you for your enthusiasm, ideas, and dedication. I have learned so much from you. To Gatwiri and Daniela: thank you for your fantastic support and excitement, reminding me of my love for research.

During my DPhil I was fortunate to have visited Dartmouth College and Strathmore University and have benefited immeasurably from the exchanges and conversations

during this time. To the international economics community at Dartmouth College: thank you for hosting me and making me grow as a researcher. Specifically, Meredith Startz, Nina Pavcnik, Matt Grant, Treb Allen, and Andy Bernard: thank you for sharing your time and insights on numerous occasions and walks around Hanover.
To the research community at Strathmore University: thank you for welcoming me and your invaluable support. Specifically, Mary Achieng, Tabitha Njunguna, and Mumbi Wachira: thank you for the many insightful conversations and discussions.

To the community at St Antony's College, the CSAE community, the Women in Economics community, the applied micro group, the various iterations of the infamous firms discussion group, and all the faculty members, postdocs, and DPhil students who I have shared my time in Oxford with: thank you for being my academic 'home' for the past seven years. It was a privilege to learn and grow in this inspiring community. Specifically, Simon Quinn, Hamish Low, Christian Meyer, Sarah Clifford, Solomon Owusu, Farshad Ravasan, Muhammad Meki, Nathan Lane, Pramila Krishnan, Sanjay Jain, Julien Labonne, Cheryl Doss, Kate Orkin, and Gabriel Ulyssea: thank you for feedback, support, and inspiration. To Rose Page and Sue Hyatt: thank you for helping me turn research proposals into projects.
To past and present postdocs and DPhil students from the economics community who I shared this journey with, Mahreen, Emma, Rocco, Lisa, Sugandha, Vanessa, Shihang, Tatjana, Giulio, Binta, Margaryta, Lukas, Marc, Michael, Katherine, Ashley, Ben, Hannah U., Artur, Diala, Marta, Henriette, Wenzhang, Juliette, and Amelie: thank you for the discussions, the learning, and many cheerful memories.

To my friends from the wider Oxford and Edinburgh community, specifically Julia, Lena, Kristina, Christian, Lily, Sophie, Blen, Fig, Dominic, Katarina, Gunjita, Allan, Khudadad, Lukas, Soha, Mia, and Lea: thank you for your friendship, the dinners, coffees, and inspiring conversations. To the Wolfson College Boatclub women's squad: thank you for the perfect escape from the econ bubble - memorable bump campaigns, summer rowing, and lock trips. To my friends from my school days, specifically Emre, Mao Mengjing, Sun Ruoyu, Franziska, Magdalena, and Sarah: thank you for unwavering friendship.

To Hannah, Vatsal, Raphael, Sanghamitra, Luke, and Raushan: thank you for making 23 Jeune Street the best home I could have asked for during the past six years. Our friendship has carried me through the highs and lows of my DPhil journey. I will thoroughly miss our board game nights and staircase conversations. Thank you for the joy and happiness we shared, for inspiring me, and for being a source of knowledge and wisdom I could I always count on.

To my family and relatives, uncles, aunts, and cousins, Anna, Luca, and Wolfgang: thank you for your support throughout all those years. To Ludger, Birgit, Daniel, and Julia: thank you for being my cheerleaders. To my godfather Klaus: thank for always being there as a voice of reason and a support I could count on. To my mother Claudia: thank you for inspiring me to be curious, to follow my interests, and for supporting me in all of my pursuits. You made this possible. Thank you.

To Raphael: thank you for cheering me along each step of the way, not just throughout this DPhil, but ever since we first met. Your insatiable curiosity about the world and its wonders continues to expand my horizons each day. Here's to many more adventures together.

Verena Wiedemann
Oxford, July 2023

Contents

Introduction **1**

1 Mapping Kenya's firm network transaction by transaction **6**
1.1 Introduction 6
1.2 Data description 8
1.3 How does the content of VAT data compare with customs records and firm survey data? 10
1.4 Sector composition of the firm network 13
1.5 Fundamental properties of the network 18
1.6 Conclusion 24

Appendices of Chapter 1 **27**
1.A Additional tables 27
1.B Data processing and cleaning protocol 28

2 Spatial Inequality and Informality in Kenya's Firm Network **37**
2.1 Introduction 38
2.2 Contributions 42
2.3 Description of the administrative data 46
2.4 The spatial concentration of domestic firm-to-firm trade 48
2.5 The role and position of the informal sector 60
2.6 A model of network formation with sectors and regions 80
2.7 Spatial inequality and unobservable informal firms - a counterfactual 87
2.8 Conclusion 96

Appendices of Chapter 2 **99**
2.A Additional material on spatial trade patterns 99
2.B Additional material on informality 103
2.C Additional material on the counterfactual 110

3 Internet use and technology adoption along supply chains 113
3.1 Introduction . . . 114
3.2 Contributions . . . 116
3.3 The expansion of Kenya's fibre optic network . . . 118
3.4 Data . . . 120
3.5 The role of Internet connectivity in a search and matching framework 123
3.6 Internet use . . . 125
3.7 Distance is indeed not dead: constraints to the effectiveness of Internet connectivity . . . 134
3.8 Discussion and conclusion . . . 139

4 Supply chains in times of crisis: Evidence from Kenya's production network 142
4.1 Introduction . . . 143
4.2 Data and context . . . 147
4.3 Links to international supply chains and firm dynamics during the COVID-19 crisis . . . 151
4.4 Network position and firm resilience . . . 163
4.5 Conclusion . . . 172

Appendices of Chapter 4 174
4.A Additional material on data and context . . . 174
4.B Additional material on international supply chain linkages . . . 177
4.C Additional material on firm networks and resilience . . . 180

5 Changing mobility patterns and food price dynamic during the COVID-19 pandemic in Kenya and Uganda 189
5.1 Introduction . . . 190
5.2 Data . . . 194
5.3 How did mobility patterns shift during the pandemic? . . . 202
5.4 How did food prices evolve over time? . . . 204
5.5 Linking mobility patterns and price dynamics . . . 207
5.6 Discussion and conclusion . . . 215

Appendices of Chapter 5 219
5.A Appendix material for data and sample description . . . 219
5.B Additional results for the event study . . . 234
5.C Appendix material on the mobility and food prices nexus . . . 235

Conclusion 242

References 245

List of Figures

1.1 Average degree by sector 14
1.2 Sector-level trade flows between VAT-registered firms 16
1.3 Average input share and average centrality by sector 17
1.4 In- and outdegree CDFs 20
1.5 Sales and the number of firm-to-firm linkages 22
1.6 Sales and the number of domestic sales destinations 22

2.1 Composition of sales and purchases by sector 47
2.2 County-level trade flows between formal firms 51
2.3 Geographic concentration of sales based on supplier vs buyer location 53
2.4 The intensive and extensive margin of domestic trade 57
2.5 Average in- and outdegrees across space 59
2.6 Comparison of formal sector shares based on census versus administrative records 65
2.7 Value added by VAT firms vs GDP 68
2.8 The extensive margins of informality - in which sectors do informal firms operate? 70
2.9 Links of small and medium sized enterprises to large firms 73
2.10 Count-level comparison of the number of firms, employment, and value added in KNBS benchmark data with the administrative data 75
2.11 Share of formal sector employment and regional market size 76
2.12 Spatial variation in informality - a story of sectors or geographies? . . 79
2.13 Objective function for various values r 86
2.14 Actual and predicted distribution inverse CDF 87
2.15 Sector-region probabilities and formal sector shares 90
2.16 Predicted change in county-level outdegree and formal sector shares . 93
2.A1 Firm headquarter locations and population density 99
2.A2 Share of sales to final demand (non-registered entities) 100
2.A3 The higher inequality in outdegree is common across contexts 102
2.B1 Quarterly GDP in billion KES (2015-2020) 105
2.B2 Quarterly GDP in billion KES (2015-2020) 106

2.B3 Sector-level correlation of the extensive margin of informality 108
2.B4 Informality, market size, and income levels 109
2.C1 Sector-region probabilities and formal sector shares 110
2.C2 Predicted change in outdegree and formal sector shares 111
2.C3 Predicted change in county-level outdegree and baseline linking probabilities ($\sum_i p(\theta, \theta')$) 112

3.1 Number of fixed and mobile broadband subscribers between 2012 and 2020 120
3.2 Sector composition of interviewed firms and firms covered in the administrative records 122
3.3 ICT adoption of firms covered in the administrative data 126
3.4 Share of inputs from the ICT sector by firm size 127
3.5 ICT adoption of firms covered in the administrative data 128
3.6 Primary mode of Internet by firm size 129
3.7 Use of Internet by firm size 130

4.1 Weekly firm-to-firm transactions 149
4.2 The evolution of major firm-level outcomes during the COVID-19 crisis 151
4.3 Output share of firms by strength of links to international supply chains 155
4.4 Firm-level exposure to international supply chains and firm sales during the crisis and recovery 158
4.5 Supply chain diversification and firm sales during the crisis and recovery phase 167
4.6 Concentration of sales and purchases (HHI) and firm sales during the crisis and recovery phase 169
4.7 Supply chain complexity and resilience 171
4.A1 Major firm-level outcomes in 2020 and 2019 175
4.B1 Import and export shares over time 177
4.B2 Major firm-level outcomes during the COVID-19 crisis vs 2019 trends 178
4.C1 Pre-crisis inventory levels and firm sales during the crisis and recovery phase 182
4.C2 Supply chain diversification and pre-crisis inventory levels 183
4.C3 Domestic network position and firm sales 187
4.C4 Supply chain diversification and the response of firm sales to international shocks 188

5.1 Geographic dispersion of price quotes for popular items 199
5.2 Variation in prices by shop type 202
5.3 Variation in activity levels across different categories of locations . . . 203
5.4 Price dynamics for April 2020 - mid-March 2021 206
5.5 Using alternative measures of mobility 210

5.6 Exploring the role of shop and product types 213
5.A1 Mapping mobility: Grocery and Pharmacy Index 221
5.A2 Mapping mobility: Workplace Index 222
5.A3 Daily number of collected price quotes 223
5.A4 Geographic dispersion of price quotes for rice over time 227
5.A5 Population and the number of collected price quotes across space . . 228
5.A6 Population density, Internet access, and the number of collected price quotes across space 228
5.A7 Comparison of national average prices with KNBS 229
5.A8 Comparison of national average prices with FEWS and KNBS 230
5.A9 Comparison of Nairobi average prices with FEWS and KNBS 231
5.A10 Comparison of national average prices with UBoS I 232
5.A11 Comparison of national average prices with UBoS II 233
5.B1 Price dynamics and policy responses to COVID-19 234
5.C1 Kenya: Controlling for price trends in previous years 235
5.C2 Alternative specifications: Price dynamics and mobility 236
5.C3 Composition of the price basket over time 237
5.C4 Mobility and the composition of the observed product basket 238

List of Tables

1.1 Overview of administrative data sources 8
1.2 Comparison of common data features 10
1.3 Density of domestic production networks and tax reporting requirements 19
1.4 Moments of the outdegree distribution: cross-country comparison . . 21
1.5 Moments of the indegree distribution: cross-country comparison . . . 23
1.6 Degree correlation coefficients 24
1.A1 Number of indegrees by sector 27
1.A2 Number of outdegrees by sector 27

2.1 Geographic concentration of economic activity in Kenya 50
2.2 Geographic concentration of economic activity in Kenya 54
2.3 Margins of informality in firm networks 61
2.4 Overview of benchmark data 64
2.5 Measures of informality 64
2.6 Correlation of formality measures 66
2.7 Geographic concentration of economic activity by degree of formalisation 77
2.8 Changes in the dispersion of outdegrees 91
2.A1 The intensive and extensive margin of domestic trade 101
2.A2 Geographic concentration of industrial activity 103
2.B1 Share of GDP covered in the administrative records 104
2.B2 Sector-level comparison with the 2016 Census of Establishments . . . 107
2.C1 Robustness checks for changes in the dispersion of outdegrees 112

3.1 Firm characteristics of interviewed firms and firms covered in the administrative records 123

4.1 Summary statistics for firms directly and indirectly linked to international supply chains in 2018 154
4.2 Firm response to international supply and demand shocks 162
4.A1 Sector-level breakdown of total sales (2019) 174
4.A2 Summary statistics on inter-firm relationships between 2015 and 2020 174

4.B1 Shocks and shares summary statistics 180
4.C1 Firm-level characteristics and 2018 outcomes by upstream diversification 184
4.C2 Firm-level characteristics and 2018 outcomes by downstream diversification 185
4.C3 Firm-level characteristics and 2018 outcomes by inventory levels . . . 186

5.1 Summary statistics for retail prices collected via a quick-response online survey 197
5.2 Variance decomposition for food prices 200
5.3 Price dynamics and mobility 208
5.A1 Kenya - detailed summary statistics 224
5.A2 Uganda - detailed summary statistics 225
5.C1 Mobility and data collection - using variation across space and time . 239
5.C2 Mobility and data collection - using variation across time 240
5.C3 Mobility and data collection - using variation across space 241

Introduction

Globalisation has profoundly shaped the world in which we live today, transforming our economies and societies. As economists, we delve into the intricate workings of this phenomenon, exploring its implications for global trade patterns and economic development. The connectivity of international hubs like Shanghai, Nairobi, New York, and London underscores the ever-increasing interdependence fostered by globalisation. However, the benefits of this global interconnectedness, such as falling world prices of consumer goods, are not uniformly distributed. Within individual economies, the allocation of these gains is fundamentally influenced by the underlying domestic market structures and the degree of inter-connectedness among regions.
At the same time, the pace of globalisation has slowed in recent years. My research focuses on Kenya, East Africa's largest economy and one of the fastest-growing economies globally. In Kenya, similar to other low- and middle-income countries, policymakers face a limited set of opportunities for export-led growth models. In response, leveraging the potential of domestic and regional markets emerges as a promising strategy (Goldberg and Reed, 2023). Nonetheless, this potential hinges on the integration of domestic markets and the reduction of economic inequality. Questions surrounding what factors shape domestic trade patterns and the structure of firm networks, and how they impact the decision-making of firms, are what motivated this thesis.

In this thesis, I study the structure of Kenya's network of formal private sector firms. I research how the network links to both international markets and the domestic informal sector, i.e. the part of the economy we cannot observe in administrative records. Chapter 1 describes how tax returns can be used to map domestic firm networks, and it introduces the readers to the world of firm networks

in Kenya. Chapter 2 focuses on recovering the part we miss out on in formal tax records – the informal sector – and how it might shape observed spatial patterns of firm-to-firm trade. Chapter 3 uses qualitative interviews to explore how firms use Internet infrastructure to expand and manage relationships with their buyers and suppliers. Chapter 4 explores the links between Kenya's domestic firm networks and international markets.

As the COVID-19 pandemic hit the world in 2020, supply chains and resilience to shocks attracted substantial attention from a wider audience. The pandemic and its impact on economic activity not only shaped ways of collaboration and doing research during my DPhil but is also reflected in two of my DPhil chapters, which examine how pandemic-related shocks affected supply chains and spatial price dynamics. Chapter 4 studies firm outcomes and supply chain dynamics during the COVID-19 crisis, while Chapter 5 explores the link between changes in mobility patterns and food prices.

Previous research on domestic firm networks and market structures has been constrained by the lack of data with sufficient spatial granularity (Donaldson, 2015). Fortunately, the timing of my DPhil journey coincided with the arrival of new data sources and technologies that enabled unconventional data collection methods. Along the way, I crossed paths with my co-authors Peter Wankuru Chacha and Benard Kipyegon Kirui, who shared my enthusiasm for firm networks and their role in private sector development. Our joint research draws on administrative tax records from the Kenya Revenue Authority (KRA). Highly granular transaction-level information in the KRA's value-added tax (VAT) returns allows us to track Kenya's firm network over time and space. Four of my DPhil chapters focus on or partly build on this data set.

Long before researchers were able to utilise firm-to-firm transaction-level data to study domestic trade flows and supply chains, economists relied on spatially disaggregated price data to make inferences about underlying market structures (Allen, 2014; Atkin and Donaldson, 2015). Spatially disaggregated price data collected by statistical bureaus were my first point of entry into this universe of geo-referenced data. While some of the related research did not make it into this

thesis, the onset of the COVID-19 pandemic and a conversation with my co-author George Kariuki Kinyanjui revived some of those ideas. As Kenya's metropolitan areas around Nairobi and Mombasa went into lockdown and marketplaces closed down, we wondered what implications this would have for food prices. The idea was born to monitor and track prices in a large number of locations across Kenya, in order to analyse how different regions and income groups were affected. As standard face-to-face data collection was not feasible during this early phase of the pandemic, we relied on a crowd-sourcing approach to collect geo-referenced price data. While markets re-opened very swiftly in Kenya and all domestic travel restrictions were lifted by July 2020, the impact of the pandemic on mobility and consumption behaviour was felt for a sustained period of time. We expanded our data collection to Uganda, and Doreen Rubatsimbira joined our research team. This research project became the foundation for my final thesis chapter.

While my thesis primarily relies on quantitative methods and the standard tool kit of economics, qualitative interviews have very much shaped both the research questions I pursued as well as my understanding of various mechanisms that shape patterns I observed in the large-scale administrative data. At the very beginning of my DPhil, I conducted a series of 34 qualitative interviews with multinationals, manufacturers, wholesalers, and retailers in Pakistan and Ethiopia. In 2022, once the COVID-19 pandemic had relaxed its grip on our everyday lives, and now informed by the patterns I observed in the quantitative firm network data, I had the privilege of interviewing 33 private sector firms in Kenya about their supply chains.

So, what can readers expect from this dissertation?

The first chapter of my thesis discusses how administrative tax records can be used to map Kenya's firm network. Peter Wankuru Chacha, Benard Kipyegon Kirui, and I use the data to discuss fundamental properties of Kenya's firm network and how they align with firm networks in other contexts for which similar data are available.

In Chapter 2, joined by Vatsal Khandelwal, we study spatial patterns of domestic firm-to-firm trade in Kenya and explore how these patterns may be shaped by the presence of an unobserved informal sector. We document a high degree of spatial

concentration in the formal firm network, over and above the concentration of aggregate economic activity. Using data from the population census and national accounts, we further show that informality is particularly prevalent in downstream economic activities and smaller regional markets. We link our insights about the sectoral and spatial composition with a network formation model to investigate how accounting for informal firms affects spatial inequality in firm-to-firm trade. Our counterfactual exercise shows that merely subsidising entry of informal firms in regions and sectors with a high incidence of non-VAT paying (informal) firms does not reduce this inequality. Rather, changes in firm-to-firm linking probabilities are required. Our results suggest that data on the formal firm network likely underestimates the connectivity, and consequently also the vulnerability, of smaller regions to shocks that pass through hubs such as Kenya's capital Nairobi.

In Chapter 3, I discuss a potential pathway for reducing search and matching frictions across space, namely the expansion of Kenya's fiber optic network and its potential to foster domestic market integration. I conduct and analyse a series of qualitative interviews to study the use of Internet infrastructure by private sector firms in five Kenyan counties. I find that Internet connectivity indeed lowers the cost of firm-to-firm communication in existing buyer-supplier relationships. At the same time, firms are constrained in their ability to use Internet infrastructure to search for new buyers and suppliers. Such constraints arise as firms in particular lack the capacity to make complementary investments that are needed alongside the use of the Internet. These complementary investments include the expansion of production capacity, hiring employees, and investing in transportation.

The results from Chapter 2 further showed that counties with a high degree of informality are more connected to metropolitan areas, the core of the spatial firm network, than administrative records of formal firms would suggest. This implies that they are likely more vulnerable to adverse international trade shocks, which tend to disproportionately affect firms located at the core of the network. Chapter 4 focuses on the exposure of Kenya's domestic firm network to international markets and implications for firm dynamics during the COVID-19 pandemic. Due to our ability

to observe domestic firm-to-firm transactions, we are not only able to quantify firms' direct reliance on imports and exports, but also further indirect exposure through domestic supply chains. While only a few firms have strong direct or indirect linkages to international supply chains, these firms play an important role in the economy. Peter Wankuru Chacha, Benard Kipyegon Kirui, and I document how Kenyan firms responded to shocks in international markets during the COVID-19 pandemic. We show that favourable shocks to the import markets of Kenyan firms crowded in domestic purchases and supported the recovery from the initial COVID-19 downturn.

Continuing with the question of the economic impact of the COVID-19 pandemic, the final chapter asks how changes in mobility patterns affected food price dynamics during the pandemic and the subsequent recovery in Kenya and Uganda. George Kariuki Kinyanjui, Doreen Rubatsimbira, and I use a crowd-sourcing approach to collect highly granular, geo-referenced price data. We find that prices increase as mobility declines and vice versa. While we lack data to robustly pin down the mechanisms behind this finding, the direction of the effect points towards challenges of the retail sector to keep businesses afloat.

CHAPTER 1

Mapping Kenya's firm network transaction by transaction

with **Peter Wankuru Chacha and Benard Kipyegon Kirui**

1.1 Introduction

There has been renewed interest in understanding the implications of supply chain structures for consumer prices (Grant and Startz, 2022; Atkin and Donaldson, 2015) and the role of firm networks in the propagation of shocks (Carvalho, 2014; Bernard and Moxnes, 2018). Transaction-level administrative data have become an important data source for studying empirical firm networks and answering these questions. The purpose of this chapter is not to ask or answer questions relating to the implications of the structure of Kenyas' firm network, but rather to take the necessary first step of understanding the features and characteristics of said network.

We use transaction-level value-added tax (VAT) records of close to 57,500 firms to examine the structure of Kenya's formal firm network and to illustrate how its fundamental network properties align with other contexts where similar data have been used. We observe the buyer-seller relationships, transaction volume and frequency, sales outside the domestic network, and participation in international supply chains (Chacha et al., 2024), at a monthly frequency. Validating the credibility of the data set, we find that in line with an elementary model of production networks

(Bernard and Zi, 2022) and other empirical contexts (Alfaro-Urena et al., 2018; Bernard et al., 2019, 2022; Cardoza et al., 2023), Kenya's firm network is sparse and exhibits so-called negative degree assortativity — firms with fewer relationships (degrees) tend to link with firms with many relationships.[1] The latter emphasises that well-connected, larger firms matter for policies aimed at the less-connected group of often smaller firms (KNBS, 2016).

With our mapping of Kenya's firm network we are able to add insights from East Africa's largest economy to a growing empirical literature on domestic production networks in different contexts, including Belgium (Dhyne et al., 2015; Bernard et al., 2022; Dhyne et al., 2021), Chile (Huneeus, 2018; Huneeus et al., 2020; Arkolakis et al., 2023; Grigoli et al., 2023), Cost Rica (Alfaro-Urena et al., 2018, 2019, 2022), the Dominican Republic (Cardoza et al., 2023), Ecuador (Brugués, 2023; Adão et al., 2022), India (Panigrahi, 2022; Fujiy et al., 2022; Gadenne et al., 2022; Böken et al., 2023), Japan (Bernard et al., 2019; Carvalho et al., 2021; Kawakubo and Suzuki, 2022; **?**), Rwanda (Spray and Wolf, 2018), Uganda (Spray, 2019; Almunia et al., 2022), the United States of America (Atalay et al., 2011; Lim, 2018; Ghassibe, 2021; Iyoha, 2022), and Turkey (Coşar et al., 2022; Demir et al., 2022, 2023). We further complement the literature on firm-to-firm trade in Kenya, which also draws on administrative data, mainly focused on international trade and relationships of Kenyan firms with trading partners abroad (Macchiavello and Morjaria, 2015; Chacha and Edwards, 2017, 2018, 2019; Majune et al., 2020; Türkcan et al., 2022).

The rest of the paper is structured as follows: Section 1.2 outlines the tax records we use to map Kenya's firm network of formal private sector firms. In Section 1.3 we discuss how the structure of VAT data sets compares to other commonly used sources of data to study private sector development, namely customs records and firm survey data. Section 1.4 provides an overview of the sector composition of the data. Relying on the transaction-level nature of the data, Section 1.5 then replicates key characteristics of firm networks predicted by an elementary model of production

[1]The literature on domestic firm networks has yet to fully converged on standard measures and descriptive statistics of network properties. We, therefore, restrict our focus to properties we can compare to some of the other contexts for which similar data is available.

networks (Bernard and Zi, 2022) and discusses implications of the basic firm network structure.

1.2 Data description

To map Kenya's firm network, we draw on a series of firm- and transaction-level tax records collected by the Kenya Revenue Authority (KRA) between January 2015 and March 2021. From the tax records, we compile two monthly panel data sets: a relationship-level panel and a firm-level panel. Basic firm characteristics such as the sector of operation, start date, headquarter location, and business type are compiled from tax registration forms. Table 1.1 provides an overview of the underlying tax records. All data sets can be linked using anonymised firm identifiers. We discuss our processing and cleaning protocol of the raw data in Appendix 1.B.

Table 1.1: Overview of administrative data sources

Name	**Frequency**	**(Dis-)Aggregation**	**Key indicators**
Tax registration forms	Static	Firms	Sector, location, start date
Value-added tax records	Daily	Transactions	Trading partners, volume, product description
	Monthly	Firms	Sales, purchases
Pay-As-You-Earn	Monthly	Firms	Payroll, # employees
Customs	Daily	Transactions	Volume, quantity, HS code, partner country
Corporate Income Tax	Annual	Firms	Assets, liabilities

All data are collected by the Kenya Revenue Authority.

Of all the tax returns listed in Table 1.1, the VAT returns stand out as the most crucial. As part of the VAT returns, firms report transaction-level details on sales to and purchases from other VAT-registered entities. The transaction-level records allow us to track the supplier and buyer for each firm-to-firm transaction, including transaction dates, unstructured product descriptions, and transaction volumes. Sales to and purchases from non-registered parties (e.g., exempt parties, non-registered businesses, final consumers) are recorded as an aggregate monthly figure.

VAT is applicable to individuals and firms with a turnover exceeding KShs five million ($40,500 as of December 2022). Once a firm becomes VAT-registered and crosses the threshold of KShs five million, they are mandated to continue filing VAT returns, even in years with lower turnover. However, firms below this threshold, as

well as those offering financial and education services, and to a large extent, those dealing with a range of unprocessed agricultural commodities and pharmaceuticals are exempt from VAT. These firms are not required to register for VAT or submit monthly returns. The category of sales to non-registered parties encompasses sales to firms below the threshold and exempt firms. Similarly, the category of purchases only includes reporting of purchases of VAT-exempt goods, leading to the exclusion of various potential inputs from the non-registered sector. While this introduces a bias in the representation of economic activity in the data, we use firm survey data collected by KNBS (2016) in a companion paper (Chacha et al., 2023) to suggest that this bias might not be as severe, as small firms tend to purchase more from larger VAT-registered firms than the other way around. Throughout this paper, we refer to private sector entities that are either not VAT-registered due to their size or exempt from VAT payments due to the products and services they sell as "non-VAT registered" firms or sectors.[2] These constraints are most important for the interpretation of any network statistic on the agricultural sector. In 2019, the agricultural sector contributed 19% of Kenya's GDP. The VAT data, however, only capture large commercial farms and aggregators only. Our analysis based on VAT data is therefore representative of these types of farms only.

Due to our focus on private sector development, we filter the data set to include only entities that identify as private companies or partnerships in their tax registration forms. This process involves excluding all government-owned firms, government agencies, international organizations, NGOs, trusts, and clubs. Additionally, we omit firms operating in the financial sector, mainly comprising banks and insurance companies. Our analysis is then focused on firms with annual purchases greater than zero and annual sales of KShs five million or more in at least one year, as observed in the data. We also apply this threshold to exclude firms that registered for VAT solely to bid for tenders but were barely active otherwise. Across all the years the filtered

[2]Broadly speaking, Kenya has a three-tier system for businesses: (i) businesses that are VAT-registered, (ii) businesses that have a license but are not VAT-registered, and (iii) businesses that do not have either. The latter category corresponds to the classic definition of informal businesses, usually household-run micro businesses. According to KNBS (2016), there were 5.85 million unlicensed businesses operational in Kenya in 2016, with 94% of them reporting a monthly revenue of less than KShs 50,000 (about $400).

dataset comprises 57,482 firms, forming 4,117,880 supplier-buyer relationships. In each year, we observe 45,500 active VAT-registered firms on average between 2015 (33,000 firms) and 2020 (54,000 firms). This figure is strikingly close to the over 44,000 firms observed in the Dominican Republic by Cardoza et al. (2023) (see Table 1.3). Despite the Dominican Republic having less than a quarter of Kenya's population and a GDP per capita 4.3 times larger than Kenya's, the overall size of the two economies remains remarkably similar in terms of total GDP.

1.3 How does the content of VAT data compare with customs records and firm survey data?

In this section, we discuss how the VAT data compare to more commonly used customs records and survey data. We present an overview of common features of the three categories of data sets in Table 1.2.

Table 1.2: Comparison of common data features

Feature	**VAT data**	**Customs records**	**Surveys/census***
Units of observation	Transactions	Transactions	Firms/products
Geographic scope	National	International	Any
Frequency	Daily, monthly	Daily	Annual
Goods/services	Several sectors	Tradable products	Any, often sector-specific
Tax status	Formal	Formal	Formal, informal
Network information	Two-sided	(Mostly) one-sided	Two-sided
Prices and quantities	No	Yes	Yes

*Firm survey data are often collected with a specific research question in mind and context-specific. The table assumes the structure of widely used firm surveys collected across the world, such as the Large and Medium Scale Manufacturing and Electricity Survey in Ethiopia (Gebrewolde et al., 2022), the Annual Survey of Industries in India (Boehm and Oberfield, 2020), and the Manufacturing Survey of Large and Medium-Sized Firms in Indonesia (Amiti and Konings, 2007; Kassem, 2021).

1.3.1 Comparison with customs data

While customs data map international trade transactions, VAT data focus on domestic firm-to-firm transactions. The two types of data are, however, very similar in spirit and in many contexts collected by the same government agency. An advantage shared by both sources of data is that they track firm-to-firm trade at a

high frequency, usually detailing the exact invoice date.[3] A shared disadvantage of the data is that they, by definition, only capture formal sector activity of tax-paying firms. In the following, we focus on the distinguishing features between each.

VAT records have two main advantages over customs records. First, for researchers interested in studying network and relationship dynamics of firm-to-firm trade, VAT data provides detailed information on both the supplier and the buyer involved in a transaction. With some rare exceptions (e.g., customs records from Norway (Bernard et al., 2018; Bernard and Zi, 2022)), customs data do not capture the identity of the trade partner abroad. Studies on relationship dynamics using customs data tend to use less fine-grinned definitions of trading relationships, such as classifying a relationship at the firm-country-product level. Even in contexts where the trade partner is identified, researchers can only observe the full firm size distribution on one side of the trading relationship. For example, the Norwegian customs data, only capture South African firms trading with Norway rather than the entire distribution of South African importers and exporters.

Second, due to the comprehensive mapping of the firm network, VAT data allow researchers to trace indirect firm-to-firm links. For example, most firms ultimately rely in part on foreign inputs, but customs data only capture direct importers and exporters. This data feature is a particular constraint when studying importer and exporter characteristics in contexts where a lot of international trade takes place via intermediaries. Ahn et al. (2011) partly circumvent this problem by using firm names to identify firms operating as intermediaries rather than producers in Chinese data. Their analysis shows that the export share of intermediaries correlates negatively with the destination country's GDP. While Ahn et al. (2011)'s approach makes an important contribution to the literature by putting the spotlight on wholesalers, it still misses firms trading with those intermediaries — the indirect importers and exporters themselves. Dhyne et al. (2021) show that especially on the import side indirect exposure to international supply chains is highly prevalent among firms in Belgium. In a companion paper to this one, Chacha et al. (2024)

[3]This feature is not shared across all VAT data sets. Some like the one discussed in Huneeus (2018) for Chile, only capture annualised information on firm-to-firm trade.

show that while only a few Kenyan firms ultimately source more than 25% of their imports from abroad, these firms account for almost half of the output of Kenya's formal sector. Lastly, customs data often omit detailed information on the actual location of the firm and largely focus on manufactured goods, while omitting services.[4]

Most VAT records have two key disadvantages relative to customs data. First, most VAT data sets do not contain any information on the products and services traded, e.g., in Belgium (Dhyne et al., 2015) and Turkey (Demir et al., 2023). If they do, the product descriptions come in the form of unstructured text like in Uganda (Spray, 2019) and Kenya. Second, with the notable exception of a small subset of data from Ecuador (Brugués, 2023), VAT data only captures the volume of firm-to-firm transactions, but not the quantity of the goods traded. Therefore, it is not possible to compute prices from most of the VAT data sets.

1.3.2 Comparison with firm survey data

Firm surveys are the most widely used, but also the most heterogeneous source of data to study private sector activity. Such surveys can range from tailored surveys by governments and researchers, to standardised, regular (often annual) surveys with national coverage like the Annual Survey of Industries in India and the Large and Medium Scale Manufacturing and Electricity Industries Survey in Ethiopia. The exact comparison with VAT data of course depends on the nature of the survey at hand.

As discussed, VAT data present a unique source of detailed information on suppliers and buyers of a firm. Designing a survey instrument that collects a detailed set of information on each supplier of a firm could be extremely challenging, especially for large firms that often have several hundred or even thousands of suppliers. In addition, it is even harder to do so for ten of thousands of firms in an attempt to map the firm network in a comprehensive manner.[5]

[4]There are of course data sets that track service sector trade, e.g., for the UK, but it rarely features in customs records by default.

[5]While providing highly valued and frequently used data sets to researchers, data providers like *Compustat*, for example, capture only trading partners with a sales share of 10% or more for listed US firms (Atalay et al., 2011). Between 1977 and 2016 the *Compustat* data captures about 18,000 firm-to-firm relationships between close to 8,500 firms in the United States (Iyoha, 2022).

Administrative data presents firm-level data at a relatively lower cost in terms of time and resources. Implementing firm surveys with national coverage can be very costly. While many countries conduct a census of production with national coverage at regular intervals, some even annually, these censuses are usually not available at higher frequencies. For example, Kenya's Census of Production is held after every five years. On the other hand, VAT data are often available at an annual, quarterly, monthly, and even daily frequency. Monthly frequency is beneficial for tracking the responses to shocks and accounting for cases where firms' sales are highly seasonal. Administrative data provides large coverage of firms at a longer horizon, making them suitable for dynamic analysis. Surveys, in turn, are usually either conducted at infrequent intervals over a longer time horizon or during a short window of time at a higher frequency.

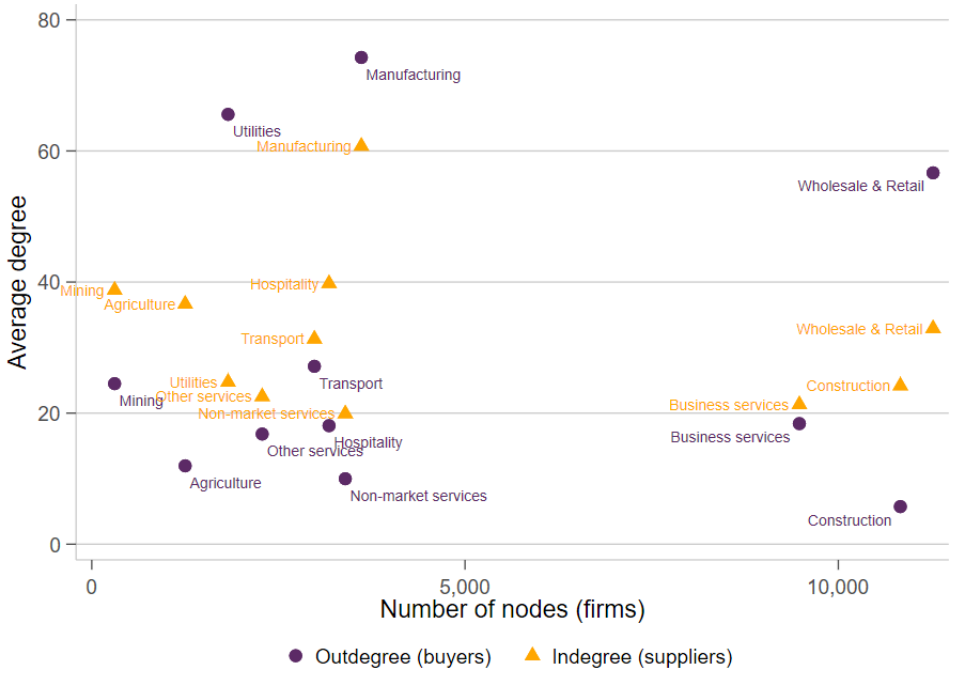
The main disadvantage of VAT data over firm survey data is that they do not cover informal firms. In the case of Kenya, sales to non-VAT paying firms are included in aggregate monthly sales figures, but cannot be distinguished from sales to final consumers. Firm survey data often contain a more comprehensive set of information that can be tailored towards a specific research or policy question, some even capturing transaction-level information (Startz, 2021). In a nutshell, VAT data usually contain a narrow set of information, but with a wider coverage than most firm surveys. A downside of the Kenyan VAT data relative to other VAT data sets (see for example Alfaro-Urena et al., 2018; Dhyne et al., 2021) is that we are unable to map affiliations between firms in the network.

1.4 Sector composition of the firm network

We start out by looking into the sector composition of Kenya's VAT firm network. The most common sectors of operation for Kenya's VAT firms are wholesale and retail, construction, and business services. In 2019, we observe just over or under 10,000 active VAT-paying firms in each of them. Business services are compromised of professional services, information and communication, and administrative services (see Figure 1.1). In terms of sales volumes, the manufacturing, and wholesale and

retail sectors jointly account for nearly 50% of the overall sales observed in the VAT data. This sector composition of VAT sales is not unique to Kenya and similar to the one documented for Chile (Huneeus, 2018), Costa Rica (Alfaro-Urena et al., 2018), and the Dominican Republic (Cardoza et al., 2023). In the latter two cases, wholesale and retail, and manufacturing sectors account for over 60 and 70% of aggregate sales, respectively. Trade between the two sectors is the largest of all inter-sectoral trade flows mapped in Figure 1.2.

Figure 1.1: Average degree by sector



The above graph plots the sector-level average in- and outdegree degree against the number of firms that operate in each sector.

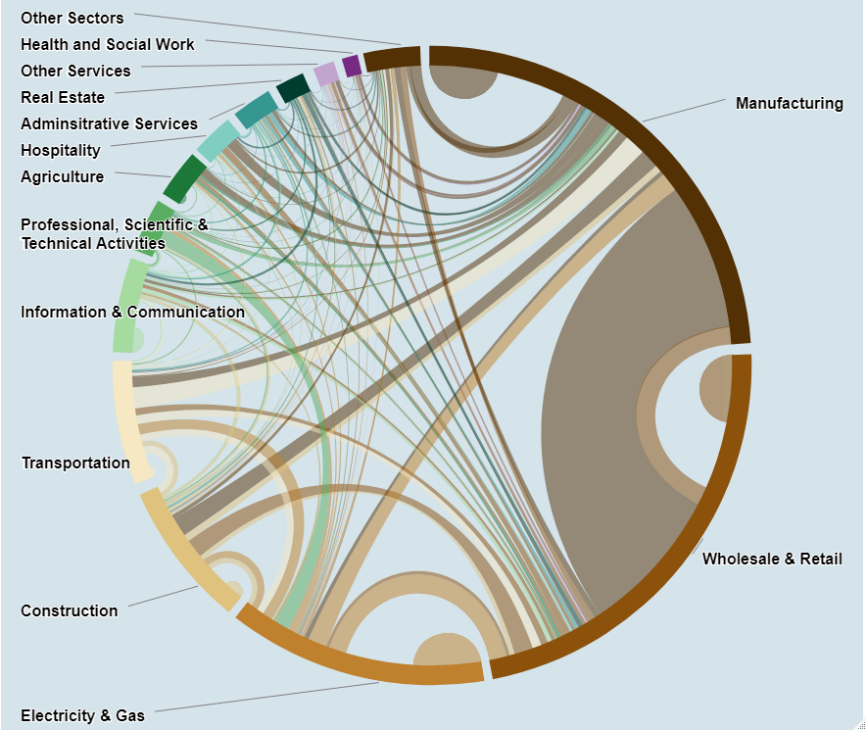
The prominence of the wholesale & retail and manufacturing sectors in the firm network is also reflected in the high average outdegree of firms in these sectors. Figure 1.1 shows that the average manufacturing firm has an outdegree of 74, while wholesale and retail firms sell to 57 other formal firms on average. While fewer in their numbers, firms in the utilities sector also have a high outdegree with 68 customers on average. This is not surprising given they provide essential inputs like electricity, gas, and water. The three sectors each account for 15% or more of sector-level inputs on

average across all sectors. In other words, for each sector, we compute how much of both its own and each other sector's inputs it accounts for and then take the average across all sectors. We plot this average input share against the average Bonacich-Katz centrality of firms in each sector in Figure 1.3.[6] Intuitively, based on the Bonacich-Katz centrality firms are considered central if they are linked to other central/well-connected nodes (Acemoglu et al., 2012; Sargent and Stachurski, 2022). Transportation and information & communication firms emerge as two other sectors with relatively more central firms in the network, despite the volumes they supply to others being smaller.[7]

[6]We group sectors into centrality deciles as centrality as a continuous measure does not bear a particularly insightful interpretation.

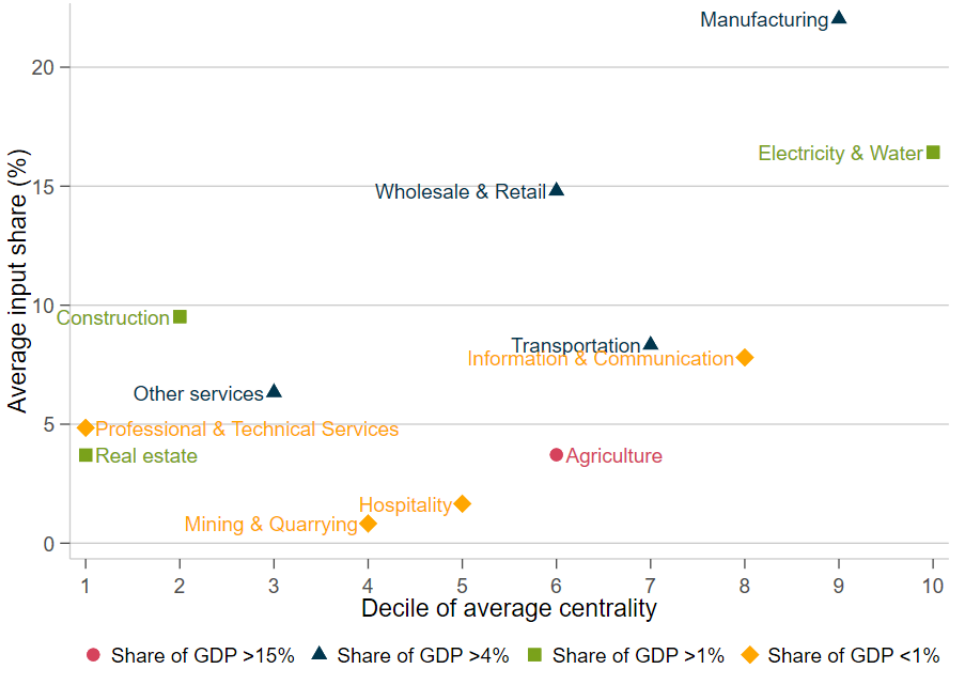
[7]For this particular graph we break up business services to highlight the relatively more central role of information & communication sector firms.

Figure 1.2: Sector-level trade flows between VAT-registered firms



This figure shows inter-firm trade flows aggregated at the sector level. The size of each node (segment) is proportional to the sector's share of purchases and sales relative to the aggregate volume of firm-to-firm trade between formal firms in Kenya. The colour of the edges (links between segments) indicates the direction of the trade flow. They take the colour of the supplying sector (e.g., goods and services provided by the transportation sector to the manufacturing sector takes the colour of the transportation sector's segment). The width of each edge (links between segments) is proportional to the share of the trade flow with respect to the aggregate volume of trade flows in the VAT data.

Figure 1.3: Average input share and average centrality by sector



Each marker represents a sector. The different colours and shapes of the sector markers reflect their contribution to Kenya's aggregate GDP.

1.5 Fundamental properties of the network

We now turn to the smallest unit of the administrative records, firm-to-firm transactions, to look at the fundamental properties of Kenya's firm network.

To aid comparison we focus on properties of firm networks that have been widely documented in other contexts:[8] sparsity,[9] a positive correlation of firm sales and both firm-to-firm links and geographic scope, a skewed in- and outdegree distribution, and negative degree assortativity (firms with fewer links match disproportionately to firms with many links). Bernard and Zi (2022) show that these widely documented empirical regularities can be expected to hold in a world with firm size heterogeneity, but without putting any particular structure on firm-to-firm matching behaviour.[10] Put differently, conditional on an underlying firm size distribution, a random allocation of buyer-supplier transactions will yield these very basic patterns. This has two implications for our purposes: First, the model serves as a useful benchmark for the empirical structure of Kenya's firm network and the list of properties to look at for validating the data. Second, the properties we are going to discuss do not yield any particular insights into firm-to-firm matching behavior (this is true in any setting, not just that which we discuss here). In a nutshell, the following explorations are a useful exercise, but researchers will need to look beyond those if we want to learn more about the search and matching behavior of firms.

[8]This includes firm-to-firm matching patterns derived from both customs records (see among others Carballo et al., 2018; Bernard et al., 2018) and domestic network data (see among others Dhyne et al., 2015; Alfaro-Urena et al., 2018; Cardoza et al., 2023; Bernard et al., 2022)

[9]In an "informal" sense (Newman, 2018), i.e. not formally tested.

[10]This framework is part of the family of balls-and-bins models, where each finite number of discrete transactions in the firm-to-firm data is interpreted as a ball held by a buyer. Crucially, sellers and buyers vary in their size. Sticking with the balls-and-bins analogy, sellers correspond to bins of different sizes, while buyers vary in the number of balls they hold. Each ball then gets randomly assigned to a bin. The elementary model only makes a statement about the matching process, not the fundamentals of the firm size distribution. Crucially, Bernard and Zi (2022) show that the elementary model outperforms many alternative models that specifically target some of the moments found in firm-to-firm trade data. Negative degree assortativity, for example, motivated researchers to write models with relationship fixed cost (see Bernard and Moxnes, 2018, for a review of the literature). Bernard and Zi (2022) show that fixed costs are not a necessary condition for negative degree assortativity. A downside of the balls-and-bins model is that it provides little insight into network formation itself and its success in pinning down network moments primarily stems from the fact that it takes the firm size distribution as given.

1.5.1 Density of Kenya's production network

Like other firm networks, Kenya's firm network is sparse. While firm networks share sparsity as a characteristic, the network density varies substantially across contexts from $9.5e-7$ in India (Panigrahi, 2022) to $6e-5$ in Cost Rica (Alfaro-Urena et al., 2018) (see Table 1.3).[11] Kenyan firms are, on average, linked to only 30 other firms. This means only 1 in 876 potential relationships is realised. The range in average degrees across different contexts (4 to 56) can partly be explained by differences in reporting requirements of firm-to-firm transactions with higher average degrees found in contexts with no specified reporting threshold for individual transactions (see Table 1.3).

Table 1.3: Density of domestic production networks and tax reporting requirements

Country	**# of nodes** in '000	**Avg degree**	**Density**	**Reporting threshold** ann. transaction volume	**Reference**
Costa Rica (2014)	33	8	$6.0e-5$	$4,800	Alfaro-Urena et al. (2018)
Dominican Republic (2012-19)	44	56	$4.5e-5$	$0	Cardoza et al. (2023)
Kenya (2019)	52	30	$3.9e-5$	$0	this paper
Chile (2011)	294	22	$6.8e-6$	$385,000 in sum*	Huneeus (2018)
Japan (2005)	950	4	$2.1e-6$	up to 24 connections**	Bernard et al. (2019)
India (5 states, 2015)	2,107	11	$9.5e-7$	N/A	Panigrahi (2022)

The above table reports the observed network's density, the number of nodes (firms), and the average degree found for domestic firm networks across a range of contexts relying on similar sources of data. The network's density can be approximated by the average degree divided by the number of nodes (Newman, 2018). Except for Japan, all papers utilise VAT data to map firm-to-firm relationships. A key difference between countries is reporting thresholds for firm-to-firm transactions captured in column 4. *In Chile, firms that have ever reached a threshold of $385,000 annually for their purchases from intermediaries have to report their transactions. ** In Japan, firms are asked to report up to 24 buyers and suppliers, irrespective of the transaction volume. The statistics reported in Panigrahi (2022) for India captures the five states Gujarat, Maharashtra, Tamil Nadu, Odisha, and West Bengal. All figures represent a cross-section for the year reported after the country name.

1.5.2 Heterogeneity in firms' degrees

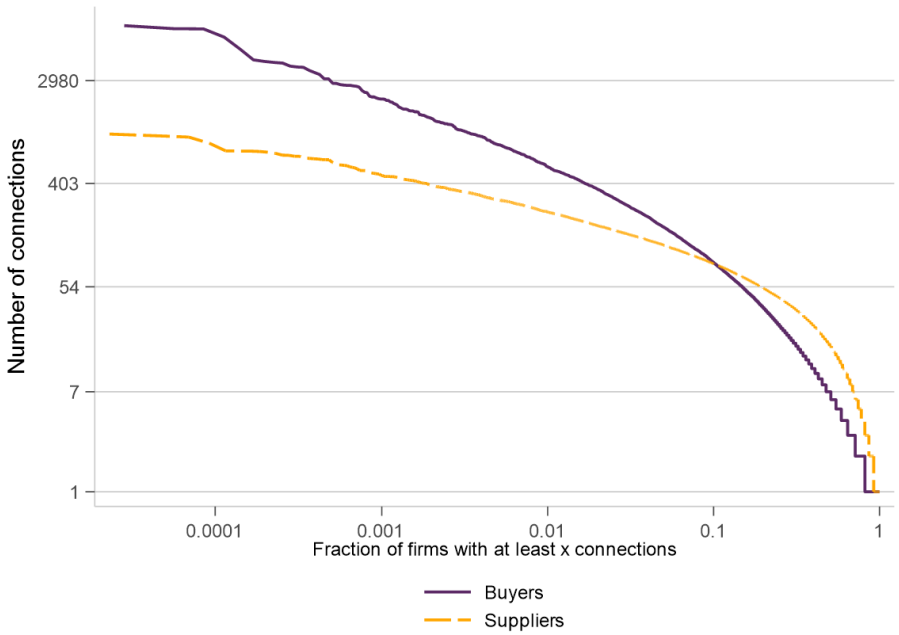
Firms vary substantially in the number of buyers (out-degree) and suppliers (in-degree) they have. Figure 1.4 plots the inverse CDF of both the in- and out-degree distribution. The median firm in Kenya has 3 buyers and 13 suppliers. In contrast, firms in the 99th percentile have several hundred up- and downstream linkages (Table 1.4 and 1.5).[12] The inequality in in- and outdegrees varies by sector. We find the most equal indegree distribution in manufacturing and hospitality (see Table

[11] Network density ρ can be approximated by dividing the average degree c by the number of nodes N (Newman, 2018).

[12] The distribution of supplier and buyer-linkages (in- and out-degrees) is very stable over the six years observed in the data.

1.A1 and Table 1.A2). For comparison, Tables 1.4 and 1.5 further lists moments of the in- and out-degree distribution for Belgium (Bernard et al., 2022), Costa Rica (Alfaro-Urena et al., 2018), the Dominican Republic (Cardoza et al., 2023), and five states in India (Panigrahi, 2022) reported in the literature.

Figure 1.4: In- and outdegree CDFs



The figure shows the inverse cumulative distribution function (CDF) for the in- and out-degree. Note the log scales on both the y- and the x-axis.

A firm's sales volume correlates positively and strongly with both the number of suppliers and buyers (Figure 1.5) as well as the geographic scope of those relationships (Figure 1.6).[13] Figure 1.6 shows that firms with higher sales also have a larger geographic scope, i.e. they are selling to and scouring from a larger number of sub-counties (Kenya's 2nd administrative layer). Both empirical regularities have been documented in other contexts (Alfaro-Urena et al., 2018; Bernard et al., 2019; Cardoza et al., 2023). Conditional on the heterogeneity in the firm size distribution, Bernard and Zi (2022) show that random allocation of transactions can explain this

[13]The pattern holds if we control for the sector and firm location. We chose to show the unconditional graph to facilitate comparison with the simulated data. In the simulation described below, we do not put any constraints on which firms can match with which sectors.

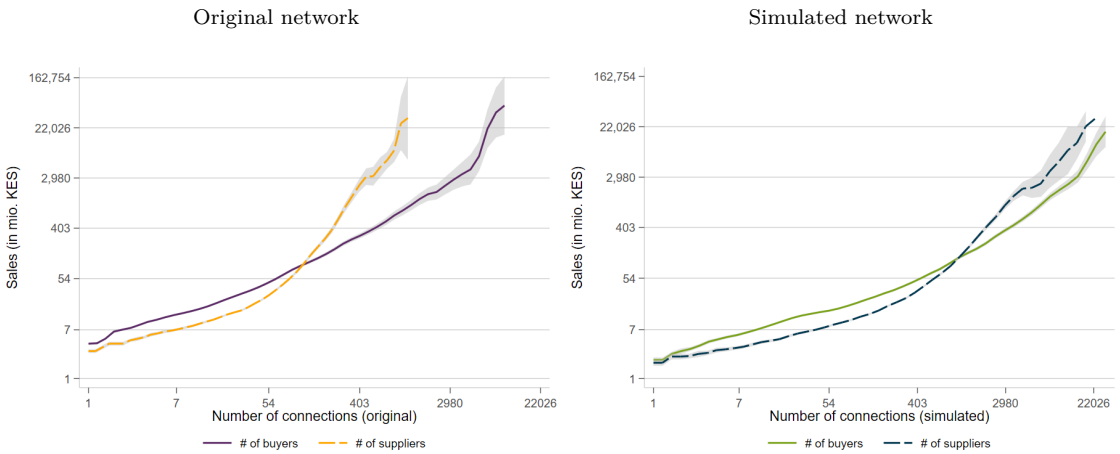
Table 1.4: Moments of the outdegree distribution: cross-country comparison

Country (year)	Number of suppliers	Number of buyers per firm							
		Mean	Sd	10th	25th	50th	75th	90th	99th
Kenya (2019)	51,749	30	151	0	0	3	14	56	456
Firms with min. one buyer									
Kenya (2019)	37,508	41	177	1	2	6	24	83	552
Costa Rica (2015)	24,741	21	141	1	2	4	11	33	290
India (5 states, 2015)	1,200,900	-	-	1	2	5	15	44	-
Firms with min. one employee									
Kenya (2019)	23,019	61	222	1	4	12	44	134	764
Dominican Republic (2017)	32,112	57	364	1	3	9	28	94	769
Belgium (2014)	94,334	123	801	3	-	26	-	152	1297

This table compares the distribution of firm-level links to buyers for Kenya, Costa Rica, the Dominican Republic, and Belgium. For Costa Rica we take the statistics from Alfaro-Urena et al. (2018), for the Dominican Republic from Cardoza et al. (2023), and for Belgium from and Bernard et al. (2022). We group the statistics according to restrictions imposed on the sample of firms in the respective papers.

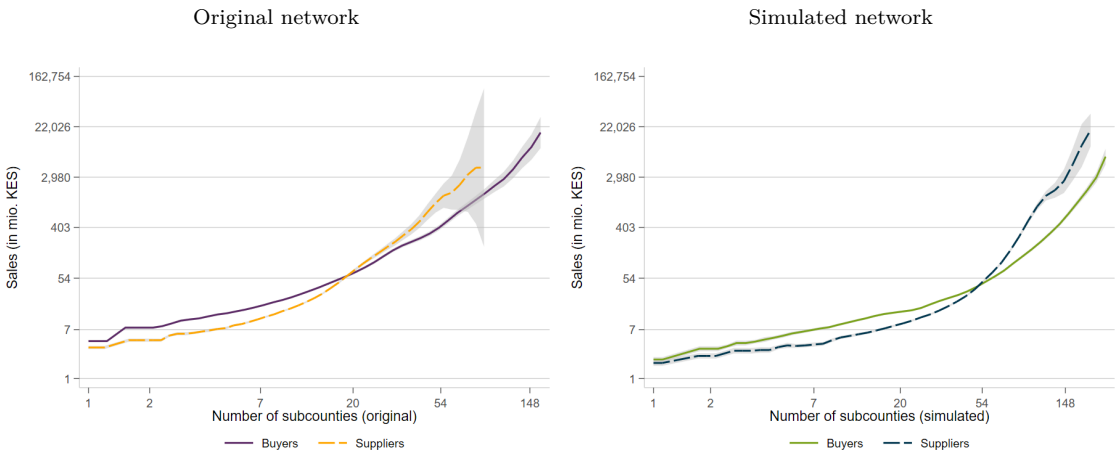
regularity. Following this notion, we simulate the matching of firms by randomly allocating transactions to suppliers. The number of transactions each buyer holds and each supplier receives is determined by the data, i.e. the underlying firm size distribution. The right graphs in Figures 1.5 and 1.6 highlight (i) how closely random matching can replicate the actual matching process and (ii) highlights the quality of the underlying data. While the positive sign of the correlation is the same, some qualitative differences, like the shorter tail of the empirical in-degree degree distribution (fewer suppliers for a given sales volume), are noteworthy. Once a certain threshold is reached, large firms do not expand their supplier base in proportion to their buyer base. At the same time, smaller firms have more suppliers in the empirical relative to the simulated data. The pattern reflects that most firm activities require a minimum number of inputs. At the same time, demand channels can vary more substantially across firms, with some selling mainly to other businesses, some selling to export markets, and others to final consumers, the public sector, or informal firms. We will return to this aspect in the next chapter of the thesis (Chacha et al., 2023).

Figure 1.5: Sales and the number of firm-to-firm linkages



The figure shows the local polynomial regression of log sales on the number of connections within the firm network. The left graph plots firm sales against the observed number of links in the original network data, while the right graph plots firms sales against the number of buyers and suppliers based on the simulated network data. The solid lines show the number of buyers, and the dashed lines show the number of suppliers.

Figure 1.6: Sales and the number of domestic sales destinations



The figure shows the local polynomial regression of log sales on the number of sub-counties a firm has linkages to. Sub-counties represent the Kenya's administrative layer. The left graph plots firm sales against the observed number of links in the original network data, while the right graph plots firms sales against the number of locations based on the simulated network data. The solid line shows the number of sub-counties the firm sells to, and the dashed line shows the number of sub-counties the firm buys from.

Table 1.5: Moments of the indegree distribution: cross-country comparison

Country (year)	Number of buyers	Number of suppliers per firm							
		Mean	Sd	10th	25th	50th	75th	90th	99th
Kenya (2019)	51,749	30	48	1	3	13	37	77	220
Firms with min. one supplier									
Kenya (2019)	46,769	33	50	2	5	16	41	82	227
Costa Rica (2015)	31,841	14	39	2	3	6	13	29	32
India (5 states, 2015)	1,285,300	-	-	1	3	8	21	44	-
Firms with min. one employee									
Kenya (2019)	26,008	48	60	4	12	30	61	110	282
Dominican Republic (2017)	37,986	48	68	6	14	30	59	105	311
Belgium (2014)	94,334	79	109	20	-	53	-	152	466

This table compares the distribution of firm-level links to suppliers for Kenya, Costa Rica, the Dominican Republic, and Belgium. For Costa Rica we take the statistics from Alfaro-Urena et al. (2018), for the Dominican Republic from Cardoza et al. (2023), and for Belgium from Bernard et al. (2022). We group the statistics according to restrictions imposed on the sample of firms in the respective papers.

1.5.3 Disassortative mixing by degree in firm networks

Firm networks have been found to diverge from social networks when it comes to matches between well-connected and less-connected firms. Social networks feature so-called assortative mixing (Newman, 2018), i.e. popular members tend to befriend other popular network members. For firm networks, the reverse is true - a phenomenon described as disassortative mixing by degree, often referred to as negative degree assortativity in the literature on empirical production networks (Bernard and Moxnes, 2018).[14] Less well-connected buyers and suppliers tend to be connected to highly connected firms. Highly connected firms, in turn, are connected to firms across the board, but their buyers/suppliers have, on average, fewer connections. To confirm the quality of the tax records, we compute the normalised covariance of firm's degrees, the so-called associativity coefficient (Barabási, 2014; Newman, 2018). In a directed firm network, we can look at four types of correlations: the correlation of a firm's indegree with its suppliers' indegrees (in-in) or alternatively the suppliers' outdegree (in-out), as well as the firm's outdegree with its buyers' indegrees (out-in) or outdegrees (out-out) (Barabási, 2014). We summarise the coefficients from the data

[14] Newman (2018) discusses how negative assortativity by degree is likely the norm rather than the exception across different types of networks. The positive assortativity by degree observed in social networks is likely driven by group structures that arise in such networks (Newman, 2018).

and the simulated network in Table 1.6.[15] All four coefficients are negative. However, the in-out and out-out correlations are barely negative. The correlation coefficients for the simulated network, which tends to be denser than the real-world data, are substantially larger in absolute terms. The relatively small negative coefficients in the empirical data could point towards potential matching frictions between large and small firms, i.e. while the two tend to link with one another, they link to a lesser extent than predicted by random matching.

Table 1.6: Degree correlation coefficients

Source	In-In	In-Out	Out-In	Out-Out
Simulated matching	-0.13	-0.09	-0.21	-0.15
Original network	-0.06	-0.03	-0.08	-0.04

Irrespective of the exact underlying matching process that gives rise to negative degree assortativity in firm networks, negative degree assortativity has important implications for policy considerations aimed at small and medium-sized enterprises: Smaller firms are almost always linked to larger firms. They rely on them for both inputs and outputs. What are the consequences? Various shocks affecting larger firms, who tend to be more likely linked to international markets (the focus of Chapter 4 of this thesis (Chacha et al., 2024)), might propagate through to smaller firms. Similarly, new developments like digitisation can both drive wedges between smaller and larger firms as larger firms have advantages when it comes to adoption (Foster et al., 2018), but likewise, being embedded in a supply chain with larger firms can facilitate technology adoption by smaller ones (Foster et al., 2018; Awori et al., 2022).

1.6 Conclusion

In this chapter, we discussed how we use transaction-level records from VAT data to study Kenya's firm network of private sector firms. In line with domestic firm networks in other contexts, Kenya's production network has a low density and

[15] Alternatively, one can plot the average degree of buyers or suppliers against the firm's own degree. As noted by Barabási (2014), graphical evidence can be hard to compare across networks. We put a number on the correlation to allow for comparison with other contexts in the future.

exhibits so-called negative degree assortativity — firms with few network links tend to link with firms that are very well connected. The latter property highlights that even in a context where promoting small and medium-sized businesses is a priority since they employ the majority of people in the labour force, the role of large firms in their supply chains deserves some attention. It, for example, illustrates the need to consider the eco-system of enterprises as a factor for decisions like technology adoption of smaller firms (Foster et al., 2018; Awori et al., 2022).

The chapter has largely neglected two important aspects of Kenya's firm network: First, we have not touched on the spatial dimension of the network. Second, we have largely ignored the fact that we are studying a network of formal firms in a context with a substantial informal sector (Schneider and Enste, 2000). The spatial dimension and the informal sector will be the focus of the upcoming chapter.

Acknowledgements for Chapter 1 We thank the Kenya Revenue Authority (KRA) for the outstanding collaboration. Romeo Ekirapa, Simon Mwangi, and Benard Sang provided excellent technical support and advice. A related project on tax incentives done jointly by the authors for KRA with assistance from the data support team gave birth to this work. We thank Daniela Villacreces Villacis for excellent research assistance during the data processing and cleaning stage. This chapter has greatly benefited from extensive conversations and discussions with Andrew Bernard and Yuan Zi. We thank Andrea Bacilieri, Raphael Bradenbrink, Douglas Gollin, Vatsal Khandelwal, Lisa Martin, Luke Heath Milsom, Sanghamitra Warrier Mukherjee, Solomon Owusu, Alexander Teytelboym, Christopher Woodruff, Hannah Zillessen, and participants of the CSAE Annual Conference, the CSAE/ISSER Workshop, and the Jeune Street Seminar Series for comments and feedback. We gratefully acknowledge financial support from the Private Enterprise Development in Low Income Countries (PEDL) programme, a joint initiative by Centre for Economic Policy Research (CEPR) and the UK Foreign, Commonwealth & Development Office (FCDO). Verena Wiedemann further acknowledges funding from the Oxford Economic Papers Research Fund (OEP) and the German National Academic Scholarship Foundation. This study has been approved by the Department of Economics Research Ethics Committee at Oxford (protocol no.ECONCIA20-21-23), and the Kenyan National Commission for Science, Technology and Innovation (protocol no.NACOSTI/P/20/5923). The views in this paper are those of the authors, and do not necessarily represent those of the KRA or any other institution the authors are affiliated with.

Appendices of Chapter 1

1.A Additional tables

Table 1.A1: Number of indegrees by sector

Sector	Number of buyers	Number of suppliers per firm							
		Mean	Sd	10th	25th	50th	75th	90th	99th
Wholesale and Retail	11,269	33	51	1	4	16	43	83	226
Construction	10,831	24	45	0	2	8	28	64	209
Manufacturing	3,611	61	72	2	11	38	81	148	344
Professional, Scientific, and Technical	3,283	23	32	1	3	12	29	57	154
Administrative and Support Services	3,254	22	33	0	2	10	29	58	147
Acommodation and Food Services	3,178	40	52	2	8	25	51	94	240
Transportation and Storage	2,982	31	45	1	4	15	39	82	205
Information and Communication	2,941	19	46	0	2	7	23	49	144
Electricity and Gas	1,564	23	46	1	2	8	26	59	188
Agriculture, Forestry, and Fishing	1,251	37	51	1	3	18	49	101	244
Total	51,749	30	48	1	3	13	37	77	220

This table shows indegree distribution by 2-digit sector for 2019.

Table 1.A2: Number of outdegrees by sector

Sector	Number of suppliers	Number of buyers per firm							
		Mean	Sd	10th	25th	50th	75th	90th	99th
Wholesale and Retail	11,269	57	233	0	1	7	38	127	739
Construction	10,831	6	37	0	0	0	2	9	94
Manufacturing	3,611	74	198	0	3	16	62	181	866
Professional, Scientific, and Technical	3,283	15	55	0	1	3	11	30	210
Administrative and Support Services	3,254	18	95	0	0	2	9	30	281
Acommodation and Food Services	3,178	18	80	0	0	3	13	37	232
Transportation and Storage	2,982	27	125	0	1	3	11	40	526
Information and Communication	2,941	23	196	0	1	2	9	33	322
Electricity and Gas	1,564	68	264	0	1	7	44	176	829
Agriculture, Forestry, and Fishing	1,251	12	37	0	0	2	7	26	188
Total	51,749	30	151	0	0	3	14	56	456

This table shows the outdegree distribution by 2-digit sector for 2019.

1.B Data processing and cleaning protocol

In this Appendix section, we discuss how each of the data sets in Table 1.1 has been processed and cleaned.

1.B.1 Value-added-tax returns

For the processing of the VAT returns we follow a number of robustness checks outlined in the literature (in particular Dhyne et al., 2015; Huneeus, 2018; Alfaro-Urena et al., 2018), and add some additional tests to account for the specific structure of tax returns in Kenya.

Basic plausibility checks

Of the VAT data sets currently used in the academic literature, the structure of the transaction-level data closely resembles the one in Uganda (see Spray (2019); Almunia et al. (2022) plus a more detailed description of the data structure in Almunia et al. (2017)) and Belgium (see Bernard et al. (2022) plus Dhyne et al. (2015) for a detailed description of the data). A key difference to the data used for research on production networks in Belgium, but also others like the VAT data from Chile Huneeus (see 2018), is that the Kenyan data are retrieved from a database that has already been screened and potential filing errors have been rectified by tax officials. This screening process together with the electronic tax reporting system employed by KRA, which automatically implements many checks and computes the tax levy, explains the small number of obvious errors we find in the raw data in comparison to Dhyne et al. (2015). We are able to show this using a series of plausibility checks largely following major steps of Dhyne et al. (2015)'s data cleaning protocol. Like Dhyne et al. (2015) we exploit the fact that we observe the VAT payment, the transaction volume as well as the tax rate applied. First, we check for sales and purchases with a transaction volume of zero (concerns only 0.14% of all transactions) and impute the transaction value using the reported tax rate if we observe a related positive VAT payment. Second, we compare the value of the transaction (net of tax) to the amount of VAT due. In less than 100 of the cases we observe a transaction value that is equal to or smaller than the amount of VAT due. For these cases, we cross-check the transaction records with those of the buyers, which reveal an error in the computation of the

amount of VAT due. We thus assume the sales value to be correct and impute the VAT. As a third robustness check, we recomputed the VAT tax levy to look for cases where the applied VAT rate and the taxable value do not match. We found this to be the case for less than 2500 transactions, with the vast majority of cases occurring for transactions with a value of less than KSh 1. For transactions of a more sizable volume, we used records of the trading partner to validate the transaction value and again traced the error back to the computation of the VAT rather than the reported transaction volume. Finally, we check for transactions with identical supplier and buyer IDs and only find 30 such cases across the two transaction-level data sets for sales and purchases.

Major cleaning steps

As a next step we implement three major cleaning steps. First, we dismiss transactions where the buyer (supplier) is a mature firm (older than 12 month), but never recorded any positive purchases (sales) in their VAT summary sheet. Second, we examine a dozen transactions that are larger than KShs 10 billion. If they are implausible based on the other records of the two parties involved, we dismiss them entirely or, where possible, impute them based on other observed transactions between the two firms. Third, we dismiss or impute purchase transactions that are larger than the overall sales of the supplier that the buyer reports on. If no other transactions between the two parties are available, the transactions are dropped; if not, we use the highest value of the remaining transactions to impute the likely faulty transaction.

1.B.2 Constructing a monthly firm-to-firm relationship panel and a firm-level panel

The goal is to construct a firm and a firm-to-firm relationship panel that is consistent with each other. We do so by combining information contained in the transaction-level data and the monthly VAT summary sheets filed by each firm. Discrepancies between the raw data can arise because of differential reporting about one and the same transaction by the two firms involved. Common issues are (i) the two counterparts record the transaction at different points in time, (ii) only one party

reports, (iii) the reported amount differs.

Harmonising sales and purchases across time: To address issue (i) we rely on the invoice date rather than the date of the tax return when the transaction was filed. However, this means that we need to adjust the total sales and purchases recorded in the summary sheets for transactions that were reported in a month that differs from the invoice date. Our first step, therefore, is to filter those transactions by comparing the date when the transaction was recorded to the invoice date. Next, we aggregate the transactions for each month and subtract them from the month in which they were recorded. Finally, we add them to the total sales and purchases in the month of the invoice date.

Aggregating transactions to a monthly relationship panel: Next we aggregate transactions to a monthly relationship-level data set where each row represents a buyer-supplier pair and summarises their aggregate monthly transaction volume. In cases where both parties report, we see deviations of less than 10,000 KES (less than 100 USD) in 75% of the cases.

Reporting discrepancies, where only one party reports, but not the other, occur because of two main reasons: First, firms have the option to bundle their sales to non-registered parties in an unconstrained cell in the tax return spreadsheet. Although mainly intended for recording sales to final consumers and other non-tax-registered entities, firms can use this cell if they want to avoid spelling out every single sales transaction and simply bundle them there. For purchases, this option is only available if firms report VAT-exempt domestic purchases. The differential reporting requirements for sales and purchases result in a more comprehensive data set for purchase transactions as opposed to sales transactions. 59% of the monthly firm-to-firm transactions in the aggregated relationship-level data set are reported by buyers only.

Second, 18% of the observed monthly firm-to-firm transactions are recorded by suppliers only. The overwhelming majority of suppliers use electronic recording machines. While the recordings in the machines are not directly forwarded to the tax return, we observe whether such a machine was used or not. We can thus attribute

some of the one-sided reporting by suppliers to the fact that suppliers report transaction-level details for dis-proportionally many business entities or individuals who have a tax identification number, but actually do not file VAT themselves (likely suppliers ask for the tax PIN, but remain ignorant on the counterpart's VAT obligations). Another minor reason for one-sided reports by suppliers are VAT exemptions on the buyer's side.

Addressing discrepancies in reporting: Our approach to address issues (ii) and (iii) — cases where only one party reports or transaction volumes differ - closely links to the source of the discrepancies. First, we use reports by the buyer as the default as buyers have less leeway to bundle purchases from registered parties in the category for transactions with non-registered parties. We only use the sales volume recorded by buyers if no information from the buyer side is available .[16] Another advantage of this approach is that it helps us to mitigate discrepancies because of credit notes, which are ubiquitous in the sales transactions.[17] It further helps us to address cases where the month of the recorded invoice date diverges and firms interact regularly. However, we currently do not have a way to address cases where a buyer-supplier pair interacts infrequently and the firms report about one and the same transaction in different months.

Time and firm size inconsistencies: We now implement two additional cleaning steps targeted at detecting outliers following Dhyne et al. (2015); Huneeus (2018). First, we check for time inconsistencies in transactions. Here we log-standardise all transactions using the mean and the standard deviation of all transactions between a pair of firms. If the absolute value is larger than 2.58 (critical value for a 2-sided t-test), we tag the transaction as a time-inconsistent (Huneeus, 2018). Among the time-inconsistent transactions, we focus on those that are upward-inconsistent and where the mean ratio exceeds 100 (Dhyne et al., 2015). For this sub-set of

[16] An exception is the first five months of 2018 when a lot of purchase invoices were dismissed during a large-scale audit. The audit did not affect invoices reported as sales by suppliers.

[17] The raw data further contains credit notes with a negative transaction value, which capture canceled purchases/sales or rectify errors. 5.8% and 2.73% of all sales and purchase transactions are credit notes. If issued in the same month as the invoice the credit notes cancel out as we aggregate transactions up to the monthly level.

transaction, we now replace the inconsistent value with the highest alternative transaction between the two parties involved. Second, we check for exceptionally large transactions involving small suppliers/buyers. We define a small buyer or supplier as an entity that has sales in the 25th percentile and a large transaction as a transaction that exceeds the 99th percentile ($>$ 7.5 mio. KES). We tag these transactions as inconsistent and replace the tagged value with the value of the highest consistent transaction between the two parties.

With the cleaned relationship-level data in hand we turn back to the monthly panel of firm-level outcomes.

Computing firm-level sales and purchases within the firm-to-firm network: For each firm we use the relationship-level data (or edge list) to infer firm-level aggregate monthly purchases and sales that are made within the firm-to-firm network. We then combine this information with the monthly summary sheet. Unsurprisingly we find that the sales and purchases recorded in the relationship-level data set often deviate from the original records on sales and purchases to/from other registered parties. In most cases, sales in the firm-to-firm relationship data set exceed the sales to registered parties (sometimes even total sales) in the summary sheet. Remember that it is because buyers are forced to break down their purchases by transactions, whereas suppliers have the option to declare them as sales to non-registered parties.

Harmonising monthly sales and purchases: We divide cases with discrepancies into two categories:

First, cases where the sales (purchases) in the network data, which takes the supplier's (buyer's) records into account, deviate from the sales (purchases) to (from) registered parties in the summary sheet. Here we assume the sales (purchases) to non-registered parties in the summary sheet are correct and simply add/subtract the amount by which the network sales (purchases) differ from the sales (purchases) to (from) registered parties in the summary sheet. Consequently, total sales (purchases) are corrected up- or downwards as well.

Second, cases where we observe zero sales (purchases) to (from) registered parties in the summary sheets, but positive sales in the network data. These are cases

where firms do not report any sales transactions themselves (this rarely happens for purchases). Here we assume the total sales in the summary sheet are correct and re-allocate the equivalent for the sales observed in the network data from the sales to non-registered to the sales to registered parties category in the summary sheet.
After those two steps, we remain with more substantial discrepancies in 16% of the monthly firm-level sales and 12% for monthly firm-level purchases.[18]
We now proceed as follows: If the sales to non-register parties are zero, we assume that no bunching of sales in this category occurred and that the sales observed in the network data correspond to the total sales of that firm in a given month. We conclude that partial bunching occurred, if the sales to non-registered parties are positive, and the difference between the sales in the firm-to-firm panel and the sales to registered parties is smaller than the sales to registered parties. We then subtract the difference from the sales to non-registered parties and add it to the sales to registered parties. If the sales to non-registered parties in the same month are insufficient to account for the difference, we include the sales to non-registered parties in the previous month. Finally, we are left with a few unresolved cases. Here we simply assume that the overall sales (purchase) volume as per the firm-to-firm panel is correct and add the difference to the total sales (purchases) captured in the summary sheet.
We now use this harmonised firm-level data set and merge it with other firm-level information such as firm age, sector, location, and where available, the number of employees.

1.B.3 Customs data

To study linkages to international supply chains we further use transaction-level customs data. A common feature of transaction-level customs records is standardised product codes. The first six digits of the code are based on the 6 Digits Harmonized Commodity Description and Coding System (HS code).[19] In most cases two, sometimes even four, additional digits further refine the product description to a level relevant for Kenya's tariff regime. The HS codes are frequently revised typically every

[18]For the unfiltered firm panel (see more on the filtering process below) the share is 10% and 8% respectively. The lower figures can be explained by the substantial number of entities reporting zero sales/purchases in the unfiltered database.

[19]https://unstats.un.org/unsd/tradekb/Knowledgebase/50018/Harmonized-Commodity-Description-and-Coding-Systems-HS

five years. The last two revisions took place in 2012 and 2017. As our data set includes transactions recorded before the 2017 revision was implemented, we harmonise all recorded product codes to align with the HS 2012 revision.[20]

The East African Community's (EAC) Single Customs Territory[21] rules allow for firms to choose in which country (either destination or origin) they want to report their trade transaction. The revenue authority recording the transaction then shares the information with their EAC counterpart. Such information is, however, not shared at the firm level and can thus not be mapped into our data set. Nevertheless, a Kenyan firm reporting their export to Uganda with the Uganda Revenue Authority is still obliged to report their exports in their VAT records. We, therefore, rely on the VAT records to measure a firm's monthly imports and exports. For any analysis that requires product, origin/destination-level information, we have to exclude international transactions within the East African Community.

1.B.4 Corporate income and pay-as-you-earn tax returns

In both the annual corporate income tax returns and the monthly pay-as-you-earn (PAYE) returns, we check for firm-level outliers in key variables such as the number of employees, salary payouts, sales, and local purchases. We impute values based on historical records in a handful of cases that we are able to identify as obvious errors. Between 10-15% of the firms filing CIT returns have a reporting cycle that does not align with the calendar year, e.g., a firm reports their annual tax return for the period April 2019 to March 2020. For variables that capture the value of stocks like assets and liabilities, we attribute them to the calendar year in which the last month of the return falls into. To compare the annual statistics of flow variables across firms, we harmonize the values of such variables to proxy reporting in line with the calendar year (the level at which we undertake most of our analysis).Examples of variables related to cash flows are sales, purchases, and employment expenses. To do so, we compute the length of the time periods covered by the return, which is typically 12 months. We then assume cash flows to be equally distributed across the months and split their values between the two years covered by the return. The share attributed to

[20]The relevant correspondence tables can be found here: https://unstats.un.org/unsd/trade/classifications/correspondence-tables.asp.

[21]Member states are Burundi, Kenya, Rwanda, South Sudan (since 2016), Tanzania, and Uganda.

each year corresponds to the number of months observed in each year. For example, if the return of a company is filled for July 2016 to June 2017, we attribute half of the sales to 2017 and the other half to 2016. If the return spans the months of October 2016 to September 2017, we attribute only ¼ of the sales to 2016 and the remainder to 2017.

1.B.5 Other data sets

We further use taxpayer registration forms to identify the 4-digit sector of operation, as well as the county (1st admin layer), and sub county (2nd admin layer) of operation. Firms typically report their sector of operation at the 4-digit level. Most of the sector classifications observed in the database align with the UN's International Standard Industrial Classification of All Economic Activities (ISIC).[22] We harmonise classifications in case of deviations.

We use the firm address noted in the registration form, to extract the county and sub-county (1st and 2nd administrative level) of operation. Typically firms report the address of the firm's headquarters or the location of their main administration. 66% of the firms in the filtered data set report an address in Nairobi County (in the unrestricted data set 60% do).

We use information on the business type as well as the tax station where the return is filed to identify public entities, NGOs, international organisations, trusts, membership organisations, and clubs. Government agencies, but also any other entity whose files are handled by the public sector tax unit are classified as a public entity. We then classify the remaining companies and partnerships as private firms. This includes special economic zones, export processing zones, and foreign firms.

Firms have the option to register multiple subsidiaries under the same tax PIN. For a small number of firms we therefore observe multiple sectors of operation and multiple firm locations. However, these firms make up less than 1% of all registered entities that file either VAT or income tax. For the case of the roughly 6,000 firms that report multiple sectors of operation, we unfortunately do not observe, which of the sectors is the most important one in terms of overall sales. Meanwhile, we cannot rule out the possibility that large corporations use different tax PINs for their affiliates. In contrast

[22]The official classification is available here: https://unstats.un.org/unsd/classifications/Econ/Download/In%20Text/ISIC_Rev_4_english_structure.Txt

to other papers in the literature working with similar data (see for example Alfaro-Urena et al., 2018; Dhyne et al., 2021), we are unfortunately unable to determine affiliations among firms from the data sets available to us.

1.B.6 Firm definition

We filter the data set for firms that identify themselves as a private company or partnership in their tax registration form. In doing so we exclude all government-owned firms, government agencies, international organisations, non-governmental organisations, trusts, and clubs. We further exclude firms operating in the financial sector (most of which are banks and insurance companies) as they file based on slightly different reporting requirements. Finally, we limit our sample to firms with annual sales of KShs 5 million or more and non-zero purchases in at least one year between 2015 and 2020. In 2019, we observe information on 236,729 tax-paying entities in the VAT data. Our sample restriction reduces the number to 51,749 firms. The sales of those firms, however, accounts for over 90% of the total sales observed in 2019. Overall, the filtered data set for the period 2015 up until March 2021 contains 57,482 firms, which form 4,117,880 supplier-buyer relationships among each other.

CHAPTER 2

Spatial Inequality and Informality in Kenya's Firm Network

with **Peter Wankuru Chacha, Vatsal Khandelwal, and Benard Kipyegon Kirui**

Abstract The spatial configuration of domestic supply chains plays a crucial role in the transmission of shocks. This paper leverages transaction-level tax records to study spatial patterns of domestic firm-to-firm trade in Kenya and explores how these patterns may be shaped by the presence of an unobserved informal sector. First, we document stylised facts about formal firms in this setting, revealing a high degree of spatial concentration in the network, over and above the concentration of aggregate economic activity. 90% of the cross-regional variation in trade volumes can be attributed to the extensive margin of trade, the location of firms and the number of firm-to-firm relationships they form. Using data from the population census and national accounts, we further show that informality is particularly prevalent in downstream economic activities and smaller regional markets. We link our insights about the sectoral and spatial composition with a network formation model to investigate how accounting for informal firms affects spatial inequality in firm-to-firm trade. Accounting for informal firms increases the outdegree of firms in regions with the highest level of informal activity. Our results suggest that not accounting for informal firms underestimates the connectivity and vulnerability of smaller regions to shocks, especially those that pass through hubs such as Nairobi.

2.1 Introduction

Leveraging domestic markets and reducing inequalities within them are gaining traction as ways to promote economic development in low- and middle-income countries (Goldberg and Reed, 2023). A more limited scope for export-led growth models has led to an increased focus on domestic supply chains (Atkin and Donaldson, 2015; Grant and Startz, 2022). The structure of these networks can influence economic development by affecting the dispersion of welfare gains from trade shocks, infrastructure investments (Atkin and Donaldson, 2015; Arkolakis et al., 2023), the impact of industrial policies (King et al., 2019; Liu, 2019; Lane, 2023), the propagation of adverse shocks (Acemoglu et al., 2012; Grassi et al., 2017; Baqaee, 2018; Baqaee and Farhi, 2019), and the exposure of the network to systemic risk (Erol and Vohra, 2022). However, studying domestic firm networks is empirically challenging due to lack of granular and comprehensive data.

A rapidly growing literature has turned to using value-added tax (VAT) returns to track firm networks (Dhyne et al., 2021; Bernard et al., 2022; Coşar et al., 2022; Alfaro-Urena et al., 2022; Adão et al., 2022; Demir et al., 2023), thus far focusing largely on high- and upper middle-income economies with relatively lower levels of informality.[1] By definition, transaction-level tax records do not capture details on informal firms and their role in inter-firm trade, which could be crucial for our understanding of how shocks pass-through across the entire economy or how welfare gains of government projects are dispersed across space. We address this concern by drawing on transaction-level data from close to 57,500 formal firms and over four million supplier-buyer relationships in Kenya and complementing them with survey and census data on the regional and sectoral composition of overall economic activity by the Kenya National Bureau of Statistics (KNBS). As the use of administrative data to map firm networks is becoming more and more popular, this paper provides insights on how including informal firms might alter the observed structure of the firm network across space.

[1]Notable exceptions are working papers on India (Panigrahi, 2022; Gadenne et al., 2022; Fujiy et al., 2022), Uganda (Spray, 2019) and Rwanda (Spray and Wolf, 2018).

We ask the following questions: What are the spatial patterns of trade flows among formal firms in Kenya? How are these patterns affected when informal firms are included? What does this mean for spatial inequality and the vulnerability of different regions to shocks? The Kenyan context is particularly well-suited to answer these questions. First, informal economic activity accounts for 25-35% of Kenya's GDP (Schneider and Enste, 2000; Elgin et al., 2021). Second, as East Africa's largest economy, Kenya boasts a substantial domestic market with vast geographic and socio-economic cross-regional heterogeneity. Third, a series of granular data sets allow us to estimate economic activity both at the sectoral and regional level – features that can otherwise be difficult to quantify in many contexts as statistical bureaus tend to focus on national aggregates (Chacha, 2019).

We proceed as follows. Firstly, we establish three stylized facts concerning the spatial patterns of domestic trade among formal firms in Kenya. Subsequently, we utilize census and national accounts data to document five facts related to the informal sector. Based on these insights, we estimate a network formation model with preferential link attachment to investigate whether taking into consideration informal firms alters the spatial inequality in firm-to-firm links documented in the tax records.

With regard to our findings on firm-to-firm trade among formal sector firms: First, we find that trade among formal private sector firms is substantially more concentrated around Kenya's metropolitan areas than both population and aggregate economic activity. For example, the Pareto exponent for the spatial dispersion of trade flows within the network of formal firms is 57%-76% lower than the exponent for regional GDP.[2] Firms from the country's capital Nairobi and the port city Mombasa participate in as much as 88% of all firm-to-firm transactions as either buyers or suppliers. Second, these aggregate trade flows are a result of spatial inequality along the extensive margins of the firm network. In fact, 90% of the variation in aggregate trade volumes across counties can be attributed to the extensive margin: the location of firms and a stark inequality in the number of firm-to-firm links across regions.

[2]A lower Pareto exponent indicates a higher degree of inequality.

The intensive margin, i.e. the number of transactions and average trade flows per shipment, plays a minor role. Third, upstream linkages (to suppliers) are more equally distributed across space than downstream linkages (to customers). That is, regions are more similar in their purchase behaviour than in the composition of their downstream sales channels.

Next, we ask whether the observed patterns are a result of our limited view due to the spatial selection of firms into the administrative data or whether it is reflective of the economy's underlying structure. We document five stylised facts about the informal sector. First, the VAT-paying sector accounts for 34% of Kenya's GDP. Second, informal firms are usually located downstream of large (formal) firms. Third, informality decreases as regional market size and income levels increase. Fourth, the spatial concentration in economic activity is largely a result of the concentration of formal sector activity. Fifth, we use a data-driven accounting exercise to show that the sectoral composition of regions explains little of the spatial inequality in the level of informality. Instead, it results from differences in the level of formalisation within the same sector and across locations.

In summary, we document a higher share of informal activity in smaller regional markets that are dominated by downstream economic activities such as wholesale, retail, personal services, and construction. VAT self-enforcement mechanisms are the lowest in more consumer-facing activities (Naritomi, 2019). We might therefore systematically undercount the number of outlinks of formal firms in these smaller markets and downstream sectors, which in turn could explain some of the observed divergence in patterns between in- and outlinks in the formal firm data.

To more formally, and quantitatively study the role of informal firms, we apply a network formation model with heterogeneous node types by Bramoullé et al. (2012) to our setting. In our adaptation of the model, we classify firms into types based on their sector of operation and location. A new firm first chooses a specific type of firm to link to in accordance with its own "bias". This bias can be reflective of the firm's underlying production technology or geographic location. Then, it forms

links with firms of this type either via undirected search[3] or preferentially i.e. it chooses a certain proportion of suppliers independent of its network environment, but the remainder from the pool of the suppliers of these suppliers. This search behaviour can be rationalised by underlying information frictions about the quality of potential suppliers (Chaney, 2014). The model provides predictions for the number of links between firms of different sector-location types. We first estimate this network formation model to predict the Kenyan firm network as it is. We find that new firms choose 42% of their suppliers through undirected search and the remaining 58% of suppliers are found via existing suppliers. In comparison, Chaney (2014) finds that only 40% of all relationships of French exporters with international trade partners are formed via preferential attachment. Our estimate of 58% of links being formed as a result of preferential attachment hence suggests that information frictions are potentially even more binding for firms in Kenya's domestic firm network.

We then simulate a counterfactual network that accounts for informal firms by combining the model with real-world data on the sectoral and regional composition of the informal sector. We use the counterfactual to answer the question of interest: keeping constant the probability with which firms interact across sectors and locations, how do spatial patterns of trade change when informal firms are accounted for? The counterfactual results confirm that regions with the highest levels of informality and fewer ex-ante outlinks are predicted to be more spatially integrated if we account for unobserved informal firms. At the same time, these firms source disproportionately with the largest trading hubs in the country, including Nairobi and Mombasa. This in turn boosts the expected outdegree of these hubs. Therefore, accounting for the spatial and regional incident of unobserved informal firms increases spatial inequality across counties. This has implications, for example, when we think about the pass-through of shocks across space. By considering the formal network only, smaller regions with a higher degree of informality appear less exposed to shocks to network hubs like Nairobi. A limitation of our setting is that we implement the counterfactual under the assumption that conditional on their sector and region, informal and formal firms link with other types in a similar fashion.

[3]The network formation literature refers to this search pattern as "random" search (Jackson and Rogers, 2007; Bramoullé et al., 2012; Chaney, 2014).

Further, we do not model the decision of firms to enter as formal or informal firms.

Our paper is structured as follows. In Section 2.2, we discuss our contributions to the existing literature. In Section 2.3, we describe the content of the administrative data used to map the firm network. We document the spatial patterns of Kenya's formal firm network in Section 2.4.3. We discuss the role of the informal sector and map its spatial and sectoral composition in Section 2.5. Finally, we tie the two together in the Section 2.6 where we discuss and estimate a network formation model with preferential attachment (Bramoullé et al., 2012). We present the results of the counterfactual in which we include informal firms in Section 2.7. Section 2.8 concludes.

2.2 Contributions

Our paper contributes to the literature on firm networks, informality, and the dispersion of economic activity within countries.

Our paper's first contribution is to describe the spatial mapping of the formal firm network in a context with a sizable informal sector, to show that the informal sector is relatively more important in smaller regional markets and downstream sectors, and to highlight the spatial patterns that we can miss if we do not account for these firms. Despite the large size of the informal sector in various contexts, informality along supply chains remains understudied to date. Exceptions to this include De Paula and Scheinkman (2010); Böhme and Thiele (2014); Zhou (2022) who use firm survey data and structural models to look at the implications of a VAT regime for informality along supply chains and Gadenne et al. (2022) who are able to observe firm-to-firm linkages between VAT-paying and non-VAT paying, but registered, firms. We complement their papers by exploring the relevance of informality for observed spatial supply chain structures. Methodologically our paper sits in between the two groups of papers in that we are also able to observe the formal firm network at a very granular level (Gadenne et al., 2022), but do not observe linking probabilities between formal and informal firms and therefore rely on a network formation model and complementary data sources to look at the sensitivity of spatial concentration in firm-to-firm trade to accounting for informal firms.

Thereby, we also contribute to a sizeable literature on estimating the size of the informal sector (Schneider and Enste, 2000; La Porta and Shleifer, 2014; Elgin et al., 2021). Reyling on cross-country regressions, this literature has documented that the relative size of the formal economy increases as income levels rise (Brandt, 2011; La Porta and Shleifer, 2014; Ulyssea, 2018). We show that this pattern extends to Kenya's domestic economy: the formal sector share correlates with income levels across regions within the country. Our finding that formal sector activity is concentrated in the Kenya's metropolitan areas mirrors Zárate (2022)'s finding from Mexico City, which exhibits a similar formal-core, informal-periphery structure. Our findings also align with a literature on the link between the size of markets and the firm size distribution (Kumar et al., 1999; Laeven and Woodruff, 2007; Gollin, 2008; McCaig and Pavcnik, 2015).

Second, the empirical facts on Kenya's firm network presented in our paper build on and extend the seminal paper by Bernard et al. (2019), who were among the first to describe key features of domestic production networks. Our detailed mapping of the spatial network structure complements recent studies by Panigrahi (2022); Miyauchi (2023); Arkolakis et al. (2023), who draw on the similar data from five Indian states, Japan, and Chile respectively. These papers focus specifically on modelling and micro-founding the endogenous network formation process. Our focus instead is a more granular mapping and quantification of the spatial patterns in firm networks with specific attention on the potential bias that arises due to informality.

Third, we also contribute to a small but growing literature on domestic trade patterns and the role of supply chains in shaping cross-regional welfare outcomes in low- and middle-income countries (Allen, 2014; Atkin and Donaldson, 2015; Bergquist et al., 2021, 2022; Grant and Startz, 2022). Due to the granular transaction-level data, we are able to decompose the so-called intensive margin of trade into transactions and average trade volume per transactions.[4] The spatial decay in cross-regional trade is fully accounted for by the decline in the number of firm-to-firm relationships

[4]The result adds nuance to the more general literature on domestic firm-to-firm trade, which has thus far mainly estimated the gravity equation for domestic trade at the relationship rather than transaction level and thus found the average trade volume per relationship to decline with distance (Arkolakis et al., 2023).

and transactions with distance. This concurs with findings on the key role of the extensive margin of firm sales, its outdegree, for firm size (Bernard et al., 2022) and subsequently how it governs international trade flows (Chaney, 2014; Eaton et al., 2022). Our finding that the average trade volume mildly rises with distance underscores the relevance of economies of scale in shaping (domestic) trade patterns (Coşar and Demir, 2018; Grant and Startz, 2022; Bergquist et al., 2021).

Finally, urban primacy, disproportionate agglomeration of economic activity in one or two primate cities, is a long-studied phenomenon (as published in Jefferson (1989), Jefferson, 1939; Ades and Glaeser, 1995; Henderson, 2002; Soo, 2005).[5] While the role of cities as hubs for the exchange of information (Pred, 1980) is a global phenomenon, Memon (1976); Alonso (1968) argue that considerable concentration of economic activity is expected in settings with fast changing economic and political environments where market participants attach a considerable price to up-to-date information. Locating in a country's primate city allows market participants to overcome information frictions. Thanks to novel data sources that have not been available to geographers and economists in previous decades, we are able to describe the individual margins of economic activity that contribute to the documented spatial concentration. While firm networks and the role of backwards and forwards linkages have long been at the core of theories around agglomeration (Venables, 1996), only the recent emergence of granular firm-to-firm trade data has allowed researchers to study them empirically (Miyauchi, 2023). Our finding on the relative importance of firm location and firm-to-firm links compared to within-relationship trade volumes is highly suggestive of a crucial role for search and matching frictions (Miyauchi, 2023). We further add to the literature on urbanisation, agglomeration and city sizes by documenting that the concentration of economic activity in Nairobi is even more a formal sector phenomenon than agglomeration of overall economic activity.

[5] Berry (1961) notes that how and why primate cities emerge is very much a function of a countries economic and political history. Given the vast heterogeneity in the history of urbanisation across the globe it is not possible to link it directly to a country's economic development.

In Kenya, the economic dominance of Nairobi and Mombasa has been a concern for geographers, economists and policy makers alike since its emergence as a result of the colonial economic system and transportation infrastructure (Memon, 1976; Obudho, 1997; Otiso, 2005). Nairobi continues to dominate despite government efforts to change this (Obudho, 1997). Moreover, Obudho (1997) notes how concentrated formal sector employment was in the Nairobi of 1992, i.e. how much Nairobi dominated the formal sector relative to other urban centres. We are able to confirm Obudho (1997)'s result using more granular and comprehensive geospatial data that covers Kenya's entire economy.

The role of information frictions in rationalising the spatial dispersion of economic activity more generally, and firm-to-firm trade more specifically, also informed our choice for a network formation model. We provide an empirical stage for Bramoullé et al. (2012)'s theoretical framework and show how it can be translated from a context of academic citation networks to firm networks. Bramoullé et al. (2012) build on the seminal paper by Jackson and Rogers (2007). Chaney (2014)'s paper on the geography of trade networks, which has become a seminal paper in its own right for the trade literature, builds on the very same framework by Jackson and Rogers (2007) and introduces geography into it. While this class of models does not micro-found firm entry decisions and is very simplistic in its approach to modelling linking decisions of network agents, it emphasises the role of preferential attachment for rationalising the power law distribution common in many networks (Newman, 2018), including firm networks (Bernard and Moxnes, 2018). Preferential attachment summarises the notion that network agents link with others (make friends, cite papers, exchange goods) through their existing links (friends, cited papers, firms). In the context of firm networks the process of searching through existing suppliers or buyers can be rationalised by information asymmetries about the quality of potential future suppliers (Chaney, 2014).[6]

[6] An additional rational could be contracting and enforcement frictions (McMillan and Woodruff, 1999).

2.3 Description of the administrative data

2.3.1 Description of data sources

Our analysis draws on mirco data from value-added and pay-as-you-earn tax returns. We further utilise the Kenya Revenue Authority's tax registry to compile basic, self-reported information on each firm, namely the 4-digit sector classification, the business type, the start date of its operations, and the headquarters location. All data sets can be linked using anonymised firm identifiers. Amongst the tax reports, the key data set are monthly value-added tax (VAT) returns. The VAT returns not only contain information on the firms' aggregate monthly sales and purchases but further include two sections in which firms report detailed information on sales and purchase transactions with other businesses and VAT-registered entities. These records allow us to identify the supplier and the buyer for each firm-to-firm transaction, the transaction date, an unstructured product description, and the transaction volume. Sales to and purchases from non-registered parties (e.g., exempt parties, non-registered businesses, final consumers) are recorded as an aggregate monthly figure and are hence not tracked at the transaction-level. We deflate all variables denoted in monetary terms using the monthly Consumer Price Index (CPI) published by the Kenya National Bureau of Statistics (KNBS).

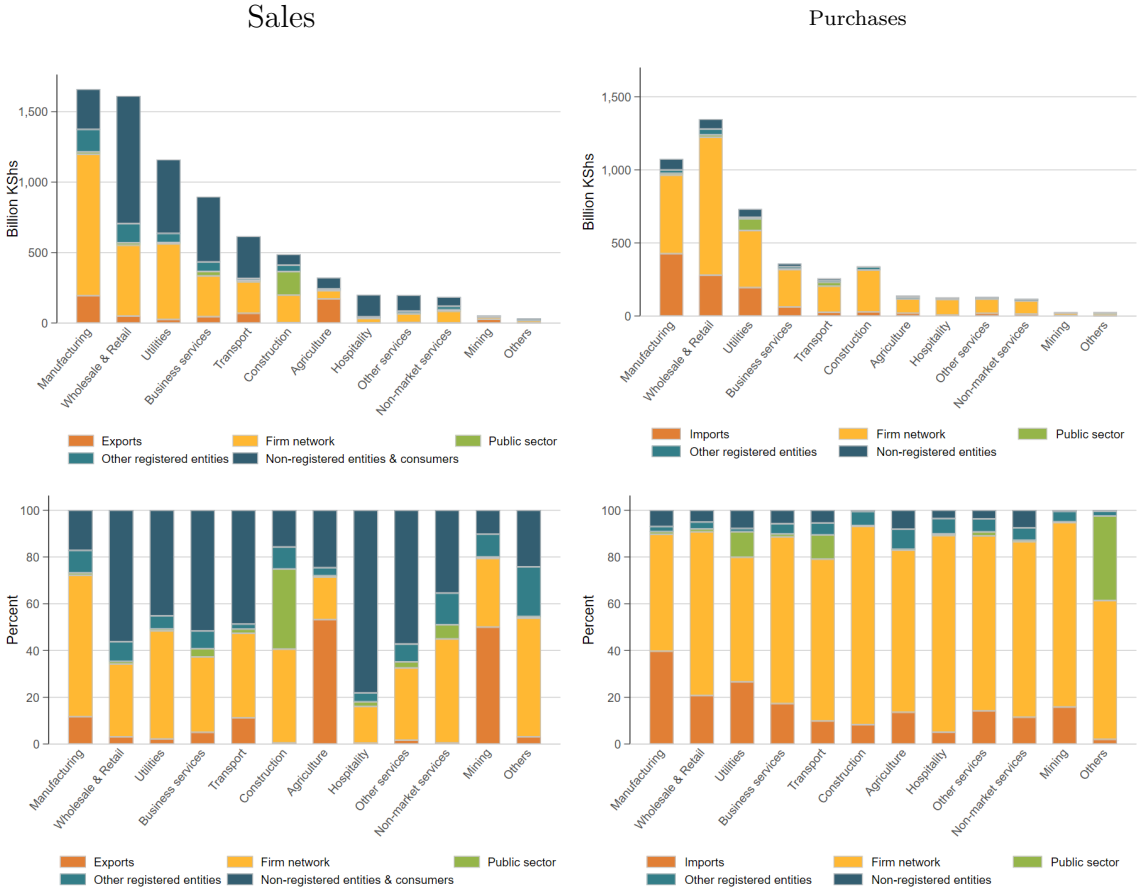
VAT applies to individuals and firms with an annual turnover of KShs five million and above ($40,500 as of December 2022). Once a firm is VAT-registered and has crossed the threshold of KShs five million, they are required to continue filing VAT returns in years with lower turnover.

We filter the data set for entities that identify as private companies or partnerships in their tax-registration form. In doing so, we exclude all government-owned firms, government agencies, international organisations, NGOs, trusts, and clubs. We exclude firms operating in the financial sector (most of which are banks and insurance companies). We restrict our analysis to firms with annual purchases greater than zero and annual sales of KShs five million or more in at least one year, which we observe in the data. We apply the VAT threshold to exclude firms that registered for

VAT to bid for tender but were never operational. In 2019, we observe information on over 236,000 tax-paying entities in the VAT data. Our sample restriction reduces the number to 51,749 firms. However, the sales of those firms account for over 90% of the total sales observed in 2019.

Figure 2.1 plots the sector composition and the respective sales and input channels of firms covered in the administrative records. Manufacturing and wholesale and retail firms account for almost half of the sales we are able to track in the tax records.

Figure 2.1: Composition of sales and purchases by sector



The figures in the first row show sector-level aggregate sales (domestic + exports) and purchases (domestic + imports) for 2019. In the second row we plot the sales to and purchases from registered vs non-registered parties as a percentage of total sector-level sales and purchases.

2.3.2 Trade with informal firms in tax records

Firms below the threshold, those offering financial and education services, and to a large extent, firms dealing in agricultural goods and pharmaceuticals, are exempt from VAT. They are not required to register for VAT and do not submit a monthly return. Sales to firms below the threshold and exempt firms are thus both captured in the category of sales to non-registered parties. The same category exists for purchases; however, it only allows for the reporting of purchases of VAT-exempt goods and hence excludes a lot of potential inputs from the non-registered sector. While this leads to a bias in what type of economic activity is captured in the data, we use firm survey data collected by KNBS (2016) in Section 2.5 to show that the bias is potentially not as severe given small firms are more likely to purchase from larger VAT-registered firms rather than the other way round.

Throughout this paper, we use the term "non-VAT paying" firm to refer to private sector entities that are either not VAT-registered due to their size, exempt from VAT payments due to the products and services they sell or do not comply with the tax law.

2.4 The spatial concentration of domestic firm-to-firm trade

We begin by examining the geography of Kenya's firm network. Firstly, we quantify the level of spatial concentration within the network. Next, we break down trade within the network into its individual components, identifying firm locations and firm-to-firm relations as the primary margins driving spatial concentration. Lastly, we demonstrate that spatial inequalities are particularly pronounced in downstream trade flows. When discussing spatial patterns of economic activities, our primary focus lies in measures disaggregated at the county level, which represents Kenya's first administrative layer. In some instances, we also consider the second layer, sub-counties.

2.4.1 Urban primacy in Kenya's firm network

Kenya's firm network is strongly concentrated around its metropolitan areas Nairobi and Mombasa.[7] Nairobi and Mombasa first emerged as Kenya's primate urban centres during the establishment of a European colonial economic system. Both locations, and Nairobi in particular, were strategically developed as entrepôts along the Kenya-Uganda railroad[8] and the region's communication network (Memon, 1976; Obudho, 1997).[9] In 1960, 49% of the wholesale sector's turnover was generated by Nairobi-based firms, who in turn also employed 46% of this sector's workforce (MoF, 1963, as cited in Memon (1976)).[10] Today, as much as 68% of the sales volume within the network of formal firms is generated by Nairobi-headquartered firms. Notably today, as in the 1960s, the city's role in the firm (or trade) network is disproportionate relative to its population and even aggregate GDP.[11] In 2019 (1960), as little as 9% (3%) of Kenya's population lives in Nairobi County and the city contributes 37% of Kenya's GDP outside the agricultural sector (see Table 2.7). A potential concern is that the observed spatial concentration is driven by the fact that we only observe firm headquarter locations, which in turn are more likely to be based in Nairobi or Mombasa. In Appendix 2.A.2 we use micro-data from the 2010 Census of Industrial Production (KNBS, 2010) to compare the spatial concentration of sales and firm locations with and without multi-establishments. We find that the excess spatial concentration introduced by multi-establishments cannot explain the aggregate concentration patterns of formal private sector activity.

[7] Although, no exact figures for comparison are reported in the respective papers, Huneeus (2018) and Cardoza et al. (2023) find a stark geographic concentration of trade flows around metropolitan areas in Chile and the Dominican Republic.

[8] The railroad followed existing caravan routes. Mombasa and Nairobi then gradually replaced Zanzibar as the major trading hub of the region (Memon, 1976).

[9] As predicted by Memon (1976), Nairobi's disproportionate growth as an economic and urban centre would start to accelerate and the previously bi-polar system would effectively become a Nairobi-centric one.

[10] Mombasa accounted for 27% of employment and 35% of turnover in the wholesale sector.

[11] We map firm headquarter locations and population density in Appendix Figure 2.A1.

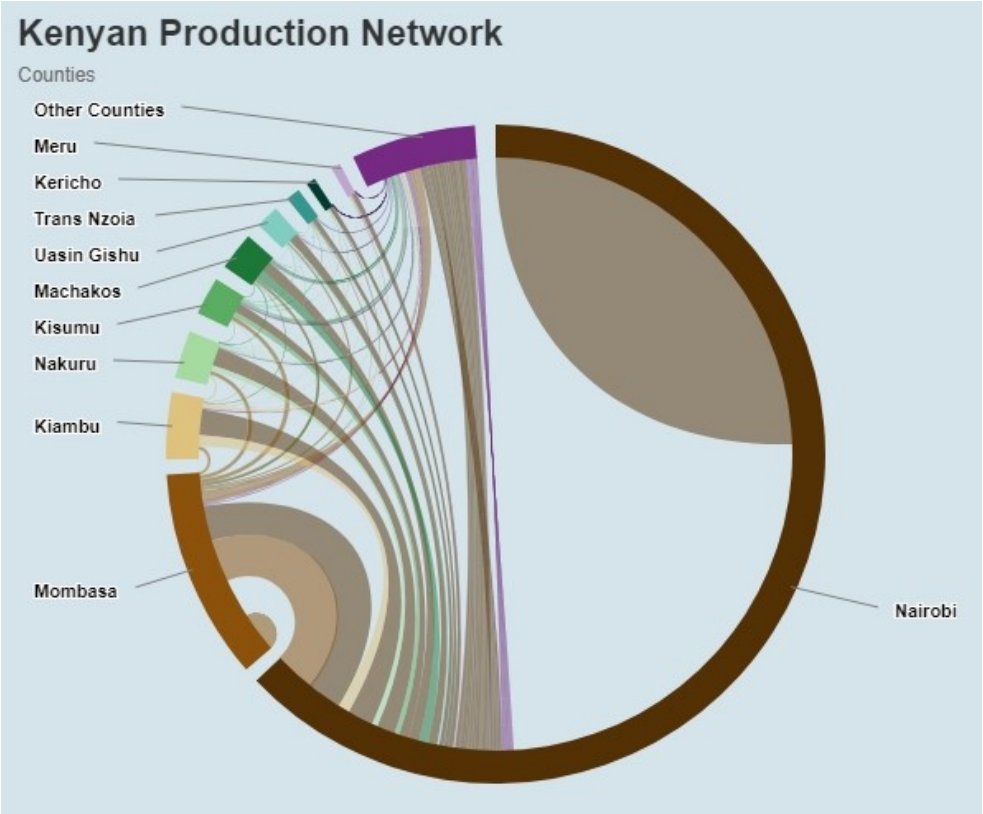
Table 2.1: Geographic concentration of economic activity in Kenya

	Nairobi	**Mombasa**	**Pareto exponent**	
	in %		α	SE
Population overall	9	3	1.29	0.18
Population of cities & towns	31	9	0.85	0.01
GDP	29	5	0.98	0.07
GDP w/o agriculture	37	6	0.95	0.06
GDP w/o non-market services	24	5	0.91	0.08
No. VAT firms	64	9	0.61	0.04
Employment in VAT firms	62	8	0.35	0.03
Value added of VAT firms	72	10	0.37	0.03
Network sales	68	13	0.24	0.02
Network purchases	60	9	0.43	0.02

The columns for Nairobi and Mombasa report their share of the respective national aggregate figures (e.g., Nairobi's contribution to Kenya's GDP). The Pareto exponent α is the estimated coefficient from a county-level regression of each county's rank (log) on the respective measure x (log): $\log \text{rank} = \log A + \alpha \log x$.

The county-to-county trade flows plotted in Figure 2.2 underscore the primacy of Nairobi and Mombasa. The size of each segment of the pie is proportional to the respective county's sales within the network and the colouring of the trade flows aligns with the county of origin. 88% of the 21 million firm-to-firm transactions in 2019 involved at least one firm based in Nairobi or Mombasa. Moreover, trade between Nairobi-based firms themselves accounts for 45% of the total trade volume. The graph further reveals that trade flows out of Nairobi and Mombasa are larger than inflows into the metropolitan areas. Such asymmetries in trade flows can exacerbate spatial frictions as logistics companies need to haul back empty containers along trading routes with a stark directional imbalance in trade flows (Ishikawa and Tarui, 2018; CAK, 2019; Wong, 2022).

Figure 2.2: County-level trade flows between formal firms



The figure shows inter-firm trade flows aggregated at the county level. The size of each node (segment) is proportional to the county's share of purchases and sales relative to the aggregate volume of firm-to-firm trade between formal firms in Kenya. The colour of the edges (links between segments) indicates the direction of the trade flow. They take the colour of the supplying county (e.g., goods and services provided by firms in Nakuru to firms in Nairobi take the colour of the segment for Nakuru). The width of each edge (links between segments) is proportional to the share of the trade flow with respect to the aggregate volume of trade flows in the transaction-level administrative data. To improve readability, we only separate the trade flows for the ten counties with the largest aggregate amount of transactions within the domestic production network. We bundle the trade flows for the remaining 37 counties.

Moving beyond Nairobi and Mombasa, how concentrated is economic activity if we consider the entire distribution? The distribution of both firm and city sizes is often well-approximated by a Pareto Distribution (Gabaix, 2009). Under this premise, the Pareto exponent can be considered a measure of inequality for the dispersion of population and economic activity (Gabaix, 2009; Soo, 2005; Gabaix and Ioannides, 2004). In Table 2.1, we compare the Pareto exponent α for the regional distribution of population and gross value added (KNBS, 2022) to a series of measures derived from the administrative data. The α for each indiacator is obtained via rank-size

regressions (Gabaix and Ioannides, 2004).[12] A lower value indicates a flatter slope and hence more inequality across counties. The Pareto exponents for both total county-level GDP and aggregate income generated outside the agricultural sector is close to unity - in line with Zipf's Law, which stipulates a power law distribution with an exponent of approximately one (Gabaix, 2009; Soo, 2005). At the same time, Kenya's population is more evenly distributed across counties than economic activity. Turning to the firm network, we find αs that are substantially lower than one, indicating a high degree of spatial inequality. An exponent of 0.61 suggests that the number of VAT-paying firms is still fairly evenly distributed across counties - despite the concentration of firms in Nairobi. Meanwhile, the αs for employment, value added, sales and purchases are 57%-76% lower than the exponent for overall economic activity (GDP aka the Gross County Product). Comparing the α for network sales, i.e. trade flows out of a county (0.24), versus network purchases, i.e. trade flows into a county (0.43), shows that a smaller number of counties supply disproportionate amounts of inputs to the rest of the country.

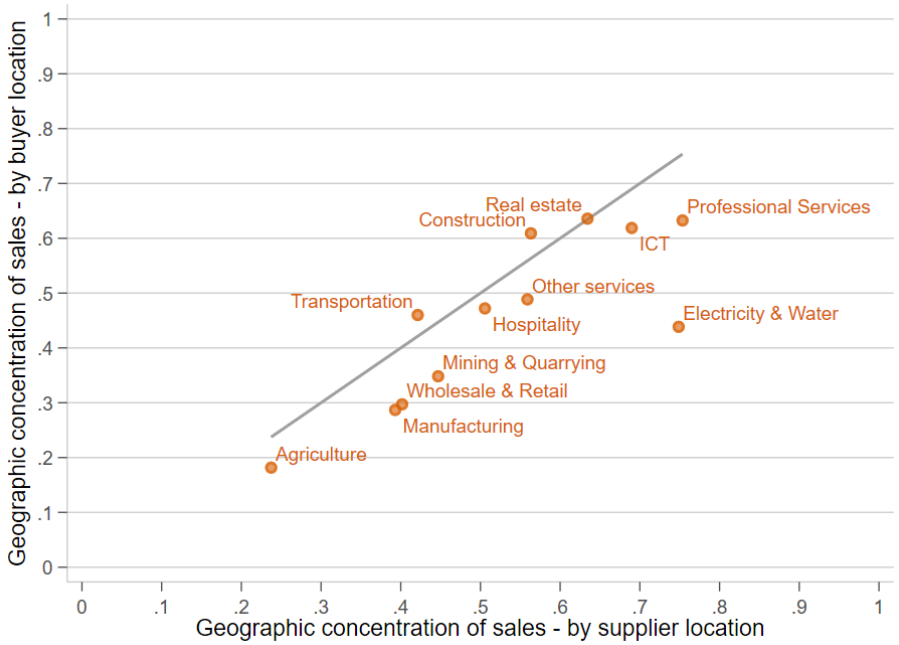
This pattern aligns with our previous finding that trade flows out of Nairobi and Mombasa are larger than inflows. In Figure 2.3, we directly compare the spatial dispersion of sales relative to purchases.[13] If the geographic concentration of sales origins and destinations are the same, sector markers will lie directly on the 45° line added to the graph. As expected, the marker for most sectors lies well below the line, indicating that buyers of network inputs are more geographically dispersed than suppliers. This pattern is particularly pronounced for business services and the utilities sector, which are largely concentrated in Nairobi.[14]

[12] I.e. a county-level regression of each county's rank (log) on the respective measure x (log): $\log \text{rank} = \log A + \alpha \log x$.

[13] Our measure of dispersion is a Herfindahl-Hirschman Index computed from county-level trade flows and takes the value zero if sales are highly dispersed and one if they are highly concentrated.

[14] A caveat of this approach is that we draw on administrative boundaries to measure the geographic dispersion instead of actual geographic proximity. This becomes evident in the manufacturing sector, whose sales appear to be equally distributed. However, most manufacturing firms are based in counties bordering Nairobi, including Machakos and Kiambu.

Figure 2.3: Geographic concentration of sales based on supplier vs buyer location



The above graph plots the geographic dispersion of sales within the network based on the supplier's location against the dispersion based on the buyer's location. The dispersion measure is a Herfindahl-Hirschman Index and takes the value zero if sales are highly dispersed and one if they are highly concentrated. The solid blue line indicates the 45° line. To compute the geographic concentration of sales, we aggregate sector-level sales at the county level based on firm locations.

To quantify which margins of aggregate trade flows drive this spatial concentration, we decompose them into sub-components as a next step.

2.4.2 Firm location and relationships drive spatial concentration in trade flows

The extensive margins of the firm network, firm location and firm-to-firm relationships, account for 70-90% of the variation in aggregate trade volumes. Using the granular transaction-level data, we are able to distinguish between four different sales margins: the number of firms N, the number of relations R per firm, the number of transactions c per relationship, and the trade volume v per transaction. In a nutshell location o's sales to the firm network τ can be summarised as:[15]

$$\tau_o = N_o \times \frac{R_o}{N_o} \times \frac{c_o}{R_o} \times \frac{v_o}{c_o} \tag{2.1}$$

[15]The same is true for purchases.

Table 5.2 summarises the share of the variance attributed to each term in both upstream (purchases) and downstream (sales) trade flows.[16] The number of firms operating in each county alone accounts for 67% of the variance in purchases across counties.[17] The number of relationships with suppliers of the county accounts for yet another 22%, leaving a little over 10% of the variance to be picked up by the intensive margins for trade, i.e. the number of transactions between firm pairs and the average transaction volume. Turning to downstream trade flows, i.e. the decomposition of the variance in sales across (sub-)counties, the location of firms plays a slightly less important role. Instead the number of firm-to-firm relationships now accounts for one third of the variance in network sales. The variance decomposition is a useful exercise to track the respective margins of trade. It, however, falls short of allowing us to identify the relative importance of selection of entrepreneurs into certain regions versus the place effect of a region on an entrepreneur's ability to form relationships.

Table 2.2: Geographic concentration of economic activity in Kenya

Purchases

Aggregation	No. firms	No. relationships/firm	No. transactions/relation	Avg. volume/transaction
County	0.67	0.22	0.14	-0.04
Sub-county	0.53	0.29	0.16	0.06

Sales

Aggregation	No. firms	No. relationships/firm	No. transactions/relation	Avg. volume/transaction
County	0.60	0.31	0.12	-0.00
Sub-county	0.39	0.34	0.15	0.16

We conclude this section by studying how each of the trade margins behaves in relationship to distance. Instead of considering trade flows in and out of a county, we turn to bilateral trade between two counties. In line with the classic gravity model of trade (Head and Mayer, 2014), trade flows within Kenya decline with distance and

[16] Our decomposition follows Klenow and Rodriguez-Clare (1997); Eaton et al. (2011); Panigrahi (2022).

[17] This includes purchases the firms of a respective county make within their own county or from firms outside the county.

travel time. Consistent with the variance decomposition, this spatial decay in trade can be traced back to the decline in firm-to-firm relationships, the extensive margin of trade.

Total trade flows between two locations can be decomposed into the number of firm-to-firm relationships R_{od}, the average number of transactions per relationship $\bar{c}_{od}$, and the average trade volume per transaction $\bar{v}_{od}$:

$$\tau_{od} = R_{od} \times \bar{c}_{od} \times \bar{v}_{od} \tag{2.2}$$

Essentially, the above Equation 2.2 is the bilateral equivalent of Equation 2.1. In Figure 2.4, we aggregate trade flows to the county level and regress the total trade volume and each of its sub-components on travel time between locations. Conditional on two pairs of counties trading with each other,[18] as much as 81% of the elasticity of overall trade flows with respect to travel time can be explained by differences in link formation between firms.[19] Put differently, the decline in trade flows over space is almost entirely driven by the number of firm-to-firm relationships declining. The elasticity result for the extensive margin of trade closely aligns with findings by Arkolakis et al. (2023) for Chile. Our granular transaction-level data allow us to go a step further and unpack what constitutes the intensive margin of trade in Arkolakis et al. (2023) - the average trade volume per relationship. We find that the spatial decay in the intensive margin is entirely driven by the decline in the number of transactions. The average trade volume per transaction on the other hand is zero or if anything increases with travel time between locations.[20] Rising average trade volume results are consistent with the role of economies of scale in governing trade flows and spatial price gaps (Grant and Startz, 2022).

We further document that social connectedness (Bailey et al., 2021) between counties is an important determinant of spatial trade flows over and above distance. The trade

[18]The results can be found in table format in Appendix Table 2.A1. Our results are robust to using a Poisson pseudo-maximum-likelihood approach to estimate the gravity equation, which allows us to account for zero trade flows between two pairs of counties. During the time period captured in our data, 75% of all possible county combinations engage in bilateral trade (1,652 out of 2,209).

[19]Using distance rather than travel time does not change the result. Using the cross-section for 2019 rather than pooling firm-to-firm data across all years yields a slightly lower percentage of the overall trade elasticity that is explained by the number of firm-to-firm relationships (78%)

[20]Depending on the specification the coefficients are small, but positive, albeit often not significant.

volume between any two counties increases with the strength of their social links. Social connectedness is measured at the county level and captures the probability of two random people drawn from two different locations being friends on Facebook.[21] Social connectedness can be a proxy for a number of different factors such as the number of transportation options between two localities or the strength of the trust network between potential trading partners. The finding backs up our result on the importance of the extensive marging. For the subset of trading counties, we find domestic trade flows in Kenya to be more than twice as sensitive to social connectedness than the international trade flows on a global level considered in Bailey et al. (2021).[22]
To study the importance of the number of buyers and suppliers in the origin and destination county, we drop the county fixed effects in an alternative specification (see Appendix Table 2.A1). Instead, we include a series of other origin and destination characteristics like distance to Nairobi or Mombasa and population density alongside bilateral travel time and social connectedness. We find that including the number of suppliers and buyers in the origin and destination county plays an important role in governing overall trade flows. The R^2 increases by 70% if we include the two regressors (column 1 versus 2).

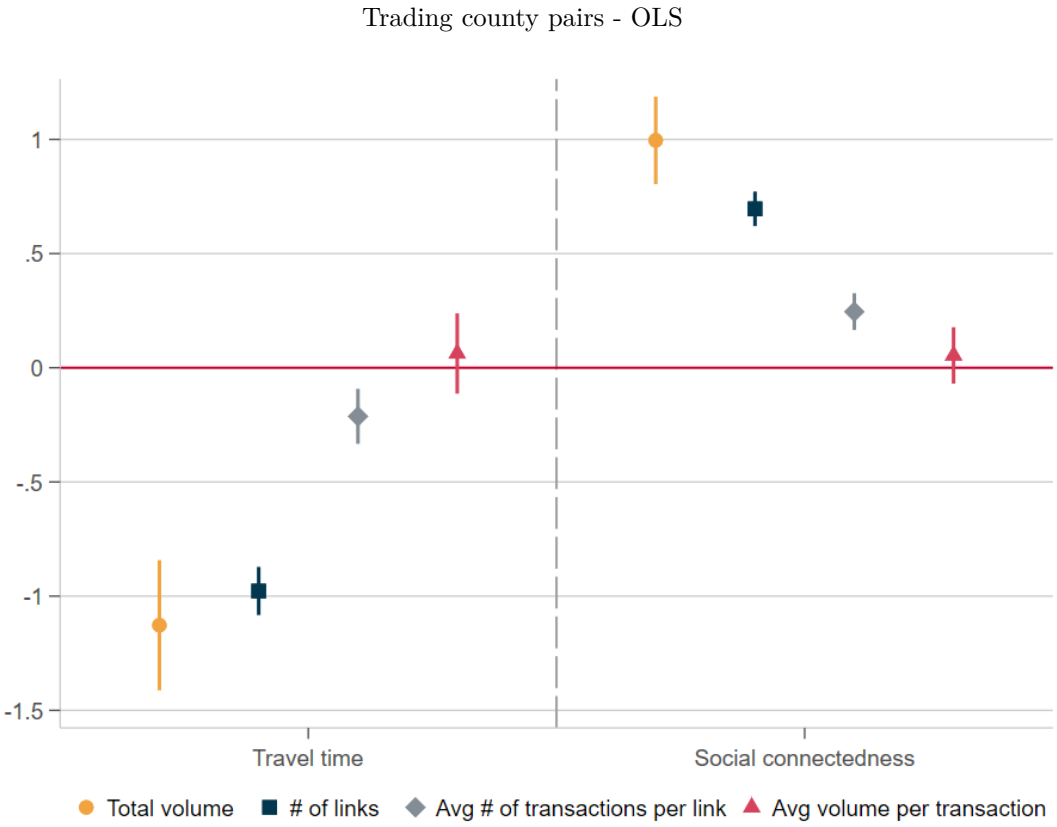
2.4.3 Upstream linkages are more equally distributed than downstream linkages

Our earlier observation that firm-to-firm relationships are a more important margin for cross-county network sales rather than purchases is consistent with the pattern

[21] The measure is called the Social Connectedness Index and captures the total number of Facebook friendship connections between two counties, divided by the product of the number of Facebook users in each county (Bailey et al., 2021).

[22] Bailey et al. (2021) report a coefficient of 0.454 for social connectedness and -1.027 for distance, if both are included in the same regression with origin and destination fixed effects. Similar to their paper, we find that controlling for both social connectedness and travel time reduces the elasticity of each, indicating that they capture overlapping channels (see Table 2.A1). At the same time each independently contributes to explaining variation in trade flows between counties.

Figure 2.4: The intensive and extensive margin of domestic trade



All variables are aggregated to the county level. The outcome variables for the OLS regressions are in log terms. Travel time is measured in log minutes. Social connectedness and travel time are included in the same regression. Social connectedness is measured using the Facebook Social Connectedness Index, which captures the total number of Facebook friendship connections between two counties, divided by the product of the number of Facebook users in each county (Bailey et al., 2021). The total number of possible county pairs is 2,209. We pool data from all available years to minimise the incident of zero trade flows. The regression results can be found in table format in Appendix Table 2.A1.

that links to suppliers are more evenly distributed among firms than links to buyers.[23] In other words, there is greater heterogeneity in the sales channels utilised by firms than in their input channels. While virtually all business models require some form of material input, firms can have diverse customer bases, including other businesses, final consumers, or the public sector (see Figure 2.1). Recent research on the origins of firm heterogeneity has highlighted the importance of the number of firm-to-firm relationships with buyers, the extensive margin, in determining firm size (Bernard et al., 2022).

We show that this pattern replicates across space. Nairobi-based firms on average have 32 buyers with those based in the central business district reaching an average of 140 formal sector buyers. At the same time, firms in the median county (sub-county) have an average of 13 (5) buyers.[24] Turning to the average indegree (suppliers), Nairobi and other larger Kenya cities and towns no longer stand out as much. The left map in Figure 2.5 shows a much more equal distribution of the average indegree across space. Firms in the median county (sub-county) have 22 (20) suppliers, Nairobi-based firms have 30 suppliers on average. This pattern also aligns with the higher level of spatial inequality found in network sales relative to the location of firms in Table 2.1.[25]

Understanding why the number of outdegrees varies substantially across locations is particularly important in a context with high informality. Figure 2.A2, for example shows that firms outside the metropolitan areas on average sell a larger proportion of

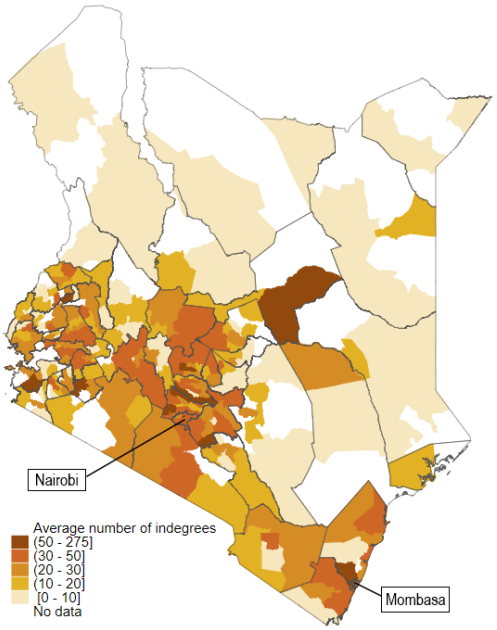
[23]Under the premise that the in- and outdegree distribution follows a power law, we can again estimate the Pareto exponent, which is 1.85 for the indegree and 1.24 for the outdegree distribution. The outdegree distribution hence features more inequality in the number of links observed in the top end of the outdegree distribution. The point estimates are very close to what has been found in the Japanese (Bernard et al., 2019) and the Costa Rican (Alfaro-Urena et al., 2018) firm network (see Figure 2.A3). The firm network in Chile (Grigoli et al., 2023) and the Dominican Republic (Cardoza et al., 2023) are also within the same range, but feature less inequality in their upper tail. Note that there are some discrepancies between contexts that can be attributed to differences in the way the Pareto exponents are estimated and reported. Bernard et al. (2019) follow the broader literature on power laws in economics (Gabaix, 2009) and estimates the Pareto exponent α by regressing the inverse CDF (log) on the degree (log). Grigoli et al. (2023); Cardoza et al. (2023); Alfaro-Urena et al. (2018) estimate the Pareto exponent by regressing the log degree on the inverse CDF (log), which yields an estimate that is close to $\frac{1}{\alpha}$ and report the resulting regression coefficient as the Pareto exponent. Looking at the inverse, all estimated exponents fall within the range 1.24 and 3.5.

[24]The contrast is even starker of course if we also exclude Mombasa and Machakos County from the average. Machakos is coloured in dark green in Figure 2.5 as it serves as an industrial hub with many large manufacturing firms.

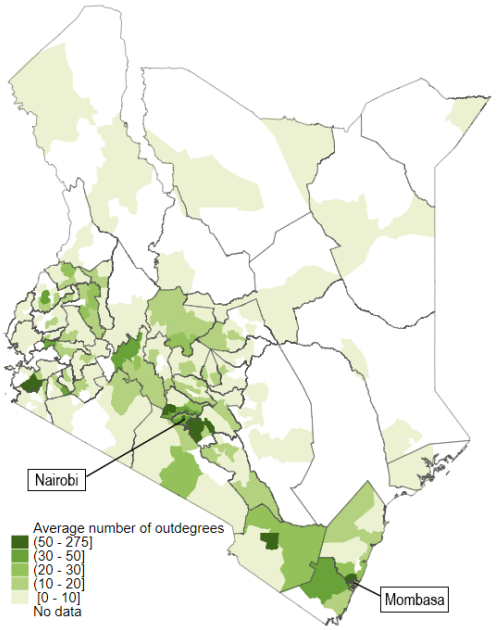
[25]In Table 2.1, the Pareto exponents for the number of firms suggest that firm locations are more evenly distributed than sales volumes and the number of firm-to-firm relationships within the network.

Figure 2.5: Average in- and outdegrees across space

Average indegree



Average outdegree



The above map plots the average in- and outdegree of firms each sub-county. County borders are outlined in grey.

their sales to non-registered entities. This result could well be an outcome of differences in sales channels of firms in the metropolitan areas compared to those outside. At the same time, the steep decline in average outdegrees outside the metropolitan areas could be driven by informality along supply chains. Recall that sales to consumers are lumped together with sales to non-VAT paying firms in the monthly tax returns. Similarly, one might wonder whether links between the metropolitan areas and the rest of the country are captured in a complete manner. This question matters for the spatial pass-through of shocks that are more likely to affect firms in Nairobi, e.g., international trade shocks. Erol and Vohra (2022) show that in networks with a core-periphery structure, systemic risk is closely linked with the size of the core.[26] While we cannot directly recover the importance of the informal firms for the supply chains of VAT-paying firms, we use third party data and a network formation model to narrow in on the informal sector and its position within the firm network.

2.5 The role and position of the informal sector

In the forthcoming section, we utilise third-party data collected by the Kenya National Bureau of Statistics (KNBS) to assess the scale of the informal sector beyond the VAT-system and to analyze the sectoral and geographic distribution of informality in Kenya. Our approach is structured as follows. First, we establish a clear definition of informality within our context. Then, we detail and examine the additional data sources employed to study informality. Finally, armed with a precise definition and relevant data, we present five essential stylised facts that will inform how we account for unobserved informal firms in our subsequent section that presents the model. The purpose of the model-based counterfactual is to study the relevance of unobserved firms for the structure of Kenya's firm network and the spatial inequality in firm-to-firm links.

[26]Core-periphery networks are characterized by a core subset of nodes with strong links among each other and a periphery subset of nodes that maintain connections with the core but demonstrate relatively weaker connections with each other. Periphery nodes can also form clusters among themselves. These clusters are, however, then only loosely linked with other periphery clusters.

2.5.1 Margins of informality in firm networks

Many plausible definitions of informality exist and can be applied even within the same setting (La Porta and Shleifer, 2014). Firms can be formally registered entities (extensive margin of informality), but engage in informal activities (intensive margin of informality). Ulyssea (2018) documents an intensive margin of informality for employment i.e. informal employees in otherwise formal, registered firms. A wholesaler we interviewed in Nairobi's Central Business District[27] explains how the notion of an extensive and intensive margin of informality extends to firm-to-firm transactions:

> *"All firms purchase from manufacturers and importers paying input VAT. They even have an interest in getting purchases that have VAT on it to inflate the input VAT. What they do to mitigate the VAT levy, they downplay their output VAT (i.e. sales). Some customers will purchase with receipt and output VAT on it. Some customers will purchase without a receipt."*

Table 2.3 summarises four different margins of informality that can occur in firm networks: an extensive margin at the firm-level and an intensive margin at the transaction-level. Within each category, informality can occur due to either non-compliance or simply because a unit is too small to be taxed.[28]

Table 2.3: Margins of informality in firm networks

	Extensive	**Intensive**
Below tax threshold	Small firms	Small transactions
Above tax threshold	Non-compliance	Non-compliance

The extensive margin helps identify who the informal firms are. These can be either (i) small firms who never crossed the annual revenue threshold for VAT or (ii) larger non-reporters, i.e. firms with revenues above the VAT threshold, but which do not file VAT. If either of the two trading parties is informal, we do

[27] The firm's customers cover the whole range of potential sales channels: other wholesalers, retailers, public institutions, and some individual consumers.

[28] Depending on the tax code not all of them arise in every setting. Further, VAT exemptions can be a legal reason why firms or transactions above the VAT threshold are not captured in administrative tax records.

not observe transaction-level information on their interaction. In our setting, as discussed in Section 2.3, we observe sales of formal firms to non-registered entities, including non-VAT businesses, but only as a monthly aggregate figure and not at the transaction-level. For the purpose of this paper, we exclusively focus on whether or not firms pay national taxes like VAT. Some of the firms we classify as informal might be formal according to alternative definitions of informality. Specific to Kenya's legal context, a lot of firm we miss out in the national tax records do pay sub-national business license fees. These license fees are collected by county governments. VAT-paying firms are thus a sub-set of the universe of firms with a county license.

The intensive margin considers informality at the transaction-level conditional on both parties being VAT-registered firms. Here informality in administrative records can either occur because (i) transactions fall below a reporting threshold specified in the tax code (for transactions rather than for firms) or (ii) non- or under-reporting of transactions that firms are required to declare. The first aspect is not a concern in the Kenyan context. The Kenya Revenue Authority requires firms to record transactions of any size, conditional on both parties being VAT-registered.[29] Omission of transactions between two formal firms or the misreporting of trade volumes on the other hand remains a concern for us.[30] We are able to recover some of the omitted transactions or under-reported trade volumes by relying on information from a firm's trade partner when processing the data. Any residual misreporting feeds into the data as informal trade flows along the intensive margin.

[29]The application of transaction thresholds varies substantially across contexts. Like the Kenyan Revenue Authority, the Dominican Republic and Uganda do not have a volume-dependent threshold either (Cardoza et al., 2023; Almunia et al., 2022). In Belgium, Costa Rica, and Turkey a trading pair of firms does not need to report their transactions if their annual bilateral volume falls below 250 €, $4800, and $2650 respectively (Dhyne et al., 2015; Alfaro-Urena et al., 2018; Demir et al., 2022). Note that transaction thresholds are different from firm size thresholds that consider a firm's overall annual revenues (see Section 2.3).

[30]It is partly mitigated by asymmetric incentives of buyers and suppliers to report transactions correctly. Buyers might want to overstate purchases to claim refunds for input VAT, while suppliers have an incentive to downplay the volume of their sales to reduce their output VAT liability. Almunia et al. (2022) show that despite this built-in VAT enforcement mechanism firms in Uganda still misreport trade volumes, sometimes even against their own interest.

2.5.2 Measuring informality

We now turn to the measurement of the informal sector outside the VAT system. Table 2.4 provides an overview of the data sources we draw on, while Table 2.5 summarises how we compute the various measures of informality. All our proposed measures represent formal sector shares, i.e. the proportion of overall economic activity that can be traced back to the formal sector. As discussed in the previous section, we will focus on a definition of informality that refers to economic activity by non-VAT paying firms. We use the administrative data as a measure for the size of the formal sector and a corresponding figure for the aggregate economy including both formal and the informal economic activity from KNBS.[31] However, we will also rely on measures of informality that only rely on mirco data and national accounts data published by KNBS.[32] Why do we need this second group of measures? The measures based on the administrative data will tell us what part of the economy we miss out on by studying the Kenyan economy through the lens of VAT data. The comparison with measures that only rely on third party data then shapes our interpretation of those gaps between the administrative records and overall economic activity. Are these gaps between the VAT data and aggregate economic activity, for example, reflective of economy wide dynamics between the formal and informal sector?

Throughout the remainder of this section, we draw on three different indicators for economic activity, namely employment figures, the number of firms, and value added (sales - purchases). Our default measure of informality will be an employment-based measure that draws on the 2019 population census (KNBS, 2019b). The two alternatives based on the number of firms and value added rely on estimates of the universe of businesses in KNBS (2016)[33] and on estimates of the regional economic

[31] While the criterion for a firm to be considered formal in our setting is whether or not it pays VAT, we prefer to refer to the underlying data set as administrative data for the simple reason that some of the variables like employment, sector, location, and age are collected from tax records outside the VAT system.

[32] Important for our purposes is that none of these data sources are explicitly based on tax records. A list of tax filing businesses informed the data collection of the Census of Establishments (KNBS, 2017). However, this is the source of data we will rely the least on. Tax revenues might inform estimates of the Gross County Products (KNBS, 2022).

[33] KNBS (2016) obtain information on the number of licensed businesses from county governments and estimate the number of unlicensed businesses based on household survey data.

Table 2.4: Overview of benchmark data

Name	Year	(Dis-)Aggregation	Key indicators
Small & Medium Sized Enterprises Survey (MSMEs)	2016	firm-level	main input source and buyer
Census of Establishments (CoE)	2017	sector OR county	# of formal sector establishments
Gross County Product (GCP)	2019	sector AND county	gross county product
Population & Housing Census (Census)	2019	sector AND (sub)county	formal & informal employment

All data are collected and published by the Kenya National Bureau of Statistics. **Sources:** Small & Medium-Sized Enterprises Survey https://statistics.knbs.or.ke/nada/index.php/catalog/69 KNBS (2016); Census of Establishments https://www.knbs.or.ke/download/report-2017-kenya-census-establishments-coe/ KNBS (2017); Gross County Product https://data.humdata.org/dataset/kenya-gross-county-product-gcp-by-economic-activities-per-county and KNBS (2022); 2019 Kenya Population & Housing Census https://www.knbs.or.ke/publications/# *2019 Kenya Population and Housing Census Volume IV: Distribution of Population by Socio-Economic Characteristics* KNBS (2019b).

size captured by the Gross County Product (KNBS, 2022) respectively. While the employment-based measure is likely to primarily capture the extensive margin of informality, the value-added-based measure can be considered as an aggregation of all margins. The number of firms will provide a more nuanced picture on the extent of potential non-compliance by firms on the extensive margin. The key advantage of the population census data is that they allow us to dis-aggregate employment records both at the sectoral and regional level at the same time. This aspect is missing for measures based on the number of firms. Further, we can distinguish between employment in the private compared to the public sector. None of the other measures allows for this distinction.[34]

Table 2.5: Measures of informality

Unit	**Numerator** (formal sector)	**Denominator**	KNBS	Use admin data
Employment	No. formal priv. sector employ.	Working population	Census	✗
Employment	No. employ. in licensed firms	No. employ. in all firms	MSMEs	✗
Employment	No. employ. VAT firms	No. employ. in licensed firms	MSMEs	✓
No. firms	No. licensed firms	All firms	MSMEs	✗
No. firms	No. VAT firms	All firms	MSMEs	✓
Value added	Value added VAT firms	Gross County Product	GCP	✓

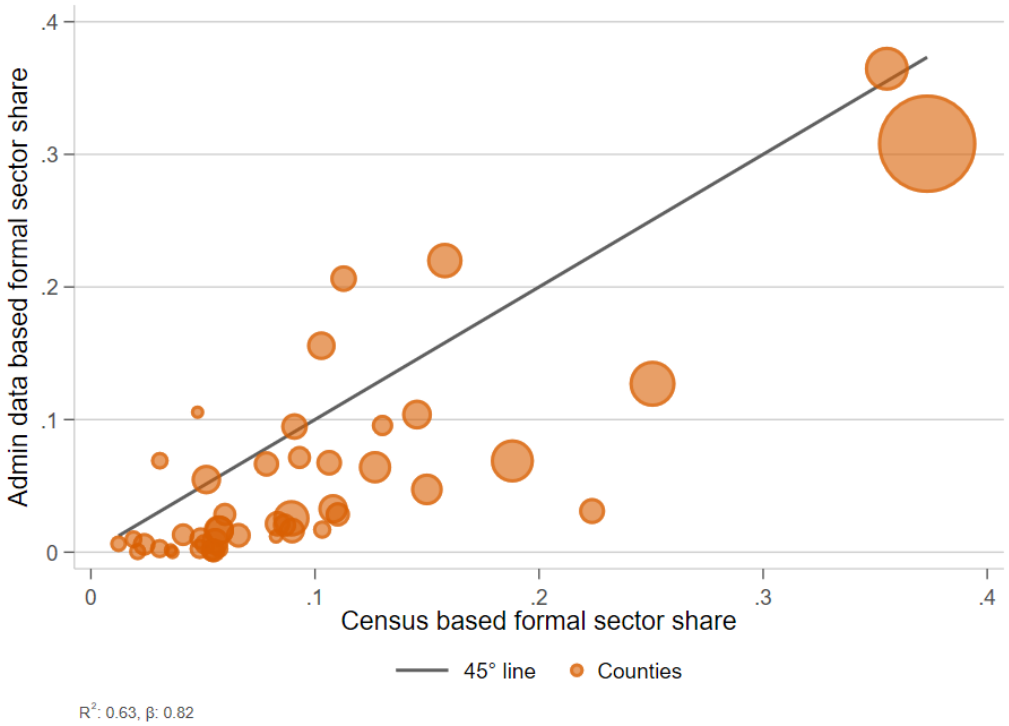
For details on the data sources by KNBS see Table 2.4. The term "all firms" refers to both licensed and unlicensed businesses based on KNBS (2016) estimates and county government records.

We exclude any employment or firms in the agricultural sector or non-market services

[34]Sometimes the lines are not clear cut. The Census of Establishments (KNBS, 2017), for example counts a lot of private and public schools that have some form of low-revenue, but for-profit activity happening at their site. These establishments account for almost a third of the firm count in this census.

when measuring informality at the regional level.[35] In both cases, the tax records cover a small and very specific sub-population of firms and employees only. In the case of agricultural firms, the administrative data cover large-scale commercial agriculture only. The vast majority of these firms are export oriented (Chacha et al., 2024). For non-market services, we mostly capture a small number of firms which operate in sectors dominated by non-profit organisations and the government. In addition, the majority of for-profit firms in this sector enjoys VAT exemptions and hence does not appear in the network.

Figure 2.6: Comparison of formal sector shares based on census versus administrative records



The above graph correlates the share of the formal sector computed using employment figures from the administrative records with the share of formal private sector employment as per the 2019 population census (KNBS, 2019b). Each market represents a county. The size of each marker is proportional to the economic size of the county, i.e. its Gross County Product.

Our preferred measure of informality considers formal sector employment as per the 2019 population census. It correlates strongly (ρ= 0.83, Table 2.6) with its

[35] We do so wherever possible. The county-level statistics in the MSME report (KNBS, 2016) and the Census of Establishments (KNBS, 2017) do not allow us to abstract from those sectors.

counterpart based on the administrative records (also see Figure 2.6).[36] In the final section of the paper, we will use the employment- and population census based measure of overall economic activity by sector and region to explore counterfactual network patterns that consider the presence of informal firms. Table 2.6 summarises the correlation coefficients of the alternative informality measures. The two employment-based KNBS measures correlate well with all measures based on the administrative data. The measure capturing licensed businesses as a share of the universe of businesses in Kenya (including micro-enterprises) in contrast only correlates weakly with them. This likely reflects the fact that many of the licensed firms are very small themselves and their geographic dispersion does not correlate as strongly with the tax records. Employment in licensed businesses (second row) is likely to be concentrated in the large firms of this population and hence aligns more strongly the estimates based on the administrative data.

Table 2.6: Correlation of formality measures

	Formality measures based on admin data		
KNBS measures	No. firms	Employment	Value added
Employment (census)	0.78	0.83	0.80
Employment (licensed MSMEs)	0.58	0.69	0.65
No. firms (licensed)	0.20	0.16	0.13

The above table shows the correlation coefficients of different measures of the formal sector share. Each measure represents a share, i.e. captures the proportion of economic activity that can be attributed to the formal sector. The labels indicate the underlying unit of measurement and the source of the data. All measures are aggregated at the county level.

2.5.3 Five stylized facts on informality in space

We document five stylised facts about the informal sector.

[36]To avoid mechanical correlation between the two measures we use total employment in licensed firms as the denominator for the administrative data. The KNBS estimate for employment in licensed firms is based on micro data that is distinct from the population census. Alternatively, one could use total employment in all MSMEs, which, however, includes many self-employed people. The correlation results are very similar for both alternatives.

Fact 1: The VAT-paying sector accounts for 34% of Kenya's GDP

The gross value added generated by VAT-paying firms on average corresponds to 34% of Kenya's annual GDP for the period 2015-2020 (see Appendix Table 2.B1).[37] This is substantially lower than for high income economies like Chile, where 80% of the country's GDP can be attributed to VAT-paying firms (Huneeus, 2018). However, not all of this gap can be attributed to informality.

The differences arise for two reasons: The first reason is differences in the tax code, in particular the treatment of financial services, non-market services, and agriculture.[38] The second reason is informality. In the following, we first discuss the influence of sectors that are not well-represented in the data in explaining the gap, then informality and its different margins.

If we exclude sectors that are to a large extent exempt from VAT (non-market services,[39] agriculture) or have special reporting rules applied to them (financial services), the VAT sector accounts for 64% of residual economic activity, implying an overall informal sector share of 36%.[40] This estimate suggests a larger informal sector compared to the 26% estimated by Elgin et al. (2021) for 2018 and the 29% estimated by Hassan and Schneider (2019) for 2013. Both of these studies utilised model-based approaches to estimate aggregate informality as a share of GDP. In addition to our relatively more simple approach to arrive at 36%, we also focus on the VAT sector only. By doing so we apply one of the most stringent possible

[37]The value added recorded in the VAT data fluctuates between 40% (2016) and 24% (2020) of Kenya's GDP. Figure 2.B1 plots the quarterly time series for both GDP and value added. We discuss related time trends in Appendix Section 2.B.1.

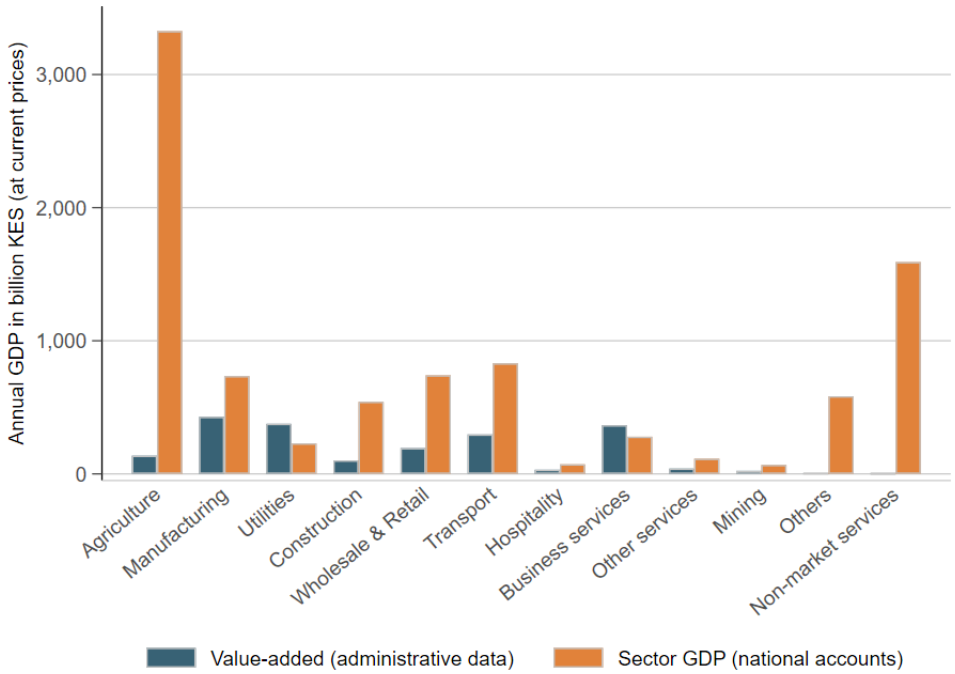
[38]Financial service firms, for example, are the biggest contributor to value added by VAT firms in Chile Huneeus (2018), although they are small in numbers. But, we do not consider financial services in Kenya due to differences in reporting requirements relative to other sectors.Alfaro-Urena et al. (2018) exclude them for similar reasons in Costa Rica.

[39]Non-market services contribute another 22-24% to Kenya's GDP, but are barely represented in the VAT data as most of the entities operating in these sectors are VAT exempt, not-for-profit, or the underlying sector's size in the national accounts being estimated using non-market prices (see penultimate column of Appendix Table 2.B1). Non-market services include education, health, public administration and real estate (Herrendorf et al., 2022). Figure 2.7 highlights another sizeable gap for "others", which includes international organisations, unclassified firms, and financial services.

[40]To arrive at this number, we exclude financial services, non-market services, and agriculture from total GDP. Appendix Table 2.B1 details how the GDP share of the VAT sector changes if each of them is added or removed sequentially.

definitions of informality for firms.

Figure 2.7: Value added by VAT firms vs GDP



This figure compares the sector-level contribution to national GDP to the value added (sales - purchases) of firms covered in the administrative tax records for 2019.

The most relevant excluded sector is agriculture, which generates 20-22% of Kenya's GDP. While part of the sector receives special tax treatment due to exemptions of mainly unprocessed agricultural commodities, some of the GDP gap can also be attributed to informality in the classic sense due to the prevalence of small holders in the sector. Figure 2.7 shows that only a fraction of the sector's GDP is captured in the VAT data.

We now turn our full attention to sectors where gaps arise due to informality rather than tax exemptions. Comparing the value added of each sector (based on the administrative data) to the sectors' contribution to Kenya's GDP (based national accounts) in Figure 2.7, shows that manufacturing and professional services (a sub-component of business services) are best represented in the administrative

data.[41] This pattern aligns with the fact that both sectors are well-connected in the firm network and mostly buy from and sell to other formal firms.[42]

The gap between value added and GDP is an aggregate of all margins of informality discussed in Section 2.5.1.[43] We are unable to quantify the extent to which each margin contributes to aggregate informality. However, we are able to study the importance of each margin across sectors. To quantify the prevalence of informality on the extensive margin, we compare the number of VAT-paying firms to a number of alternative firm counts by KNBS.

The first type of extensive margin informality arises due to non-reporting firms with revenues above the VAT threshold. Figure 2.8 plots both the number of VAT paying firms and firms in the Census of Establishments (CoE) (KNBS, 2017) with revenues of KShs 5 million and above in 2016. The distribution of firms across sectors mirrors each other quite well. Manufacturing, ICT, and professional services show the least deviation between the two data sources (see Table 2.B2 for the same comparison in tabular form).[44] Wholesale & retail stands out as the sector with the largest number of non-compliant firms.[45] This is unsurprising given VAT self-enforcement is weakest for consumer facing firms (Naritomi, 2019), a theme we explore further below.

While the VAT-paying sector accounts for a substantial share of GDP, VAT paying firms are easily outnumbered by a large volume of small businesses with annual revenues below the VAT threshold. In 2016, KNBS (2016) estimated 7.4 million businesses to operate across Kenya. This includes licensed businesses registered with

[41] The ranking of sectors and variation in the formality share observed in Kenya is closely aligned with the findings of Murthy (2019) in India.

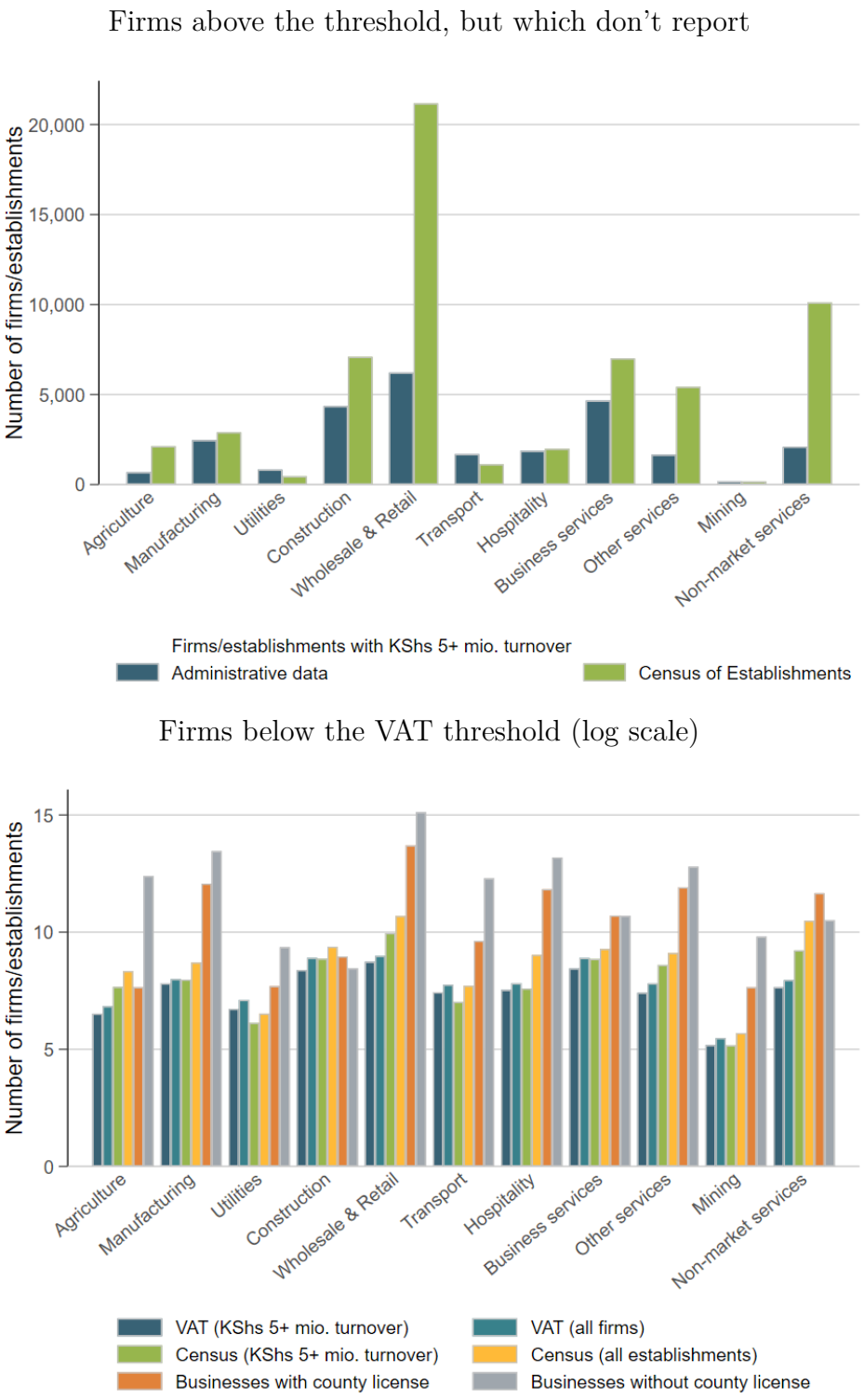
[42] For professional services, as much as 80% of their transaction volume is attributed to trade with other businesses (see Figure 2.1).

[43] The most standard measure for informality in the literature is employment. Private sector firms observed in our sample only employ 0.92 million employees. This corresponds to 4.6% of the overall working population in the 2019 population census. The number improves to about 10.8% in the un-filtered raw data, which covers parts of the government sector, most of the non-profit sector, as well as the financial sector in the tax returns.

[44] The number of firms in the administrative records can be higher because the CoE tends to under-count firms without a highly visible establishment.

[45] In line with the discussion above, the gap for education & health is mainly a result of VAT exemptions that apply to most firms in this sector. The CoE, for example, includes public sector establishments such as public schools that offer educational services.

Figure 2.8: The extensive margins of informality - in which sectors do informal firms operate?



The top graph compares the number of firms covered in the administrative data and had a annual revenue of over KShs 5 million in 2016 to the number of firms with annual revenues above KShs 5 million in the 2016 Census of Establishments (CoE) (KNBS, 2017). The bottom graph compares the two groups of firms to all firms in the VAT data and the CoE, irrespective of their performance in 2016, and the number of licensed and unlicensed businesses reported KNBS in KNBS (2016).

county governments, as well as unlicensed establishments, mostly street vendors and other forms of micro enterprises.[46] Only 1 in 5 of the businesses is licensed. Of the licensed businesses the CoE and the VAT data capture 9% and 2.5% respectively. Figure 2.8 plots the business count by data source and sector. For agriculture, utilities, and construction, the overall number of licensed businesses aligns closely with former two sources. In all other sectors, the number of licensed businesses is substantially higher. As mentioned above, the definition of what constitutes a business and the inclusion of public sector establishments varies by data set. The comparison is thus primarily useful for a big-picture overview. For both cases of firms that do not show up in the administrative data, firms in the wholesale & retail sector show the highest incident of informality.[47]

Turning to the intensive margin of informality, the prevalence and possibility of informal transactions involving registered firms becomes evident when considering the construction sector. As little as 20% of the construction sector's contribution to GDP is reflected in the VAT records. At the same time, the majority of construction sector firms is registered for VAT (see Figure 2.8). Therefore, most of the observed value added gap is a result of informality on the intensive margin, i.e. mis-reporting in VAT returns.

Having established that we are missing out on 36% of the Kenyan economy by looking at the formal sector only begs the question - where to find those informal firms?

Fact 2: Informal firms are usually located downstream of large firms

We find that informal firms are located mostly in consumer facing roles and downstream of large formal firms. In other words, large firms provide inputs for informal firms, while informal businesses in turn often take on the role of distributors in the

[46]The majority of unlicensed businesses is not permanent. 94% have a monthly revenue of less than KShs 50,000 (about $400). KNBS (2016) estimates they contribute about 2.4% of Kenya's GDP.

[47]We do, however, not find any other notable pattern in the data when we correlate the sector-level deviation of the administrative data from firms above the VAT cut-off and the total number of firms in Kenya (see Figure 2.B3).

economy with consumers as their main source of demand (Böhme and Thiele, 2014).[48]

While the high number of non-VAT paying firms in the wholesale & retail sector already is a good indicator for the relative downstream positioning of informal firms along the supply chain, we show that this pattern also emerges from survey data on trading partners of micro, small and mid-sized enterprises (MSMEs) by KNBS (2016). The survey asks firms for the type of entity that best describes their main source of inputs as well as their main customer. Only 2.3% of all MSMEs sell to large firms, while 14.5% purchase inputs from them.[49] Figure 2.9 shows that the pattern holds across sectors.[50] Our results confirm findings by Böhme and Thiele (2014); Zhou (2022) who document similar linking patterns for formal and informal firms in Benin, Burkina Faso, Côte d'Ivoire, Mali, Sénégal, Togo (Böhme and Thiele, 2014) and West Bengal, India (Zhou, 2022) respectively. While a dual economy view on the interaction between the formal and informal sector is widespread (La Porta and Shleifer, 2014), the lines between two are often blurred. Böhme and Thiele (2014), for example, find that trade between formal and informal firms varies with the degree of informality. More formalised firms (unregistered firms with high captial stocks) purchase as much as 25% of their inputs from formal industrial enterprises and thus a much higher proportion than the average informal firm. The hierarchical structure of formality suggests that the vast majority of informal firms ultimately sources formal sector inputs through both indirect and direct linkages.[51] The point estimates for direct linkages between the formal and informal sector can therefore

[48] Cordaro et al. (2022), for example, show how microenterprises subsidise the distribution of fast-moving consumer goods of a multinational in Kenya.

[49] KNBS (2016) defines non-MSMEs/large firms as entities with more than 99 employees. Other survey response options for inputs are: MSMEs, farmer, direct import, individual supplier, government. The corresponding options for customer types are: MSMEs, direct exports, individual consumer, government.

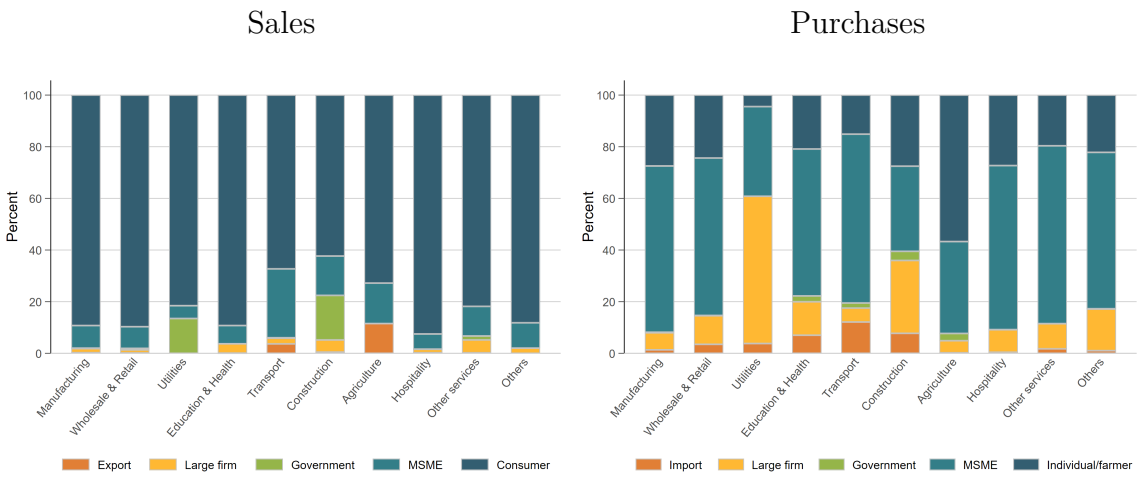
[50] The survey responses can be interpreted as a lower bound on the interaction between the VAT-registered and non-registered sector. The main trading partner of MSMEs are other MSMEs and the relevant survey question does not distinguish between the three types of enterprises - micro, small, and medium - summarised under MSME. Kenya's Micro and Small Enterprises Act No.55 of 2012 defines small enterprises as firms with up to 50 employees and up to KShs five million annual turnover (KNBS, 2016). Medium sized enterprises thus cross the VAT threshold of five million and any link with them would count as a link with a formal firm.

[51] Think of a small retailer selling soap. While the retail might have purchased the soap from an informal wholesaler, the wholesaler itself or its suppliers will have sourced the soap from a formal manufacturer or importer further up the chain.

be interpreted as a lower bound on the degree of integration between the two types of firms. Nevertheless, despite large and small firms indeed interacting with each other, the VAT system itself creates wedges in firms' supply chains (De Paula and Scheinkman, 2010; Gadenne et al., 2022; Zhou, 2022). A VAT-registered retailer in a third tier town in Machakos County said:

> *"We don't buy from the local distributor as they are not filing tax returns and hence do not issue receipts. So we cannot transact with them. [..] To make it easier for us to buy from manufacturers and access trade credit, we opted for our shop to be VAT-registered."*

Figure 2.9: Links of small and medium sized enterprises to large firms



The figure draws on data from the 2016 Small and Medium Enterprises (MSME) Survey by the Kenya National Bureau of Statistics (KNBS, 2016). The survey asks each firm for their main input sources and their main customer type. We restrict the sample to participating firms with an annual revenue below the VAT registration cut-off. Note that the category "MSME" also contains medium sized firms which can include formal tax-registered firms. The percentage captured in the "Large firm" category thus represents a lower bound on linkages between small non-VAT registered businesses and large VAT-registered private sector firms. KNBS (2016) defines non-MSMEs/large firms as entities with more than 99 employees.

The two findings that the extensive margin of informality is particularly prevalent in the wholesale & retail sector, as well as informal firms being more likely to purchase from larger firms rather than vice versa, are well in line with the underlying enforcement structure of VAT systems. This enforcement mechanism incentivizes downstream firms to ask their suppliers for receipts in order to claim input VAT they can then deduct from the output VAT they have collected on behalf of the revenue

authority. The weak link of any such system are consumers or VAT-exempt entities, who are not eligible for VAT refunds and hence do not have an incentive to ask for a receipt Naritomi (2019). Put differently, we expect a larger share of economic activity to take place outside the VAT system in more downstream sectors.

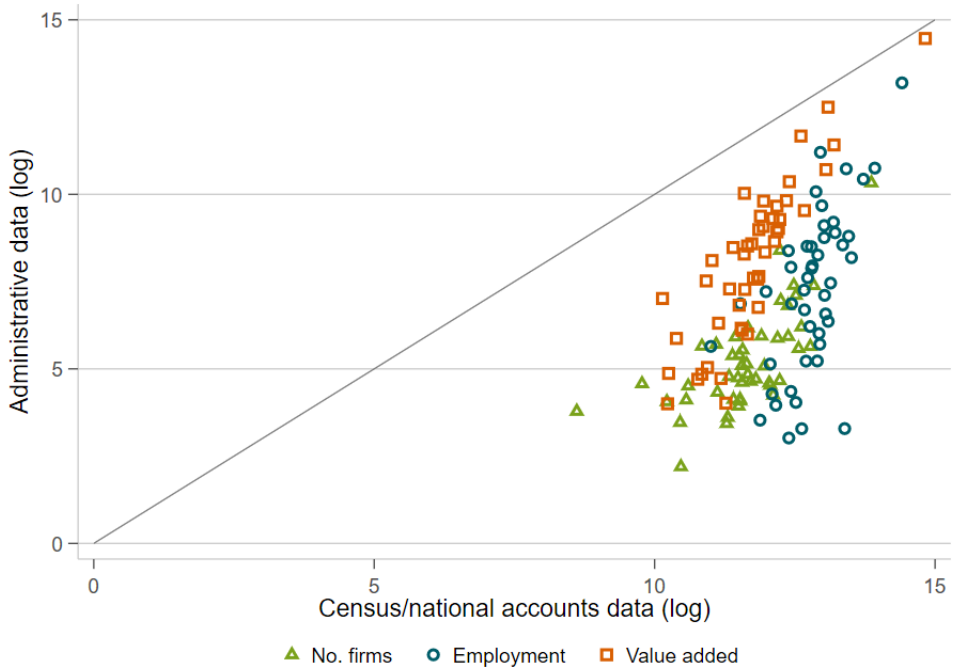
Fact 3: The VAT-sector share correlates positively with regional economic size and income levels

Moving on from sectors, where do we find informal firms in space? Here the answer is: in smaller markets.

We find a strong correlation of both the overall economic size and income levels (Gross County Product per capita) with the county's formal sector share. A county's economic size explains between 35% and 52% of the variation in the the county's formal sector share (see Figure 2.B4). We visualise the pattern in the top graph of Figure 2.10 where we plot the aggregate number of firms (green), value added (orange), and employment (turquoise) against their respective KNBS benchmark statistics. Each marker in the scatter plot corresponds to one of the 47 counties. In case of a 1:1 mapping the markers would lie on the 45-degree line (in blue). For cases where the KNBS benchmark exceeds the corresponding figure from the administrative data, the scatter lies below the 45-degree line and vice versa. Larger counties lie closer to the 45-degree line than smaller ones. This pattern is observed across all three of the available indicators we are able to benchmark against KNBS data: employment, value added, and the number of establishments. We note, however, that the markers for employment stay further away from the 45° line for economically larger counties. This finding is in line with the general notion previously shown in cross-country studies that informality of output declines less steeply with income levels than the share of informal workers (Kose et al., 2019).

To test whether the positive correlation of market size and the formal sector share is an artefact of the administrative records, we also correlate three other employment-based formality measures with the Gross County Product in Figure 2.11. While the slope becomes flatter for measures that apply a more stringent definition of informality, with the VAT-based measure being the most stringent one, the R^2 barely changes. This suggests that the variation in county-level informality that is explained by the economic size of the county is very similar for different measures of informality.

Figure 2.10: Count-level comparison of the number of firms, employment, and value added in KNBS benchmark data with the administrative data

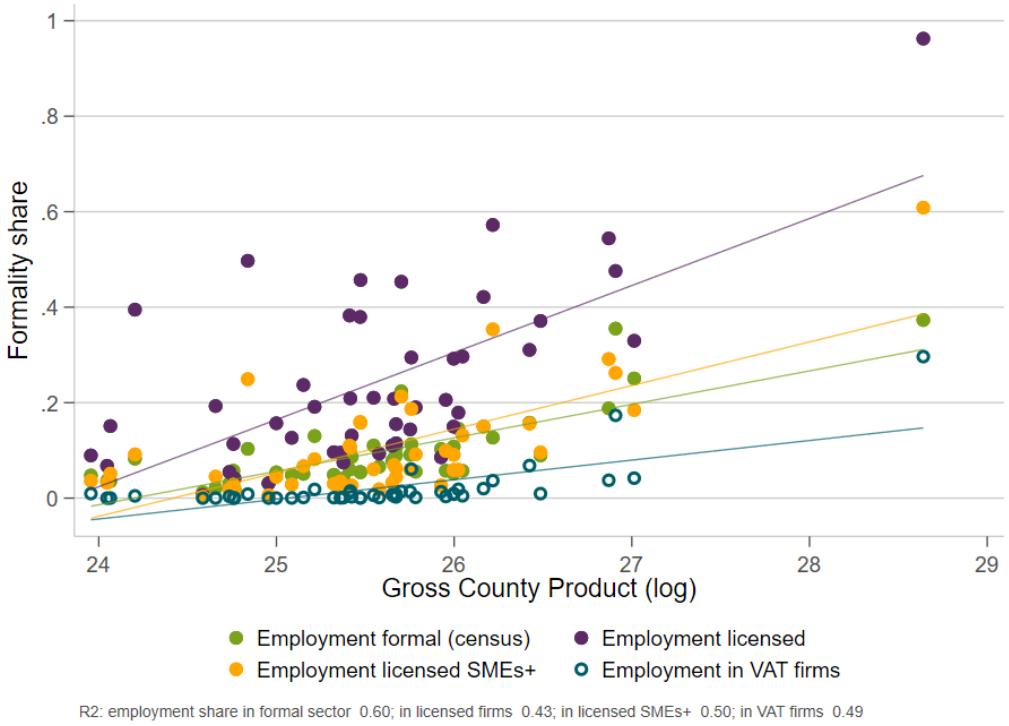


This figure plots county-level indicators derived from the administrative firm data to other county-level statistics published by the Kenya National Bureau of Statistics. Each marker represents one of Kenya's 47 counties.

Fact 4: Kenya's spatial concentration of economic activity is predominantly a feature of the formal sector

As a next step, we revisit the question of spatial concentration of economic activity. To achieve this, we expand Table 2.1 from Section 2.4.3 with additional measures of economic activity, presented in Table 2.7. We observe that spatial concentration becomes more pronounced as we move from less formal to more formal economic activities. The universe of both unlicensed and licensed businesses (KNBS, 2016) exhibits a more even dispersion across space compared to licensed businesses alone. In turn, licensed businesses show a more equal distribution than formal entities engaged in industrial production (KNBS, 2010), many of which were likely VAT-paying firms in 2010. This pattern aligns with Obudho (1997)'s discussion of spatial concentration in economic activity back in 1992 when Nairobi accounted for 73% of formal sector employment in Kenya.

Figure 2.11: Share of formal sector employment and regional market size



The first measures uses the formal sector employment share according to the 2019 population census, the second measure considers the number of employees in licensed businesses, the third uses the same measure but disregards mirco-enterprises, and the forth measure considers employment in the tax records. Each measure represents a share, i.e. captures the proportion of economic activity that can be attributed to the formal sector. For an exact definition of each measure see Table 2.5

Regarding the role of information frictions in spatial concentration, one interpretation of this finding is that these frictions have a greater impact on formal sector firms. Formal firms tend to be less transient than micro-firms (KNBS, 2016), and their owners have likely incurred higher fixed costs to establish their businesses. Consequently, they may seek to safeguard against the risk of exiting due to rapidly changing market conditions. In a similar spirit, studies by Kumar et al. (1999) and Laeven and Woodruff (2007) suggest a positive correlation between the quality of local institutions, particularly the judicial system, and firm size. Considering that proximity to the capital often ensures better access to the judicial system, we would expect larger firms to concentrate in Nairobi and nearby counties.

Table 2.7: Geographic concentration of economic activity by degree of formalisation

	Nairobi	**Mombasa**	**Pareto exponent**	
	in %		α	SE
Population overall	9	3	1.29	0.18
Population of cities & towns	31	9	0.85	0.01
GDP	29	5	0.98	0.07
GDP w/o agriculture	37	6	0.95	0.06
GDP w/o non-market services	24	5	0.91	0.08
No. MSMEs	14	3	0.86	0.17
Employment in MSMEs	19	3	0.78	0.13
No. licensed MSMEs	18	3	0.73	0.09
Employment in licensed MSMEs	28	3	0.67	0.07
No. SMEs	37	3	0.58	0.06
Employment in SMEs	36	3	0.60	0.05
No. census establishments	36	4	1.10	0.12
No. firms census of industrial production	48	6	0.54	0.02
Sales census of industrial production	61	7	0.32	0.03
No. VAT firms	64	9	0.61	0.04
Employment in VAT firms	62	8	0.35	0.03
Value added of VAT firms	72	10	0.37	0.03
Network sales	68	13	0.24	0.02
Network purchases	60	9	0.43	0.02

The columns for Nairobi and Mombasa report their share of the respective national aggregate figures (e.g., Nairobi's contribution to Kenya's GDP). The Pareto exponent α is the estimated coefficient from a county-level regression of each county's rank (log) on the respective measure x (log): $\log \text{rank} = \log A + \alpha \log x$.

Fact 5: Within-sector informality variation drives spatial disparities in aggregate informality

Having considered both the difference in informality levels across sectors as well as regions within Kenya, we ask which of the two margins matters more for spatial inequality in formal sector activity. In other words, to what extent does inequality in formal sector activity arise solely due to differences in sector compositions across space? Or does it arise because the share of formal activity within a sector varies across counties?

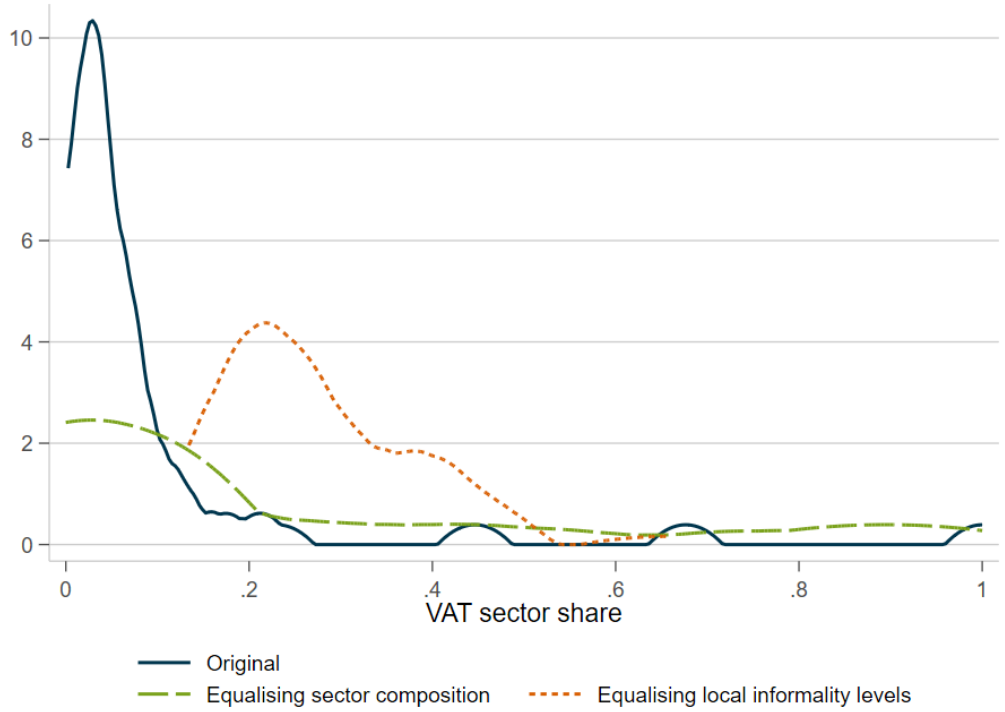
We rely on an accounting exercise with two counterfactual scenarios to study this question. We use the value-added based measure of informality for the purpose of this exercise. We start by noting that the aggregate formal sector share f_c is each county c is the sum of each sector s's degree of formality f_{sc} weighted by the respective sector's contribution to the county economy s_{sc}:

$$f_c = \sum s_{sc} \times f_{sc} \tag{2.3}$$

In the first counterfactual scenario, we alter the sector composition of each county's economy to align with Kenya's sector composition on a national level, i.e. we equalise s_{sc} for each sector to align with its national share s_{sKenya}. In the second counterfactual scenario, we keep the sector composition unchanged but harmonise the degree of formality of each sector f_{sc} to align with the national degree of formality f_{sKenya}. We plot the baseline dispersion in county-level formal sector shares in Figure 2.12 alongside the results from the two counterfactuals. First, note that the majority of counties have a formal sector share f_c of less than 20%. Only three counties have an above average degree of formality. The unweighted average formal sector share is 9%. By design, neither of the counterfactuals notably change weighted aggregate formality levels.[52] They do, however, affect the spatial dispersion. Altering the sector composition s_{sc} to s_{sKenya} increases the formal sector share in counties with the lowest levels of informality. As a result the average formal sector share increases from 9% to 11%. This moderate impact pales in comparison with the counterfactual that alters the formal sector shares f_{sc} to align with the national level of formality f_{sKenya}, but keeps the sector composition s_{sc} of each county constant. In this scenario, all counties' construction sectors, for example, get assigned the same degree of formalisation. As a result the (unweighted) average formal sector share f_c jumps up to 28%. As the dashed orange density plot in Figure 2.12 shows, the counterfactual pushes the formal sector share f_c above 20% in virtually all counties. The variance in outcomes shrinks drastically.

[52]While counties with low levels of informality will improve if national standards are applied, the top counties are worse off.

Figure 2.12: Spatial variation in informality - a story of sectors or geographies?



The above graph plots the pdf for the dispersion of formal sector shares across Kenya's 47 counties. The solid blue line captures the original dispersion in informality levels across counties. The long-dashed green line keeps local informality levels constant, but equalises the sectoral composition of the counties' economies by weighting each sector based on its share of the national economy. The short-dashed orange line instead equalises the local level of informality to be the same across all counties and to correspond with the respective sector's national formal sector share.

What have we learned from the five stylized facts? The level of informality varies drastically across space. Informal firms can be found in downstream activities, which in turn are relatively more important in smaller markets. We would therefore expect that not observing informal firms would results in us underestimating the connectednness of remote counties, especially in terms of the number of outlinks of their firms. To explore this hypothesis we rely on a network formation model that allows us to run counterfactuals.

2.6 A model of network formation with sectors and regions

We apply the network formation model proposed by Bramoullé et al. (2012) to our setting. The purpose of the model is to study how the spatial inequality in trade links would look like if we were able to observe informal firms. Bramoullé et al. (2012) is particularly well-suited for our purposes for three reasons: First, it allows us to easily incorporate two key dimensions of firm heterogeneity - sectors and geography. Sectors reflect the underlying input-output structure, while the geographic dimension allows us to study the question of spatial inequality. Second, the framework is explicitly tailored towards studying a network formation process of networks whose degree distribution follows a power law. In this class of models, the power law distribution is generated by so-called preferential attachment, i.e. link formation via existing links rather than via search mechanisms that do not rely on the network environment. The framework allows us to estimate the share of firm-to-firm links formed via existing network links versus undirected search (often referred to as *random search* in the networks literature (Jackson and Rogers, 2007; Bramoullé et al., 2012; Chaney, 2014)). The motivation at the core of this type of set up is that searching via existing suppliers allows firms to overcome information asymmetries about future supplier's quality type (Chaney, 2014). Using a model allowing for preferential attachment to overcome information frictions also links back to theories on the emergence of urban centres as "communication nodes and loci of information exchange and accumulation" (Memon, 1976). Lastly, we opted for a simple tractable framework to stay focused on our main goal to estimate the spatial dispersion in firm-to-firm links. We abstract from complexities like endogenous firm entry and exit or the decision of a firm to formalise. The underlying dynamic network formation process gives rise to the widely documented extreme heteroegeneity in outdegrees across firms (Bernard et al., 2019; Panigrahi, 2022; Bernard et al., 2022; Bernard and Zi, 2022; Demir et al., 2023; Arkolakis et al., 2023). We first present the dynamic network formation model proposed by Bramoullé et al. (2012) below and then discuss how we estimate it.

2.6.1 Model Setup

Consider an economy with a set of firms N. Each firm $i \in N$ is of a given type $\theta_i \in \Theta$ where Θ is the set of all possible types. In our application, we specify firm types as unique sector-county pairs.[53] i.e. all firms in the same sector and county are classified as the same type.

The network formation process is as follows. In every period t, a new buyer firm of type θ enters with probability $p(\theta)$. In order for its operations to be viable it needs to source inputs from suppliers through a fixed number of links m. It first chooses a sector-county pair (i.e. a type) with probability $p(\theta, \theta')$ for all θ' in Θ. Then, it forms m links with firms of the chosen type(s). The probabilities $p(\theta, \theta')$ represent the firm's bias in terms of sectors and regions it wants to link with. Having chosen the sector-region type it wants to link with, the firm now relies on two different search technologies to form its m links: first, undirected search (aka random search). Here, the new firm "randomly" links to other firms of the chosen type. It forms a fraction r of its total m links in this manner. Second, preferential attachment. The new firm forms the remaining fraction $1 - r$ of its m links to suppliers by searching among the existing suppliers it acquired via undirected search. In other words, once the buyer firm forms links to the first set of suppliers, it then "randomly" links with the suppliers of its suppliers. The second step of this process is preferential in that suppliers that are more connected are more likely to be chosen. This process continues for several time periods and the network evolves accordingly.

Let us note two important aspects of this process. First, while a firm's number of buyers evolves over time, the number of suppliers that a firm chooses is fixed to m and does not change as the network grows. While this is a strong assumption that we will maintain, we can also imagine this to reflect a fixed production technology that the firm needs to operate. It is further consistent with the third stylized fact from Section 2.4.3, documenting that the number of inlinks is more evenly distributed across firms and localities relative to the number of outlinks. Second, the model

[53] In the original paper by Bramoullé et al. (2012) the authors apply the framework to a setting where nodes correspond to academic papers, links to citations, and types to research fields.

includes "biases" in that the probability that a buyer of type θ finds a supplier of type θ' may not necessarily be equal to the probability of θ' in the firm population. This aspect of the model captures the core features of the context that we are interested in as these biases can reflect production technologies or homophilous preferences arising out of search costs and information frictions. Firms in a location θ might find it easier to link to firms in location θ' that is close to them as opposed to firms in location θ'' that is far. Likewise, firms in sectors that supply services like electricity or telecommunication, which almost every firm requires as inputs, might find themselves with linking probabilities $p(\theta, \theta')$ that exceed their entry probability $p(\theta)$.

At the aggregate level, we are interested in the outdegree of each sector-county type. To this end, consider a matrix B where each row and column represents a type $\theta \in \Theta$. Its $\theta\theta'$'th entry is then equal to $p(\theta)\frac{p(\theta,\theta')}{p(\theta')}$. Bramoullé et al. (2012) rely on $\mathbf{B}$ to derive the following matrix whose ij'th entry shows the number of directed links at time t between buyers of type i and suppliers of type j which are born in t_0:

$$\pi^t_{t_0} = m\frac{r}{1-r}(f(t, \mathbf{B}) - \mathbf{I}) \tag{2.4}$$

Here, t refers to the time period, $\mathbf{I}$ is the identity matrix, and f is a scaled geometric series of the matrix $\mathbf{B}$ defined as follows:

$$f(t, \mathbf{B}) = \sum_{\mu=0}^{\mu=\infty} \frac{((1-r)\log(t)\mathbf{B})^\mu}{\mu!}$$

Newly entered buyers form m inlinks in every period.[54] As a result the outdegree of existing firms, i.e. the suppliers of the newly entered firms, evolves over time. Thus, the matrix $\pi^{\mathbf{t}}_{\mathbf{t_0}}$ gives the expected outdegree (i.e. number of buyers) of each column node born in time t_0 to a row, computed at time t. The purpose of the dynamic network formation process is to rationalise the heterogeneity in outdegree. At the same time, the framework's intention is not to study the dynamics themselves, but rather consider the network's steady state properties.

[54]Note that we diverge from Bramoullé et al. (2012) here. In their setting newly entered nodes (scientific papers) form outlinks (citations) with the implication that everybody's inlinks evolve over time.

2.6.2 Estimation strategy

Given the granular data on the empirical firm-to-firm network, we are able to obtain the majority of the model parameters from the data. These include all probabilities $p(\theta) \in \Theta$ that a firm enters in a given sector and county as well as all interaction probabilities $p(\theta, \theta')$ between sector-county types. As a result, we only need to estimate the parameter r, i.e. the fraction of input links a firm obtains via undirected search independent of the network environment. We use the cross-section from 2019, the last pre-COVID year of our panel, to obtain the $p(\theta)$s and $p(\theta, \theta')$s.[55]

First, we classify firms into types defined by unique sector-location combinations where the location is given by the county in which the firm is located. For example, all firms in the manufacturing sector in Nairobi are classified as the same type. Next, we compute the probability that a type exists for all types in Θ. We do so by dividing the number of formal firms of a sector-county type by the total number of formal firms in the economy. The interaction probabilities $p(\theta, \theta')$ then represent the fraction of a sector-county type θ's inlinks that it forms with type θ'.[56] We compute the above probabilities for all possible combinations of types and use them to construct the matrix $\mathbf{B}$. Moreover, we follow Jackson and Rogers (2007) and define m as the average indegree of the network. The variable t denoting time is found by dividing the total number of links in the 2019 network with the average indegree and is equal to 50,897. This is because the model predicts that m links are formed by buyers in every period implying that the total number of links in the network must be mt.

Using the parameters from the empirical data, we are able to predict the matrix of type-to-type network links $\pi(r)$ for different choices of $r \in [0, 1]$. However, we face two concerns. First, note from Equation 2.4 that $\pi^t_{t_0}$ only tells us the expected outdegree of types born in t_0 evaluated at time t. Since a new firm is born in every period up until period t, we need to aggregate these matrices across all time periods

[55] In particular, we use all firms and their linkages in the year 2019. We exclude a small proportion of firms that do not report buying from any other firm. This is because the model requires all entering firms to form m buying links with existing firms. Including firms who never buy from any other firm does not align with this feature of the model.

[56] The model also allows for self links. Wholesale firms in Nairobi, e.g., are able to buy from other wholesalers in the city.

leading up to t to get the type-by-type adjacency matrix of the network as a whole. The matrix of connections at time t is given by $\pi_t = \sum_{t_0}^{t} (\mathbf{p} \cdot \pi_{t0}^{t}{}')'$ where $\mathbf{p}$ is a column vector with the probability that each type is born. For example, we compute the probability that a node of a certain type is born in time t_0 and its expected links in time t with every other type to get $\mathbf{p} \cdot \pi_{t0}^{t}$. Then, we repeat the process again to compute the probability that a node of a certain type is born in time t_{0+1} and it's expected degree in time t to get $\mathbf{p} \cdot \pi_{t0+1}^{t}$. We have to undertake this exercise for all time periods leading up to t. In other words, we must compute t such matrices and add them up to give us the type-by-type degree distribution at time t. It is computationally difficult to compute π^t for $t = 50897$ in every iteration while looping through different candidate values of r during computation. As a result, in every iteration, we only compute π_{t0}^{t} for 500 representative time periods over which we then aggregate to obtain π_t. We space the sample of 500 time periods equally out between our first period $t_0 = 1$ and our final period $t_0 = 50879$. This approach ensures that do not disproportionately sample from either older or younger nodes and hence bias our results. For example, sampling from nodes born in the first 500 periods will lead us to predict the type-by-type outdegree distribution only for firms at the right tail of the firm degree distribution if the observed network happens to exhibit preferential attachment since preferential attachment results in older nodes having a higher chance of being more connected. This is because older nodes are likely to be more connected. This can bias our estimation of r as we will be matching the predicted distribution of such firms with all firms observed in the data.

As a result, we compute $\pi_t = \sum_{t_0=1:100:50897} (\mathbf{p} \cdot \pi_{t0}^{t}{}')'$. This implies that we will under-predict the average degree of the network as our model ignores firms born between specific time periods. At the same time, it ensures that our estimate of r is not dependent on including or excluding specific types of older or younger firms. Even if the network is scaled down in terms of number of firms, the features of the network are kept intact.

Second, based on Bramoullé et al. (2012)'s formula for predicting π_t also requires us to compute a geometric series of matrix $\mathbf{B}$.[57]

[57]For ease of computation, we restrict this to the first five entries of the geometric series as the matrix has very small entries afterwards.

In addition to the predicted version of the matrix π, we also observe the actual π in the data where the ij'th entry of π is just the number of links between types i and j. We match the model predicted matrix and the matrix in the data using the method of moments procedure to obtain r^*. Each moment is weighted by the probability with which we observe a specific sector-region type in the data.[58] In particular, r^* is defined as follows:

$$r^* = \arg\min \sum_{\theta} \sum_{\theta'} p(\theta')(\pi_{model}(\theta, \theta'; r) - \pi_{actual}(\theta, \theta'))^2$$

r^* is obtained by minimising the distance between the model predicted matrix of type-by-type interactions and the corresponding matrix obtained from the data. We estimate r using simulated annealing. Having to only estimate a single parameter comes with the advantage that we can plot the above objective function for various values of r to ensure that our estimated value is indeed the global minimum (see Figure 2.13).

2.6.3 Estimation results

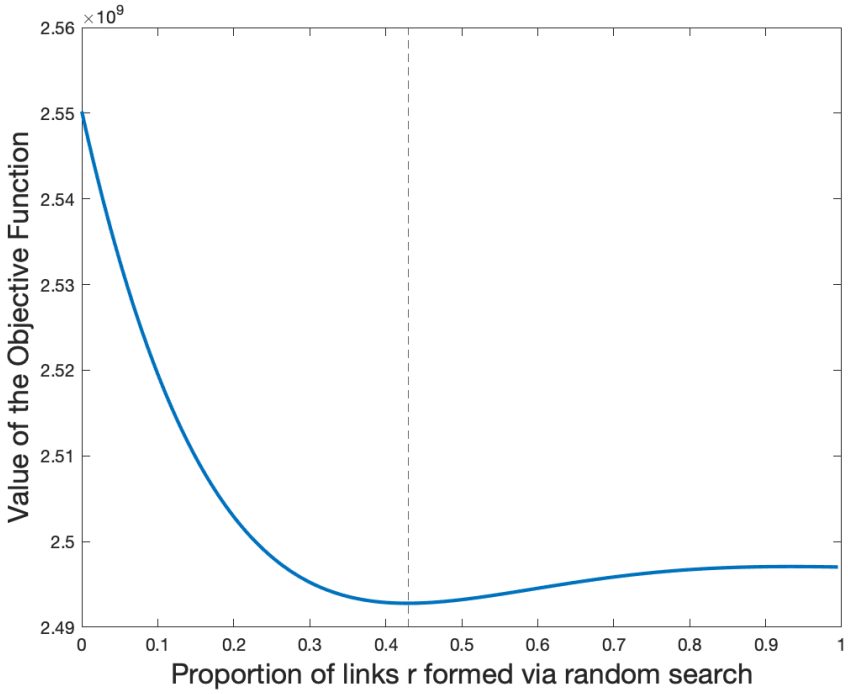
Our estimation strategy yields a result of $r^* = 0.42$. It suggests that a newly entered firm chooses 42% of its m suppliers randomly, and the remaining 58% among the suppliers of its existing suppliers. A network with 58% of all links being formed via preferential attachment suggests a prominent role for information frictions. It aligns with previous research documenting the importance of relational contracts in Kenya and neighbouring economies (Fafchamps, 2003). It is further backed by our own finding in Section 2.4.3 that social connectedness explains substantial variation in cross-regional trade flows, over and above distance. In the empirical section we found the elasticity of domestic trade flows with respect to social connectedness to be more than twice as high than for international trade flows (Bailey et al., 2021). In a variant of this model, Chaney (2014) estimates $r = 0.6$ for French exporters forming links with trade partners abroad, which also

[58] In doing so, we assign greater weight to more common sector-regions types whose probabilities often tend to more stable over time.

suggests a substantial, but not quite as prominent role for information asymmetries.[59]

In order to validate the estimated value of r, we plot the objective function in Figure 2.13 and show that it reaches a global minimum as $r = r^* = 0.42$.

Figure 2.13: Objective function for various values r



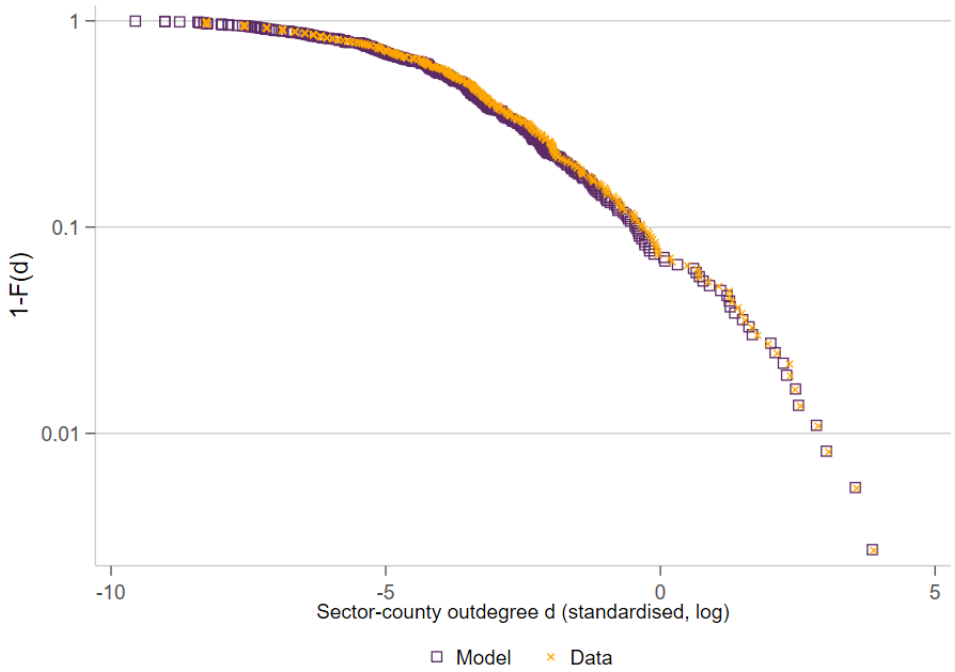
This figure plots the sum of the squared difference between each element of the model predicted interaction matrix π and the matrix π directly observed in the data, for various values of the parameter $r \in [0, 1]$. The figure shows that $r^* = 0.42$ obtained via simulated annealing minimises the objective function.

Next, we plot the degree distribution (i.e. total number of outlinks) of each sector-county type as observed in the data and as predicted by the model. As discussed, our predicted network will have a lower average degree than the real-world data. To compare the predicted degree distribution to the degree distribution in the data, we therefore standardise the outdegrees by dividing them by the mean of the respective

[59]From a welfare perspective, Chaney (2014) shows that in the case of highly substitutable goods, increasing r boosts aggregate welfare as it ensures more equal access to a variety of goods across buyers. However, aggregate welfare gains from overcoming preferential attachment dynamics are more ambiguous if goods are less substitutable. Boosting the average indegree m, i.e. network density, on the other hand is always beneficial.

degree distribution. Figure 2.14 shows that the key properties of the outdegree distribution are replicated by the model's predictions. Model and data match particularly well in the right tail of the distribution and hence the part that is specifically targeted by the preferential attachment framework. Estimating the Parto exponent for both degree distributions, we obtain an α of 0.3599 from the model and 0.3628 from the data. The standard errors are 0.0069 in both cases and we can hence not reject the null hypothesis that both coefficients are equal.

Figure 2.14: Actual and predicted distribution inverse CDF



This figure plots inverse CDF for the actual and model-predicted total number of outlinks for each type (i.e. sector-county pair). The number of outdegrees is standardised. Note the log scale on both the x- and the y-axis.

2.7 Spatial inequality and unobservable informal firms - a counterfactual

With r in hand, we are now able to tackle the question of informal firms and spatial inequality in network links. Our proposed thought experiment is the following: suppose we were to observe informal firms. What would happen to the outdegree distribution of various types θ? In model terms, our counterfactual shifts the

probabilities $p(\theta)$ with which we observe nodes of certain sector-region types θ to be born. Against the background of the stylized facts presented in Section 2.5, we would expect our current estimates for $p(\theta)$ based on the administrative data to under-account for downstream firms in counties with smaller local markets. We will therefore use our population census-based measures of informality from Section 2.5 to update our estimates for all $p(\theta) \in \Theta$. Knowing r^* and our updated $p(\theta)$s, we can then once again predict the matrix π, keeping everything else constant.

The proposed counterfactual hinges on a strong assumption. We assume that the type-by-type linking biases i.e. $p(\theta, \theta')$ do not change even after informal firms are included in the model. Put differently we assume away an endogenous relationship between $p(\theta)$ and $p(\theta, \theta')$. For example, Nairobi's manufacturing sector, which has high interaction probabilities with wholesalers in Garissa county continues to do so even as a wave of new informal wholesale firms enters in Garissa. Holding the interaction probabilities of the two sector-region types constant can be interpreted as a reflection of the economy's fundamental input-output structure and geography. Further, anecdotal evidence suggests that smaller retailers often benefit from being located in the vicinity of larger establishments who receive deliveries from larger distributors or manufacturers. At the same time, however, the argument can be made that informal firms in Garissa are unable to directly link with Nairobi-based firms thereby reducing their interaction probability. This argument would be in line with a hierarchical structure of supply chains within a sector-county cell, i.e. larger wholesalers sourcing directly from Nairobi, while selling to smaller retailers in Garissa, who in turn do not end up sourcing directly from Nairobi themselves. We will revisit the implications of keeping $p(\theta, \theta')$ unchanged when we discuss the counterfactual results.

2.7.1 Predicting the sector-county profile of non-VAT firms

To incorporate informal firms into the network, we need to update the firm-type probabilities $p(\theta)$ for each sector-county pair. We will refer to the updated, alternative versions of $p(\theta)$ that incorporate the sector-county characteristics of informal firms as $p(\theta_a)$. To update $p(\theta)$, we ideally would want to observe the number of firms N_{cs} in each sector s and county c cell – irrespective of their formality status. However,

none of the KNBS records available to us feature a breakdown of the firm count along both the sector s and the county c dimension.[60] Therefore, instead of the firm count, we rely on the number of people who work in the private sector in each sector-county cell to compute $p(\theta_a)$:

$$p(\theta_a) = \frac{\text{No. people who work in sector s in county c}}{\sum_c^{47} \sum_s^{12} \text{No. people who work in sector s in county c}}$$

The denominator is equivalent to the total number of people who work in the private sector (employed and self-employed, formal and informal) across Kenya. For both $p(\theta)$ and $p(\theta_a)$ the sector-region probabilities sum to one. This is important to keep in mind for the interpretation of the counterfactual: $p(\theta_a)$ captures a relative change in the number of firms rather than an absolute change.

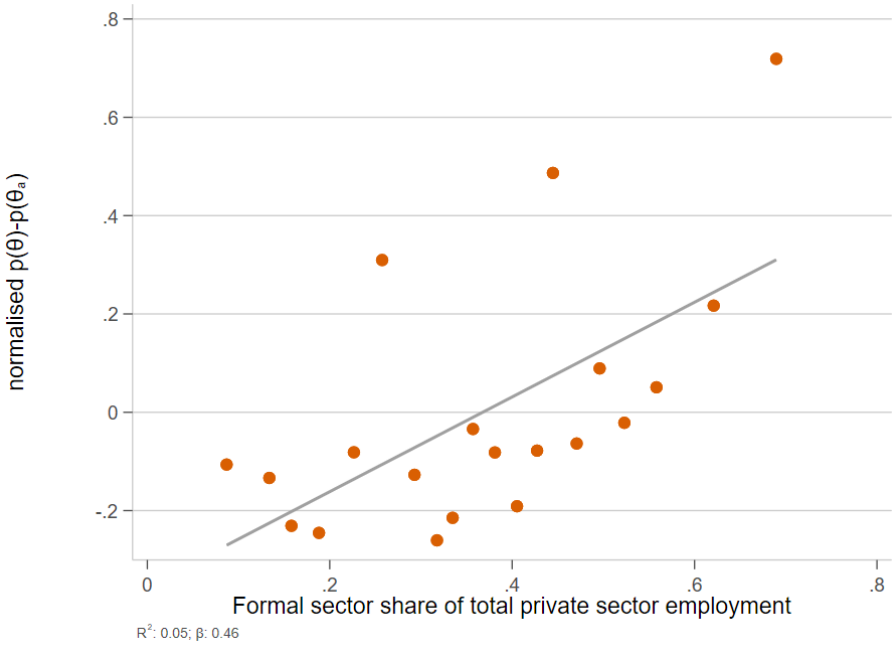
We apply an exception for agriculture and non-market services: We estimate their $p(\theta_a)$ drawing only on formal private sector employment as formal VAT-paying firms occupy a very specific niche in both cases (see discussion in Section 2.5.2).[61] Amending the $p(\theta)$s of agriculture and non-market services using records on total employment (instead of formal private sector employment only) would greatly overestimate the number of firms that are operating in these sectors and participate in a similar manner in the private sector network to their peers.

What are our expectations about the differences between the probability $p(\theta)$ that a formal firm enters in a given sector-county cell versus the alternative probability $p(\theta_a)$ that takes informal ones into account? Sector-county types with a high degree of formality will have a $p(\theta)$ that is larger than the new $p(\theta_a)$. Put differently, their importance in the economy is overstated by the administrative data. For sector-regions with high levels of informality we thus expect $p(\theta_a)$ to be larger than $p(\theta)$. In Figure 2.15, we plot the share of total private sector employees who have a formal sector

[60] 12 and 47 refer to the 12 aggregate sectors and 47 counties respectively.

[61] Alternatively, we could have excluded them entirely from both the counterfactual and the model estimation. However, this would entail disregarding valid participants in the network. Especially commercial agriculture businesses are highly interlinked with local firms for non-agricultural inputs and services. Firms in these sectors participate in 10% of the observed firm-to-firm relationships.

Figure 2.15: Sector-region probabilities and formal sector shares



The graph plots each sector-regions formality share against the normalised difference between the baseline $p(\theta)$ and the augmented version $p(\theta_a)$ that takes into account informal firms. $p(\theta)$-$p(\theta_a)$ is reported in terms of standard deviations. For a version of the same graph as a local polynomial with confidence intervals and including two alternative measures see Figure 2.C1.

job[62] against the difference between the baseline $p(\theta)$ and the augmented $p(\theta_a)$s.[63] A 10 percentage point increase in formality leads to an increase of $p(\theta)$-$p(\theta_a)$ by one percentage point (0.46 standard deviations).[64]

By using simple employment shares to compute $p(\theta_a)$, we rely on the assumption that the mapping of employees to firms is the same across all sectors and regions. This assumption does not hold empirically. Manufacturing firms, for example, tend to be larger than businesses in the hospitality sector. Nairobi hosts larger firms than Mandera County in Kenya's north. We therefore propose two alternative approaches to compute $p(\theta_a)$, which entail re-scaling the number of employees using the average firm size in each sector-county cell. In the first alternative we use the administrative

[62]We use total private sector employment rather than the overall working population as the denominator to abstract from employment in public sector entities.

[63]We normalise the difference such that the y-axis can be interpreted in terms of standard deviations.

[64]To estimate the slope in Figure 2.15, we exclude five sector-county pairs which are adjusted by more than two standard deviations. All of the five sectors are Nairobi-based. The slope becomes twice as steep if the five sectors-county pairs are included.

data to compute the average number of employees for each sector-county. For the second one, we use the 2016 Medium, Small and Mirco Enterprise survey to compute the average number of employees (KNBS, 2016).[65] Given both measures of firm size by employment do not cover the entire firm size distribution (the survey covers the left tail, while the administrative data cover the right tail), we rely on the plain-vanilla employment shares as the default measure and consider the other two robustness checks. We plot the default measure and the two alternative measures against the respective sector-region formality shares in Figure 2.C1. Plotting them as the results of a local polynomial with confidence intervals reveals that the three measures behave similarly.[66]

2.7.2 Counterfactual results

We find that accounting for informal firms increases the variation in outdegrees across counties by 21% (see Table 2.8). Meanwhile, the standard deviation to mean ratio in outdegrees across all sector-county types increases by 17%. The dispersion of outdegrees across sectors on the other hand falls by 7%. We look at the coefficient of variation as the key metric. Adjusting for the mean accounts for the fact that the number of outlinks predicted by the model needs to be looked at in relative rather than absolute terms (see Section 2.7.1). Rather than thinking of the counterfactual as adding new firms, we adjust the weights of each sector-region type.

Table 2.8: Changes in the dispersion of outdegrees

Aggregation	Δ sd/mean (in %)
County outdegree	20.8
Sector outdegree	-6.9
Sector-county outdegree	17.1

The above table reports the result from our baseline counterfactual that adjusts entry probabilities of nodes in sector-county cells using employment figures from the 2019 population census. Results excluding the degrees of Nariobi and Mombasa or adjusting our estimates for the entry probabilities by two different proxies for the firm size distribution are reported in Table 2.C1.

[65] Medium-sized enterprises in the survey can have up to 100 employees. Using the enterprise survey comes with the caveat that the data is only representative at the county, but not necessarily sector-county level.

[66] The only notable difference is that correcting for firm size by relying on the MSME survey yields significantly higher entry probabilities in sector-county cells with the lowest formal sector share.

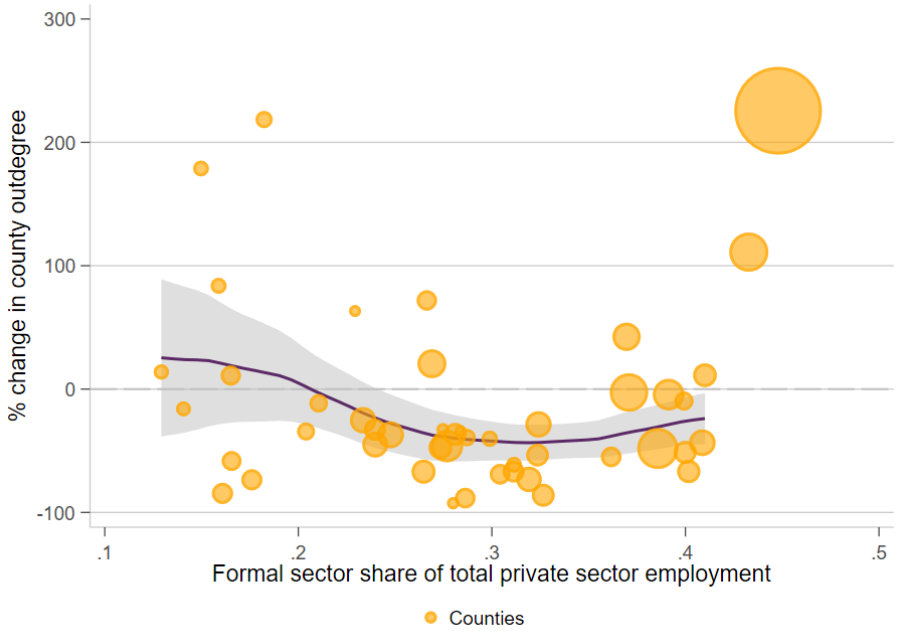
The increase in inequality of outlinks across counties is driven by sectors and counties that the newly accounted for firms are more likely to link to. Nairobi and Mombasa-based firm types are 12 percentage points and 4 percentage points more likely to supply to firm types whose augmented entry probabilities are higher than before. Firm types in almost all remaining counties have a lower connection probability to these previously unaccounted for firms. We re-compute the county-level variance in outlinks by excluding Nairobi and Mombasa in Table 2.C1 as these counties have the highest average interaction probabilities. We find an increase in the variance to mean ratio of only 8% rather than 21% across counties.

The aggregate patterns of course mask substantial heterogeneity across counties. Counties with the largest percentage increment in outlinks are, in decending order, Nairobi, West Pokot and Garissa. Garissa and West Pokot are counties with low levels of formality. Garissa is a geographical gateway for some of Kenya's northern counties, the region with the highest concentration of counties with a low GDP per capita and high levels of informality. Overall, the top ten[67] counties in terms of gains are counties with either high levels of informality (Marsabit, Elgeyo Marakwet, West Pokot, Mandera, and Isiolo) or because they serve as a service or distribution hub either nationally or for their respective regions (Nairobi, Garissa, Mombasa, Uasin Gishu, and Kisumu). All else constant, ignoring informal firms underestimates the importance of the ex ante well-connected counties as hubs and the exposure of the ex ante less-connected counties to the rest of the economy. This pattern is reflected by the U-shape correlation between the change in the number of outlinks and a county's formal sector share shown in Figure 2.16.[68] The U-shape disappears (see Figure 2.C2) once we control for the probabilities with which new firms want to link withe the county across all county-sector types ($\sum_i p(\theta, \theta')$). In Figure 2.C3 we plot the change in outdegrees against the interaction probabilities themselves. The steep slope for counties popular with previously under-counted buyer types mirrors the

[67]The top ten are: Nairobi, West Pokot, Garissa, Mombasa, Marsabit, Elgeyo Marakwet, Isiolo, Uasin Gishu, Kisumu, and Mandera.

[68]The U-shape pattern is also present if we look at the change in outlinks across sector-region types rather than aggregating outlinks to the county level in Figure 2.C2. Moreover, it is less pronounced if we exclude Nairobi and Mombasa, but there is still a noticeable rise due to gains in outlinks by other counties that serve as hubs such as Kisumu, Machakos and Nakuru.

Figure 2.16: Predicted change in county-level outdegree and formal sector shares



Each marker represents a county. The size of each marker is proportional to the respective county's Gross County Product. The local polynomial is, however, an unweighted version treating each county equally.

upward slope in the U-shape graph, i.e. the rise in outdegrees for counties with a high formal sector share.

At the sector level, business services (which includes ICT), construction, mining,[69] and utilities sectors gain the most links in relative terms. Note our earlier observation in Figure 2.3 of Section 2.4.3, which showed that business services and utilities are the sectors with the highest degree of spatial concentration. Their gains therefore align with the increase in outdegrees of counties that act as hubs in the network and host a disproportionate number of firms from these sectors. The result for the construction sector can be explained by the fact that construction sector firms have particularly strong linkages to or within counties with otherwise high levels of informality. Formalisation of construction sector links is often incentivised by the participation of the government in construction projects. The wholesale and retail sector, i.e. the sector that gained the largest number of new firms in

[69]The observed relative increase in links for the mining sector is influenced by small numbers since the administrative data include only a few hundred mining firms. The population census adds small-scale mining activities.

the counterfactual exercise, is the median sector in terms of its increase in outdegrees.

Our suggested counterfactual accounts for informal firms being born into the network based on their sector-region profile. At the same time, our exercise keeps linking probabilities between sector-region types $p(\theta, \theta')$ constant. We thereby treat $p(\theta, \theta')$ as a fundamental production technology. The U-shaped response to changes in outdegrees suggests that spatial inequality actually increases if we account for informal firms and that linking probabilities play a key role in shaping spatial inequalities in firm-to-firm links across regions. If a policy goal was more regional economic integration, examples of policies that might shift these linking probabilities include the expansion of communication infrastructure like the fiber optic network, the strengthening of transportation infrastructure between secondary cities and towns or the introduction of a physical address system that reduces the cost of shipping goods. The contextual relevance of our result is highlighted by the fact that all counties among those that gain relatively more links and feature a lower ex ante formal sector share, are currently targeted by a 100,000 km extension of Kenya's fiber optic network.[70] At the same time our set-up is not able to speak to the optimal spatial allocation from an aggregate welfare point of view. Given the current market frictions, agglomeration and a strong hub-based network could indeed be optimal for private sector development.

The counterfactual exercise brings us closer towards the question of the firm network's structure if we were to observe informal firms, but it abstracts from the up- and downstream dynamic of supply chains within sectors, in particular transportation, wholesale and retail. To give an example, a formal wholesaler in West Pokot County might be more likely to source directly from a manufacturer in Nairobi than a smaller, informal firm. Using the formal wholesaler's linking probabilities in the counterfactual might therefore overpredict how many more links between West Pokot and Nairobi we are missing out on. In reality, Nairobi might obtain relatively fewer links. At the same time, we are likely to underestimate downstream links to other businesses

[70]The programme targeting relatively underserved counties is part of Kenya's 2022 Digital Master Plan (Ministry of ICT, Innovation and Youth Affairs, 2022). https://www.connectingafrica.com/author.asp?section_id=816&doc_id=784428

rather than consumers as we rely on observed linking probabilities of formal firms and cannot distinguish between their sales to final consumers and sales to non-VAT-registered businesses. We might therefore underpredict the number of outlinks West Pokot itself actually gains.

While we advise against taking our stylized model too literally, it serves its main purpose - to test our intuition on the role of the informal sector based on the stylized facts presented in the empirical part of the paper. Our initial intuition that the spatial inequality in network links would reduce if one were to observe informal firms, potentially requires an additional assumption about the manner in which informal firms form links with the formal sector. The most plausible is that informal firms mostly link with other local firms due to their size.

2.7.3 Limitations of the model and counterfactual exercise

In the following we will revisit and discuss four shortcomings of the model and the counterfactual.

First, a limitation of our counterfactual is that we are only able to adjust entry probabilities rather than linking probabilities between different sector-county pairs. While the model would allow us to implement such a counterfactual, we lack information on how these probabilities change with the addition of informal firms. Future extensions of this work could use shocks to formalisation to estimate the linking probabilities of newly entered firms. Such shocks include the opening of new tax offices or business registration centres, the introduction of an automatic validation system, and the announcements of new withholding agents. Another possible extension of our analysis could look into how the observed versus unobserved spatial inequality affects the propagation of shocks across regions.

Second, using the outlined model, we are unable to test the implications of policies aimed at increasing formalisation rates among informal firms (e.g. lowering entry cost or more stringent enforcement policies) similar to Ulyssea (2018). This would require us to micro-found and endogenise the probabilities $p(\theta)$ that firms of a certain sector-region type are born. We currently treat $p(\theta)$ as exogenous and compute it directly from administrative and census data.

Third, in the model we assume that the number of inlinks of a firm remains constant over time. While this assumption might seem restrictive, we for now rationalise it with the fact that we observe less heterogeneity in the number of inlinks across firms. Future iterations could test this assumption in a more rigorous manner. Importantly, however, the model achieves its main objective in predicting our main target moments the number of outlinks across sector-region types, with great fit (see Section 2.6).[71] Fourth, our model set up does not allow us to make a statement about optimal policy and in particular the optimal level of spatial concentration.

2.8 Conclusion

Memon (1976) discusses how the debate on the economic dominance of the metropolitan areas in Kenya was as much of a concern, but to some extent also an accepted status quo, back in the 1970s as it is today:

> *"A recent document on the country's urban development strategy, for instance, rationalises the dominant position of Nairobi and Mombasa in relation to other towns as a demonstration of "the role primate cities play in a transitory economy" (MoF, 1974, as cited in Memon (1976)). Acceptance of urban primacy as a necessary evil is also reflected in the "large-city bias" of the policies governing resource allocation for urban infrastructure."*

He goes on to show that the same policy report by the Ministry of Finance outlines the following urban planning goals:

> *"... to encourage the expansion of several large towns in addition to Nairobi and Mombasa in order to promote a geographical dispersal of benefits arising from urban development amongst a number of centres." (MoF, 1974, as cited in Memon (1976))*

We use granular transaction-level records to document a high degree of spatial concentration of trade amongst formal private sector firms in Kenya. This concentration can be traced back to the fact that the share of the formal sector increases

[71] Many purpose-built firm-to-firm models or alternative networks models like the balls and bins model are either unable to directly predict this moment or substantially over- or underpredict the number of links in the tail (Bernard and Zi, 2022).

with regional market size. We use a network formation model with preferential attachment (Bramoullé et al., 2012) to show that accounting for informal firms in the network increases the relative number of outlinks in counties with the highest level of informality. At the same time, counties that serve as trading hubs also gain. This is because informal firms are particularly prevalent in sectors and regions that ex ante have a high probability to link with those hubs. This result aligns with Blanchard et al. (2021)'s finding that people who travel from more rural to more urbanised areas in Kenya favour visits to Nairobi and Mombasa over secondary cities and towns. They are able to rationalise this pattern with a model where people place a premium on variety in amenities, which in turn tends to be higher in cities. In the firm context this can translate to a higher density of potential trading partners, a greater variety in available inputs, and access to complementary services like transportation. In a nutshell, our counterfactual results suggest that data on the formal firm network likely underestimate the connectivity and thereby also vulnerability of smaller regions to shocks which pass through hubs such as Kenya's capital Nairobi.

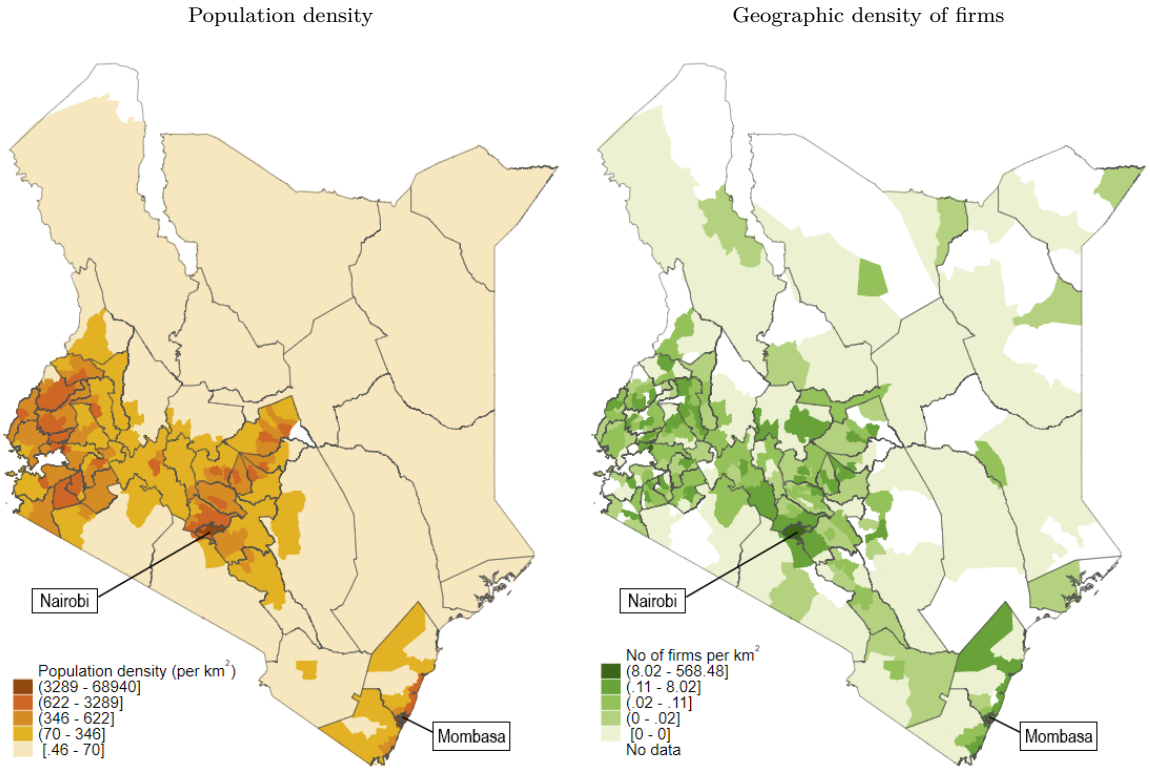
Acknowledgements for Chapter 2 We thank the Kenya Revenue Authority (KRA) for the outstanding collaboration. Romeo Ekirapa, Simon Mwangi, Benard Sang, and Micah Kiprono provided excellent technical support and advice. We thank Elizabeth Gatwiri and Daniela Villacreces Villacis for excellent research assistance. We thank Andrea Bacilieri, Raphael Bradenbrink, Douglas Gollin, Luke Heath Milsom, Sanghamitra Warrier Mukherjee, Solomon Kweku Owusu, Nina Pavcnik, Simon Quinn, Alexander Teytelboym, Christopher Woodruff, Hannah Zillessen, and participants of the Oxford Firms Discussion Group, the CSAE research workshop, and the Jeune Street Seminar Series for comments and feedback. We gratefully acknowledge financial support from the Private Enterprise Development in Low Income Countries (PEDL) and the Structural Transformation and Economic Growth (STEG) programmes, both of which are joint initiatives by the Centre for Economic Policy Research (CEPR) and the UK Foreign, Commonwealth & Development Office (FCDO). Verena Wiedemann further acknowledges funding from the Oxford Economic Papers Research Fund (OEP) and the German National Academic Scholarship Foundation. This study has been approved by the Department of Economics Research Ethics Committee at Oxford (protocol no.ECONCIA20-21-23), and the Kenyan National Commission for Science, Technology and Innovation (protocol no.NACOSTI/P/20/5923). The qualitative data collection was approved by the Department of Economics Research Ethics Committee at Oxford (protocol no.ECONCIA21-22-48), the Strathmore University Institutional Scientific and Ethical Review Committee (protocol no.SU-ISERC1480/22), and the Kenyan National Commission for Science, Technology and Innovation (protocol no.NACOSTI/P/22/20556). The views in this paper are those of the authors, and do not necessarily represent those of the KRA or any other institution the authors are affiliated with.

Appendices of Chapter 2

2.A Additional material on spatial trade patterns

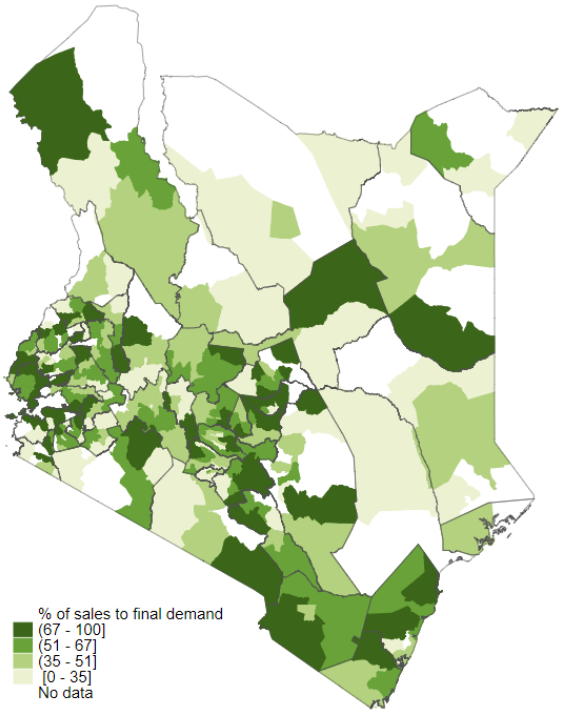
2.A.1 Additional graphs and tables

Figure 2.A1: Firm headquarter locations and population density



The right map shows the density of firm headquarter locations at the sub-county level in number of firms per km^2. The left map shows the population density - also at the sub-county level. Sub-counties represent the second administrative layer. Their size varies between 3 and 19,837 km^2 with a median size of 1,738 km^2 and an average size of 421 km^2. We therefore chose to map the density of firms rather than absolute numbers. Sub-counties are much more comparable in terms of population. The median sub-county has a population of 143,156 people, while the average sits at 129,263. The borders of Kenya's 47 counties, the first administrative layer are outlined in grey.

Figure 2.A2: Share of sales to final demand (non-registered entities)



The above map plots the average share of sales firms in each sub-county sell to non-registered entities. These are mostly consumers, but can also be non-VAT paying firms. County borders are outlined in grey.

Table 2.A1: The intensive and extensive margin of domestic trade

All trading county pairs with origin and destination FE

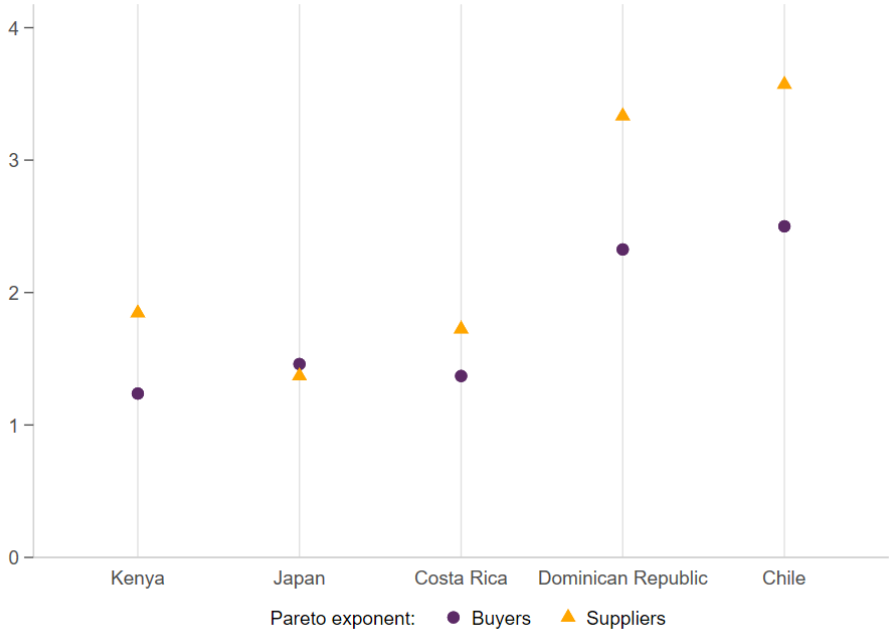
	Total trade volume		# of links		Avg # of transactions per link		Avg volume per transaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Travel time	-2.353***	-1.137***	-1.906***	-1.012***	-0.448***	-0.176***	0.001	0.051
	(0.101)	(0.138)	(0.050)	(0.056)	(0.035)	(0.051)	(0.053)	(0.082)
Social connectedness		1.042***		0.767***		0.233***		0.042
		(0.086)		(0.040)		(0.032)		(0.055)
No. observations	1,652	1,652	1,652	1,652	1,652	1,652	1,652	1,652
R2	0.739	0.759	0.890	0.914	0.423	0.439	0.242	0.242
Origin FE	✓	✓	✓	✓	✓	✓	✓	✓
Destination FE	✓	✓	✓	✓	✓	✓	✓	✓

All trading county pairs with origin and destination characteristics

	Total trade volume		# of links		Avg # of transactions per link		Avg volume per transaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Travel time	-0.878***	-1.317***	-0.702***	-1.055***	-0.308***	-0.356***	0.132*	0.094
	(0.162)	(0.125)	(0.111)	(0.053)	(0.048)	(0.049)	(0.077)	(0.078)
Social connectedness	0.876***	0.823***	0.735***	0.691***	0.106***	0.102***	0.035	0.031
	(0.111)	(0.075)	(0.077)	(0.036)	(0.031)	(0.031)	(0.053)	(0.052)
Pop density (destination)	0.643***	0.222***	0.375***	0.023	0.138***	0.101***	0.130***	0.098***
	(0.056)	(0.047)	(0.034)	(0.019)	(0.016)	(0.017)	(0.027)	(0.030)
Pop density (origin)	0.667***	-0.034	0.467***	-0.077***	0.075***	-0.015	0.124***	0.059*
	(0.053)	(0.050)	(0.032)	(0.019)	(0.015)	(0.017)	(0.027)	(0.033)
Distance to NBO (destination)	-0.505***	0.248***	-0.537***	0.105***	-0.016	0.041	0.048	0.102*
	(0.108)	(0.094)	(0.074)	(0.038)	(0.030)	(0.036)	(0.051)	(0.062)
Distance to NBO (origin)	-1.096***	0.297***	-0.926***	0.153***	-0.212***	-0.029	0.041	0.174***
	(0.103)	(0.092)	(0.071)	(0.038)	(0.030)	(0.036)	(0.047)	(0.059)
Distance to MSA (origin)	-1.005***	-0.174***	-0.679***	-0.038	-0.210***	-0.100***	-0.115***	-0.035
	(0.080)	(0.063)	(0.053)	(0.026)	(0.024)	(0.026)	(0.038)	(0.042)
Distance to MSA (destination)	-0.184*	0.241***	-0.308***	0.055*	0.005	0.038	0.118**	0.149***
	(0.098)	(0.078)	(0.061)	(0.033)	(0.026)	(0.027)	(0.053)	(0.056)
No. buyers (destination)		1.195***		1.000***		0.103***		0.092**
		(0.056)		(0.023)		(0.022)		(0.036)
No. suppliers (origin)		1.709***		1.328***		0.220***		0.161***
		(0.056)		(0.022)		(0.021)		(0.038)
No. observations	1,652	1,652	1,652	1,652	1,652	1,652	1,652	1,652
R2	0.396	0.674	0.490	0.880	0.226	0.286	0.037	0.053
Origin FE	✗	✗	✗	✗	✗	✗	✗	✗
Destination FE	✗	✗	✗	✗	✗	✗	✗	✗

All variables are aggregated to the county level. The outcome variables for the OLS regressions are in log terms. Travel time is measured in log minutes. Social connectedness is measured using the Facebook Social Connectedness Index (SCI), which captures the total number of Facebook friendship connections between two counties, divided by the product of the number of Facebook users in each county (Bailey et al., 2021). The total number of possible county pairs is 2,209. We pool data from all available years to minimise the incident of zero trade flows. The estimated equation is $\log \text{outcome} = \alpha \log \text{time} + \beta \log \text{SCI} + \gamma X + \delta_o + \delta_d$, where X represents a matrix of control variables included in the regressions in the bottom table, δ_o represents the origin FE, and δ_d the destination FE. The latter two are excluded in the bottom table. *; **; and *** denote significance at the 10; 5; and 1 percent levels respectively.

Figure 2.A3: The higher inequality in outdegree is common across contexts



The figure plots the pareto exponents for Kenya (this paper), Japan (Bernard et al., 2019), Costa Rica (Alfaro-Urena et al., 2018), Chile (Grigoli et al., 2023), and the Dominican Republic (Cardoza et al., 2023). The latter three papers estimate and report the inverse of the pareto coefficient, which we invert here for the purpose of comparability.

2.A.2 Exploring the robustness of spatial concentration with respect to multi-establishment firms

A potential concern is that the observed spatial concentration is driven by the fact that we only observe firm headquarter locations, which in turn are more likely to be based in Nairobi or Mombasa. We use micro-data from the 2010 Census of Industrial Production (KNBS, 2010) to compare the spatial concentration of sales and firm location for all firms, including those with multiple establishments, to the spatial concentration in single establishment firms in Table 2.A2. Firms covered in the Census of Industrial Production overlap closely with the group of VAT-paying firms we observe in the tax records. A 1:1 mapping is not possible due to the anonyomised nature of the data sets. However, the overall number of industrial firms observed in each of the two data sources aligns closely. In 2015, we observe 3,960 VAT-paying firms[72] in mining, manufacturing and utilities, while KNBS (2010) covered 2,252 firms five

[72]2015 is the earliest year for which the VAT records have been fully digitised.

years earlier.

48% of all firms involved in industrial production are located in Nairobi County generating as much as 61% of total sales in 2010. These figures are very similar to concentration of formal manufacturing firms reported by Storeygard (2016) for Tanzania.[73] Restricting the census data for Kenya to single establishments only, the overall concentration of firm locations does not change. The concentration becomes even slightly more unequal once we consider sales instead of purely counting the number of firms. We, however, overstate the concentration of sales in Nairobi by six percentage points if multi-establishment firms are in the sample, but we aggregate their sales geographically based on headquarter information only. Nevertheless, the difference is not sufficient to explain the relatively higher spatial concentration in the VAT-paying sector relative to overall economic activity.

Table 2.A2: Geographic concentration of industrial activity

	All firms		**Single est. firms**	
	in %	α	in %	α
Census of Industrial Production (2010)				N = 2252
No. firms	48	0.54	48	0.54
Sales	61	0.32	55	0.30
Industrial firms in admin data (2015)				N = 3960
No. firms	64	0.51	-	-
Sales	69	0.21	-	-

The columns for Nairobi report their share of the respective national aggregate figures (e.g., the share of industrial establishments located in Nairobi). The Pareto exponent are the estimated coefficients from a county-level regression of each county's rank (log) on the respective measure (log): $logrank = logA + log\alpha$. The Census of Industrial Production was carried out by KNBS (2010).

2.B Additional material on informality

2.B.1 Fluctuations of the VAT-paying sector as a share of GDP over time

Figure 2.B1 and Table 2.B1 illustrate that the value added generated by the VAT sector has been declining over time as a proportion of GDP. This downward trend

[73]Dar es Salaam, Tanzania's primate city, accounts for 8% of its population (Storeygard, 2016), as very similar figure to Nairobi's population share in Kenya (KNBS, 2019b).

in value added can be attributed to two factors. Firstly, the introduction of a fuel tax in September 2018, which was previously VAT exempt, has led to a reduction in value added.[74] Secondly, certain sectors that have significantly contributed to Kenya's growth over the years, such as agriculture, real estate, financial services, and public administration, are not well captured in the VAT data. This is highlighted by the rapidly rising GDP for non-market services in Figure 2.B2 (orange line), where we present the trends for each sector.

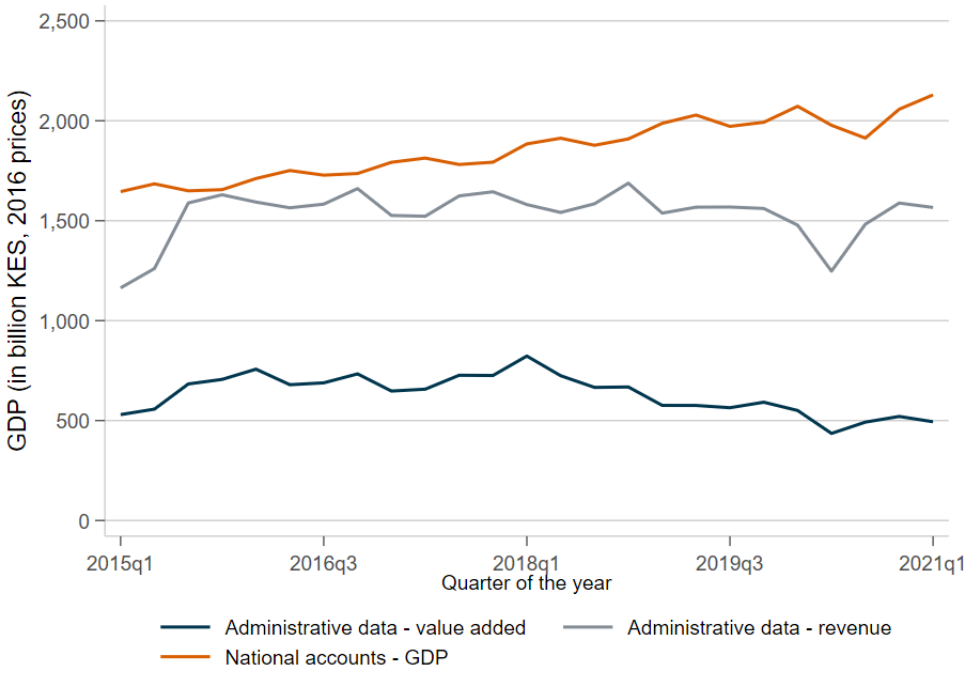
Table 2.B1: Share of GDP covered in the administrative records

	Share of GDP (%)						
Year	All	ex Fin.	ex NMS+Fin.	ex Agri.	ex NMS+Fin.+Agri.	NMS	Agri.
2015	36	39	51	43	69	22	22
2016	40	43	56	46	74	22	21
2017	37	40	53	44	71	23	20
2018	37	40	53	43	70	23	20
2019	28	30	40	33	53	24	20
2020	24	26	35	28	46	24	21

Fin. refers to financial services. NMS refers to non-market services, i.e. education, health, public administration, and real estate (Herrendorf et al., 2022). Agri. refers to the agricultural sector. The first five data columns report the proportion of GDP capture by value added of the VAT-paying firms. The final two columns report the GDP share of non-market services and agriculture respectively. GDP figures are based on national accounts data published by the Central Bank of Kenya: https://www.centralbank.go.ke/statistics/national-accounts-statistics/.

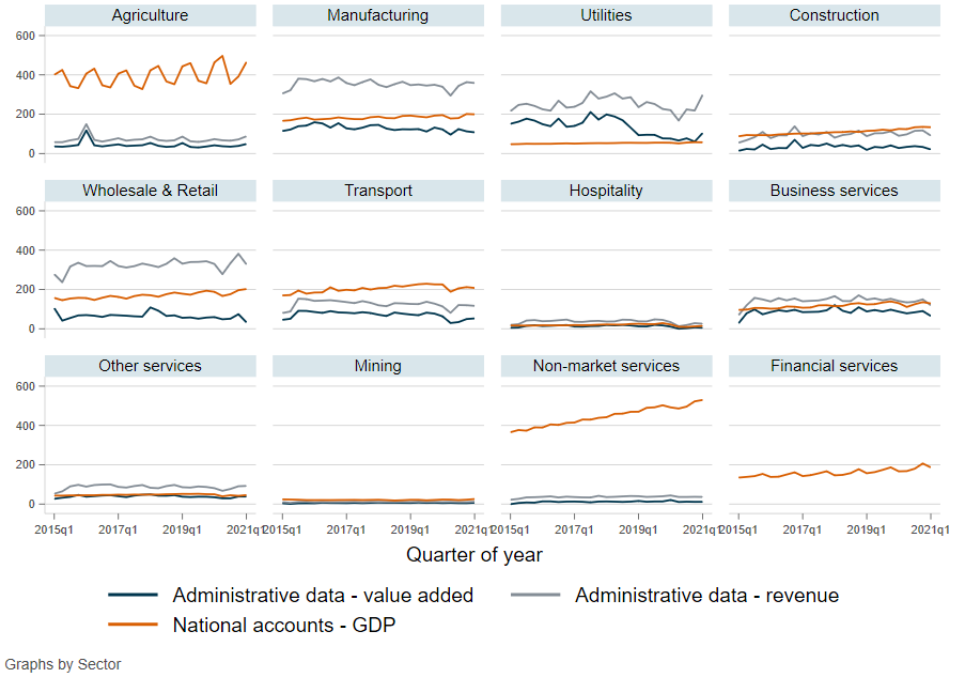
[74]The impact of this tax is particularly evident in the plot for utilities (electricity, gas, and water) in Figure 2.B2. However, this sector alone cannot fully explain the overall downward trend and kink in the data.

Figure 2.B1: Quarterly GDP in billion KES (2015-2020)



This figure compares the quarterly GDP in 2016 prices to the quarterly value added (domestic sales + exports - domestic purchases - imports - salaries) from the administrative data (deflated). The quarterly GDP time series was downloaded from the Central Bank of Kenya's database: https://www.centralbank.go.ke/statistics/national-accounts-statistics/

Figure 2.B2: Quarterly GDP in billion KES (2015-2020)



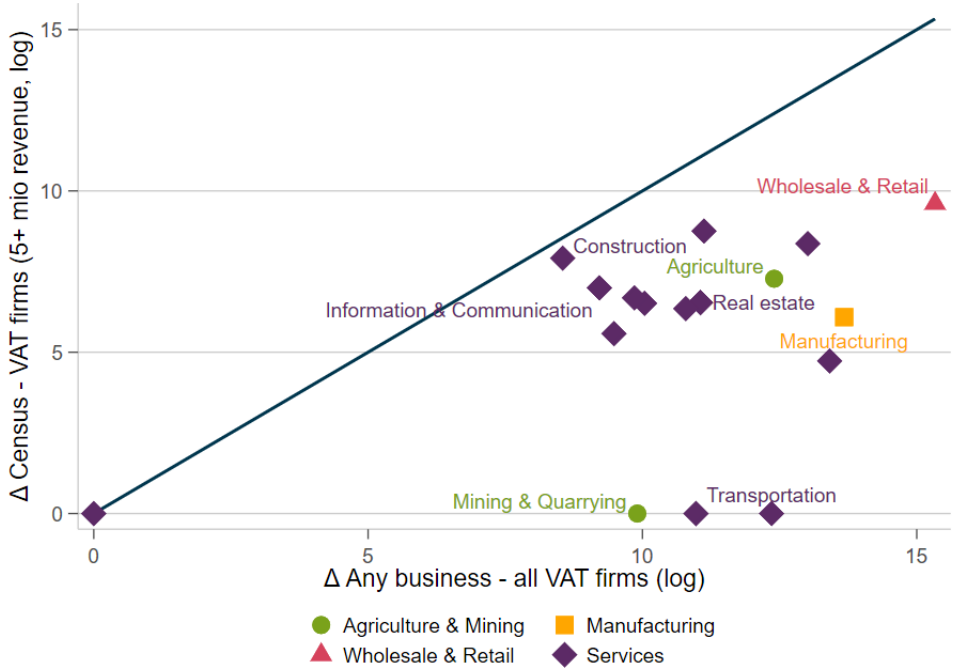
This figure compares the quarterly sector-level GDP in 2016 prices to the quarterly value added (domestic sales + exports - domestic purchases - imports - salaries) from the administrative data (deflated).

Table 2.B2: Sector-level comparison with the 2016 Census of Establishments

Sector	**Admin data**	**Census of Establishments**		Δ in %
		all	5+ mio. KES turnover	
Agriculture, Forestry, & Fishing	931	4,189	2,128	-56
Mining & Quarrying	239	297	176	36
Manufacturing	2,965	6,038	2,904	2
Electricity & Gas	1,031	84	67	1,438
Water supply	189	598	391	-52
Construction	7,351	11,846	7,096	4
Wholesale & Retail	7,991	44,023	21,175	-62
Transportation & Storage	2,333	2,241	1,123	108
Hospitality	2,474	8,364	1,982	25
Information & Communication	2,349	3,480	2,620	-10
Real estate	2,122	3,680	2,322	-9
Professional, Scientific & Technical	2,579	3,669	2,256	14
Administrative & Support Services	2,386	3,622	2,130	12
Public Administration	55	307	201	-73
Education	203	29,582	6,478	-97
Human Health & Social Work	470	2,427	1,109	-58
Arts, entertainment & recreation	433	575	215	101
Other services	1,339	8,440	5,174	-74
Households as employers	698	46	39	1,689
Total	38,138	133,508	59,587	-36

The last column captures the percentage deviation of the firm count in the administrative data (column 1) from the number of establishments with a turnover of more then five million KES (column 3) as per Census of Establishments.

Figure 2.B3: Sector-level correlation of the extensive margin of informality



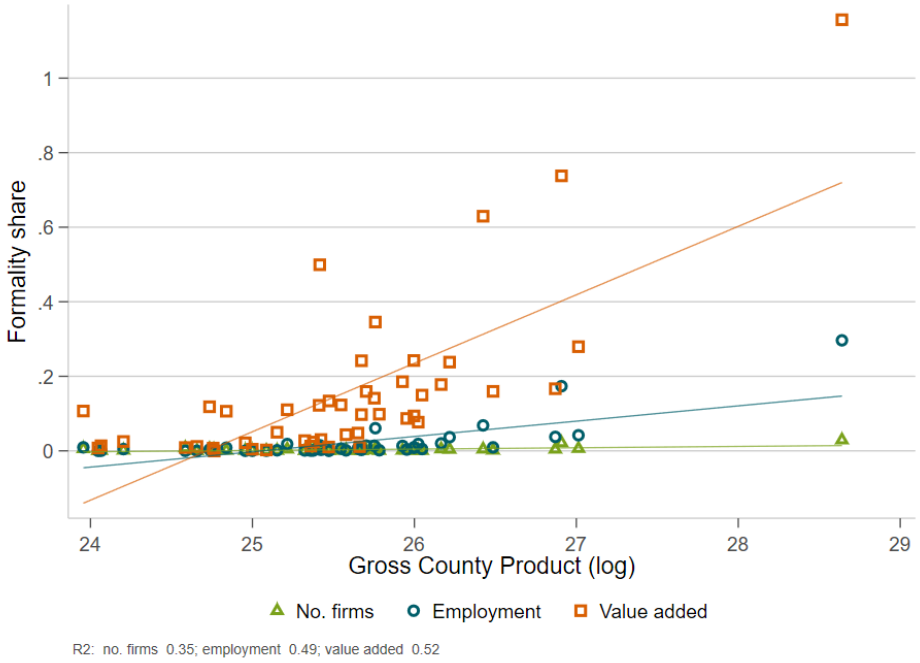
The above graph correlates the two types of extensive margin informality at the sector level. On the y-axis we put the gap between the number of firms with more than KShs 5 million of revenue from the Census of Establishments (KNBS, 2017) and the administrative data. The x-axis shows the gap between the total number of businesses (KNBS, 2016) and the number of VAT firms in each sector.

2.B.2 Informality, market size, and income levels

Figure 2.B4: Informality, market size, and income levels

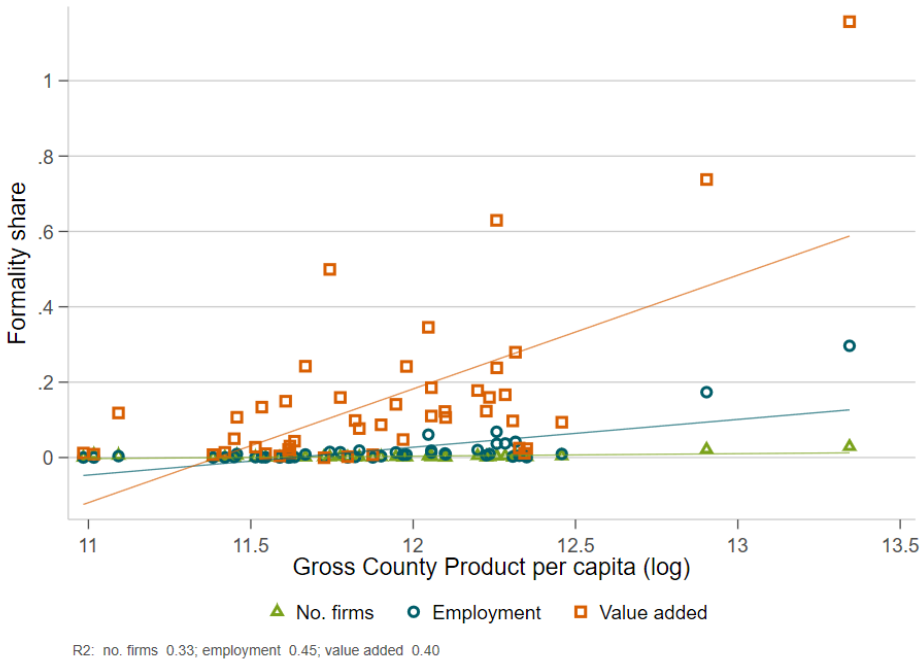
Correlation of the formal sector share and ...

... Gross County Product



R2: no. firms 0.35; employment 0.49; value added 0.52

... Gross County Product per capita

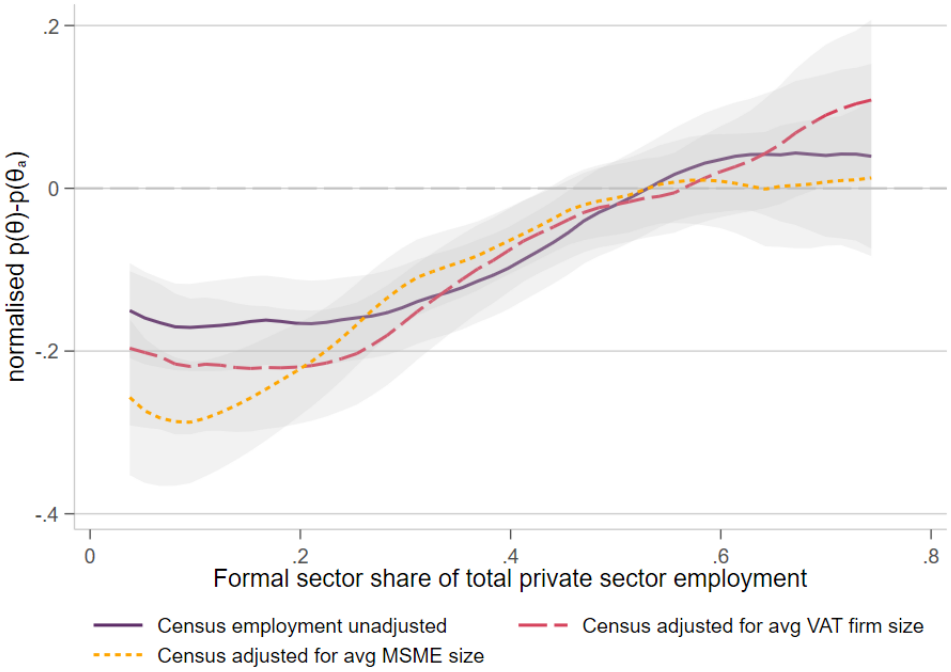


R2: no. firms 0.33; employment 0.45; value added 0.40

The two graphs plot the correlation of the formal sector share with the Gross County Product in absolute and per capita terms respectively. Each marker represents one of Kenya's 47 counties.

2.C Additional material on the counterfactual

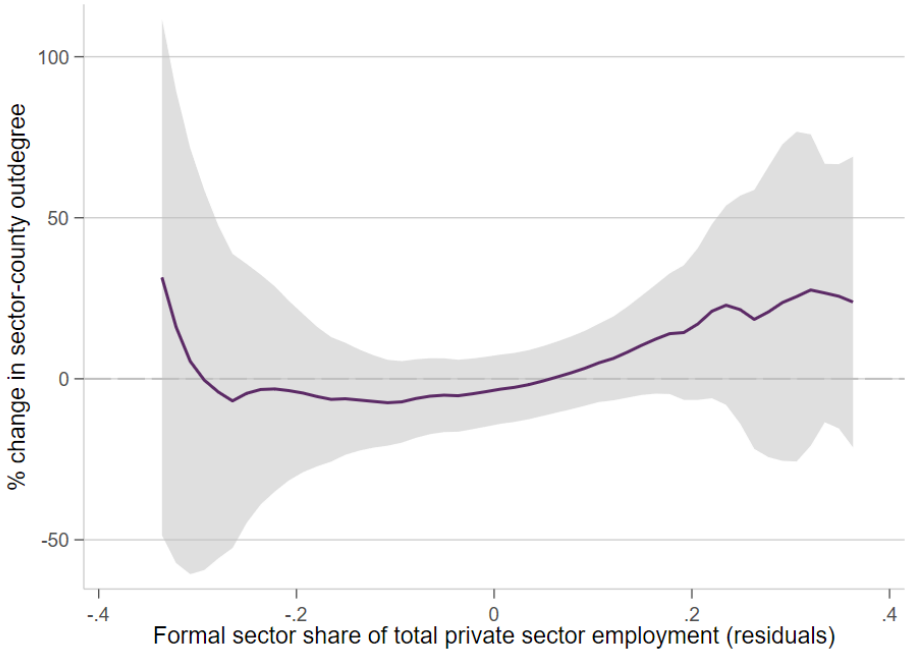
Figure 2.C1: Sector-region probabilities and formal sector shares



The graph plots each sector-regions formality share against the normalised difference between the baseline $p(\theta)$ and the augmented version $p(\theta_a)$ that takes into account informal firms. $p(\theta)$-$p(\theta_a)$ is reported in terms of standard deviations. In addition to the default $p(\theta_a)$, the graph plots two additional versions, which account for differences in the firm size distribution across sector-county cells drawing on the average firm size based on the administrative data and a survey of medium, small and micro enterprises (KNBS, 2016) respectively.

Figure 2.C2: Predicted change in outdegree and formal sector shares

By sector-county type

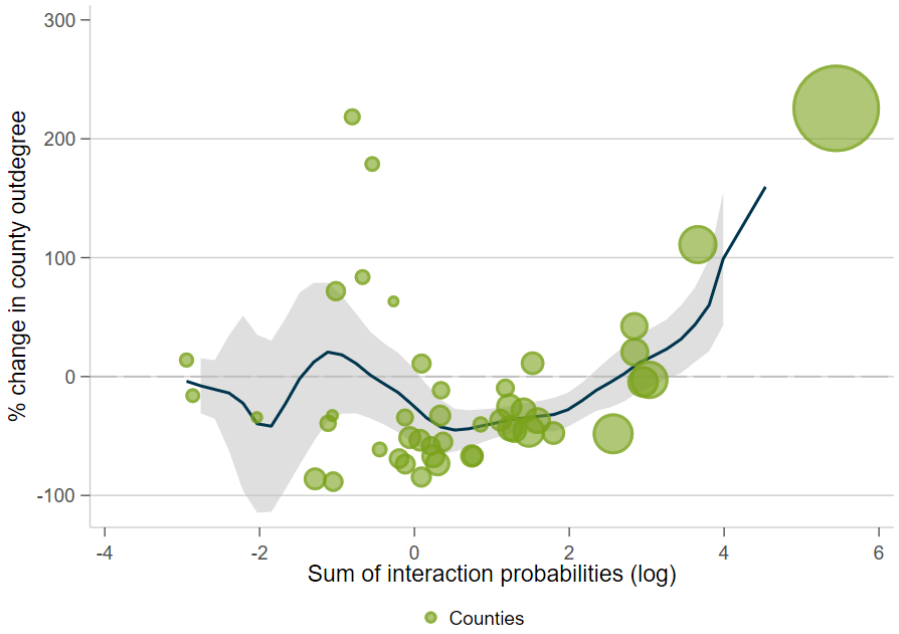


Controlling for baseline linking probabilities ($\sum_i p(\theta, \theta')$)



We plot the percentage change in outdegrees at the sector-county level in the top graph and at the county level in the bottom graph. Relative to Figure 2.16 we control for the linking probabilities of a county at baseline in the bottom graph, i.e. the aggregate probability sector-county types link with the county ($\sum_i p(\theta, \theta')$). Each marker represents a county. It's size reflects the county's Gross County Product. The local polynomial is unweighted.

Figure 2.C3: Predicted change in county-level outdegree and baseline linking probabilities ($\sum_i p(\theta, \theta')$)



Each marker represents a county. Its size reflects the county's Gross County Product. The local polynomial is unweighted.

Table 2.C1: Robustness checks for changes in the dispersion of outdegrees

County outdegree	Δ sd/mean (in %)
Baseline	20.8
Baseline without Nairobi & Mombasa	7.7
Adjust for firm size (admin data)	25.5
Adjust for firm size (MSME survey)	18.5

The above table reports robustness checks for the baseline findings reported in Table 2.8. The first two lines capture the results from the baseline counterfactual except for in the second row we exclude the degrees of Nairobi and Mombasa when we compute the coefficient of variation. Row three and four report the results for alternative distributions of $p(\theta)$ that correct for differences in the firm size distribution across sector-county types.

CHAPTER 3

Internet use and technology adoption along supply chains

Abstract Communication infrastructure, such as Internet connectivity, can reduce information frictions and transaction costs among market participants. However, the realisation of these benefits hinges on the effective use of Internet infrastructure by firms. Using administrative data on private sector firms in Kenya, I show that larger firms in the top quartile of the firm size distribution are nearly three times more likely to purchase inputs from information and communication technology (ICT) providers. I draw on in-depth qualitative interviews with 33 firms to examine the reasons behind heterogeneity in adopting Internet-based technologies and their implications for domestic firm-to-firm trade. I find that firms' ability to adopt Internet-based technologies is influenced by their capacity to make complementary investments alongside ICT usage, with financial constraints and demand uncertainties disproportionately affecting small firms. Further, despite some realised advantages of the use of Internet connectivity, in-person interactions and market frictions in the transportation sector imply that advantages from geographic proximity prevail.

3.1 Introduction

> *"Integration, that is, a growing intimacy and firmness of the connections between the separate parts of the industrial organism, shows itself in such forms as the increase of security of commercial credit, and of the means and habits of communication by sea and road, by railway and telegraph, by post and printing-press."*
> Marshall (1890, p. 201)

Digital infrastructure has emerged as a crucial element in 21st century policy agendas focused on fostering inclusive growth and economic development (Adam and Mubila, 2017). Dating back to the days of the telegraph (Marshall, 1890), a prevailing notion is that physical connectivity infrastructure lowers information frictions and the cost of transactions between market participants, thereby enhancing trade and overall welfare (Aker, 2010; Allen, 2014; Juhász and Steinwender, 2018; Malgouyres et al., 2021; Akerman et al., 2022). Nevertheless, a critical precondition for realising these benefits is the effective utilisation of Internet infrastructure by private sector firms. At the same time, previous research has documented substantial heterogeneity in the ability of market participants to leverage Internet infrastructure (Foster et al., 2018).

This mixed-methods paper combines insights from administrative firm data with in-depth qualitative interviews with 33 private sector firms to study why heterogeneity in adoption can occur and what this implies for the prospects of leveraging Internet infrastructure as a tool to foster domestic market integration in Kenya. Based on the administrative data, I first document that firms in the top quartile of the firm size distribution are almost three times more likely to purchase inputs from information and communication technology (ICT) providers. The qualitative interviews then allow me to study how firms use the Internet infrastructure available to them for their interaction with both existing and potential buyers and suppliers as well as firm-internal operations. I find that heterogeneity in adopting Internet-based technologies is driven by a firm's capacity to make complementary investments alongside the use of ICT inputs for the latter to become viable. A firm's ability to invest in complementary transportation and production capacity and the need

to hire employees, in turn, hinges on overcoming financial constraints and demand uncertainties. Both factors can be more binding for small firms. With the firm size distribution shifted to the left in smaller markets (Kumar et al., 1999; Gollin, 2008) farther away from metropolitan areas (Chacha et al., 2023), these constraints could have consequences for the spatial dispersion of welfare gains from improved Internet infrastructure.

With this paper, I join a lively literature debating what role distance plays in social and economic interactions in a world with rapidly improving communication technologies (Goldenberg and Levy, 2009; Mok et al., 2010; Bailey et al., 2020; Akerman et al., 2022). The insights gained from my analysis of the qualitative interviews suggest a two-part answer. First, communication tools that rely on Internet connectivity have led to substantial improvements in communication between firms of all sizes. Large firms, in particular, use digital tools to expand their geographic scope. Second, however, the effectiveness of these tools in enhancing a firm's ability to search for and connect with new buyers and suppliers still depends heavily on geographic proximity. This result seems to be driven by the advantage of face-to-face interaction in building trust and gathering market information as well as underlying market frictions in the transportation sector. These two mechanisms align with Bailey et al. (2020)'s findings on how social online networks in New York City are shaped by the city's underlying public transport system.

The paper is structured as follows: I start by outlining the paper's contribution to the literature. In Section 3.3, I provide some contextual background by discussing the recent expansion of Internet infrastructure in Kenya. In Section 3.4, I introduce both the quantitative and qualitative data I draw on for my analysis. Section 3.5 discusses how Internet connectivity might affect firm performance and network linking decisions in a search and matching framework. Moving on to the empirical core of the paper, Section 3.6 draws on administrative data to document the heterogeneity in ICT use by firm size and the qualitative data to document which specific digital applications firms use Internet connectivity for. This is followed by an in-depth analysis of how those applications affect interaction cost within the firm's existing network,

search cost for new buyers and suppliers, and firm internal operations. Throughout, I emphasise heterogeneity by firm size. Finally, Section 3.7 looks into constraints that prevent firms from utilising Internet infrastructure even more explicitly to expand their supplier and buyer networks.

3.2 Contributions

This paper contributes to the literature on the role of communication technologies in reducing information barriers and search cost, which, in turn, shape trade patterns within and across countries (Allen, 2014; Juhász and Steinwender, 2018; Arkolakis et al., 2023).[1] Benefits from lowering such search cost by means of communication technologies, for example, include a smaller dispersion of prices across localities within an economy (Jensen, 2007; Aker, 2010; Allen, 2014). While many such outcomes are generally viewed as beneficial for communities, it is not necessarily the case that the resulting gains are equally distributed. Who gains will ultimately depend on the market participants' capacity to leverage technologies (Allen, 2014; Foster et al., 2018), the level of technology adoption within their ecosystem (Awori et al., 2022; Bailey et al., 2022), and the underlying structure of competition (Iacovone et al., 2017).

With this study, I further contribute to the literature that examines the relevance of Internet infrastructure for private sector development. Upgrades to Internet infrastructure have been shown to improve firm productivity through both, enhanced efficiency of internal firm operations (see for example Chen et al., 2020; Tian, 2021; Jiang, 2022), as well as changes to the firm sourcing decisions and improved market access (see for example Fort, 2017; Malgouyres et al., 2021). Due to the lack of suitable data, most research in this area has focused on the first channel. In addition, most papers that study the latter aspect concentrate on international sourcing decisions (Malgouyres et al., 2021; Akerman et al., 2022). A notable exception is Fort (2017), who demonstrates that information technology helps US-based firms reduce costs in coordinating with suppliers. The channel of domestic sourcing and market

[1] Juhász and Steinwender (2018) also highlight how the impact technologies can have on alleviating search and matching frictions might vary greatly across technologies.

access is particularly relevant for contexts where search frictions and monitoring cost of firms are typically high and often thought to limit greater domestic market integration (Atkin and Khandelwal, 2020). In the aggregate, productivity gains from Internet infrastructure can shift the firm size distribution to the right (DeStefano et al., 2023). In the UK, DeStefano et al. (2023) find this effect to be particularly pronounced in urban areas and knowledge-intensive industries. An earlier literature on the role of ICT in enhancing firm productivity documents the relevance of complementary investments for ICT to translate into productivity gains. They emphasise the need for investments in intangibles, human capital, and alterations to organisational structures (Brynjolfsson and Hitt, 2000; Bresnahan et al., 2002; Jorgenson et al., 2008). Mukherjee (2023) shows how learning costs are a particularly binding constraint for the decision of small firms to adopt digital technologies. Building on their findings, this paper focuses on complementary investments that interact specifically with firm size and the question of the geographic scope of supply chains.

I want to highlight three studies that are particularly closely related to this paper: Foster et al. (2018), Awori et al. (2022), and Cirera et al. (2022b). All three papers study ICT adoption in the context of Kenya. The findings of this paper, which show substantial variation in ICT adoption based on firm size, align with the study conducted by Cirera et al. (2022b), who survey over 1,300 firms in Kenya to study technology adoption more broadly. The researchers draw on novel and highly comparable measures of technology adoption that allows them to compare technology use across a wide range of countries (Cirera et al., 2022a). Discussing barriers to adoption, they ask firms about key perceived obstacles to the adoption of various types of technologies, including ICT, and identify demand uncertainty and lack of finance as the most commonly named challenges. My paper complements theirs in that I am able to explore firm adoption choices and their constraints in greater depth adding layers of complexity that are more challenging to capture in quantitative surveys. I show that complementary investments in production capacity, transportation, and expanding employment are a key link between technology adoption decisions and the named constraints – in addition to a potential direct channel.

Foster et al. (2018) and Awori et al. (2022) also adopt a qualitative approach to tackle the question of Internet-based digital technology adoption by private sector firms in Kenya. Awori et al. (2022) emphasise the role of the socio-economic context for technology adoption, while Foster et al. (2018) focus on the potential of Internet connectivity to enable participation in global value chains. Awori et al. (2022) focus on the adoption of digital technologies of firms during the COVID-19 pandemic. Foster et al. (2018) focus on the period from 2010 to 2014, which saw the rapid expansion of Internet access through the mobile network across Kenya (2G and 3G) and the fibre optic network in metropolitan areas and large towns.

In this paper, I am able to consider a more recent time period that saw an expansion of the fibre optic network across all of Kenya, and with it, a more reliable and higher-speed mobile broadband based Internet (4G) and fixed wireless. I will discuss the usage patterns and differences of the various technologies in more detail in Section 3.6.4. Due to improved reliability and speed, particularly upload speed critical for two-way communication (Koutroumpis et al., 2022), this wave of Internet infrastructure has increased potential to enable productive firm-to-firm interactions. While important as a transformative force, the focus of the academic literature and policy debate on mobile broadband and Internet access via mobile phones on the African continent, places emphasis on consumption rather than production-focused technology. A digital economy, however, requires both (Augustine, 2022).

3.3 The expansion of Kenya's fibre optic network

The question of Internet infrastructure and domestic trade is of great relevance for countries across the world, as governments are investing billions into the expansion of fibre optic networks or are incentivising private sector providers to do so.[2] Studying this question is particularly exciting and relevant in the context of Kenya, where the government's 2022 Digital Master Plan outlines the goal of deploying 100,000 km of last mile fibre (Ministry of ICT, Innovation and Youth Affairs, 2022),[3] which would

[2]https://thegeopolity.com/2021/05/04/the-global-fibre-optic-network/

[3]The latest update of the plan stipulates a completion of the roll-out within the next five years: https://www.businessdailyafrica.com/bd/economy/president-ruto-eyes-100000km-fibre-optic-network-3992434

more than double the length of the country's fibre optic network.[4] To put this number into context, between 2008 and 2022, the Kenyan government deployed 8,900 km of backbone fibre as part of the National Optic Fibre Backbone Infrastructure Project (NOFBI). More than half of the cable has been rolled out in recent years since 2014. This government infrastructure is complemented by capacity leased by energy and water providers who maintain their own fibre optic networks to monitor the soundness of infrastructure. The extensive geographic scope of the public sector network, in turn, enabled private sector providers to substantially expand the geographic coverage of their service. As of early 2023, Safaricom, Kenya's largest telecom provider operates close to 11,000 km of fibre.[5] As a result, broad-band subscription numbers have increased severalfold in recent years (see Figure 3.1).

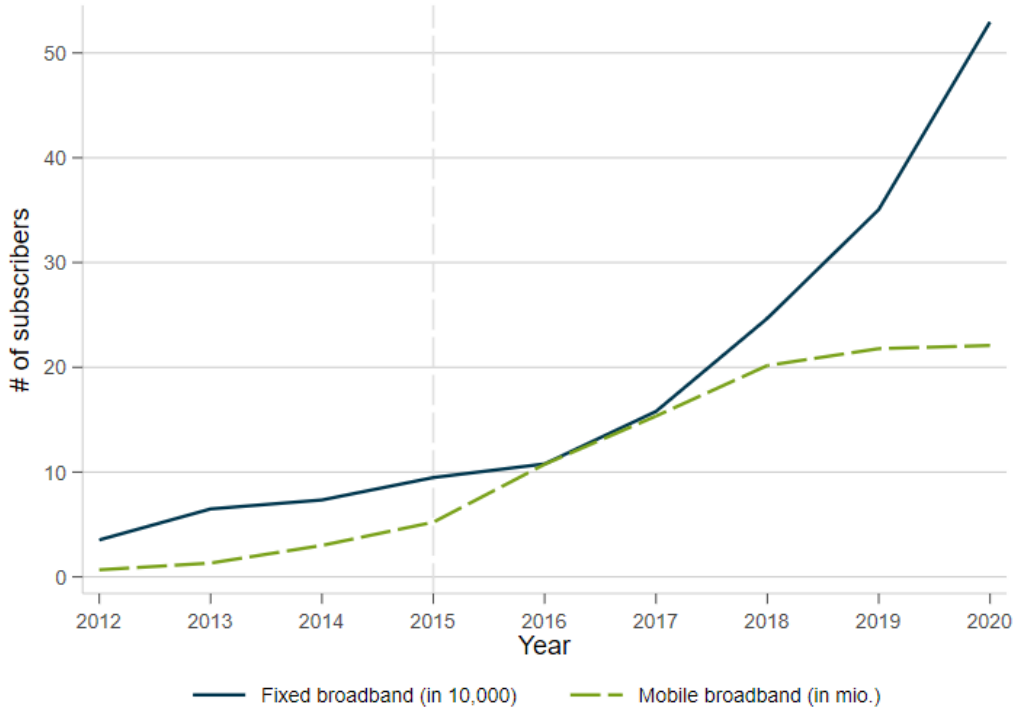
Highlighting the relevance of this paper's research question regarding whether communication infrastructure aids the integration of periphery markets, the two government authorities that oversee the process announced that the fibre-optic expansion plan would specifically target previously unserved or underserved counties (Kenya's second administrative layer).[6] The vast majority of these counties belong to the list of counties that are the least integrated into the formal private sector network (Chacha et al., 2023).

[4]The exact length of all available fibre optic infrastructure is hard to assess as many public records focus on backbone fibre ('the highways') only. At the same time, length is of course only one relevant metric to assess overall coverage.

[5]https://www.africabandwidthmaps.com/?p=6872

[6]These counties are Turkana, West Pokot, Baringo, Elgeyo Marakwet, Samburu, Marsabit, Mandera, Tana River, Lamu, Kilifi, Kwale, Kitui, Laikipia, Isiolo, Garissa, Wajir, Makueni, Kajiado and Narok. https://www.connectingafrica.com/author.asp?section_id=816&doc_id=784428

Figure 3.1: Number of fixed and mobile broadband subscribers between 2012 and 2020



The graph is based on figures reported in the 2021 Economic Survey by the Kenya National Bureau of Statistics (KNBS, 2021).

3.4 Data

3.4.1 Quantitative data

To track the link between Kenya's ICT sector and other formal private sector firms, I draw on administrative tax records collected by the Kenya Revenue Authority. The administrative data details firm-to-firm transactions that take place between over 4 million firm pairs formed by close to 57,500 formal sector firms. Crucial for the objectives of this study, the data detail the characteristics of firms on both ends of the transaction, including their sector of operation. The value added generated by the firms corresponds to 34% of Kenya's GDP (Chacha et al., 2023).

3.4.2 Qualitative data and methods

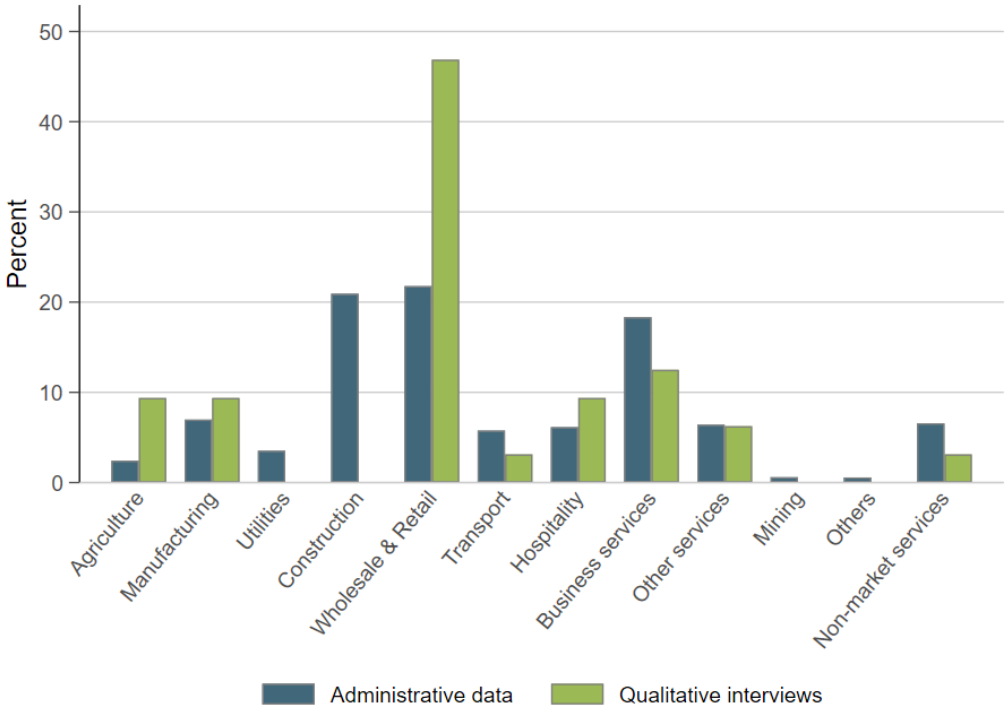
The qualitative data was collected through semi-structured interviews in October and November 2022. A research assistant and I jointly conducted 32 interviews covering 33 firms.[7] In all but six cases, we interviewed the main manager or the owner of the firm. In three cases, middle management was interviewed. In the case of two wholesalers and one licensed distributor, sales personnel were interviewed. The interviews lasted between 30 minutes and two hours. During the semi-structured interviews, both the research assistant and I took written notes. After each interview, we compiled those notes and compared them.

We visited 17 towns in five different counties: Kiambu, Kirinyaga, Kajiado, Machakos, and Nairobi. While Kiambu, Machakos, and Kajiado all border Nairobi County, the main sites for firm interviews were chosen such that they were well outside the metropolitan area in these counties. Most of them were 50 to 100 km outside of downtown Nairobi (one to two hours' drive outside of rush hours). The idea behind the sampling was to interview a selection of firms that reflects the size distribution and sector composition of firms represented in the administrative data. As a result, the main criterion for approaching a firm to be interviewed was that their annual revenues are close to or exceeding KShs 5 million, the threshold for value-added tax (VAT) in Kenya. Further, the aim was to achieve a sector composition that was roughly representative of the overall sample of firms in the administrative data. We did not approach construction sector firms, given that their demand structure relies heavily on the public sector, which was outside the scope of this study. Most firms were recruited via visits to their premises.[8] In most cases, we visited one or two firms for which contact details were obtained through the researcher's or the research assistant's network. We would then sample other firms within a 30-minute drive of those firms.

[7] One of the interviewees owns and manages two independent businesses. There is no overlap of the industries the businesses operate in.

[8] E-mailing firms based on contact lists of industry associations was not successful.

Figure 3.2: Sector composition of interviewed firms and firms covered in the administrative records



The figures for the administrative data rely on a cross section for 2019.

Figure 3.2 shows the sectors of operation of the interviewed firms next to the sector shares of firms included in the administrative data. Table 3.1 summarises key characteristics of the firms interviewed and compares them to the average of the respective indicators from the administrative data. The interviewed firms are of a similar age to firms in the administrative data but are more likely to export or import directly. We slightly over-sampled firms outside Nairobi in order to understand whether geography-dependent constraints pose a challenge for firms outside the metropolitan area. While the average firm size is larger than in the VAT data (see Table 3.1), it would have been preferable to sample a few manufacturers of consumer goods who operate upstream of wholesalers. The intention was to learn more about their supply chain management. Unfortunately, this group of firms turned out to be a challenging population to interview.

Table 3.1: Firm characteristics of interviewed firms and firms covered in the administrative records

Source	Age	No. employees	Nairobi based	Exporter	Importer
Qualitative interviews	11	50	41%	16%	22%
Administrative data	10	18	64%	4%	14%

The figures for the administrative data rely on a cross section for 2019.

Each interview proceeded as follows: The firms were first asked about basic firm characteristics including the firm's age, the sector of operation, the number of employees, and their primary mode of Internet access. We then asked firms an open-ended question about how they use their Internet access. Then, we prompted a list of use cases based on the interview guidelines. Furthermore, firms were asked why they have or have not adopted certain tools and how the type of Internet access they have has changed their business. Moreover, we asked firms about the type of suppliers they source their inputs from and the types of customers they sell their products and services to. For each type of buyer (supplier), we asked how they first met their customer that is located closest to their premises and farthest away. To this, we would add questions on how often they engage with them, for what reason, and through what mode of communication.

The interview transcripts were subsequently coded to identify recurring themes.

3.5 The role of Internet connectivity in a search and matching framework

Internet infrastructure can be used across a variety of business functions by firms. In this way, the Internet affects different aspects of a firm's production function and consequently profit maximisation problem (Hjort and Tian, 2021). Recent papers that rely on a search and matching framework to model firm-to-firm linking decisions provide a useful structure for thinking about the relevant channels in which the decision making and performance of firms are affected by the use of the Internet (Miyauchi, 2023; Demir, Fieler, Xu and Yang, 2023; Arkolakis, Huneeus and Miyauchi, 2023). In this type of framework, firms post advertisements to search for buyers and suppliers. In equilibrium, an endogenous fraction of those advertisements is converted

into successful matches. Matching with buyers boosts the firm's revenues, while supplier matches affect the cost of intermediate inputs. In a version of this model that incorporates a spatial dimension (Arkolakis et al., 2023), the Internet can affect four parameters,[9] namely firm location-specific search cost, origin-destination-specific search cost, origin-destination-specific matching technologies, and the firm's productivity draw. First, the Internet can affect the location-specific cost of search. In their model, this is equivalent to increasing the productivity of the local advertisement sector and thereby lowering the cost of advertising for local firms. Second, the cost of search not only depend on local productivity, but crucially also on the bilateral search cost between the searching firm's location and the destination of its search. The difference between local cost and bilateral cost can arise for various reasons. The simplest one is that firm-to-firm interactions often require two-way communication in which case those bilateral search costs depend on the firm's own connection, but also the connection of their counterpart. Third, origin-destination dynamics also play a role in bilateral matching technology. This parameter governs the number of successful equilibrium matches. While the underlying framework is static in nature, the matching technology can be interpreted as an indicator of ease of communication or location-dependent contracting frictions. Successful matches could become more likely if firms think the relationship can be sustained and governed, for example, by reducing the number of disputes via more effective communication that is supported by the Internet. I will return to the role of the bilateral nature of the search cost and matching technologies when discussing potential constraining factors to leveraging the potential of the Internet for network formation and firm productivity more broadly. Fourth, the Internet can shift the firm's own productivity by improving the efficiency of internal processes (Jiang, 2022). I will return to each of the four aspects when discussing Internet use by firms and the constraints that interact with it.

[9]These parameters are treated as exogenous in Arkolakis, Huneeus and Miyauchi (2023), but could be become endogenous in an extension that models the firm's adoption choice.

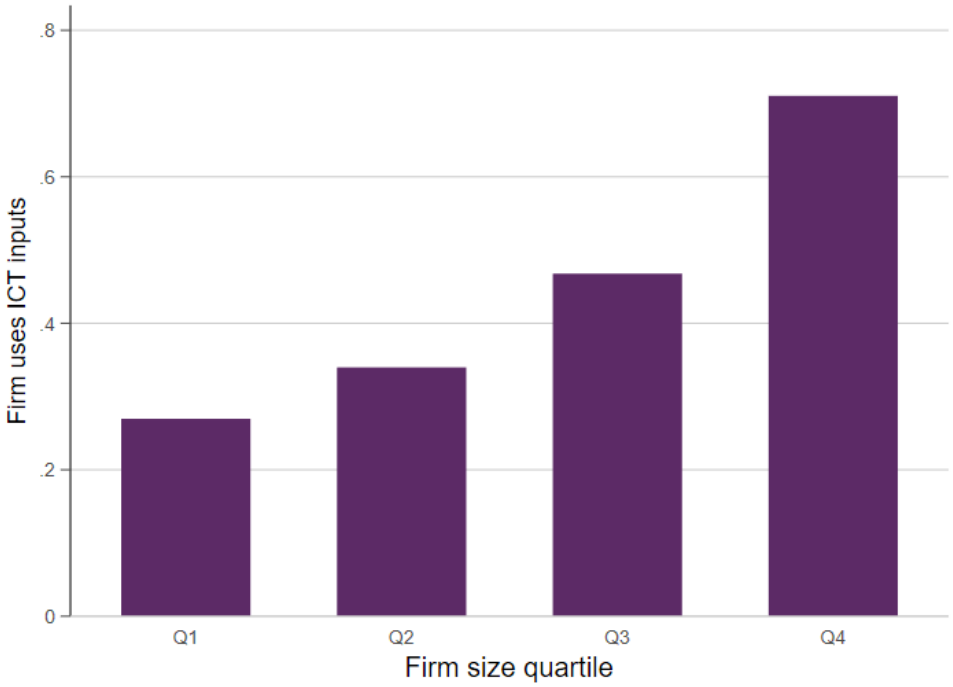
3.6 Internet use

3.6.1 Firm size and ICT use: motivating evidence from administrative records

I will now draw on the administrative data to motivate the paper's research question and the subsequent analysis of the qualitative interviews. While it is not possible to directly measure Internet adoption in the administrative data, the firms' input purchase transactions serve as an indicator of whether firms have purchased any inputs from suppliers in the ICT sector. The administrative records show that the use of ICT inputs varies greatly by firm size. Firms in the top quartile of the firm size distribution (based on sales) are almost three times as likely as firms in the bottom quartile to purchase ICT inputs.[10] ICT inputs, especially those related to Internet access and maintenance of ICT infrastructure, account for a larger input share of small firms than they do for larger firms (see Figure 3.4). For firms in the fourth decile and above, ICT inputs account for 60-80% less of their overall inputs than for small firms. This suggests that ICT can, to some extent, be interpreted as a fixed cost rather than a variable cost.

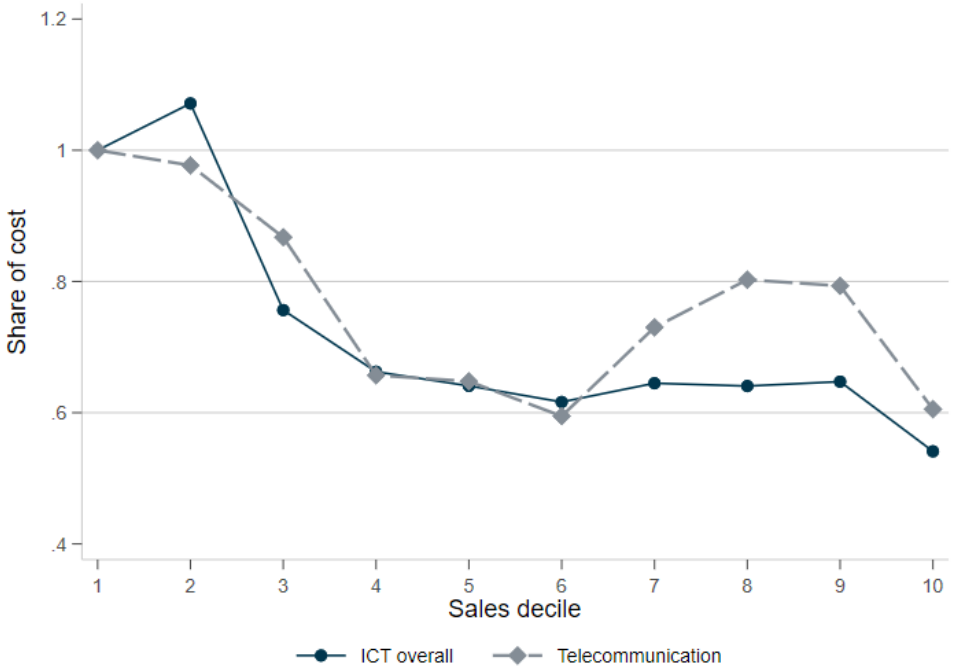
[10]The percentage of ICT use for the bottom three quartiles is lower than the 60% of firms with Internet access at their premise found by Cirera et al. (2022b). Based on the available descriptive statistics in Cirera et al. (2022b) it is not possible to precisely pin down how the size, sector and geographic composition in their sample compares to the population of firms captured in the tax records (i.e., the data used in this paper). Further, a number of firms interviewed for the qualitative interviews noted that they rely on smaller Internet Service Providers, which might not be VAT registered. This prevents firms from claiming input VAT for these transactions.

Figure 3.3: ICT adoption of firms covered in the administrative data



This graph plots the share of firms in each size quartile (based on sales) that purchases inputs from a firm in the ICT sector. Here I rely on input transactions from the 2019 cross-section.

Figure 3.4: Share of inputs from the ICT sector by firm size

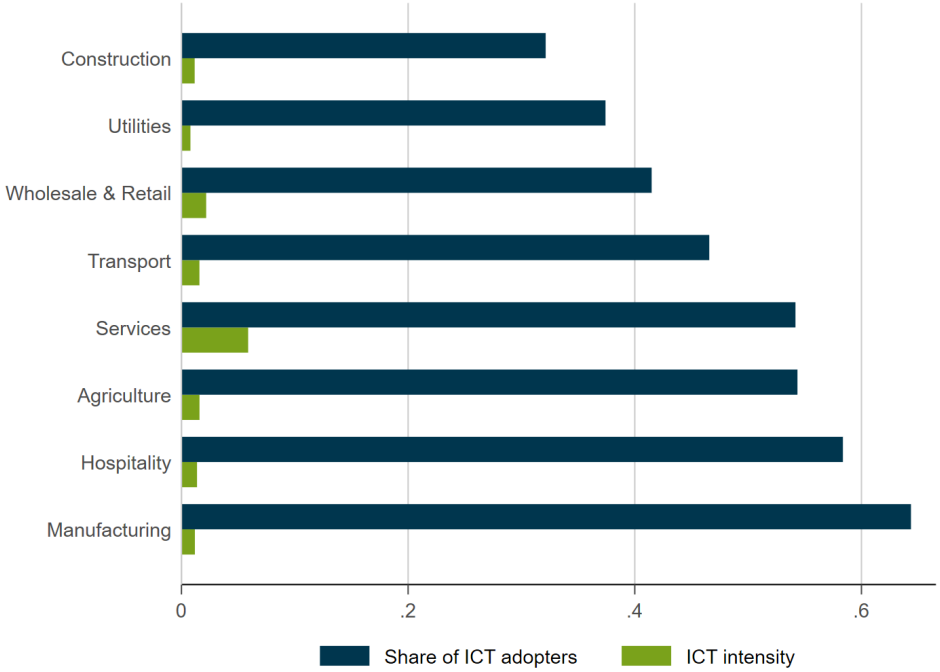


This graph relies on input transactions from the 2019 cross-section. The input share for the first decile is normalised to one, while all other deciles are plotted relative to the first.

I plot the share of firms that purchase inputs from an ICT supplier and the ICT input share dis-aggregated by sectors in Figure 3.5. At the top end of the graph, over 60% of all manufacturing firms buy from the ICT sector, while only just over 30% of construction sector firms do so. Service sector firms (excluding wholesale, retail, transportation, and hospitality) have by far the highest ICT input share. However, on average, the intensity of use is still less than 10%. The reported incidence of ICT input purchases likely represents a lower bound for the share of firms that access any form of Internet. Many sole proprietors use their phones for access and do not claim input VAT for those smaller transactions.[11] I am unfortunately unable to identify the exact nature of the ICT inputs purchased by firms. This includes, for example, the type of internet connection the firms opted for.

[11]If firms rely on pre-paid SIM cards, it can be cumbersome to claim input VAT from these purchases. For much of the study period, the Kenya Revenue Authority would only allow firms to claim input VAT on purchases where the supplier has already registered them as official sales in their transactions.

Figure 3.5: ICT adoption of firms covered in the administrative data



This graph relies on input transactions from the 2019 cross-section. Services comprise business services and any consumer services that do not fall under transportation, hospitality, wholesale & retail.

The above insights motivate the central question I explore in the qualitative interviews: Why does ICT use vary by firm size? And what implications does the documented heterogeneity in ICT use have for the potential of Internet infrastructure to foster domestic market integration?

3.6.2 Internet use: an in-depth exploration using qualitative interviews

I begin this section by examining the types of Internet access utilised by firms before delving into the firms' operational usage of the Internet connection. To investigate the significance of firm size, I classify firms with ten or fewer employees as small firms, while firms with more than ten employees are considered large firms. Out of the 33 firms surveyed, 18 (55%) belong to the small firm category.[12]

[12]Six firms in the sample have more than 50 employees. The remaining nine firms have between 11 and 50 employees.

3.6.3 Mode of Internet access

Figure 3.6: Primary mode of Internet by firm size



The underlying data were collected as part of 32 semi-structured interviews with 33 firms.

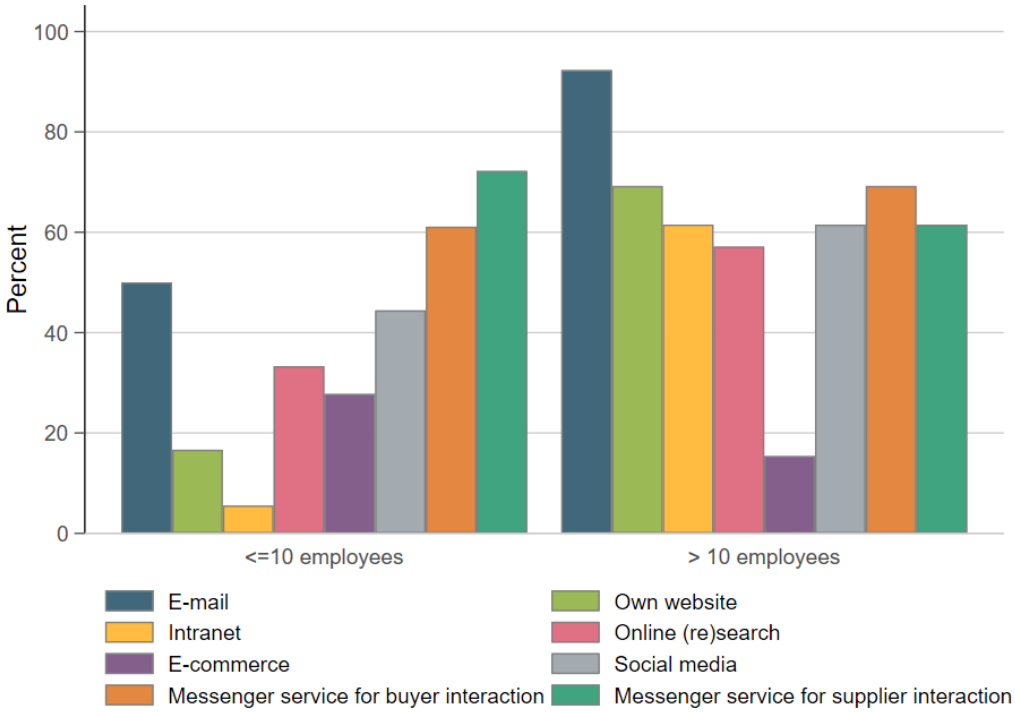
The primary mode of Internet access for small firms is mobile broadband. Most of the firms' managers, many of whom own the firm, access the Internet solely through their mobile phones. Most of the small firms that opted for a more stable form of Internet access were either based in Nairobi with easy access to the fibre network or were particularly reliant on Internet for their business model and hence opted for FTTB.[13] Larger firms, on the other hand, do not primarily rely on mobile broadband for their Internet access. Instead, they opt for fixed wireless or fibre connections to their premises.[14] Fixed wireless connections provide a relatively cheaper option for obtaining a stable connection, which has significantly improved connectivity for

[13]This for example includes firms in the hospitality sector that offer WiFi to their customers.

[14]The two chosen modes for Internet access are also the most popular with households and the remaining universe of broadband subscribers in Kenya.https://www.spglobal.com/marketintelligence/en/news-insights/research/kenya-fixed-broadband-ftth-fixed-wireless-improving-access-speeds

many firms that previously relied on microwave connections.[15] The expansion of fixed wireless has been made possible by the growth of 4G technology. Among the interviewed firms, fixed wireless is particularly popular with smaller firms in urban areas and larger firms in more rural areas where deploying fibre to their premises would involve substantial costs.[16]

Figure 3.7: Use of Internet by firm size



The underlying data have been collected as part of 32 semi-structured interviews.

3.6.4 Applications of Internet connectivity: Contrasting usage patterns between small and large firms

In line with the framework in Section 3.5, I distinguish between three different use cases of Internet connectivity: communicating and transacting with existing customers, searching for new buyers and suppliers, and improving firm operations. Using

[15]Modern microwave connections can in some cases be at least as reliable and powerful as fibre optic connections. However, this is not always the case, especially for more antique versions, which are sensitive to weather patterns.

[16]While fibre connections typically offer higher upload and download speeds and are less likely to come with speed restrictions, firms are required to bear the costs associated with deploying fibre to their premises. This cost can be significant if the closest fibre connection point is located far away.

the language of the model, these three aspects will affect match quality, search cost, and the firm's productivity, draw respectively.

Communication and transaction cost with existing buyers and suppliers

Consumer and B2B oriented businesses alike mentioned how digital tools such as social media and messenger services help them to communicate with their existing customer base more effectively, improving both the quantity and quality of interactions.

A food and beverage producer, whose distribution otherwise takes place via a network of retailers and wholesalers, remarked how public profiles on social media platforms have opened up a direct channel of communication with final consumers. To them, it is critical to obtain feedback on their existing products and to launch new ones. Businesses of all sizes and product categories, for example, ask employees to post product information on their WhatsApp statuses for their existing trade partners to see.[17] A retailer in a smaller town noted:

> *"Basically, online gives you more visibility."*

Messenger services emerge as a common way for all firms, but small ones in particular, to use their Internet connection, as shown in Figure 3.7. Almost all small firms stated that they rely on messenger services to communicate with their buyers and/or suppliers. From the interview responses, there is evidence that the use of messenger services improved the match quality and efficiency of firm-to-firm relationships. Firms note that prior to messenger services they primarily relied on phone calls and text messages. The stated key advantage of using messenger services is that they allow firms to send photos of goods and/or invoices. This, in turn, enables them to communicate more effectively about more complex product specifications and to avoid disputes over the same. A retailer of construction materials and hardware who has been in the business for over 20 years, for example, noted that sending photos results in fewer mistakes, such as sending a product with the wrong specifications. A common use, although not necessarily specific to messenger services, is that firms

[17] While it might be less surprising that employees of beverage distributors post information in their WhatsApp statuses, the interviews revealed that the phenomenon of advertising through status updates extends even to firms selling industrial input materials.

use digital correspondence as evidence of what was ordered, thereby substantially reducing the incidence of disputes with their trading partners.
Owner and managers of firms which have been in business for over a decade mentioned how online communication via messenger services and e-mail saves them time that they would have previously spent on visiting suppliers to place orders or to visit customers to discuss the specifications and fulfilment of orders. A small business offering printing services in Nairobi explained:

> *"[We] have a lot of efficiency gains from not having to travel through town to communicate with clients. Doing [our] work is now easier. [...] In person visits are usually for clarifications, e.g., if the client has a specific type of print in mind and we want to show them a sample."*

The applications thus far suggest that Internet services can serve as an improvement to the match quality of firms of any size with their buyers and suppliers.

Searching for buyers and suppliers

At the same time, smaller firms are less able to utilise digital tools to search for customers or suppliers. One exception is mapping tools that allow firms to publish a pin of their location in the public domain. For many, it is a search tool their customers rely on.[18] Another exception is those firms whose primary business model relies on interactions in an online sphere. These include providers of digital services, such as online courses, or firms that create digital content for others, such as websites or videos. These businesses primarily rely on online advertisements, such as influencer marketing[19], and their own online presence on social media and their websites to expand their customer base.
For large firms, the most commonly mentioned and most transformative application of Internet-based technologies are improvements to the firm's sales operations. Firms use them to broaden the geographic scope of their sales by penetrating additional domestic (and sometimes also international) markets. Here the interviews suggest that digital tools allow firms to maintain a more comprehensive and up-to-date database of their

[18]In one instance, a businesse had a pin with customer reviews attached in a widely used application, but without being aware of it.

[19]Influencer marketing is a common tool even among smaller businesses that do not use online engagement otherwise.

customers. Furthermore, tools that allow managers to monitor and track the work of their sales personnel were popular use cases. A food and beverage company's use case for a Customer Relationship Management (CRM) system illustrates the relevance of both aspects. The firm's managers noted how the system allows them to collect sales information in real time, keep track of sales personnel in the field, and, as a result, target customers more effectively. They also explained how the introduction of an integrated sales system allowed them to expand their sales operations and achieve a greater geographic scope within Kenya:

> *"Once the market reaches a certain stage, you can hand it over to the distributor. [Our own] sales team can then be re-directed to new markets."*

Furthermore, virtually all large firms use e-mail communication and 70% have their own website with the main purpose of providing contact and product information. In most cases, the website is designed to provide information for potential buyers. In some instances, firms mentioned that potential suppliers had reached out to them after coming across their websites as a result of an online search. In the majority of cases, this applied to firms that import directly and have international suppliers.

Internal operations and firm productivity

Large firms in particular mentioned that they use Internet-based tools to expand the scope of their internal operations. This includes the introduction of their own intranet (Jiang, 2022) or cloud storage for document and data sharing as quintessential for expanding their team of employees. This allows them to bring on more employees, middle, and top-level managers.

While some companies rely on dedicated sales applications to monitor their employees, others noted that maintaining a digital inventory system or deploying a point-of-sales system with a digital paper trail allows them to reduce instances where employees sell products under the counter. CCTV cameras connected to the Internet are used for remote monitoring and to collect information on customer-employee interactions. CCTV cameras connected to the Internet, inventory, and point-of-sales systems are also used by firms with less than ten employees.

Finally, two-thirds of the small firms pointed out constraints that prevent them from using the available Internet infrastructure more extensively and more effectively. I discuss these constraints in the next section.

3.7 Distance is indeed not dead: constraints to the effectiveness of Internet connectivity

The focus of the previous section was to explore the ways in which firms use Internet connectivity to enhance their interactions with potential and existing buyers and suppliers and their internal operations. At the same time, 19 out of the 30 interviewed firms pointed out constraints that prevent the more effective use of their Internet connection. Of these 19 firms, 13 have ten employees or fewer. Up to now, my analysis has focused on documenting and describing heterogeneity in the use of Internet connectivity amongst firms of different sizes. Studying factors that limit the extent to which firms can leverage Internet connectivity brings us closer to the question of why this is the case. Based on the qualitative interviews, I identify preferences for face-to-face interactions and the need to undertake complementary investments as the two key factors.

3.7.1 The importance of personal relationships and face-to-face interaction

Two main reasons as to why online tools were used less frequently for customer acquisition and supplier search emerged from the interviews. First is the importance of personalised relationships with buyers and suppliers, and second, in a related manner, is the importance of face-to-face interactions to gather information.

Business owners and managers consider developing a personal relationship with buyers and suppliers to be essential. A software developer described the feared lack of a personal touch as the major concern of business clients, who are considering the use of chat bots to enhance their customer services. In line with the findings of Awori et al. (2022), who study the adoption of technology by small businesses in Kenya, the developer noted that the personalisation of messages was considered essential for establishing a rapport with trade partners. Concurrently, firms felt this personal aspect

was harder to replicate in online interactions where the two parties had never met in person. A travel agency described how the use of WhatsApp groups to exchange information with hotels, travel packages, and tour providers domestically and abroad has become important to them. Nevertheless, they emphasised that even with many of their trading partners based abroad, they are required to meet someone in person before adding them to their network:

> *"There are companies that reach out to use via those WhatsApp groups or WhatsApp in general and they will send fantastic offers. But we will ignore them until we have met someone from the company in person."*

Overall, two main reasons why firms emphasise personalised forms of interaction were raised. The first is to overcome contracting and information frictions through trust building and long-term relationships (Johnson et al., 2002; Macchiavello and Morjaria, 2015). Several companies pointed out that a lack of trust on the buyer's side impedes online sales from taking off. For example, companies dealing with a wide range of products, from beverages to beauty products and electronics, noted that customers' primary concerns are related to the quality of the product and counterfeit products in circulation. The second reason for emphasising personalised interaction was the notion that it allows firms to better understand and consequently tailor products or services to their customers' needs and remain competitive. Smaller firms were more likely to mention the aspect of personalisation and the need to stay closely aligned with the local community.[20] Relative to larger firms, small ones are not able to rely on a more universally known brand reputation. This constraint is more binding for downstream businesses, which tend to be smaller and dominate markets outside Kenya's metropolitan areas (Chacha et al., 2023). These businesses primarily rely on final consumers as their customer base, who in turn might be less well-versed in assessing product quality.

[20]One retailer described doing business as a "family transaction". They and another retailer noted that building a business based on personal relationships can also come with the downside of operating on tighter margins as repeated customers are more likely to ask for discounts. The move towards online sales and purchases was seen as a way to engage on a more transactional rather than relational level. The anonymity of online transactions was stated by one of the owners as an enabling factor to take a more "behind the scenes" role.

Closely related to the aspect of personalised interactions is the relevance of face-to-face meetings. Businesses mentioned a wide range of cases in which face-to-face meetings are considered an advantage over or a complement to online interactions. Most crucial for my overarching research question, to establish a new business relationship, firms rely on information that many say could only be acquired through face-to-face meetings.

For example, larger upstream firms use in-person visits to their buyers to assess their creditworthiness. Many of these firms extend trade credit to their customers. Firm size and inventory volumes, in particular, are considered good indicators for a buyer's creditworthiness and a smaller probability of default in the near future. Firms across the size distribution rely on in-person visits to locations where they expect potential buyers and suppliers to be located. Firms explained that this in-person form of market research allows them to obtain contact information that might not be available online, learn about new products, obtain a more complete overview of available varieties (Startz, 2021), and gather general market information (Memon, 1976). Firms that operated in less dense markets were more likely to note that offering a wide variety of choices and maintaining a high level of availability allows them to compete more effectively. They also mentioned how variety is seen as a pre-condition for expansion. In the words of a retailer of fast-moving consumer goods:

> *"Every product adds a customer."*

These aspects motivate their repeated search and in-person visits to Nairobi, which gives them access to a wide range of wholesalers.[21] Nevertheless, once a relationship is established, fewer in-person interactions are necessary. Exceptions arise when a deeper understanding of business needs is required, disputes need to be resolved, or newly entered businesses feel as if they are not able to attract sufficient attention from potential suppliers to have their demands met with a reasonable turnaround. For example, an agri-processing firm located in Central Kenya mentioned that suppliers of non-agricultural inputs would take a four–hour car ride from Nairobi to visit their premises:

> *"Suppliers also come and visit us to understand the operation and needs of the firm."*

[21] Many would still choose to buy from the same subset of long-term trade partners.

Common among large firms are a tight networks of distributors and sales agents who pay regular visits to customers to take orders. One of those customers, a mid-sized supermarket, noted that suppliers pay them weekly or bi-weekly visits in the hope of boosting their sales, often with success:

> *"The quantities we order are larger if sales people are in the shop, sales people will nudge us to buy larger quantities and more products."*

Some of the firms mentioned how the COVID-19 pandemic has shifted norms around meeting suppliers in person in advance of the first transaction. Awori et al. (2022) discuss how small businesses leveraged the existing social networks of their employees and referrals to expand their customer base at a time when the world moved online, and face-to-face interactions were no longer an option. However, since social networks are often shaped by physical distance (Bailey et al., 2020), reliance on pre-existing social networks works against Internet connectivity as a potential means to overcome spatial frictions and foster domestic market integration.

3.7.2 Complementary investments

Section 3.6.4 discussed how, albeit with constraints, social media and other forms of online presence allow firms to enhance their visibility in relation to potential and existing customers. In parallel, however, a greater degree of visibility and the attached shift in demand were mentioned by several smaller to medium-sized businesses as one of the reasons why they have decided against pursuing digital solutions for customer interaction. The business owners were concerned about their capacity to handle an increased volume of customer interactions. Depending on the type of business, effective use of digital technologies would require either (i) hiring one (or more) employee(s), (ii) expanding existing production capacity, or (iii) making complementary investments in transportation capacity. Financial constraints and uncertainty over the volume and stability of the resulting demand informed firms' decisions to scale their operations.

Firms noted that expanding their business would require being able to offer delivery to their buyers' door-step. In the absence of cost-effective, reliable, flexible, and easy-to-use delivery services, this would require an investment in transportation capacity directly owned by the firm. A wholesaler of fast-moving consumer goods in

a suburb of Nairobi, who does not own a vehicle, explained that transport capacity and the pace of delivery were their most important criteria for selecting suppliers (besides product prices). A wholesaler of construction materials and hardware in central Kenya pointed out that:

> *"There is nothing that is technology-related that hinders us from selling more. The main obstacle is that we don't have a vehicle. Transport is the main constraint to our business. [...] At the moment we cannot afford to buy one as it requires a substantial amount of capital."*

Several firms with existing in-house transport capacity explained that they would only be able to expand their sales if new customers were located along transportation routes they already serve with their existing fleet. Small and large firms alike had apparently explored the possibility of moving into online sales, but pointed out that it would only be feasible if they were to offer a complementary delivery service. At the same time, digital tools have resulted in the recent entry of disruptive business models into Kenya's transportation industry. Their digital platforms aim at enabling easier and more reliable matches of idle transportation capacity, allowing firms to avoid having to back haul empty lorries after delivering products to customers. A manufacturer mentioned how the reliable access to load for their lorries via such platforms reduced the downside risk of expanding their fleet and resulted in an expansion of their sales.

In a related manner, but emphasised less often, firms noted that using online tools would only make sense in conjunction with an expansion of their production capacity. A mid-sized bakery noted:

> *"Scaling sales would also mean scaling up production. We currently have one oven, but would then need a second oven. We lack the capital to buy a second oven."*

A flour manufacturer pointed out that their firm does not struggle to access capital, but their concern is that expanding too fast without solidifying their demand exposes them to large downside risks in the event of idle capacity. This notion is reflected in Bassi et al. (2022)'s finding that the indivisibility of investments in machinery and

production often do not allow for small incremental expansions and can as a result become a binding constraint for firms.

Expanding production capacity is linked to hiring more employees. Asked about their decision to adopt digital technologies, successful adopters emphasised their reliance on consultants, newly hired employees, or knowledgeable family members who helped them set up and maintain the respective technology. At the same time, the need to hire new employees prevented some companies, especially smaller ones, from adopting Internet-based technologies. Owners and managers noted that they either lacked the skills required for successful implementation or felt they did not have the bandwidth to oversee their business and implement new technologies at the same time. This includes tasks such as maintenance, managing the expected additional number of customer queries, and providing up-to-date content for their online presence. Smaller businesses, in particular, expressed uncertainty over future demand and concerns that the business would be unable to sustain more employees.

3.8 Discussion and conclusion

What do the findings imply for the potential of Internet connectivity to foster a greater degree of domestic market integration? Both the reliance on face-to-face communication to search for new buyers and suppliers, and market frictions in the transportation sector, mean that the effectiveness of Internet connectivity still interacts with distance. As a result firms with closer physical proximity maintain their advantage over potential competitors.

Moreover, the choice to adopt Internet-based information- and communication technologies is closely linked with complementary decisions over transport, production capacity, and hiring. This conclusion aligns with discussions in earlier work by Jorgenson et al. (2008), who emphasise the role of complementary investments in skills, organisational structure and intangible capital as a prerequisite for the productivity-enhancing use of information technologies. In the context of Kenya, small firms are more likely to be financially constrained than large firms, which in turn limits the feasibility of making concurrent adoption and expansion decisions. Small firms are, in turn, more common in smaller markets, which tend to be located farther away from

metropolitan areas and are thus the main target areas for policies aimed at fostering the integration of domestic markets. The outlined constraints hence provide options for policymakers to support the private sector in leveraging the full potential of Internet infrastructure. This could include policies aimed at reducing frictions in the transportation sector, such as the introduction of the National Addressing System that is currently being implemented in Kenya.[22]

[22]https://www.ca.go.ke/index.php/national-addressing-system-kenya

Acknowledgements for Chapter 3 First and foremost I thank the interview respondents for their time. This research would not have been possible without them. I thank Elizabeth Gatwiri for excellent research assistance. This chapter has benefited from extensive conversations with public and private stakeholders and experts in the ICT sector. I thank Geoffrey Mwenda, Kagonya Awori, Alex Njihia, Benjamin Muoki, Miriam Mutuku, Thomas Luti, Fred Ong'aro, Caroline Koech, James Njunguna, and Nashon Muok for their time and valued input. The data collection for this chapter took place during my visit to the Institute of Mathematical Sciences (SIMS) at Strathmore University. I thank Elias Kaburu, Mary Achieng, Tabitha Njunguna, and Mumbi Wachira and the wider community at Strathmore for welcoming me and their support. I am grateful for comments and feedback by Raphael Bradenbrink, Kristina Kämpfer, and Sanghamitra Warrier Mukherjee on this draft. I gratefully acknowledge financial support from the Structural Transformation and Economic Growth (STEG) programme, a joint initiative by Centre for Economic Policy Research (CEPR) and the UK Foreign, Commonwealth & Development Office (FCDO), the Oxford Economic Papers Research Fund (OEP), the Centre for the Study of African Economies (CSAE) and the German National Academic Scholarship Foundation. This study has been approved by the Department of Economics Research Ethics Committee at Oxford (protocol no.ECONCIA21-22-48), the Strathmore University Institutional Scientific and Ethical Review Committee (protocol no.SU-ISERC1480/22), and the Kenyan National Commission for Science, Technology and Innovation (protocol no.NACOSTI/P/22/20556).

CHAPTER 4

Supply chains in times of crisis: Evidence from Kenya's production network

with **Peter Wankuru Chacha and Benard Kipyegon Kirui**

Abstract Trading relationships between suppliers and buyers play a key role in transmitting both local and international shocks. We use transaction-level data from Kenya to study the relevance of a firm's domestic network position and links to international supply chains in determining its trajectory during the COVID-19 crisis. We document that firms with high exposure to import and export markets tend to be larger, older, and employ more workers. The specialisation of direct importers, often intermediaries, on international markets made them very vulnerable to the initial COVID-19 shock. Exporters, one-third of whom operate in primary sectors, experienced a less severe decline in sales. We find that both importers and exporters adjust their domestic supply chains in response to international trade shocks - before and during the crisis alike. Sourcing from international markets does not crowd out domestic purchases, while sales abroad and at home can act as substitutes. Diversified domestic supply chains helped firms to mitigate the impact of the COVID-19 crisis and recover more strongly.

4.1 Introduction

The COVID-19 pandemic threw a new spotlight on the role of supply chain structures in firm dynamics. Previous research has documented the key role of interdependencies that arise through firm-to-firm trade in the transmission and amplification of shocks.[1] The present paper relies on administrative firm- and transaction-level data from Kenya, East Africa's largest economy, to document firm dynamics in response to the COVID-19 crisis. The setting enables us to study the role of a firm's supply chain position in determining its trajectory during times of heightened uncertainty and recovery from a severe aggregate shock. We focus on three aspects of the firm's network position: the degree of exposure to international supply chains, diversification and complexity of its domestic network.

While only a small share of firms worldwide participates in international supply chains, non-participants are indirectly linked to them via domestic buyers and suppliers (Ahn et al., 2011; Abel-Koch, 2013; Bernard et al., 2015; Dhyne et al., 2021; Grant and Startz, 2022). Customs data, the most popular sources to trace global supply chains at the firm and transaction level, typically capture only firms that directly import and export. As a result, firms with indirect exposure remain understudied compared to their relative importance in many low- and middle-income economies (Ahn et al., 2011). A key contribution of our paper is its ability to combine domestic firm-to-firm transaction-level data with customs data, allowing us to trace the indirect exposure of Kenyan private sector firms to global value chains.

The purpose of this paper is three-fold: First, we document the degree to which formal firms in Kenya's private sector are directly or indirectly embedded in international supply chains. We find that only a few large firms have strong direct or indirect linkages to international supply chains. These firms are key to the Kenyan economy, both in terms of their share of total sales and employment. On the export side, the linkage is dominated by primary sectors, while imports are dominated by the wholesale and retail sectors.

[1] See, among others, Acemoglu et al. (2012); Gabaix (2011); Barrot and Sauvagnat (2016); Huneeus (2018); Bernard and Moxnes (2018); Carvalho and Tahbaz-Salehi (2019); Carvalho et al. (2021); Kashiwagi et al. (2021); Korovkin and Makarin (2022); Arkolakis et al. (2023).

Second, we provide evidence on the response of firms to international shocks, both during the years immediately prior to the COVID-19 crisis and during the crisis itself. We find that the COVID-19 crisis had an almost universally negative effect on major firm outcomes. Average firm-level sales dropped by 56 percent in April 2020 relative to January 2020, taking seasonal fluctuations into account.[2] The downturn was smaller for exporters, which experienced a drop of 47 percent, and firms with indirect links to import markets, which saw a decline of less than 20 percent during the peak of the crisis. We then combine the data with information on global trade flows to study how firms respond to international trade shocks. We are not only able to study how firms with indirect linkages to international supply chains respond to the shocks but also how firms adjust their domestic supply chain in response to changes in world demand and supply. We find that sourcing from international and domestic markets is complementary for Kenyan firms, while exports and domestic sales can be substitutes. More favourable shocks to world supply supported the recovery process of Kenyan firms via imports. On the flip side, we find a particularly pronounced direct pass-through of import and export shocks during the peak of the crisis. It highlights the downsides of links to international markets.
Finally, we show how firm dynamics during the COVID-19 crisis differed for firms with highly diversified domestic supply chains, i.e. a larger set of active domestic supplier or buyer relationships, relative to those with less diversification. Firms with a highly diversified domestic buyer base lost, on average, 4 percent less of their sales relative to their less diversified peers. Firms with upstream diversification (a larger supplier base) recovered more strongly.

This paper is related to and contributes to several strands of literature. First, we contribute to a large literature that shows exporting is a rare undertaking among firms, with only a small proportion participating in international trade (Bernard et al., 1995; Bernard and Jensen, 1999, 2004; Greenaway and Kneller, 2007; Wagner, 2007; Coulibaly et al., 2022).[3] Examining exporters' characteristics, the literature highlights that they differ from non-exporters in virtually all performance measures,

[2]Including the hospitality sector, the average drop was as much as 67 percent.

[3]It is a precursor to the modern theory of international trade and firm heterogeneity (Pavcnik, 2002; Melitz, 2003; Bernard et al., 2003).

even before entering export markets. Within Kenya, a number of studies have examined exporting at the firm-level (Rankin et al., 2006; Granér and Isaksson, 2009; Abala, 2013), relying on survey data and comparing attributes of Kenyan firms with findings from the rest of the world. A recent strand of papers on Kenyan exporters uses transaction-level customs data (Macchiavello and Morjaria, 2015; Chacha and Edwards, 2017, 2018, 2019; Türkcan et al., 2022) to examine dynamics of firm-level exports,[4] including survival of trade relationships and adjustment along extensive and intensive margins in international trade. Nevertheless, in many contexts, a larger pool of firms imports or exports indirectly via wholesalers (Ahn et al., 2011; Abel-Koch, 2013; Bernard et al., 2015) or even longer chains of intermediation (Dhyne et al., 2021; Grant and Startz, 2022).[5] Two complementary approaches in the literature on global value chain participation of firms in low- and middle-income countries attempt to address the lack of data on indirect participants: The first strand focuses on specific sectors, often collecting targeted quantitative and qualitative primary data (Navas-Alemán, 2011; Startz, 2021; Foster et al., 2018; Bassett et al., 2021; Hansen et al., 2022). A second strand combines firm-level data with input-output tables to study firms in sectors with high exposure to global value chains (Del Prete et al., 2017; Amendolagine et al., 2019; Ndubuisi and Owusu, 2022). Our empirical set up bridges the two approaches by studying the nature of the exposure to international supply chains at the firm level for a large range of sectors, including services.[6] Transaction-level administrative data allow us to identify firms with indirect links to international markets in the context of Kenya.[7] It enables us to compare their characteristics to firms focused on domestic markets and track their trajectory during a global crisis, the COVID-19 pandemic. In doing so, we also complement a small set of papers

[4] Majune et al. (2020) study product-level dynamics and are thereby able to consider two decades of trade dynamics.

[5] A related, very active literature studies the role of intermediaries in (mostly domestic) agricultural and food supply chains (Minten et al., 2012; Reardon et al., 2021; Bergquist and Dinerstein, 2020).

[6] These sectors tend to receive less attention relative to agriculture and manufacturing (Hansen et al., 2022).

[7] The availability of similar administrative records to the one used in this paper has enabled researchers to study a series of questions related to globalisation and domestic firm networks, mostly in high and upper middle income countries. Alfaro-Urena et al. (2022), for example, study the effects of joining the supply chain of multinationals operating in Costa Rica on local suppliers, Demir et al. (2022) look at the effect of import taxes on domestic firms, and Adão et al. (2022) examine the implications of international trade for inequality in Ecuador.

that studies the response of domestic supply chains to international shocks (Huneeus, 2018; Boehm et al., 2019; Arkolakis et al., 2023). Our empirical setting shares the limitation with this literature in that we are unable to study the role of inventory in transmitting shocks along the supply chain.[8]

This paper links to a broader literature on the role of trade and global supply chains in the transmission of the COVID-19 shock (Majune, 2020; Berthou and Stumpner, 2022; Bassett et al., 2021; Fujiy et al., 2022; Lafrogne-Joussier et al., 2023). With the exception of Fujiy et al. (2022), the majority of the papers in this series rely on sector-level input-output tables (Barrot et al., 2021) or focus exclusively on international trade flows and direct exporters and importers (Majune, 2020; Berthou and Stumpner, 2022; Lafrogne-Joussier et al., 2023). Our paper complements existing studies, which also rely on more granular information but are focused on a specific industry, such as Bassett et al. (2021), who use a detailed mapping of the supply chain of small-scale fishing industries in four countries.

Moreover, we speak to a small but growing literature that studies the relevance of supply chain diversification for firm dynamics in response to shocks (Freund et al., 2022; Kashiwagi et al., 2021; Lafrogne-Joussier et al., 2023; Boehm and Sonntag, 2023). In a paper closely related to ours, Lafrogne-Joussier et al. (2023) looked at French exporters studying the pass-through of the initial import shock triggered by the first lockdown in China in February 2020. Firms with geographically diversified international supply chains were not able to cope better with the shock than firms that solely relied on imports from China. In our setting, we find that diversification helped firms to mitigate the impact of the COVID-19 shock and allowed for a stronger recovery. Our setting complements theirs in that we provide evidence on the diversification of the domestic firm network and study a substantially broader set of firms with various degrees of exposure to international supply chains.

Finally, our paper contributes to evidence on the impact of the COVID-19 crisis on business outcomes, employment, and trade in Kenya (Shupler et al., 2020; Pape et

[8]The COVID-19 crisis led to heightened uncertainty about firm-level sales growth (Bunn et al., 2021). Greater volatility and uncertainty in forecasting demand, in turn, can exacerbate the misalignment of stock levels along supply chains, a distortion that leads to inefficiently high inventory levels and gets amplified upstream (the so-called bullwhip effect (Lee et al., 1997; Metters, 1997; Yao et al., 2021)). It would require us to observe either stocks, production output or order quantities on a monthly basis alongside sales (Yao et al., 2021).

al., 2020; Majune, 2020; Nordhagen et al., 2021; Kansiime et al., 2021; Egger et al., 2021; Kinyanjui et al., 2023; Kaberia and Muathe, 2021; Awori et al., 2022; Brooks et al., 2022). Relying on administrative records limits our analysis to the VAT-paying sector, which accounts for close to one third of Kenya's GDP (Chacha et al., 2022). We are therefore unable to trace the knock-on effects of international shocks to the informal sector.[9]

4.2 Data and context

4.2.1 Data description

We combine information from six different data sets collected by the Kenya Revenue Authority: (i) transaction-level data on domestic firm-to-firm trade from value-added-tax (VAT) records; (ii) import and export transactions from customs records; (iii) firm-level information on aggregate monthly sales, purchases, imports, and exports from VAT records; (iv) the monthly number of employees and payroll from pay-as-you-earn returns, (v) the value of end of year inventory holdings reported in annual corporate income tax (CIT) returns, and (vi) information on basic firm characteristics from registration forms. The data sets cover the period from January 2015 to March 2021 and can be linked through unique, anonymised firm identifiers. We restrict our analysis to private sector firms and firms with annual purchases greater than zero and annual sales of at least 5 million KES (approximately 40,500 USD) in at least one year that we observe in the data.

We use the transaction-level firm-to-firm data to construct a monthly supplier-buyer panel, capturing the monthly transaction volume and the number of transactions between each supplier-buyer pair. We further compile a monthly firm panel detailing sales, purchases, imports, exports, number of employees, and total payroll. We complement this data with information on firm age, firm headquarter location, and the sector of operation collected from registration forms. The sector classification

[9] Dix-Carneiro et al. (2021) show that the informal sector can act as an unemployment buffer, but not a welfare buffer in the event of negative aggregate (trade) shocks. Mobility restrictions during the COVID-19 pandemic seem to have disproportionately affected small businesses below the VAT cut-off in low- and middle-income countries (Nordhagen et al., 2021).

corresponds to the 4-digit International Standard Industrial Classification of All Economic Activities (ISIC) code. The customs data yields information on the origin/destination of imports/exports, as well as on 8-digit Harmonized System (HS) product codes,[10] quantity and value of the goods traded. Due to the reporting system implemented as part of the East African Community's (EAC) Single Customs Union, we are unable to map product-level details for trade with other EAC member states to the firm data. However, the information on total monthly imports and exports in the VAT data still captures aggregate trade with the EAC.

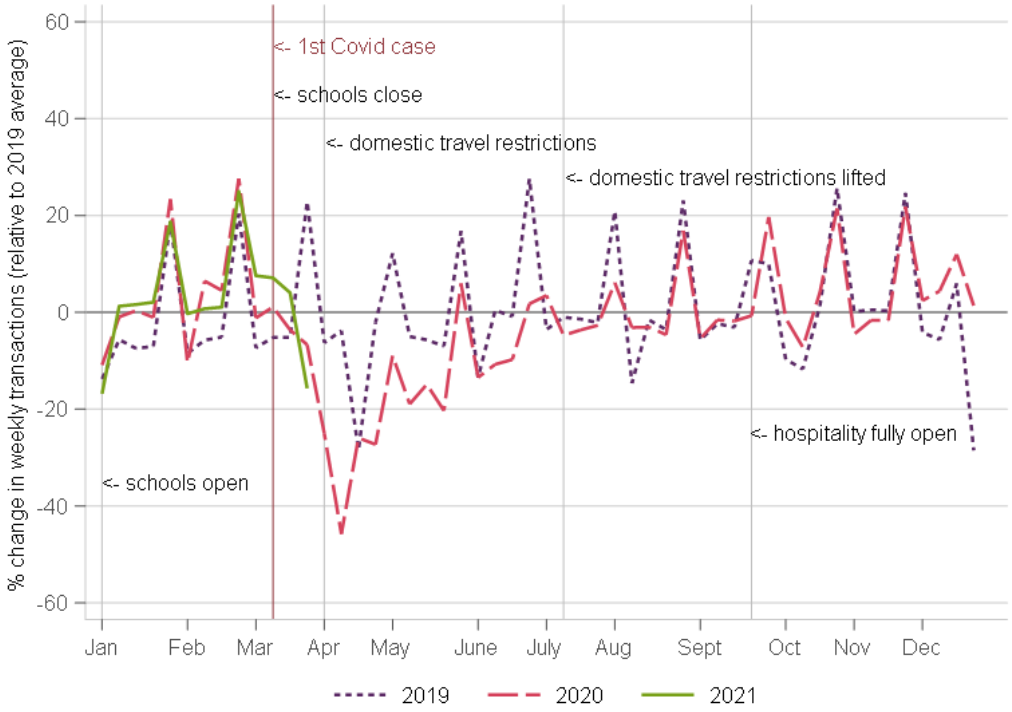
We complement the administrative data with monthly product-level customs data from UN Comtrade database to measure changes in world demand for products using 4-digit HS codes.[11] A number of important trade partners of Kenya, including China, Vietnam, and South Africa, do not report to the monthly database. Nevertheless, the export transactions that can be linked to the monthly UN Comtrade database represent 78 percent of the observed export volume. All variables denoted in monetary terms are deflated using the monthly consumer price index.

The data set covers 57,482 firms considering all the years between 2015 to 2021. For instance, in 2019, 44,888 formal firms participated in the firm-to-firm network, forming close to 1.5 million relationships among each other (see Appendix Table 4.A2). Their domestic sales aggregated to 6,806 billion KES (about 62 billion USD) and they employed 14 percent of the urban working population in 2019. Between 2015 and 2020, the value-added generated by VAT-paying firms, on average, accounted for 34 percent of Kenya's GDP (Chacha et al., 2022). The largest sectors in terms of volume are wholesale and retail, and manufacturing (see Table Appendix 4.A1).

[10]We aggregate products to the 4-digit level for the purpose of this paper and when linking the data with the UN Comtrade database below. While we lose out on some heterogeneity at the country-product level, as a result, this approach reduces the number of cases with potential mismatches due to classification errors. It further allowed for an easier cross-walk between cases where product classifications from previous HS nomenclatures were applied. We harmonise all HS codes to align with the HS Nomenclature 2012 Edition.

[11]https://comtrade.un.org/Data/bulk

Figure 4.1: Weekly firm-to-firm transactions



The figure shows the evolution of the percentage change in the number of purchase transactions relative to the weekly average number transactions in 2019. The data includes the universe of input VAT transactions reported by any tax filing entity.

4.2.2 Key events during the COVID-19 pandemic in Kenya

As a result of COVID-19-related shocks, Kenya faced its first recession in almost two decades in 2020.[12] The private sector in Kenya was first impacted by disruptions of its trade with China in early 2020.[13] The first domestic COVID-19 case was reported in early March and a series of measures to contain the spread of the virus was introduced.[14] Figure 4.1 depicts key events of the COVID-19 crisis in Kenya alongside the count of firm-to-firm transactions in 2019, 2020 and the first quarter of 2021.

Starting in mid-March of 2020 the Government of Kenya closed national borders, restricted domestic travel, banned public gatherings, closed schools, and later imposed

[12]https://citizentv.co.ke/business/kenyas-economy-slumped-into-recession-after-18-years-in-september-5068864/#:~:text=The%20Kenyan%20economy%20sunk%20to,cent%20in%20the%20third%20quarter.

[13]See for example Majune (2020) on lockdown policies and trade flows in and out of Kenya.

[14]https://www.covidlawlab.org/wp-content/uploads/2020/06/MOH-Press-Statement-on-the-Update-of-Coronavirus-in-the-Country-and-Response-Measures-as-at-20th-March-2020.pdf

a night-time curfew. The number of firm-to-firm transactions started to fall substantially in the second half of March before the recovery process started to kick in around mid-April. By early July, major domestic travel restrictions had been lifted and by September, the evolution of inter-firm transactions had largely caught up with the regular fluctuations experienced in 2019. In March 2021, Kenya experienced its third and till that date largest wave of COVID-19 cases. Another round of containment measures was imposed.[15] Overall, the transaction-level records suggest both a sharp initial downturn between March and April 2020 and a relatively swift recovery. Going forward, we define April to June 2020 as the peak months of the COVID-19 crisis and any of the subsequent months as the recovery phase.

4.2.3 Fluctuations in major firm-level outcomes during the COVID-19 crisis

To examine the evolution of key firm-level outcomes during the COVID-19 pandemic in Kenya, we first regress firm-level outcomes on a series of time dummies, firm-, and 2-digit sector-month-fixed effects.[16] The coefficient for January 2020 is normalised to zero. The results plotted in Figure 4.2 show that domestic sales of the average firm drop by 67 percent in April relative to January 2020. We further compare the trends in key firm-level outcomes to 2019 trends during the same window in Figure 4.A1 in the appendix. Despite the number of buyers and suppliers falling by as much as 20 percent at the peak of the crisis relative to January 2020, domestic firm-to-firm relationships bounce back relatively swiftly (see bottom two graphs of Figure 4.A1). Exports and imports took a much smaller hit than domestic trade flows and even rose to above pre-pandemic levels by the second half of 2020. It is important to bear in mind that the, on average, smaller dip in imports and exports is also driven by the fact that very few firms import and export directly. Therefore, we use a more nuanced measure of exposure to international supply chains in the following sections

[15] For a summary of key policy responses see, for example: https://www.imf.org/en/Topics/imf-and-covid19/Policy-Responses-to-COVID-19#K.

[16]
$$y_{it} = \beta \text{ preC19} + \sum_{k=1}^{2} \tau_{-k} \text{ jan2020}_{t-k} + \sum_{k=1}^{14} \tau_{+k} \text{ jan2020}_{t+k} + \phi_i + \eta_{sm} + \varepsilon_{it} \tag{4.1}$$

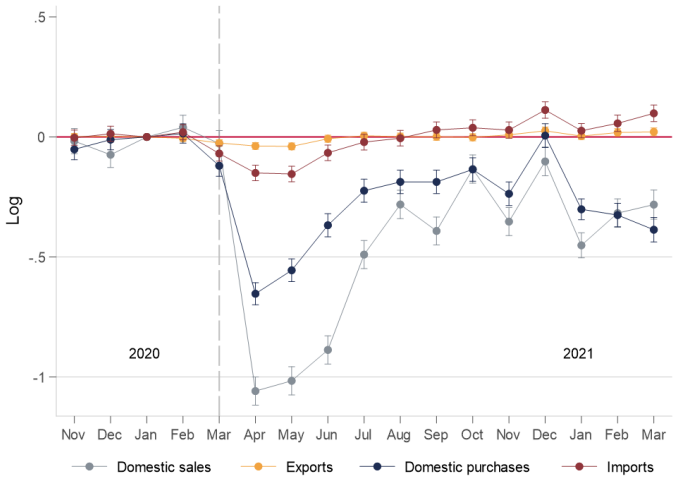
where y_{it} denotes the firm-level outcome, ϕ_i firm FEs, η_{sm} 2-digit sector-month FE, and ε_{it} the error term, which is clustered at the firm level.

of the paper to obtain a better understanding of the trajectory of firms linked to international markets.

We then exclude firms from the hospitality sector, which has been disproportionately affected by the pandemic, to address concerns regarding the impact of outliers in the hospitality sector being the key driver of results.[17] Much of the subsequent analysis further relies on being able to observe a firm's links with other firms in the domestic network. Therefore, we exclude firms that do not have at least one up- or downstream linkage to another domestic firm.

Figure 4.2: The evolution of major firm-level outcomes during the COVID-19 crisis

Sales, purchases, imports, and exports



The graph shows the results from estimating Equation (4.1), i.e. the regression of firm-level outcomes on a series of monthly time dummies. All outcomes are in log(+1) terms. Standard errors are clustered at the firm-level and the error bars show the 95 percent confidence intervals. The regressions include data from April 2015 to March 2021.

4.3 Links to international supply chains and firm dynamics during the COVID-19 crisis

In this section we take a closer at the characteristics of Kenyan firms with strong direct and indirect linkages to international supply chains, their trajectory during the

[17]Restaurants and hotels faced major restrictions up until September 2020. Their initial lack of recovery until the last quarter of 2020 therefore primarily reflects a response to targeted policy measures.

COVID-19 crisis, and their response to international shocks.

4.3.1 Measuring direct and indirect exposure to international supply chains

The data for 2018 indicate that only 11 percent of the firms import directly; however, as many as 90 percent of the firms that year buy from an importing firm. While virtually all firms in the network are linked to at least one importer or one exporter, these links tend to be very weak.[18] To get a more complete picture of the extent of indirect exposure to international supply chains, following Dhyne et al. (2021), we compute each firm's total exposure to foreign demand r_{iX}^{total} and foreign inputs r_{iM}^{total}. In other words, we consider the share of a firm's sales or purchases that is ultimately sold to or sourced from international markets. Focusing on exports for now, a firm's full exposure to foreign demand r_{iX}^{total} is the sum of its own share of total output that is directly exported r_{iX}^{direct} and the export share of its buyers, the buyers' buyers and so on ($r_{iX}^{indirect}$). Having constructed the adjacency matrix A of the domestic firm-to-firm network,[19] we use the following linear system to compute the total exposure to foreign demand:

$$r_X^{total} = (I - A)^{-1} r_X^{direct}$$

A key element of the above equation is $(I - A)^{-1}$, which can be interpreted as the firm-level version of the Leontief inverse. It captures the total impact or exposure of element i to j, both directly and indirectly (Sargent and Stachurski, 2022). i and j correspond to firms in our case rather than sectors in the classic version of the Leontief.[20] The total exposure to import markets can be computed following a

[18] This result is entirely driven by a majority of firms buying from highly connected firms which happen to import directly in one way or another. Of all firms, 88 percent buy from a firm that is both (i) a direct importer and (ii) in the 90th percentile of the outdegree distribution (i.e. firms with a lot of buyers). This pattern, in turn, is rooted in the underlying fundamental network structure of firm networks, including Kenya's firm network (Chacha et al., 2022), where less connected nodes tend to link with highly connected nodes. The resulting negative correlation between a firm's own degree and the average degree of the firm's trade partners is referred to as disassortative mixing by degree (Newman, 2018) (or sometimes negative degree assortativity in the empirical literature on production networks (Bernard and Moxnes, 2018)). Disassortative mixing by degree can be found in most other types of networks, with some notable exceptions like social networks (Newman, 2018).

[19] Element i, j of the adjacency matrix A corresponds to the share of i's output that is sold to j.

[20] $(I - A)^{-}1$ is further proportional to the Bonacich-Katz centrality $(I - \beta A)^{-1}$, which captures the firm's importance as a supplier of intermediate inputs in the network (Acemoglu et al., 2012; Sargent and Stachurski, 2022).

similar approach, except that the elements of the adjacency matrix now correspond to the inputs i purchases from j divided by i's total intermediate inputs (Dhyne et al., 2021).

4.3.2 Firm characteristics and links to international supply chains

Five percent of the average firm's sales are ultimately absorbed by foreign demand, while eight percent of its inputs were sourced abroad in 2018. In the following, we distinguish[21] among:

(i) firms with high direct exposure to exports and imports,

- i.e. $r_{iX}^{direct} \geq 0.25$ and $r_{iM}^{direct} \geq 0.25$

(ii) firms with high total exposure (direct plus indirect),

- i.e. $r_{iX}^{total} \geq 0.25$ and $r_{iM}^{total} \geq 0.25$, but $r_{iX}^{direct}, r_{iM}^{direct} < 0.25$

(iii) firms that export (import) less than 25 percent.

- i.e. $r_{iX}^{total} < 0.25$ and $r_{iM}^{total} < 0.25$

Key firm characteristics of the firms in each group are summarised in Table 4.1. While as little as nine percent of all firms are highly exposed to import markets, these firms capture more than half of the total sales volume reported in 2018. Notably, the 657 firms with high indirect imports account for 28 percent of the total sales across all firms in that year. Likewise, on the downstream side, only five percent of the firms are highly exposed to export markets but account for 13 percent of total sales. In Belgium, 10 percent of firms - twice as many as in Kenya - directly or indirectly export 25+% of their output (Dhyne and Duprez, 2015). Importantly, Kenyan firms with strong linkages to international supply chains also employ substantially more workers on average and tend to be older.

[21]The 25% cut-off aligns with the one used in Dhyne and Duprez (2015) when discussing the exposure of Belgian firms to export markets. For exports, the chosen cut-off at 25% corresponds to total export exposure r_{iX}^{total} of firms in the 95th percentile. For imports (r_{iM}^{total}), it is close to the 90th percentile.

Table 4.1: Summary statistics for firms directly and indirectly linked to international supply chains in 2018

Exports

export group	# of firms	% of sales	Age	Sales		Employment		No. buyers	No. suppliers	% sales to HH
			avg	avg	sd	avg	sd	avg	avg	avg
No/low exports	46,193	87	9	115	1,773	15	149	28	25	40
25%+ total exports	1,456	2	11	98	438	25	120	24	32	1
25%+ direct exports	863	11	15	717	2,122	147	636	19	60	28

Imports

import group	# of firms	% of sales	Age	Sales		Employment		No. buyers	No. suppliers	% sales to HH
			avg	avg	sd	avg	sd	avg	avg	avg
No/low imports	44,002	43	9	60	368	12	124	16	23	40
25%+ total imports	657	28	18	2,497	11,816	277	966	354	118	3
25%+ direct imports	3,853	29	11	436	3,311	34	131	94	45	26

The table groups firms by their exposure to international supply chains. Total export and total import exposure are the sum of both direct and indirect exposure. The category for 25%+ total exposure excludes any firms that are part of the 25%+ direct exposure group. The number of buyers and suppliers refers to domestic suppliers and buyers only. % sales to HH captures the share of domestic sales outside the network - mostly to consumers.

Primary sectors dominate Kenya's exposure to export markets. Of the large formal agricultural firms captured in the administrative data, over 70 percent of their sales are destined for export markets, mostly through direct exports, but in some cases also indirectly (see top panel of Figure 4.3). Close to half of the output of mining and quarrying firms is exported. This is followed by 20 percent of each manufacturing output and the transportation sector. Jointly those two sectors capture a relatively larger share of the formal economy (see Table 4.A1), while sourcing most of their inputs from international markets (see bottom of Figure 4.3). Figure 4.3 further highlights that exposure to import markets is less concentrated in specific sectors and that indirect linkages to import markets play a much bigger role for upstream supply chains than exports do for downstream supply chains.

Figure 4.3: Output share of firms by strength of links to international supply chains

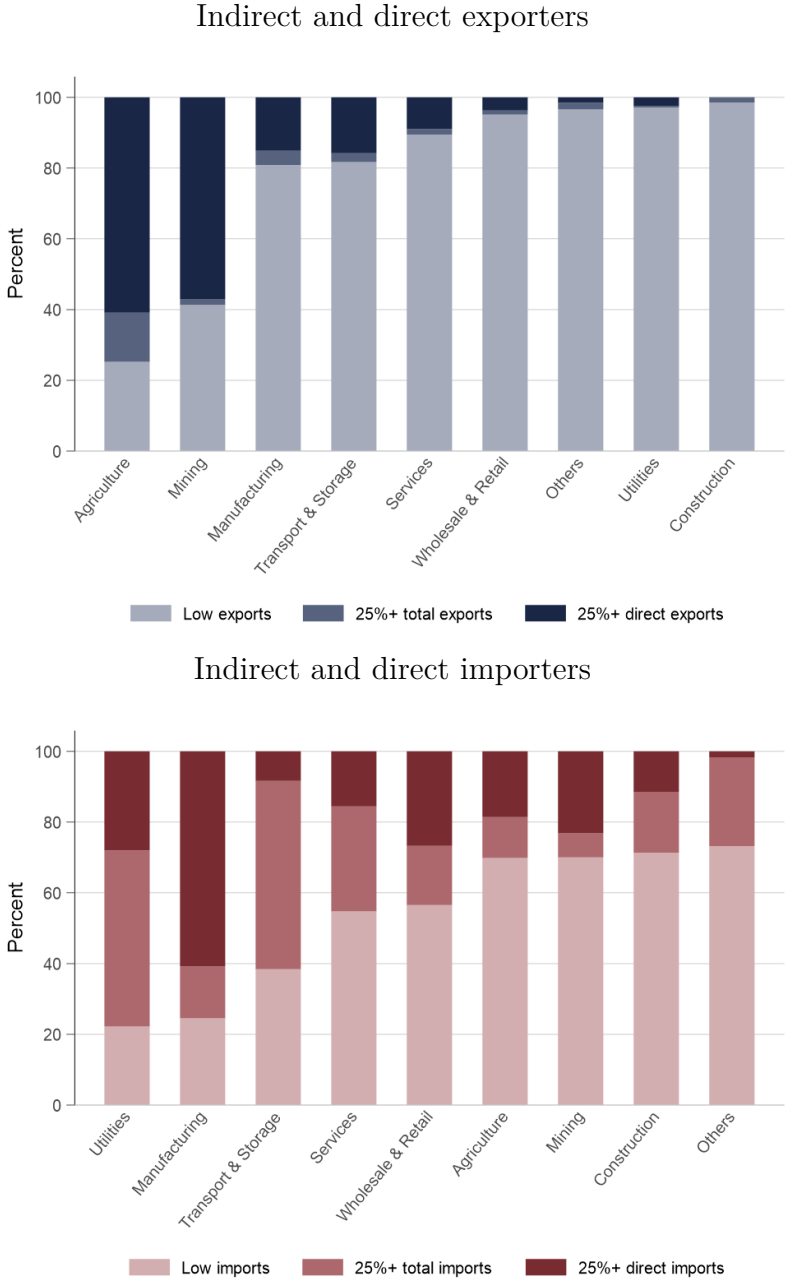


Figure 4.3 categorises firms by their degree of exposure to international supply chains and plots their sales (exports)/input (imports) share by sector in 2018. Total export (import) exposure is the sum of both direct and indirect exposure to exports (imports). The category for 25%+ total exposure excludes any firms that are part of the 25%+ direct exposure group.

4.3.3 The trajectory of firms with strong linkages to international supply chains during the COVID-19 crisis

Next, we examine the trajectory of firms with different degrees of exposure to international supply chains during the crisis. We start with the simple categorisation of firms into importers and exporters[22] and plot the evolution of their sales relative to those with no direct prior links to international supply chains in Figure 4.4a.[23] We record three main findings: First, the downturn for direct importers starts in February 2020; at the height of the lockdowns in China. Second, sales of direct importers drop by a similar margin compared to sales of firms without any direct links to international supply chains. They initially recover more swiftly but stagnate in the second half of 2020. Third, at the lowest point of their trajectory in May 2020, sales of exporters drop by only 47 percent relative to January 2020 and bounce back quickly.

In Figures 4.4b (exports) and 4.4c (imports), we now draw on the more nuanced measure to distinguish between firms that face a high direct versus high indirect exposure to international supply chains, respectively. For the case of exports, direct exporters, which specialise in international markets, hit their lowest point in May.[24] By July, these firms reach their pre-pandemic level of sales. However, firms with indirect links to export markets only reach their January sales level by the end of 2020. A dominating factor for these differential trends is the sector in which direct and indirect exporters operate.[25] Direct exports are dominated by firms exporting

[22]We classify a firm as an importer (exporter) if more than 10% of their inputs (sales) are directly sourced from (sold) abroad.

[23]We again rely on the same event study specification as above including firm- and sector-month-fixed effects as well as normalising the coefficient for January 2020 to zero.

[24]May coincides with Mother's Day in many European export destinations for Kenya's flower exporters.

[25]Other facts that could contribute to this trajectory are inventory holdings by exporters or a time lag due to already placed international orders or international shipping times. Unfortunately, we do not observe monthly inventory levels of firms in the administrative records. If we consider inventory levels at the end of the previous fiscal year as an imperfect proxy, we find that firms with a high exposure to export markets tend to hold less inventory than their peers with little exposure. At the end of 2019, the value of inventory held by direct exporters is equivalent to 38 days of their average daily sales volume. For indirect exporters it is 51 days, and 66 days for all other firms. Even within their 4-digit sector exporters (direct and indirect) are disproportionately less likely to feature in the top quartile of inventory holdings. In Appendix 4.C.1 we explore the trajectory of firms with high inventory levels in more detail and compare it to our results on supply chain diversification in Section 4.4.

agricultural products, the demand for which was barely affected by the pandemic or even increased (that is, the demand for vegetable exports). Indirect exporters operate in a wider range of sectors, with a substantial proportion in transport and logistics as well as manufacturing.

Turning to indirect importers, we find a relatively small initial downturn of less than 20 percent on average but levels of sales that remain below those of January 2020 until the end of our data set, March 2021. The descriptive statistics in Section 4.3.2 highlight the older age profile and larger size of firms with high indirect exposure to import markets. These characteristics often coincide with greater firm resilience. In addition, larger firms often have a more diversified demand profile that helped them to buffer the initial shock (also see Section 4.4 below).

Direct importers, many of whom act as intermediaries, were hit the hardest by the initial downturn. Their sales started declining substantially from February 2020 onward, falling by 20 percent in March and 60 percent in April and May relative to the average for January 2020. While their downturn was even more pronounced than for the average firm without any or with only weak linkages to import markets, their initial recovery was swifter. However, their imports again stalled in the last quarter of 2020 and remained well below pre-pandemic levels until (at least) March 2021.[26] Domestic sales of these firms largely trail their imports, which evinces their role as intermediaries.[27] Section 4.3.4 shows that the pass-through of direct import shocks to firm sales is particularly high during both the crisis peak and the recovery. These results seemingly contradict those of Figure 4.2, which shows that imports start to rise above January 2020 levels by August for the average firm (as do import volumes overall[28]). However, this increase is driven less by firms that previously specialised in imports and more by firms that have not imported before, or only to a limited extent. Figure 4.B1 shows that firms that source disproportionately less of their inputs from import markets account for an increasingly larger (but still small) share of aggregate import volumes - a trend that pre-dates the pandemic.[29]

[26]We are unable to unpack the extent to which this is driven by global shipment cost dynamics and container shortages that were prevalent around the time.

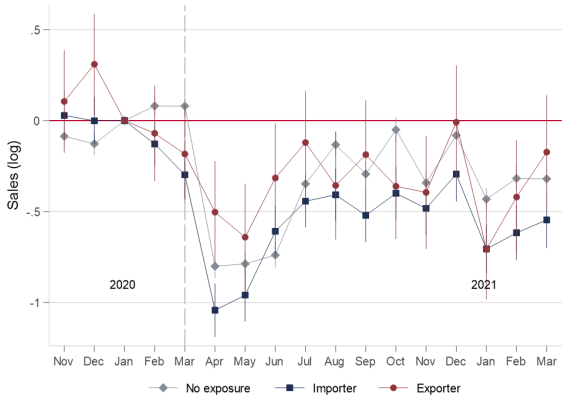
[27]Graph available on request.

[28]See https://pedl.cepr.org/publications/domestic-and-international-trade-flows-during-covid-19-pandemic-evidence-kenya%E2%80%99s

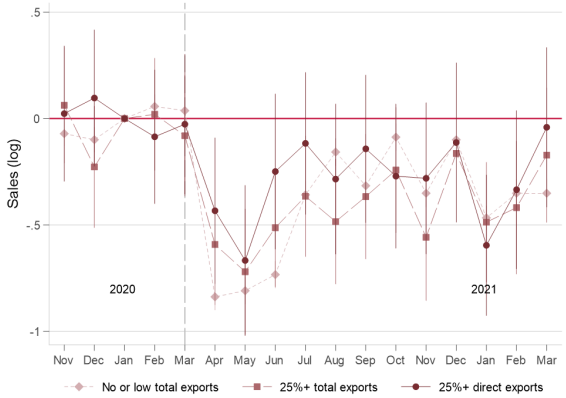
[29]This trend is also reflected in the time series of aggregate import volumes rather than shares attributed to each group of firms.

Figure 4.4: Firm-level exposure to international supply chains and firm sales during the crisis and recovery

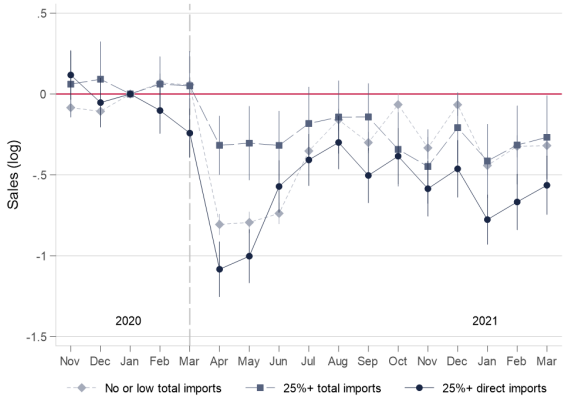
a) Importers and exporters



b) Indirect and direct exporters



c) Indirect and direct importers



The graph shows the results from estimating Equation (4.1), i.e. the regression of firm-level outcomes on a series of monthly time dummies and including firm fixed effects. It clusters standard errors at the firm level. The error bars show the 95-percent confidence intervals. The regressions for the COVID-19 period include data between November 2019 and March 2021.

4.3.4 Pass-through of shocks to import and export markets and domestic supply chains

So far we have focused on differences in firm dynamics based on their pre-COVID links to supply chains. To get a better understanding of the response of domestic supply chains to international shocks, we further exploit the fact that the COVID-19 shock was nowhere near universal. More specifically, we exploit the heterogeneity in supply and demand to and from worldwide markets across countries and product categories.

Empirical strategy

We deploy a shift-share design approach popularised by Hummels et al. (2014) and subsequently applied in settings similar to ours, such as Huneeus (2018); Dhyne et al. (2021). The measure consists of two components. The first is the shares $\omega_{i,c,p,t-12}$,[30] which capture a firm i's exports of product p to country c as a share of the firm's total revenues of the last 12 months at time $t-12$.[31] The second is, the change in demand WID for (supply WES of) product p by country c from all other countries, excluding Kenya.[32] In a nutshell, firm-level changes in foreign demand FD and foreign supply FS are defined as:

$$\Delta \log FD_{it} = \sum_{c,p} \omega_{i,c,p,t-12} \Delta \log \ WID_{c,p,t}, and$$

$$\Delta \log FS_{it} = \sum_{c,p} \omega_{i,c,p,t-12} \Delta \log \ WES_{c,p,t}.$$

Crucially, here we do not assume exogeneity of the exposure to international shocks (that is, the shares)[33] but instead seek to exploit variation in changes in demand and supply at the country-product level.[34] In our setting, we observe 12,244 unique

[30] For the computation of the shares we hold the country-product shares fixed in the same month of the previous year, i.e., for April 2020 we compute the shares based on import and export transactions between May 2018 and April 2019.

[31] Or the share of total inputs (intermediate inputs plus payroll) in the case of imports.

[32] The data to construct the shock measure was retrieved from the UN Comtrade database. Products p are defined by 4-digit HS codes (see Section 4.2.1). https://comtrade.un.org/Data/bulk

[33] A concern is that characteristics of firms selecting into certain country-product markets - e.g., better access to credit or risk management capacities (Sheffi, 2015) - correlate with their ability to respond to negative shocks.

[34] Drivers of such variation in the demand for or supply of products by specific countries can, for example, be policy measures and changes in behaviour (such as lockdowns or working from home) during the COVID-19 pandemic.

country x product pairs for exports and 29,815 pairs for imports. To compute the shocks, we consider the month-over-month changes between the current and the same months in the previous year (for example, April 2020 versus April 2019). The average export and import shocks are 2 percent and 3 percent, respectively. We summarise key properties of the underlying shares and shocks in Appendix Table 4.B1.

We follow Dhyne et al. (2021)'s estimation strategy to simultaneously estimate the pass-through of both direct and indirect import and export shocks. However, being able to rely on more granular monthly data instead of annual data, we define the relevant adjacency matrix as the network observed in the same month of the previous year, taking into account all firm-to-firm relations in the 12 months leading up to it (that is, to compute the indirect exposure for April 2020, we consider all firm-to-firm relationships between May 2018 and April 2019). While this approach is computationally intensive and involves inverting $(I - A)$ for every single month in the data set, it is crucial for holding the predictive power of the foreign demand and supply measure constant at any point. The full exposure to export shocks becomes:[35]

$$Z^X = (I - A)^{-1} r_F \Delta \log FD$$

At the firm-level, we split Z^X_{it} into its direct and indirect components and ultimately estimate the following specification:

$$\Delta y_{it} = \beta_1 Z^{X_{direct}}_{it} + \beta_2 Z^{X_{indirect}}_{it} + \gamma_1 Z^{M_{direct}}_{it} + \gamma_2 Z^{M_{indirect}}_{it} + \eta_s + \phi_{sy} + \varepsilon_{it} \qquad (4.2)$$

y_{it} denotes the change in (log) firm-level outcomes of firm i in month t relative to the same month in the previous year. We further controlled for two-digit sector-date-fixed effects. The first-difference approach is crucial to isolate the shock variation from the variation in exposure shares in the case of panel data with time variation in the exposure to shocks (Borusyak et al., 2022). While the month-over-month first difference introduces more noise on the left-hand side, the shock measure gains in predictive power. We control for firm age to address the fact that younger firms typically experience more pronounced growth patterns. Finally, we control for the contemporaneous sum of shares and their interaction with the control variable (log) firm age, a crucial addition for contexts where the shares do not sum to 1 (Borusyak

[35] Following the notation introduced in Section 4.3.1.

et al., 2022). In our case, the sum is not 1 because we do not capture imports and exports of all relevant countries in the UN Comtrade database and miss out on product-level trade with EAC member countries in the customs data. Relying on the 12-month difference for firm outcomes comes with the caveat of losing observations at the extensive margin where either the current or lagged observation is zero. This issue is particularly prevalent for the count of buyers and suppliers, where using log+1 as the outcome greatly distorts the results. We, therefore, restrict the sample to firms which have at least one supplier in both periods in order to hold the sample of firms constant. Only when looking at the number of buyers and employment will the sample size again drop due to fewer firms also having domestic buyers and formal employees.

Results

The results show that international supply and demand shocks not only impact firm-level exports and imports but are also passed on through domestic supply chain linkages. Moreover, they are associated with changes in the domestic production network (see Table 4.2). Unsurprisingly, direct import and export shocks have the strongest effects on firm-level imports and exports. A 10 percent increase in world demand results in a 3.7 percent increase of exports by Kenyan firms. A shock of the same magnitude to the world supply of products usually purchased by Kenyan firms increases imports by 8.6 percent. Note that most export and import shocks we observe in the data are much smaller in magnitude than a 10 percent shift. Only firms in the 1st and 99th percentiles experienced shocks of a magnitude well above 10 percent during our sample period. During the peak of the COVID-19 crisis, some firms saw a decline in demand for their products by as much as 82 percent, while other firms experienced a 76 percent boost. We, therefore, re-estimate Equation 4.2 interacting each of the shock measures with two dummy variables for the crisis peak and the recovery phase (see Figure 4.B2). It shows that the strong pass-through of import and export shocks is primarily driven by the large shocks that hit firms during the COVID-19 crisis.

Nevertheless, we observe a pass-through of international shocks to domestic supply chains in the pre-COVID-19 period as well. Domestic purchases of importers increase.

Firms exposed to indirect import shocks through their suppliers export more but also concentrate their inputs on fewer suppliers while keeping overall purchase levels unchanged. The latter dynamic is reversed during the recovery period from the crisis. Here, indirect import shocks result in more domestic supplier relationships. Further, indirect import shocks have a strong positive effect on domestic sales, purchases, and the number of buyers in the second half of 2020 and the first quarter of 2021. This suggests that the bouncing back of international supply chains very much supported the recovery trajectory of Kenyan firms.

Table 4.2: Firm response to international supply and demand shocks

	Imports	Exports	Domestic sales	Domestic purch.	No. of suppliers	No. of buyers	No. employees
export direct	0.266*	0.387*	0.154	0.053*	0.035*	-0.074**	0.008
	(0.147)	(0.205)	(0.138)	(0.030)	(0.019)	(0.030)	(0.019)
export indirect	0.007	0.037	0.735*	0.177	0.125*	0.047	0.006
	(0.250)	(0.227)	(0.429)	(0.132)	(0.068)	(0.056)	(0.073)
import direct	0.870***	0.002	0.248**	0.105***	0.034**	0.055***	0.023*
	(0.162)	(0.074)	(0.105)	(0.029)	(0.014)	(0.017)	(0.012)
import indirect	0.083	0.278***	-0.166***	-0.001	-0.021***	-0.001	0.002
	(0.113)	(0.087)	(0.039)	(0.017)	(0.008)	(0.010)	(0.011)
No. of observations	950,713	950,713	950,713	950,713	950,713	594,342	607,403
No. of firms	40243	40243	40243	40243	40243	27597	23711
R2	0.076	0.107	0.150	0.157	0.197	0.256	0.361
Firm FE	✓	✓	✓	✓	✓	✓	✓
Sector-date FE	✓	✓	✓	✓	✓	✓	✓

The results present the estimation of Equation 4.2 regressing the respective shock measures on the change in logged firm-level outcomes relative to the same month in the previous year as well as firm- and sector-date-fixed effects. Standard errors are clustered at the firm level. We further control for the sum of shares and firm age (log) weighted by the sum of shares. The regressions include data from April 2018 to March 2021.

A positive export shock increases upstream participation in international supply chains, leading to more imports. On average, the increase in imports does not crowd out domestic inputs. Rather, there is a crowding-in effect, as firms indirectly exposed to export shocks increase their domestic sales. These firms, in turn, increase their domestic purchases and the number of suppliers. The finding supports the interpretation that shocks to foreign demand pass through to domestic suppliers. During the peak of the COVID-19 crisis, firms exposed to less negative direct or indirect export shocks see less of a decline in their domestic purchase volume and the number of suppliers. The effect size is, however, quite small for direct export shocks, with a 10 percent increase in foreign demand leading to an increase in the number of domestic suppliers of only 0.5 percent. The estimated effect is more substantial for firms

exposed to indirect export shocks, where a 10 percent shock results in a 4 percent increase of links to suppliers.
While we do not detect a statistically significant response of domestic sales to foreign demand shocks, the number of domestic buyers declines as a result of an increase in foreign demand. We observe a similar pattern for the case of indirect import shocks, where firms start to export more but sell less domestically as a result of a positive import shock. Jointly, these results suggest – a perhaps surprising – substitution effect between sales to international supply chains and domestic supply chains, at least for the average firm.

Overall, we find that both the downstream substitution and upstream complementary patterns between foreign and domestic markets predate the pandemic. At the same time, however, the direct pass-through of both import and export shocks to major outcomes is particularly pronounced during the peak of the COVID-19 crisis. So, while domestic supply chains do benefit from favourable conditions in international markets, it also implies that they are also vulnerable to the attached downside risk.

4.4 Network position and firm resilience

Finally, we take a closer look at the relevance of a firm's domestic network position for its trajectory during the COVID-19 crisis. Here we consider diversification of domestic up- and downstream supply chains on both the extensive and the intensive margin and the number of up- and downstream sectors as a proxy for the complexity of the supply chain. We find that downstream diversification is beneficial during times of severe economic downturn. Meanwhile, upstream diversification of suppliers improves firm recovery as demand picks up, but stories of supply chain backlogs start to make headlines worldwide.

4.4.1 Measuring diversification and complexity of domestic supply chains

In the following, we again fix the firm's network in the same month of the previous year and consider all suppliers (or buyers) the firm has interacted with during the past 12 months (rolling network).

Measuring diversification

Diversification can take place along both the extensive and the intensive margins. On the extensive margin, we look at the number of suppliers (buyers) the firm has within each 4-digit sector it purchases from (sells to). For firms purchasing (or selling to) more than one 4-digit sector, we consider the weighted average number of suppliers (buyers). To account for differences across sectors, age cohorts and firm size, we first group firms based on their 2-digit sector, age cohort[36] and sales quartile.[37] We then define a firm as highly diversified if the number of suppliers (or buyers) within the sector-cohort-size bin lies above its peers in the same 4-digit sector. Even within each sector-age-sales bin, we still see sizeable differences between firms that are highly diversified on the extensive margin versus those with little diversification (see Table 4.C2 for downstream diversification and Table 4.C1 for upstream supply chains). Firms with higher downstream diversification, for example, have more employees, are more likely to be direct importers, and have larger sales. Many of the noted differences, such as firm age, are statistically significant but not sizeable. For firms with a high degree of upstream diversification, the sales and age gap to less diversified firms is even smaller, while the gap is similar for purchases (about 5 million KES in each case). While the results below are robust to alternatively grouping firms into sales tertiles or deciles, we cannot fully rule out any size effects that correlate with our measure of diversification to drive some of the results.

On the intensive margin, diversification essentially corresponds to the concentration of sales and purchases. We use the Herfindahl–Hirschman Index (HHI), which can take a value between 0 (least concentrated) and 1 (highly concentrated).

Measuring complexity

The delay or failure of a small input can often result in the break-down of an entire product (Kremer, 1993). Thus, supply chains with a large number of inputs might be particularly fragile during times of crisis when the probability of delays or shortages increases steeply. We look at the number of 4-digit sectors a firm sources from (sells

[36]The respective cohorts are < 5 years, 5-10 years, 11-20 years and 20+ years.

[37]Conditioning on 4-digit sectors would come with the caveat of having too few firms in some of the age-cohort bins.

to) as a proxy for the complexity of its supply chain. Firms sourcing from an above-median number of industries within their 4-digit sector are categorised as having complex supply chains.

Empirical strategy

With those measures at hand, we estimate an adapted version of the event-study specification in Section 4.2.3. In our preferred specification we pool the month dummies into pre-COVID (November 2019 to March 2020), COVID peak (April to June 2020), and recovery phase (July 2020 to March 2021) and interact them with the respective diversification and complexity measure N_{it}:

$$\begin{aligned} y_{it} = \quad & \beta_1 N_{it} + \beta_2 C19_{pre} + \beta_3 C19_{pre} * N_{it} + \beta_4 C19_{peak} + \beta_5 C19_{peak} * N_{it} \\ & + \beta_6 C19_{recov} + \beta_7 C19_{recov} * N_{it} + \beta_8 X_i + \delta_{st} + \phi_i + \varepsilon_{it} \end{aligned} \tag{4.3}$$

y_{it} denotes firm-level outcomes of firm i in month t in log (+1) terms. We further control for 2-digit sector-month FE (δ_{st}) and firm FE (ϕ_i). To ensure the firm's network position is not primarily driven by its proximity to final demand and/or international markets, we controlled for whether the firm is an importer or an exporter and for its shares of sales to domestic entities outside the network. We further account for firm age (log) and the average N_{it} for the firm's 4-digit sector to capture sector-level shocks - for example, due to the exit of a key supplier.

β_1 (for N_{it}) is identified by firms moving in and out of respective diversification and complexity categories over time. β_2, β_4, β_6 are identified by within-firm and calendar-month variation relative to the pre-COVID years 2015-2019. β_3, β_5 and β_7 rely on the same source of variation, but are interpreted as interaction terms. Figure 4.5 plots the respective point estimates and confidence intervals.

As an alternative specification, we again start out from the same event-study approach, but this time normalise the coefficients for each group of firms to January 2020. While the interpretation of this specification is more straight-forward and exactly mirrors the specification in Section 4.2.3, it relies heavily on the firm's relative performance in the anchor month January 2020 (see Figures 4.C3 for the results).

Finally, we are interested in the network position's relevance for the pass-through of shocks to import and export markets. Here we run the specification from Section

4.3.4 but interact the shock measures with the respective network position measure of choice N_{it}. The results are presented in Figure 4.C4.

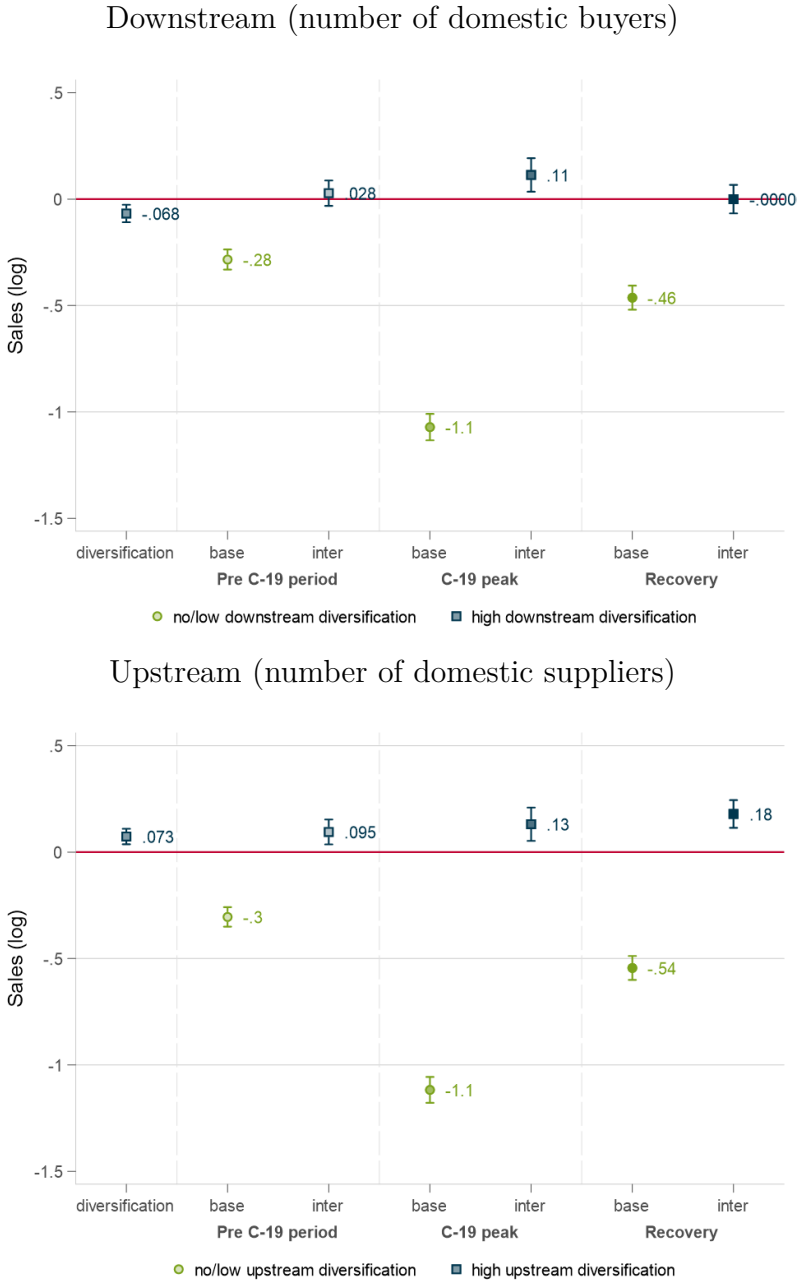
4.4.2 Results on resilience and the firm network position

Diversification

We find that firms with a greater degree of downstream diversification face a less drastic downturn during the peak of the crisis (top panel of Figure 4.5). Firms with below median diversification see their sales drop by an average of 67 percent during the peak months, while the sales of firms with a high degree of diversification drop by an average of 63 percent. Firms with more upstream diversification recover more strongly (bottom panel of Figure 4.5). The strong positive effects are primarily driven by the firm trajectory in the last quarter of 2020 and the first quarter of 2021 (see dashed blue line in the right top row panel of Figure 4.C3), when stories of supply chain backlogs started to make headlines worldwide. Average sales of firms with high upstream diversification decline by 30 percent relative to the pre-COVID period in comparison to a 42 percent decline for the less diversified firms.[38]

[38]In Appendix 4.C.1, we compare the size of the effects of the diversification of upstream supply chains to the trajectory of firms with high levels of inventory. Both are very similar in size. The estimated difference in firm-level outcomes for firms with high pre-crisis levels of inventory is, further, very close to the estimates in Lafrogne-Joussier et al. (2023). Lafrogne-Joussier et al. (2023) focus on export volumes of French exporters as they faced supply issues during the initial lockdown in China in 2020.

Figure 4.5: Supply chain diversification and firm sales during the crisis and recovery phase



The above graphs present the results from estimating Equation 4.3. We control for firm age, share of sales to final domestic demand, share of highly diversified firms in the firm's own 4-digit sector, and import and export status. The pre-COVID-19 (C-19) phase corresponds to November 2019 to March 2020, the COVID-19 peak is April to June 2020, and the recovery July 2020 to March 2021.

Both estimates hinge on the assumption that the outcomes of more and less diversified firms followed a parallel trend prior to the COVID-19 crisis. We do not

see any meaningful diverging pre-trends in the event study plots of Figure 4.C3 for either upstream and downstream diversification, and we cannot reject the null that the coefficient for downstream diversified firms is zero. However, when we estimate specification 4.3 for firms with a high degree of upstream diversification, the pre-COVID coefficient is significantly different from zero at the 5-percent level. While the effect size is almost double during the recovery phase and quite substantial in economic terms, we cannot entirely rule out the possibility of differential pre-trends driving the upstream results.

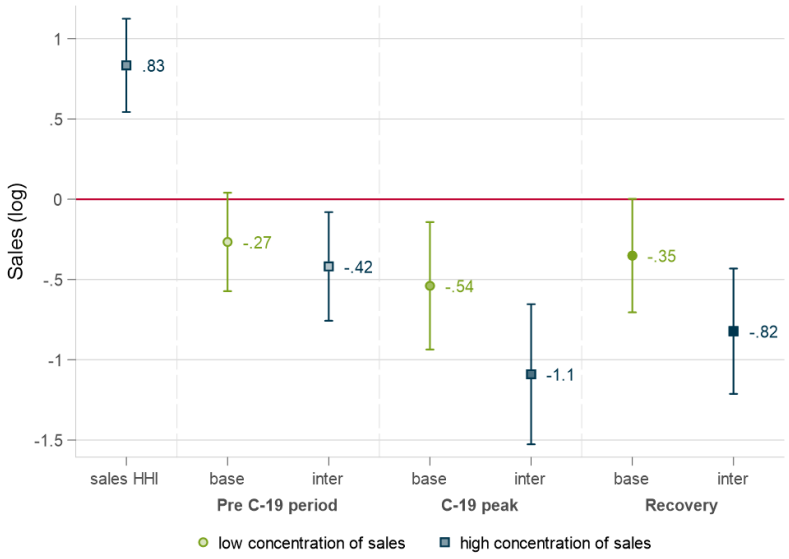
Figure 4.6 plots the results for diversification on the intensive margin. While we again struggle to rule out differential pre-trends, the results largely mirror those for diversification on the extensive margin. Conditional on the number of buyers, firms with a larger concentration of sales among their buyers faced a more drastic decline in their sales during the crisis peak.

Finally, we interacted our proposed measures for supply chain diversification with the firm's direct and indirect exposure to international shocks (see Section 4.3.4). Unsurprisingly, the response of firm sales to international shocks is more pronounced for less diversified firms than for their highly diversified peers.[39] Assuming a linear relationship, this implies that firms can benefit more from positive shocks but are also more vulnerable to negative ones. Note that the proposed specification, following Borusyak et al. (2022), controls for the firm's total exposure (the shares) to the shock at any given point (see Section 4.3.4). However, we are unable to provide clear insights as to whether the results are primarily driven by (i) a change in the composition of the sources of exposure (a firm's total import share might remain constant, but now purchases it from two suppliers instead of one) or (ii) the firm's ability to respond better to the shock as a result of having a more diversified network.

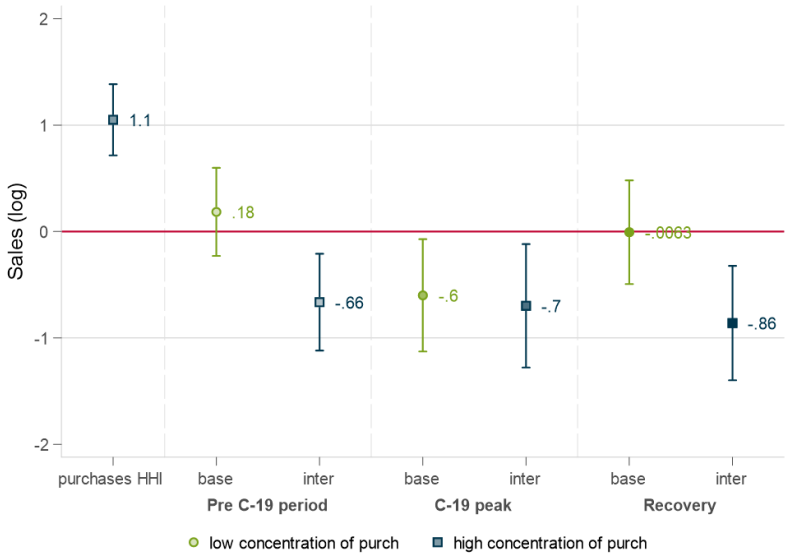
[39]Unfortunately, we lack the power to detect meaningful effects when interact all of the shock measures and network measures with COVID-period dummies.

Figure 4.6: Concentration of sales and purchases (HHI) and firm sales during the crisis and recovery phase

Domestic sales



Domestic purchases



The above graphs present the results from estimating Equation 4.3. We controlled for firm age, share of sales to final domestic demand, the average concentration of sales/purchases in the firm's own 4-digit sector, and import and export status. The pre-C-19 phase corresponds to the period from November 2019 to March 2020, the C-19 peak is from April to June 2020, and the recovery is from July 2020 to March 2021. We further control for diversification along the extensive margin.

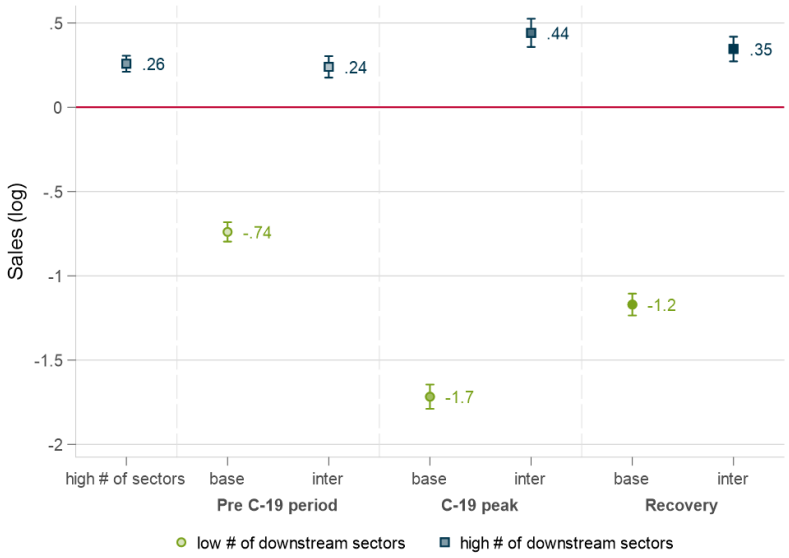
Number of upstream and downstream sectors (complexity)

Looking at the number of sectors the firm purchases from or supplies to, the coefficients plotted in Figure 4.7 suggest a very similar story for complexity relative to diversification. Firms with above-median complexity relative to their peers in the same 4-digit sector fare better during the pandemic. Crucially, however, they are already on a favourable trend in the pre-pandemic period. Therefore, the results suggest that the observed patterns are, if anything, indicative of a complex supplier network serving as a proxy for a number of relevant firm characteristics, such as high productivity and managerial capital.[40] To test for the robustness of the results, we again draw on the event-study specification and normalise all coefficients to January 2020. Using this approach the results suggest a parallel pre-trend for firms with above and below median supply chain complexity. The bottom left panel of Figure 4.C3 suggests that, especially during the peak of the crisis, firms selling to a larger number of 4-digit sectors face a smaller loss in their sales. One potential explanation is that selling to more downstream sectors serves as a form of diversification, which helps firms smooth the shock in times of a severe demand shock. While the results leave room for different interpretations, they also highlight the need to test the sensitivity of the results using different base periods.

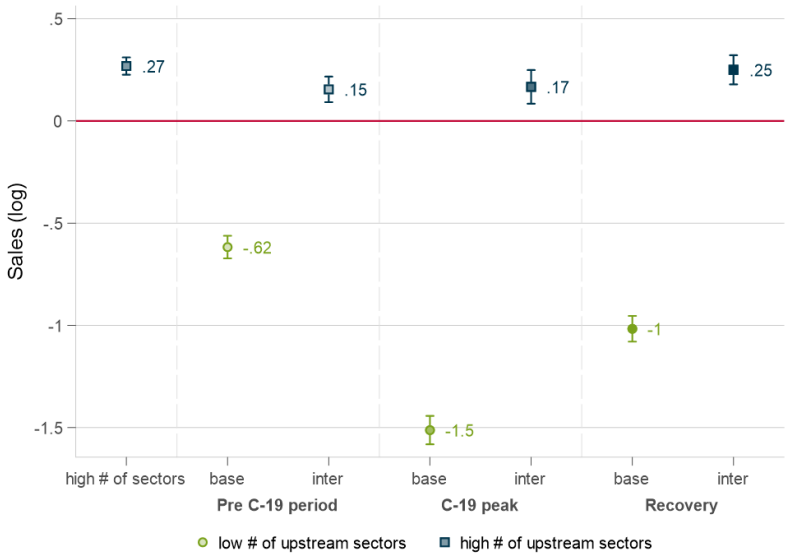
[40] Sheffi and Rice Jr (2005); Sheffi (2015) argue that more complex supply chains require firms to pay more attention to vulnerabilities and strategically invest in enhanced risk management capabilities. The results could thus reflect better preparedness of firms with more complex supply chains.

Figure 4.7: Supply chain complexity and resilience

Downstream (number of downstream 4-digit sectors)



Upstream (number of upstream 4-digit sectors)



The above graphs present the results from estimating Equation 4.3. We controlled for firm age, share of sales to final domestic demand, share of firms with above-median supply chain complexity in the firm's own 4-digit sector, and import and export status. The pre-C-19 phase corresponds to November 2019 to March 2020, the C-19 peak is within the period from April to June 2020, and the recovery is from July 2020 to March 2021.

4.5 Conclusion

We combine customs records with granular data on Kenya's domestic network of formal firms to map not only direct but, importantly, also indirect links of firms to international supply chains through domestic suppliers. Our findings reveal that only nine percent of Kenya's formal firms source a significant amount of their inputs from abroad, and only five percent sell more than a quarter to international markets - either directly or indirectly through the domestic network. However, these firms are crucial for the economy, both in terms of their turnover and employment and their connections to domestic suppliers and buyers. This finding mitigates a widespread concern that firms with links to international supply chains - or more specifically, global value chains - in lower income economies remain largely separate from domestic supply chains (Amendolagine et al., 2019; Gereffi et al., 2005). Our analysis shows that shocks to the demand and supply of goods in international markets are passed on to domestic trade partners. While favourable import shocks helped firms recover from the COVID-19 crisis more swiftly, negative shocks during the peak of the crisis had adverse effects. This highlights the fact that links to international markets come with both opportunities and vulnerabilities. Moreover, our results suggest that international trade shocks can change the composition of the domestic supply chain, with firms responding to positive import shocks by concentrating their purchases on fewer domestic suppliers in pre-COVID times. Conversely, a stronger bounce back in the world supply of imported goods during the COVID-19 crisis led to a larger number of domestic suppliers being crowded in. Our analysis focuses on the short-run pass-through. A natural extension would be to look at medium- and long-run dynamics of the participation of firms in international supply chains and how these shape their domestic network.

Acknowledgements for Chapter 4 We thank the Kenya Revenue Authority (KRA) for their collaboration. Romeo Ekirapa, Simon Mwangi, and Benard Sang provided excellent technical support and advice. A related project on tax incentives conducted jointly by the authors for KRA with assistance from the data support team led to this work. We thank Daniela Villacreces Villacis for her excellent support in processing, cleaning, and benchmarking the data. We further thank seminar and conference participants at the CSAE Research Workshop, the CSAE Annual Conference, the Oxford Applied Micro Workshop, and the Oxford Firms Discussion Group, the STEG Annual Conference, the World Bank ABCDE Conference, the WCP Virtual International Conference, the Research Conference on Sustainability in Global Value Chains, and the Jeune Street Seminar Series for comments and feedback. We received valuable guidance from Chris Woodruff and have benefited greatly from discussions with and feedback from Vatsal Khandelwal, Hannah Zillessen, Luke Heath Milsom, Sanghamitra Warrier Mukherjee, and Raphael Bradenbrink. We gratefully acknowledge financial support from the Private Enterprise Development in Low Income Countries (PEDL) programme, a joint initiative by the Centre for Economic Policy Research (CEPR) and the UK Foreign, Commonwealth & Development Office (FCDO). Verena Wiedemann further acknowledges funding from the Oxford Economic Papers Research Fund (OEP) and the German National Academic Scholarship Foundation. This study has been approved by the Department of Economics Research Ethics Committee at Oxford (protocol no.ECONCIA20-21-23), and the Kenyan National Commission for Science, Technology and Innovation (protocol no.NACOSTI/P/20/5923). The views in this paper are those of the authors, and do not necessarily represent those of the KRA or any other institution the authors are affiliated with.

Appendices of Chapter 4

4.A Additional material on data and context

Table 4.A1: Sector-level breakdown of total sales (2019)

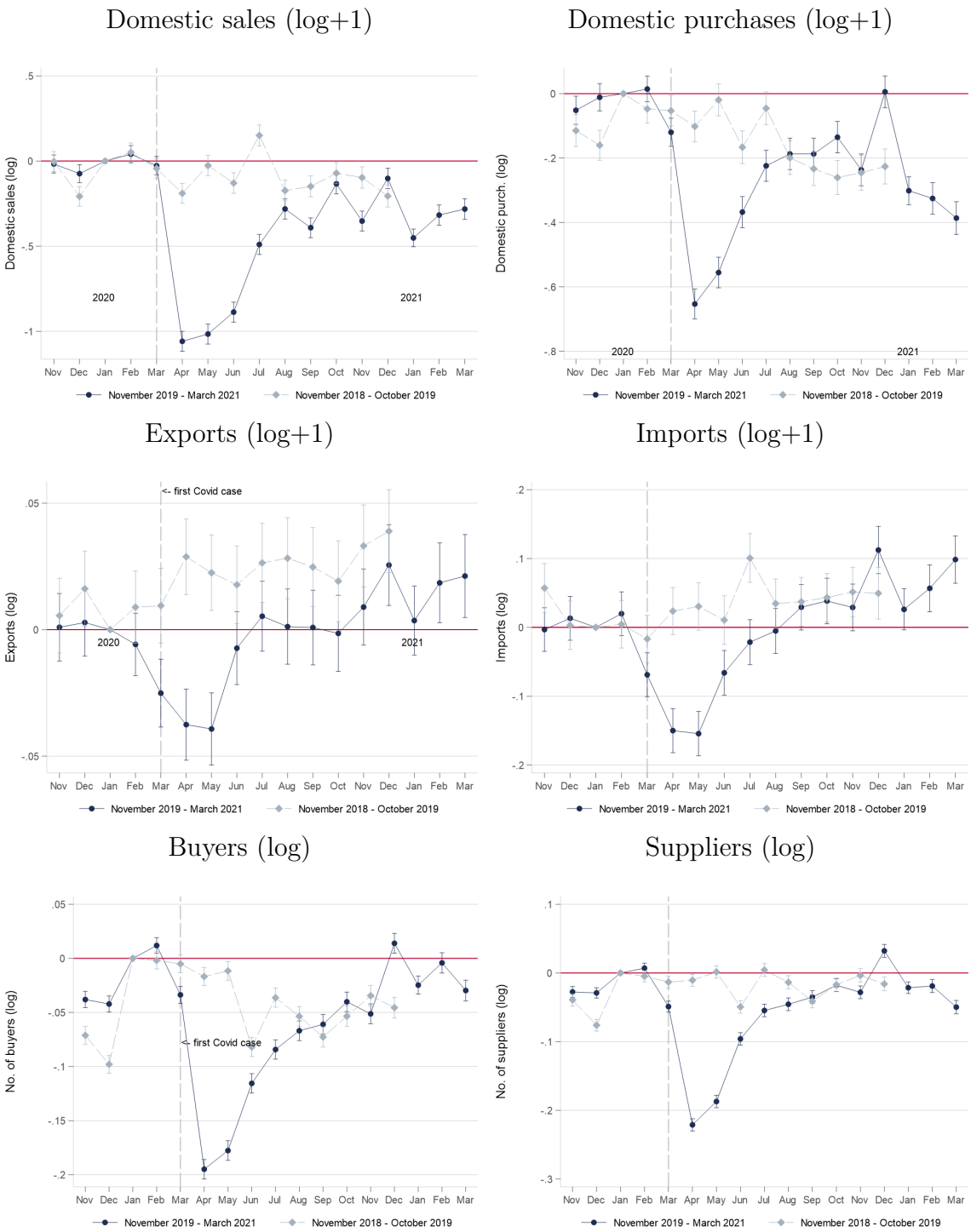
Sectors	%
Wholesale & Retail	22.88
Manufacturing	21.44
Electricity & Gas	16.37
Information & Communication	8.20
Transportation & Storage	8.00
Construction	7.07
Hospitality	2.93
Administrative & Support Services	2.32
Agriculture, Forestry, & Fishing	2.20
Professional, Scientific & Technical	1.95
Total	93.37

Table 4.A2: Summary statistics on inter-firm relationships between 2015 and 2020

Year	Firms	Total sales (in bn. KES)	Relationships	Firm-to-firm sales (in % of total sales)
2015	31,684	4,857.5	886,940	49.3
2016	36,920	5,881.1	1,134,159	49.5
2017	40,677	6,218.8	1,204,754	50.6
2018	44,997	6,594.0	1,332,150	49.0
2019	48,697	6,828.3	1,528,410	56.3
2020	49,955	6,651.1	1,528,109	60.0

This table shows the number of firms in the production network, firm-to-firm relationships, the aggregate sales volume, as well as the share of aggregate domestic sales that is linked to transactions between registered entities.

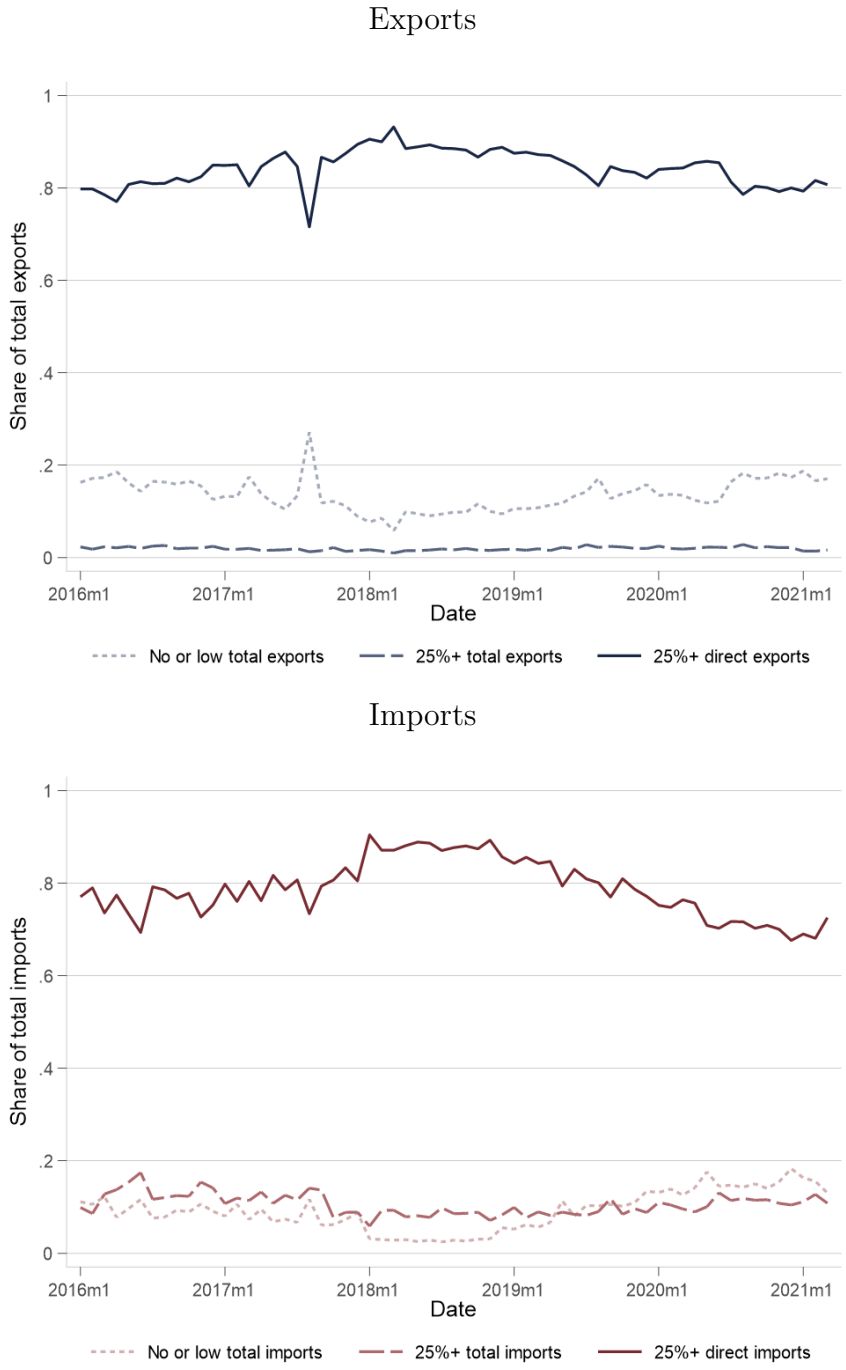
Figure 4.A1: Major firm-level outcomes in 2020 and 2019



The graph shows the results from estimating Equation (4.1), i.e. the regression of firm-level outcomes on a series of monthly time dummies. We further included firm- and month-fixed effects and normalised the coefficient for January 2020 to zero. Standard errors are clustered at the firm level, and the error bars show the 95-percent confidence intervals.

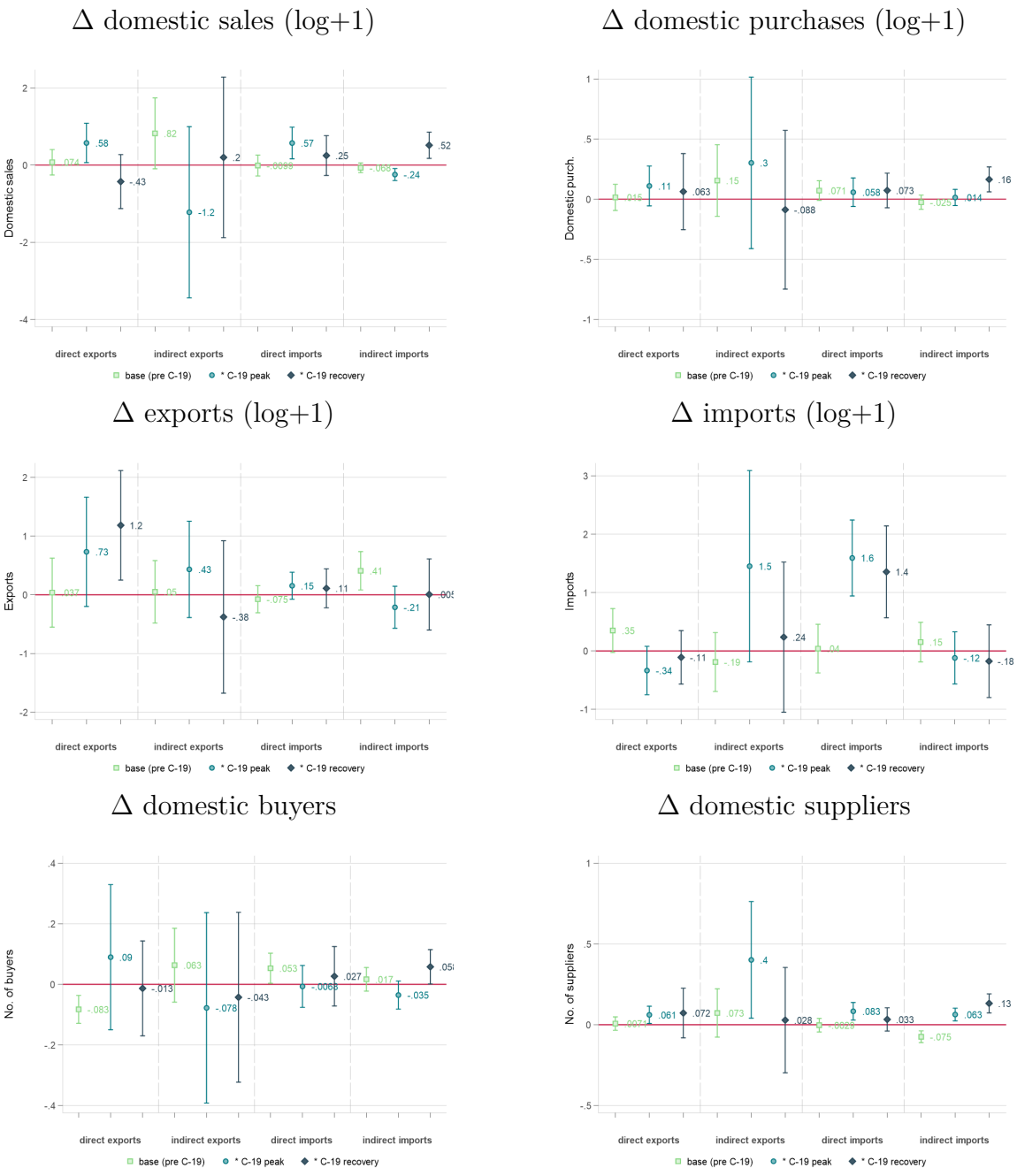
4.B Additional material on international supply chain linkages

Figure 4.B1: Import and export shares over time



The above graph plots the share of aggregate imports and exports that can be attributed to (i) firms with no or low exposure to international import/export markets, (ii) firms with high indirect exposure to international markets, and lastly (iii) firms with high levels of direct exports and imports specified in Section 4.3.2.

Figure 4.B2: Major firm-level outcomes during the COVID-19 crisis vs 2019 trends



The results present the estimation of Equation 4.2 regressing the respective shock measures on the change in logged firm-level outcomes relative to the same month in the previous year as well as firm- and sector-date-fixed effects. Here we further interact the shock measures with dummies that are equal to one for the dates April to June 2020 (C-19 peak) and July 2020 to March 2021 (C-19 recovery). The coefficients for the interaction terms thus need to be interpreted relative to the respective coefficients estimated for the pre-COVID period. Standard errors are clustered at the firm level. We further control for the sum of shares and firm age (log) weighted by the sum of shares. The regressions include data from April 2018 to March 2021.

4.B.1 Properties of world demand and supply shocks and country-product exposure shares ω_{cp}

To estimate the pass-through of international shocks to world demand and supply we rely on variation at the country-product-level. Firms in our sample export close to 1,000 products to 128 different countries and import close to 1,200 products from 130 countries. The asymptotic properties of the shift-share estimator hinge on a large enough sample size (Borusyak et al., 2022), which in a shift-share world implies (i) sufficient variation in world demand and supply shocks, and (ii) sufficiently dispersed exposure shares. Table 4.B1 summarises key properties of the empirical variation we observe for the shocks and shares during our sample period. The reported average, median and maximum shares ω_{pct} are averages across all firms within a country-product-month cell. We do not observe a high concentration in the exposure shares, which pins down our effective sample size (Borusyak et al., 2022). Unsurprisingly, the observed concentration is, however, much higher for exports than for imports with an effective sample size of 11,098 relative to 58,074. Turning to the shocks to world demand and supply, we find sufficient variation in the shock sizes. Import and export shocks are on average negative, which is driven by the large negative shocks we observe during the COVID-19 period. A concern is that these shocks are highly correlated due to the aggregate nature of pandemic-related shocks. Following Borusyak et al. (2022) we compute the intra-class correlation coefficient using a hierarchical decomposition of the within-month variation. We are primarily concerned about a high degree of within-country correlation of shocks and therefore follow a country to country-product hierarchy for the decomposition. The estimates of the model further yield the random effect for a time-invariant country-product component. We find low levels of clustering of the shock residuals at the country level.[41]

[41] As a robustness check we also look at 2020 only, but find only a moderate increase in clustering. The clustering for the time-invariant component doubles but is still sufficiently small.

Table 4.B1: Shocks and shares summary statistics

	Exports	Imports
Countries	128	130
Products	996	1,197
Country x products	12,244	29,815
Effective sample size (1/HHI of ω_{cpt})	11,098	58,074
Avg. ω_{cpt}	5.00e-06	1.51e-06
Median ω_{cpt}	1.65e-07	1.54e-07
Max ω_{cpt}	.000384	.0000975
Avg. Δ log world demand/supply	-.0205	-.0292
Standard deviation Δ log world demand/supply	1.07	.877
Interquartile range Δ log world demand/supply	.56	.448
Intra-country correlation	.0000533	.0000126
Intra-country-product correlation	.0626	.037

4.C Additional material on firm networks and resilience

4.C.1 The role of inventory levels

Similarly to diversification of a firm's supplier portfolio, inventory levels can serve as a buffer stock for firms to cope with supply chain shortages in the short-run. For a subgroup of firms, we are able to look at the relevance of pre-crisis levels and compare its relevance to the diversification of upstream supply chains.

Measurement

We distinguish between firms with a high inventory to sales ratio in the previous year versus firms with lower inventory levels. Information on inventory is not captured in the VAT data. We therefore merge the data with corporate income tax (CIT) files, which detail the value of the stock of inventory at the end of the firm's fiscal year. We classify firms as high inventory holders if their inventory to sales ratio falls into the top quarter within their 4-digit sector.[42] We exclude any 4-digit sector where

[42]In an alternative specification we use the median as a cutoff. Here we do not find any difference in the recovery of firms with above-median inventory levels relative to those with lower levels of inventory. We attribute this to the fact that inventory levels of firms that sit just above the median are not very different from firms in the lowest two quartiles for most sectors. Only those in the top quartile differ notably. Our cut-off choice is in line with Lafrogne-Joussier et al. (2023), who classify firms in the top quintile as high-inventory firms.

inventory holdings of firms in the top quartile is zero. Moreover, only a subset of VAT-paying businesses, those which are also corporate entities, file CIT returns.[43] Our sample size for this section therefore drops from over 48,000 firms to just over 30,000 firms. This subset of firms accounts for 60 percent of the annual VAT sales in 2018. Within the top quartile, the value of inventory held by the median firm is equivalent to 60 days of their average daily sales volume (in non-COVID years).[44]

Results

Firms with high levels of inventory at the end of the previous fiscal year recover more strongly from the shock. During the recovery period their monthly sales are on average 8.6 percentage points higher than for firms with lower levels of inventory in the same sector.[45] The event-study plot in Figure 4.C3 (bottom graph) suggests that this effect can primarily be attributed to a better performance of high-inventory firms in the third quarter of 2020, just after the initial downturn. However, high-inventory firms do not seem to start performing worse once their pre-crisis stock is potentially depleted, which is to be expected a few months into the crisis. A possible interpretation is that high end-of-year inventory levels not only represent a snapshot in time, but also correlate with the availability of medium- to long-term storage facilities that allow for the holding of buffer stocks. Further, we only look at firm sales as a well-measured outcome rather than productivity or measures that are able to capture potential downsides of inefficiently high inventory levels (Yao et al., 2021). Our estimated difference in trajectories for high-inventory firms is similar to the findings in Lafrogne-Joussier et al. (2023) for French exporters. We closely follow their approach in classifying high and low inventory firms. Lafrogne-Joussier et al. (2023) find that French exporters with high levels of end-of-year inventory export 5.2 percentage points more in the aftermath of the early lockdown period in 2020 in China. Returning to our own setting, the estimate for high-inventory firms is also comparable to the difference observed for firms with a diversified supplier base

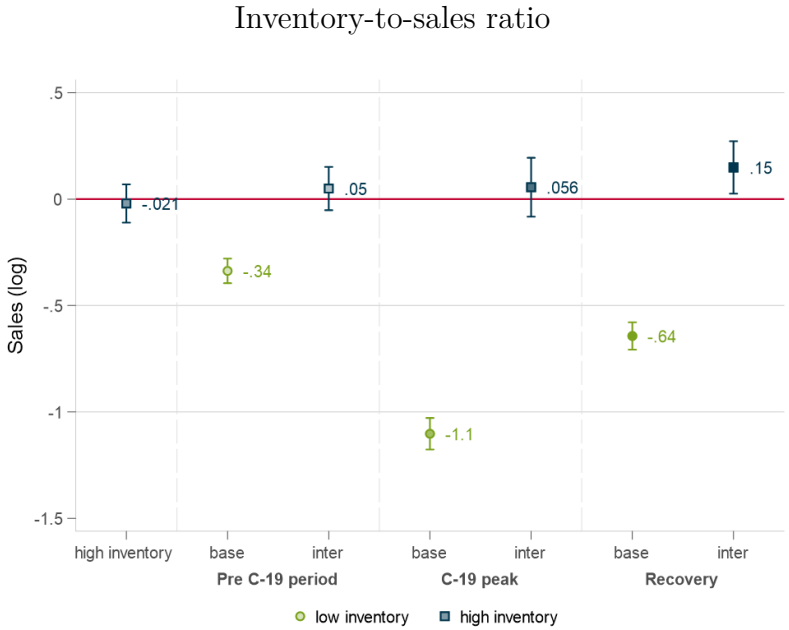
[43] VAT-paying businesses that do not file CIT are mainly sole proprietors or export processing zones.

[44] Our findings are robust to ignoring sector heterogeneity in inventory levels and simply using 60 days (=2 months) as a general cut-off for all firms.

[45] Note the log scale of the coefficients in Figure 4.C1.

relative to firms with lower levels of diversification.[46] The last row of Table 4.C3 shows that high inventory firms are also more likely to feature in the group of highly diversified firms. However, there is no 1:1 overlap. While 25 percent of the high-inventory firms are also highly diversified firms, as many as 22 percent of the firms with low diversification also have high inventory levels. In Figure 4.C2, we look at whether firms that feature in both groups drive the overall results. We do not find this to be the case. Inventory levels and diversification of the upstream supply chain seem to operate as independent channels. For the subgroup of firms for which we can look at both indicators at the same time, the effects are limited to the recovery period. This finding is again in line with our interpretation of the initial COVID-19 downturn primarily as an uncertainty and demand shock, while upstream supply chain bottlenecks become more binding during the recovery.

Figure 4.C1: Pre-crisis inventory levels and firm sales during the crisis and recovery phase

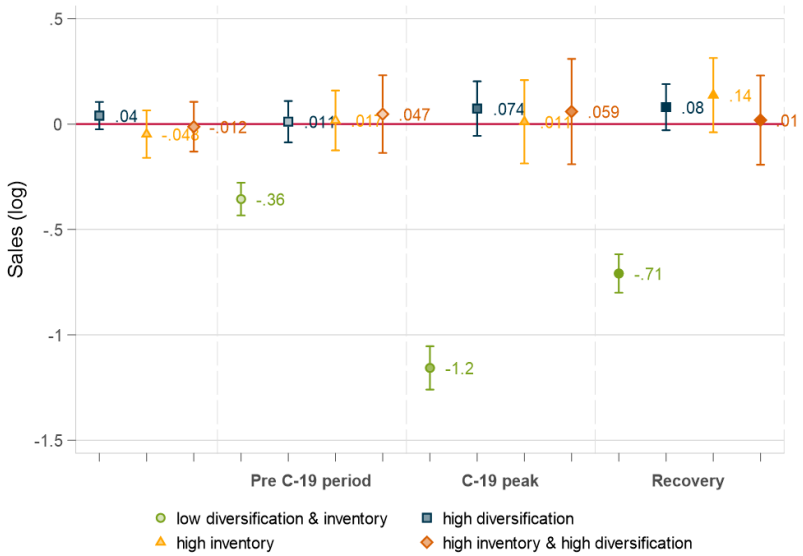


The above graphs present the results from estimating Equation 4.3. We control for firm age, share of sales to final domestic demand, share of high-inventory firms in the firm's own 4-digit sector, and import and export status. The pre-COVID-19 (C-19) phase corresponds to November 2019 to March 2020, the COVID-19 peak is April to June 2020, and the recovery July 2020 to March 2021.

[46]The effect is slightly less for diversification due to higher levels of average sales among the group of firms with lower levels of diversification.

Figure 4.C2: Supply chain diversification and pre-crisis inventory levels

Inventory-to-sales ratio



The above graphs present the results from estimating Equation 4.3. We control for firm age, share of sales to final domestic demand, share of high inventory and highly diversified firms in the firm's own 4-digit sector, and import and export status. The pre-COVID-19 (C-19) phase corresponds to November 2019 to March 2020, the COVID-19 peak is April to June 2020, and the recovery July 2020 to March 2021.

4.C.2 Additional material on network characteristics

Table 4.C1: Firm-level characteristics and 2018 outcomes by upstream diversification

	(1) low diversification		(2) high diversification		T-test Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Firm age	16828	11.745	10787	12.582	-0.837***
		(0.072)		(0.093)	
Nairobi or Mombasa based	16828	0.777	10787	0.786	-0.009*
		(0.003)		(0.004)	
Direct exporter	16828	0.042	10787	0.057	-0.015***
		(0.002)		(0.002)	
Direct importer	16828	0.141	10787	0.240	-0.099***
		(0.003)		(0.004)	
Share of dom. sales out. network	16828	0.288	10787	0.191	0.096***
		(0.003)		(0.003)	
Sales (log)	16828	16.355	10787	17.032	-0.677***
		(0.019)		(0.022)	
Purchases (log)	16828	14.977	10787	16.128	-1.150***
		(0.030)		(0.030)	
No. employees	16828	20.891	10787	37.054	-16.163***
		(1.371)		(2.474)	
Value added (log)	14208	15.837	9294	16.224	-0.386***
		(0.016)		(0.021)	
Value added per employee (log)	8579	14.578	7156	14.383	0.194***
		(0.018)		(0.017)	

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 4.C2: Firm-level characteristics and 2018 outcomes by downstream diversification

	(1) low diversification		(2) high diversification		T-test Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Firm age	17683	11.455	13905	11.764	-0.309***
		(0.070)		(0.078)	
Nairobi or Mombasa based	17683	0.773	13905	0.740	0.033***
		(0.003)		(0.004)	
Direct exporter	17683	0.037	13905	0.055	-0.017***
		(0.001)		(0.002)	
Direct importer	17683	0.167	13905	0.174	-0.007*
		(0.003)		(0.003)	
Share of dom. sales out. network	17683	0.325	13905	0.378	-0.053***
		(0.003)		(0.004)	
Sales (log)	17683	16.224	13905	16.851	-0.627***
		(0.019)		(0.020)	
Purchases (log)	17683	14.974	13905	16.187	-1.213***
		(0.027)		(0.022)	
No. employees	17683	13.565	13905	40.056	-26.490***
		(0.612)		(2.531)	
Value added (log)	14638	15.746	11720	16.124	-0.378***
		(0.016)		(0.019)	
Value added per employee (log)	8497	14.593	8600	14.403	0.190***
		(0.018)		(0.016)	

Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

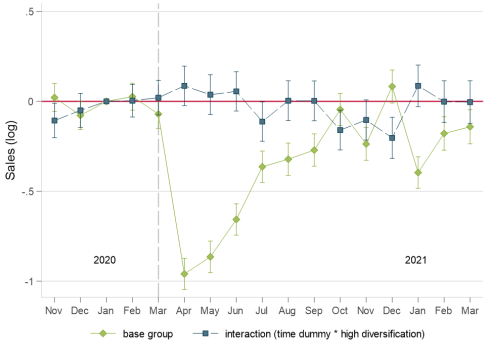
Table 4.C3: Firm-level characteristics and 2018 outcomes by inventory levels

	(1) low inventories		(2) high inventories		T-test Difference
Variable	N	Mean/SE	N	Mean/SE	(1)-(2)
Firm age	13221	10.300	4046	12.455	-2.155***
		(0.072)		(0.149)	
Nairobi or Mombasa based	13222	0.744	4046	0.728	0.016**
		(0.004)		(0.007)	
Direct exporter	13222	0.044	4046	0.045	-0.001
		(0.002)		(0.003)	
Direct importer	13222	0.184	4046	0.277	-0.093***
		(0.003)		(0.007)	
Share of dom. sales out. network	13222	0.378	4046	0.357	0.022***
		(0.004)		(0.007)	
Sales (log)	13222	16.585	4046	16.619	-0.034
		(0.022)		(0.034)	
Purchases (log)	13222	15.506	4046	15.995	-0.489***
		(0.032)		(0.045)	
No. employees	13222	19.305	4046	24.310	-5.005**
		(0.852)		(2.250)	
Value added (log)	11196	15.895	3146	15.813	0.082**
		(0.019)		(0.036)	
Value added per employee (log)	6991	14.577	2362	14.237	0.340***
		(0.019)		(0.030)	
High upstream diversification	10854	0.376	3453	0.481	-0.105***
		(0.005)		(0.009)	
High downstream diversification	12518	0.447	3915	0.499	-0.052***
		(0.004)		(0.008)	

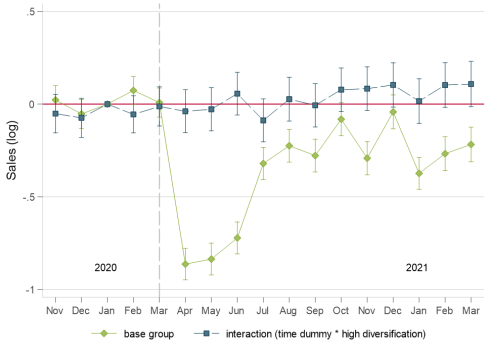
Notes: The value displayed for t-tests are the differences in the means across the groups. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Figure 4.C3: Domestic network position and firm sales

Downstream (number of domestic buyers)

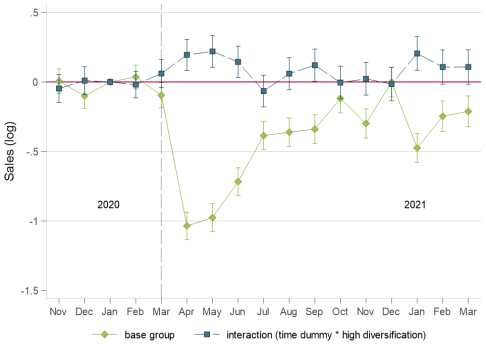


Upstream (number of domestic suppliers)

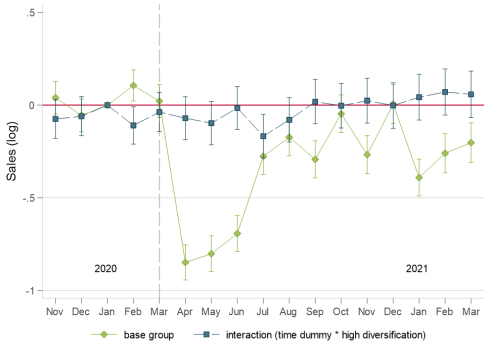


Complexity of the supply chain

Number of domestic downstream sectors

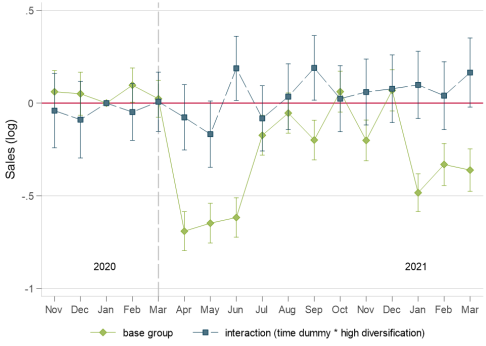


Number of domestic upstream sectors



Inventory levels

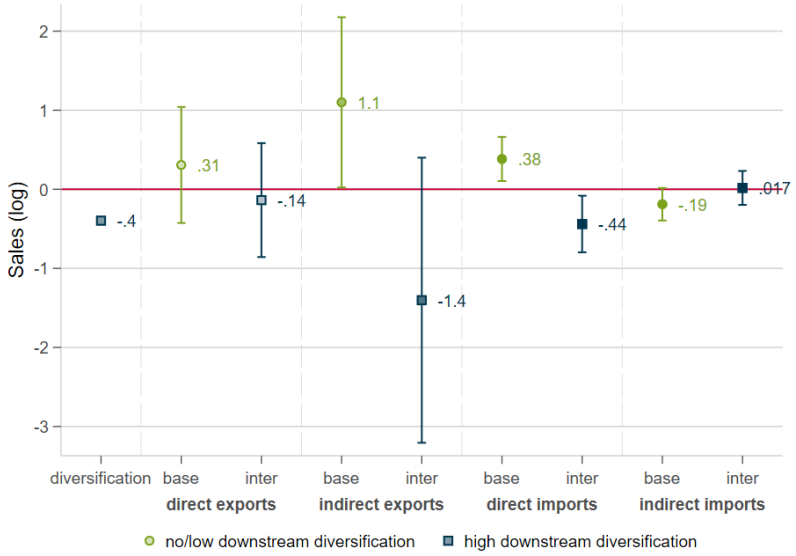
Pre-crisis inventory-to-sales ratio



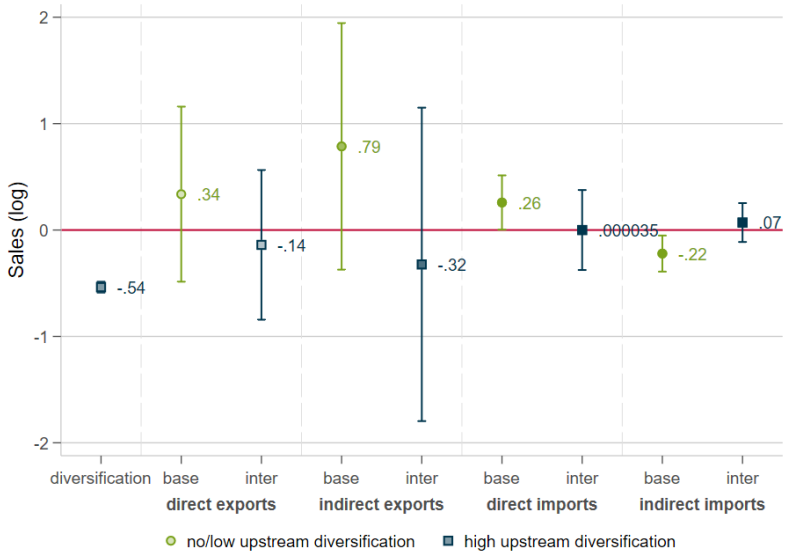
In the above graphs we regressed firm-level outcomes on a series of monthly time dummies (solid green line) in line with Equation 4.1. We further interacted the time dummies with the respective measure for the firm's network position or inventory levels (dashed blue line). We included firm and month fixed effects and normalised the coefficient for January 2020 to zero. Standard errors are clustered at the firm level and the error bars show the 95-percent confidence intervals. The regressions included data from April 2015 to March 2021.

Figure 4.C4: Supply chain diversification and the response of firm sales to international shocks

Downstream (number of domestic buyers)



Upstream (number of domestic suppliers)



The above graphs present the results from estimating Equation 4.2 while interacting the shock measures with dummies for above-median diversification of the up- and downstream supply chain. The coefficients for control variables are not included.

CHAPTER 5

Changing mobility patterns and food price dynamic during the COVID-19 pandemic in Kenya and Uganda

with **George Kariuki Kinyanjui and Doreen K. Rubatsimbira**

Abstract Real-time price data collection during crises is crucial for informing policy responses, but can be challenging due to fast-changing consumption and mobility patterns. We adopt a crowd-sourcing approach to investigate the impact of the COVID-19 pandemic on prices of essential food items in Kenya and Uganda. Combining this price data with information on changes in mobility patterns, we find that a 10 percentage point reduction in mobility leads to a 0.3% and 1.5% increase in food prices in Kenya and Uganda, respectively. Our results are robust across a variety of empirical specifications, but we cannot conclusively rule out a zero effect in Kenya. Furthermore, our findings indicate that mobility patterns continue to impact price dynamics beyond the initial shutdown phase.

5.1 Introduction

The onset of the COVID-19 pandemic triggered drastic changes in consumption and mobility patterns of people across the globe. Our paper studies the link between mobility patterns and price dynamics in Kenya and Uganda. We use a crowd-sourcing approach to collect high-frequency, geo-coded retail price data for essential food items. We then combine this granular price information with daily data on changes in visiting patterns of smartphone users at locations like workplaces, grocery stores and other retail stores.[1]

In times of crisis and fast-changing environments, an effective policy response benefits of real-time information. Hence, timely data collection through non-traditional channels can form a crucial tool to track prices in such an environment (Jaravel and O'Connell, 2020). For instance, during the COVID-19 pandemic, governments faced the task to design social protection programmes. An informed decision between cash transfers versus in-kind assistance, for example, depends on the local market environment and consumer price dynamics (Gerard et al., 2020). Cash transfers are fungible and easier to deliver, while in-kind programmes provide security in times of high price volatility (see for example Barrett, 2002; Barrett et al., 2009; Khera, 2014; Gerard et al., 2020; Gadenne et al., 2021). Therefore, it is crucial to monitor spatially dis-aggregated price dynamics, particularly in contexts with incomplete domestic market integration, e.g., due to spatial heterogeneity in available infrastructure (Barrett, 1996). At the same time, statistical bureaus faced challenges in collecting vital food price data during the pandemic. The U.S. Bureau of Labour Statistics (BLS) reported that response rates for food prices fell by 15-30% between June 2019 and June 2020, particularly for items with drastically shifted consumption and expenditure patterns.[2] Crowd-sourcing can be an effective data collection tool to collect data with wide geographic coverage, especially in the absence of alternative data sources, such as high-frequency scanner data (O'Connell et al., 2021). Crowd-sourcing as a means of data collection has traditionally been popular

[1] We rely on publicly available data from Google's COVID-19 Community Mobility Reports: https://www.google.com/covid19/mobility/

[2] https://www.bls.gov/opub/mlr/2020/article/the-impact-of-the-covid-19-pandemic-on-food-price-indexes-and-data-collection.htm

in the humanitarian sector, e.g., to map humanitarian needs after natural disasters (Butler, 2013). While it is not a perfect substitute for traditional data collection methods, the advent of the pandemic presented limited options to engage in a standard contact data collection approach.

To collect real-time data on the impact of the COVID-19 pandemic on food prices in Kenya and Uganda, we relied on a quick-response online survey, which could be accessed via mobile phones. Over 250 registered volunteers, mainly university students and staff members of non-governmental organisations (NGO), recorded prices and availability of selected food items through an online form. The form collected information on product brands, neighbourhoods, and shop characteristics. The broad geographic coverage of the collected data allows us to compare dynamics in two neighbouring economies with similar household consumption patterns but varying initial shutdown policies, providing an interesting case study. We show that changes in mobility patterns continue to drive price dynamics during and beyond the initial lockdown phase. Our findings suggest that retail prices increase as the level of activity at workplaces, grocery and retail stores declines. We estimate that a 10 percentage point reduction in activities at workplace locations corresponds to a 0.3% and 1.5% increase in food prices in Kenya and Uganda, respectively. Although the results are consistent across various empirical specifications, we cannot rule out that the effect is zero in Kenya. Our results suggest that lockdown measures are not the only driver of economic outcomes during the pandemic. Behavioural changes to reduce the risk of infection or adverse impacts on income-generating activities could equally impact consumption and mobility patterns and consequently affect price dynamics. Additionally, post-lockdown mobility patterns might drive business decisions to reopen.

Our study contributes to three strands of literature: (i) on food security during the COVID-19 pandemic, (ii) data collection methods in dynamic and crisis environments, and (iii) the interplay between policies and individual behavior during the pandemic. First, we add to the literature that addresses questions of food security during the COVID-19 pandemic, both in a broader sense and more specifically related to food

prices. Most papers in this literature focus on households on one side (Amare et al., 2020; Kansiime et al., 2021; Mahmud and Riley, 2021; Egger et al., 2021; Rudin-Rush et al., 2022; Mahmud and Riley, 2023; Tabe-Ojong et al., 2023; Abay et al., 2023) and wholesale markets on the other side (Ruan et al., 2021; Lowe et al., 2021). By studying retail price dynamics, we examine the link between the two. For instance, Tabe-Ojong et al. (2023) provide a detailed review of research documenting food security outcomes on the African continent, with a focus on papers that rely on household surveys. While a substantial proportion of this literature was published during the early days of the pandemic and thus focuses on the initial shock, Rudin-Rush et al. (2022) extend this early literature by covering an extended time period well into 2021 and examining food insecurity in Burkina Faso, Ethiopia, Malawi, and Nigeria. Our paper complements these studies by tracking price dynamics specifically, which can be a key contributing factor for food insecurity. The COVID-19 shock led to both demand and supply side shocks that simultaneously impacted prices in different ways (Jaravel and O'Connell, 2020; Baqaee and Farhi, 2022), with initially unclear implications (Jaravel and O'Connell, 2020).

In a very early study, Agyei et al. (2021) documented price dynamics of staples between March and September 2020 for 28 countries in Africa. While they are limited to studying prices at the national level, they can compare the dynamics of local and imported product varieties. Another closely related study by Narayanan and Saha (2021) compiles a large dataset containing retail and wholesale prices using administrative data from India. Like us, they focus on fast-moving food commodities. They find a substantial surge in prices right after the announcement of the national lockdown in India, with no sign of prices reverting back to pre-lockdown levels a month later. Our study builds on Narayanan and Saha (2021) by (i) adding a context with much less stringent lockdown measures and (ii) tracking price dynamics throughout a re-opening phase. Lastly, Mahajan and Tomar (2021) provide another relevant study, using data from India's largest online retailer to observe the entire supply chain. They estimate a 10% reduction in the availability of vegetables, fruits, and edible oils and only a minimal impact on prices. However, a limitation of their study is that online retailers largely target India's middle class, which has been less affected by income losses during the pandemic. While we observe fewer details on

the nature of supply chains, our approach of using quick-response online surveys filled out by a varied network of volunteers allows us to trace price dynamics relevant for a broader set of income groups.

Second, we contribute to the literature on scaling non-traditional approaches to data collection to obtain real-time food security and food price information, particularly in fast-changing environments like the pandemic, which limit face-to-face interaction. Similar studies using innovative approaches include that of Kansiime et al. (2021), who broadcast an online survey through social media platforms such as WhatsApp, Facebook, Telegram, and Twitter in Kenya and Uganda to study the implications of the coronavirus pandemic on household income and food security. Consoli et al. (2023) developed a method to track food price policies on a close-to-real-time basis. A paper closely related to ours is Adewopo et al. (2021), who also used a crowd-sourcing approach to collect real-time geo-referenced price data for two varieties of rice and two varieties of maize in Northern Nigeria in 2019 and the first half of 2020. They found that prices for maize increased by 26% and those for rice went up by 44%. A key advantage of their setting is that their data collection pre-dates the pandemic, providing them with baseline data of a similar spatial and temporal granularity. We complement their study by (i) covering a much extended time period, (ii) considering processed food items like sugar and maize flour in addition to unprocessed staple food, and (iii) collecting data for a larger geographic region covering two neighboring countries with a joint population of 98.86 million. Importantly, this geographic variation allows us to study the link between mobility patterns and food prices.

Lastly, we contribute to a literature that tries to contrast the impact of shutdown policies as a containment measure with shifts in individual behaviour to avoid contraction of the virus (Aum et al., 2021; Goolsbee and Syverson, 2021)[3]. Looking at price dynamics during both the lockdown and re-opening phase, our setting does not allow us to cleanly distinguish between the two. However, in our context the drop in household incomes due to the pandemic - driven by a complex interplay of factors

[3]For updated findings, see https://bfi.uchicago.edu/insight/research-update-drivers-of-economic-decline/.

such as global trade disruptions, local shutdowns, and changes in the nature of social and economic activities (Nechifor et al., 2021) - might be the key driver of our findings rather than extended periods with strict containment policies. Highlighting the relevance of this channel, Kansiime et al. (2021) show that two thirds of the households in their sample experienced a loss in income during the pandemic in Kenya and Uganda. Shupler et al. (2020) report that households in urban informal settlements are even more likely to have experiences a loss in income (95%) and struggle with food-insecurity (88%). Mahmud and Riley (2023) show that incomes of business-owing households in western Uganda had still not recovered from the income shock over a year into the pandemic.

5.2 Data

5.2.1 Mobility

To measure the change in mobility patterns and activity levels across space, we rely on daily movement trends published in Google's Community Mobility Reports[4]. Their mobility indices capture the change in activity levels of Android smart phone users at different types of localities relative to a 5-week baseline period between 3rd January and 6th February 2020 at the county (Kenya) or district (Uganda) level. Each index reports the daily percentage change in activity relative to the median activity level on the same weekday during the baseline period.

We rely on the Workplace Index as our default measure for changes in mobility and henceforth refer to it as simply the Mobility Index. Of all the indices published it has the most comprehensive and most consistent geographic coverage. Given our focus on retail prices of essential food items, our preferred alternatives are the Grocery and Pharmacy Index and the index for retail and recreation-related localities (Retail and Recreation Index). Unfortunately, their geographic coverage lacks the comprehensiveness and consistency of the Workplace Index (see Figure 5.A1 and Figure 5.A2). The downside of using the Workplace Index is that (i) if mobility patterns mainly capture the demand side effect of the economic shock, workplace mobility might not be as relevant, (ii) during certain periods of the year it moves counter-cyclical to retail

[4]https://www.google.com/covid19/mobility/ Google compiles the underlying data from the location history of Google users (Aktay et al., 2020).

activities, e.g., during the holiday season in December and early January. We discuss the geographic coverage of the mobility indices and changes in mobility patterns over time in more detail in the appendix (see Section 5.A.1).

5.2.2 Retail prices

To track product prices and the availability of selected essential food items, we rely on primary data from a quick-response online survey. During the survey period a network of about 250 volunteers collected over 19,000 geo-coded price quotes for 17 items in Kenya and 6,400 quotes for 11 items in Uganda.[5]

Survey design and data collection

Relative to well-established retail price data collected by national statistical agencies to compute the Consumer Price Index (CPI), our data collection emphasises the tracking of prices on a frequent basis rather than monthly and focused on a large variety of geographic areas rather than major urban centres. Our price survey asked respondents to record prices through an online form - either when or after visiting a shop in-person or getting deliveries. Embedding the survey into day-to-day activities of respondents helped us to overcome concerns about initiating or prolonging human interactions that would not occur otherwise. The online survey listed pre-selected products that are purchased by households on a frequent basis, are available nationwide in both countries, and carry substantial weight in the official consumer price index. Volunteers had the option to record prices for a total range of 40 products. However, the responses of volunteers were highly skewed towards the 17 and 11 items we focus our analysis on. The data is highly sparse for all other items. Rice emerges as the most frequently recorded item in both contexts.

The respondents consisted of volunteers, primarily university students and staff/network members of NGOs, who pre-registered for participation. The advantage of relying on university students as respondents is that many of them own smartphones, and during the pandemic, they were scattered across the country as universities shifted to remote activities and students returned to their families. We launched the pilot phase of the data collection in Kenya in late March and were able

[5]Figure 5.A3 shows the daily number of collected price quotes.

to scale it up by May 2020. The piloting in Uganda took place in June and the scale up was implemented in August of 2020.

While our sample of respondents itself is not representative of the wider population, their shopping locations overlap with the population overall. General shops and kibandas - the most ubiquitous forms of retail in both Kenya and Uganda - are well-represented and are the modal retail locality for all items (see last columns of Tables 5.A1 and 5.A2). However, our respondents are more likely to shop at supermarkets and less likely to record price information from open air markets.[6]

To track the quality of the data, we closely monitored our survey and periodically reached out to individual respondents whenever we found implausible entries. In cases of obvious typos and misunderstandings of the survey, corrections were made. Finally, we standardised prices for each item to the quantities and measurements reported in Table 5.1. For fruits and vegetables volunteers mostly found it easier to record the prices for an individual item instead of a standardised unit of measurement. Acknowledging the caveat that mango, avocado, cabbage, and tomato sizes might vary substantially, we opted for sticking with this preferred measurement.

[6]We are unfortunately unable to quantify the exact deviation using a representative household survey due to lack of data on purchasing behaviour of food items. The Kenya Integrated Household Budget Survey 2015-2016, for example, only covers shopping locations for non-food items.

Table 5.1: Summary statistics for retail prices collected via a quick-response online survey

Kenya

Item	Quantity	N	N/month	Avg.price	Sd price	#brands	#counties
Rice	1 kg	2,072	156	123	44	11	39
Bread (white)	400 g	2,061	153	49	6	15	38
Sugar	1 kg	1,556	109	112	14	8	37
Eggs	one	1,556	163	13	3	1	42
Wheat flour	2 kg	1,338	96	128	15	13	39
Maize flour (sifted)	2 kg	1,323	98	121	18	17	39
Milk (fresh, packaged)	500 ml	1,246	88	51	9	14	39
Bread (brown)	400 g	1,104	82	49	7	11	40
Tomatoes	one	1,070	111	7	3	1	38
Banana (ripe)	one	967	100	8	4	1	39
Cooking oil	1 litre	777	52	170	39	10	40
Onions (bulbs)	1 kg	770	80	66	34	1	33
Cabbage	one	672	73	46	30	1	38
Avocado	one	657	73	19	10	1	35
Mango	one	628	72	18	10	1	35
Soda/Soft drink	300 ml	576	64	30	8	1	37
Salt	1 kg	352	21	39	15	5	32

Uganda

Item	Quantity	N	N/month	Avg.price	Sd price	#brands	#districts
Rice	1 kg	1,224	264	3,740	1,751	9	58
Bread (loaf)	400 g	838	134	4,543	719	5	49
Beans (dry)	1 kg	771	132	3,587	761	2	48
Sugar	1 kg	761	190	3,357	520	7	51
Eggs	one	526	81	413	99	1	42
Avocado	one	425	70	649	323	1	41
Groundnuts	1 kg	348	45	5,700	1,190	1	37
Maize flour (sifted)	2 kg	333	76	2,020	567	11	37
Tomatoes	one	318	54	248	171	1	37
Banana (ripe)	one	192	44	337	277	1	39
Salt	1 kg	182	45	2,128	1,040	6	37

Prices are denoted in Kenyan and Ugandan shilling respectively.

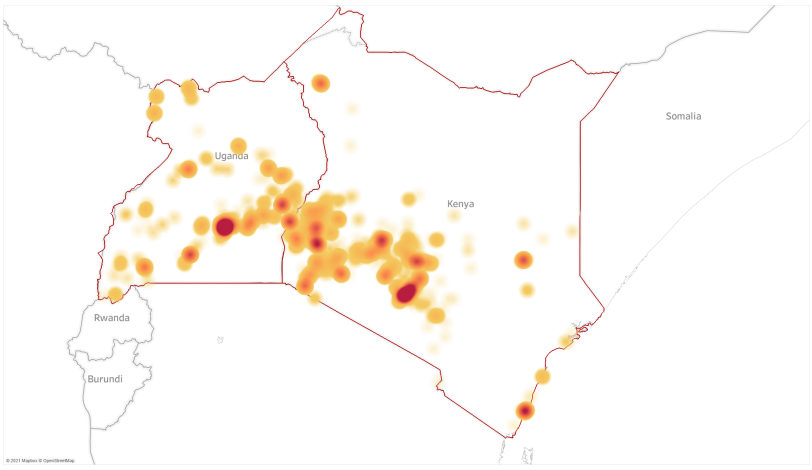
Geographic coverage of the price survey

Our data collection succeed in its primary goal to achieve a wide geographic scope and generate high-frequency output. The number of collected price quotes further highly correlates with population density. In Figure 5.1 we map the geography of the overall number of responses for rice, sugar, and eggs. Darker areas indicate a higher density of price quotes. The densest spots on the maps in Figure 5.1 can be easily mapped

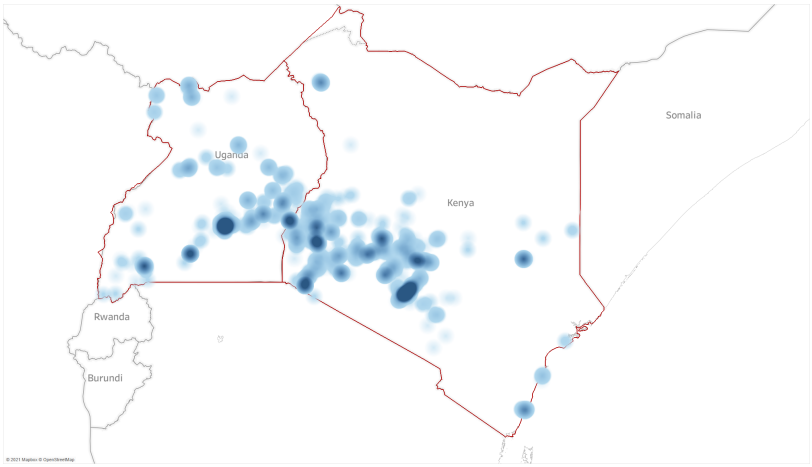
back to major towns in Kenya and Uganda. For the most popular item rice, we collect at least one response in 39 out of Kenya's 47 counties and 58 out of Uganda's 135 districts during the data collection period (see Table 5.1). Map 5.A4 in the appendix shows a consistent geographic pattern across across time. In both Kenya and Uganda the sample is skewed towards semi-urban and urban areas. In Uganda, we only cover the most populous districts (see Figure 5.A5) and 50% of the price quotes used in the following analysis have been recorded in the Greater Kampala region. We discuss the geographic coverage of the data and correlations with indicators such as population density, mobile phone coverage, and Internet access in more detail in appendix section 5.A.2.

Figure 5.1: Geographic dispersion of price quotes for popular items

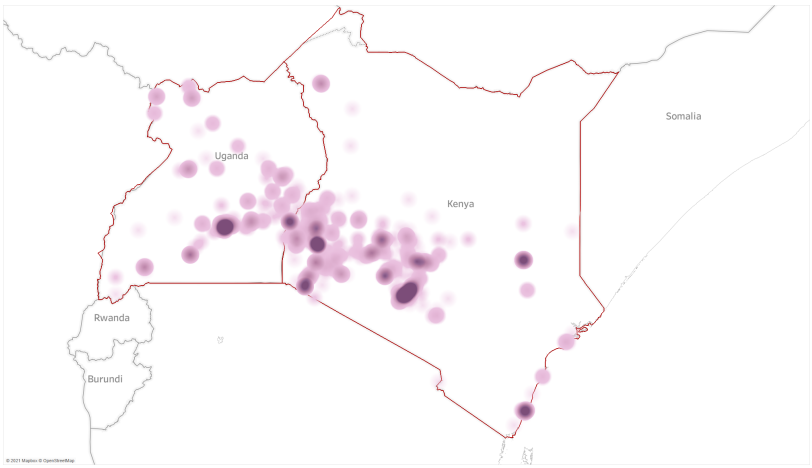
Rice



Sugar



Eggs



The above maps plot the geographic dispersion of the collected price quotes for rice, sugar and eggs (from top to bottom). Darker colours indicate a greater density of price quotes. Lighter colours indicate fewer price quotes.

Benchmarking of the survey data using national price statistics

We benchmark the data against available alternative sources. In Kenya we are able to compare them to monthly national average prices published by both the Kenya National Bureau of Statistics (KNBS)[7] and the Famine Early Warning Systems Network (FEWS)[8] in Figure 5.A7 and 5.A8. A few caveats with respect coverage of the benchmark data need to be kept in mind: (i) item-level price data from KNBS is only available with gaps, as KNBS reports national average prices for a different subset of items each months. For FEWS we are only able to obtain prices for cooking oil, maize flour, bread and milk. (ii) the detailed brand for which KNBS and FEWS collect prices are unknown and we hence compare the prices to the simple average across all observed brands in our sample, and (iii) KNBS' data collection focuses on 14 cities and towns[9] and FEWS focues on Nairobi. We hence look at the comparison with the average prices for Nairobi in Figure 5.A9. Product variety and brands explain a substantial proportion of the variation in prices, 14% in both Kenya and Uganda (see Table 5.2). The observed trend provides evidence that our price averages for maize flour, cooking oil, pasteurised milk, tomatoes, mangoes, and bread for Nairobi did not differ systematically from those reported by FEWS and KNBS.[10] As for sugar, wheat flour, and tomatoes, the trends look similar even after overlaying the 2019 KNBS national averages.

Table 5.2: Variance decomposition for food prices

	Brand	Shop	Region	Residual
Kenya	14.2%	0.2%	0.8%	84.9%
Uganda	15.0%	0.2%	2.6%	80.6%

Based on January and February 2021 prices.

In Uganda, we benchmark the average price of each of the 11 commodities from the price survey against the average price for the same commodity compiled by the

[7]We extracted the data from monthly inflation reports for the years 2019, 2020 and the first two months of 2021. See February 2021 as an example: https://www.knbs.or.ke/?wpdmpro=cpi-february-2021

[8]https://fews.net/fews-data/337

[9]https://www.knbs.or.ke/download/cpi-users-guide/

[10]We had to scale tomato and mango prices to a kilogram to make them comparable to KNBS prices - the overall price trend, however, looks the same if only one fruit is considered as the standard unit.

Uganda Bureau of Statistics (UBoS) in the same period (see Figure 5.A11). UBoS prices are collected from ten centers across the country. Albeit we are unable to compare prices at the beginning of the lockdown period, both UBoS and survey prices depict a declining trend with the easing of COVID-19 mitigation measures. Some of the discrepancies between UBoS prices and the price survey are likely driven by the bias towards more price quotes from Greater Kampala (see Figure 5.1), where prices generally tend to be higher than in the rest of the country.[11] Similar to Kenya, we overlaid historical data for 2019 from UBoS, highlighting the seasonality of some products and the generally lower price levels in the second half of the year.

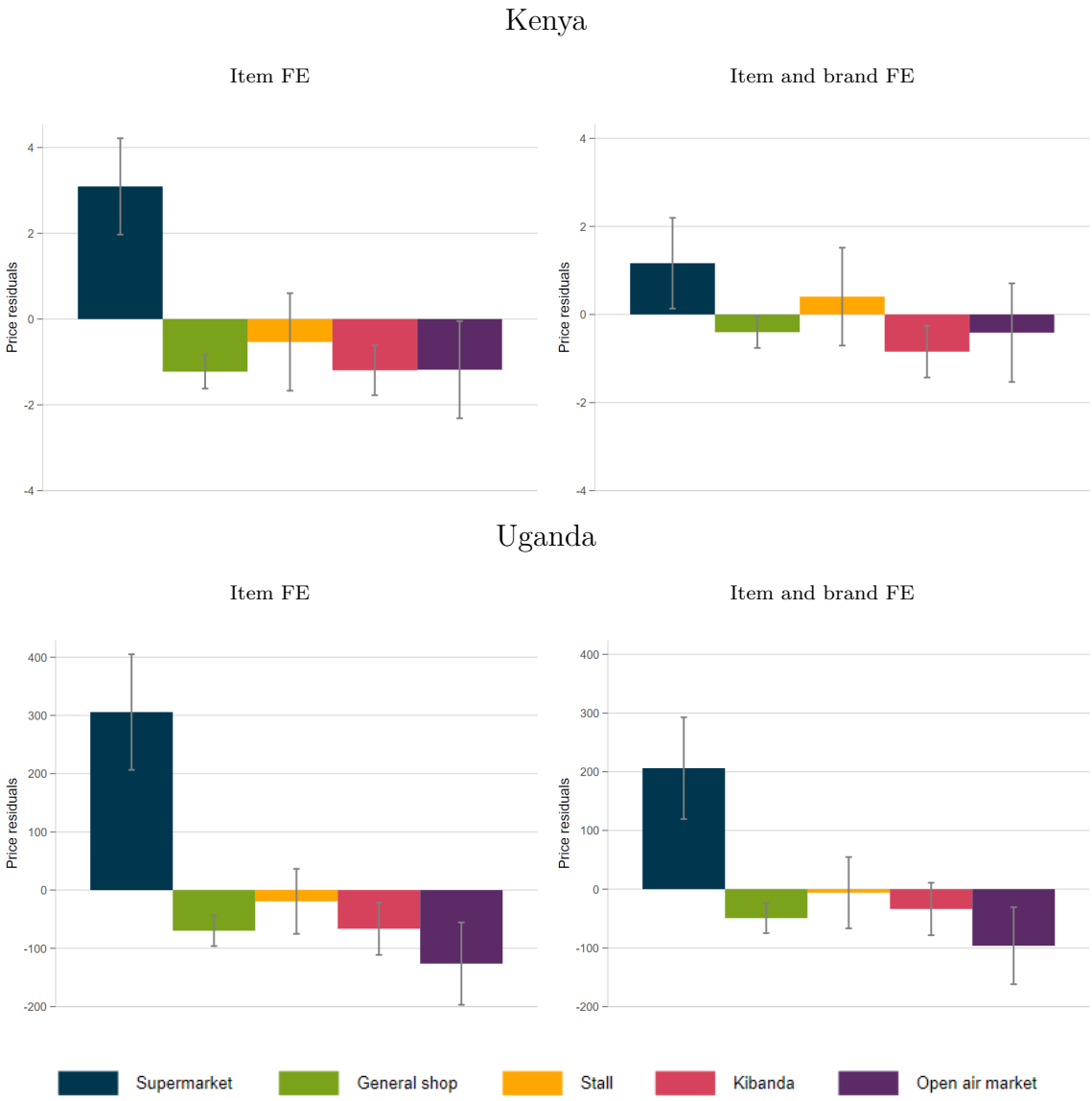
Survey prices of beans, maize flour, sugar, and tomatoes move consistently with UBoS compiled prices, indicating high prices in June and a gradual fall thereafter. Similarly, bread and avocado prices in both sources depict a falling trend and a gentle rise around September/October. Egg prices from both the survey and UBoS pick up following improved demand due to a recovery led by partial easing of the containment measures where the main consumers mainly hotels and restaurants were allowed to operate as outside caterers and as "take outs". The initial spike in rice prices in the survey data is mainly driven by the Kampala bias early on in the data collection. Higher priced, imported rice brands are more prevalent in Greater Kampala.

How do prices vary across retail categories?

As noted, our sample of respondents might be dis-proportionally likely to shop at supermarkets and less likely to shop at open air markets. We also take note that different items would more likely be sold in certain shop types. For instance, fruits and vegetables are more likely to be found in a kibanda as opposed to a general shop. In addition, some item brands and varieties are more likely to be found in certain types of shops. Figure 5.2 shows the variation in prices across different types of shops, controlling for item fixed effects in the left column and item as well as brand fixed effects in the right column. The residual variation in prices across shop types is drastically reduced once we control for brand fixed effects. In the variance decomposition in Table 5.2 we find that relative to the item's brand, little variation in prices is explained by the retail outlet.

[11] By including item-level fixed effects in all specifications below, we control for the potentially higher average price levels picked up by the survey.

Figure 5.2: Variation in prices by shop type



The above graphs show the variation retail prices by shop type. In the left column we control for item fixed effects only. In the right column we add brand fixed effects. The grey error bars show the 95% confidence intervals.

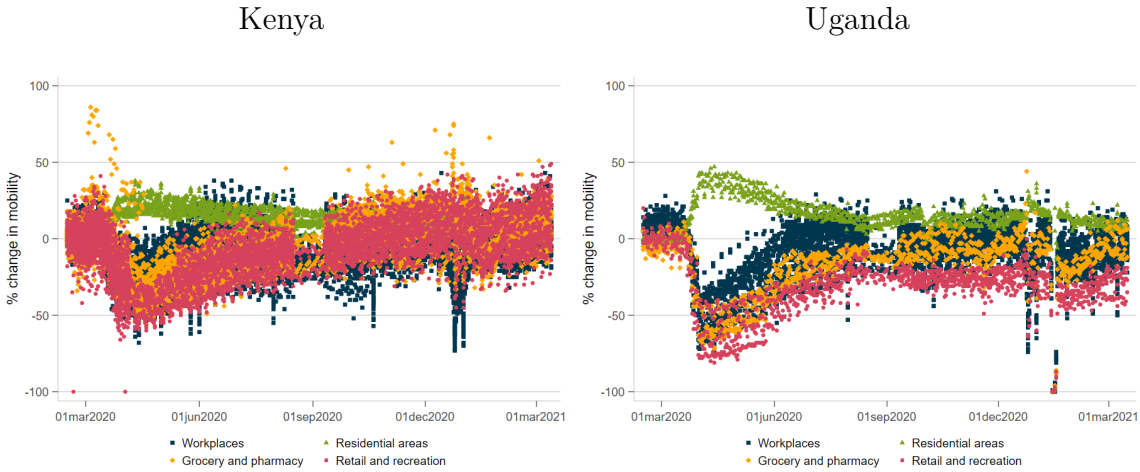
5.3 How did mobility patterns shift during the pandemic?

In mid-March, Uganda and Kenya both limited international travel from countries with ongoing COVID-19 epidemics. On 13th March 2020, the first COVID-19 case

was recorded in Kenya with Uganda following a week later on 21st March 2020. Throughout March and April national and local governments in both countries swiftly introduced a series of containment measures.

The stringency of the containment measures ranged from almost complete lockdown in Uganda to night-time curfew and city-specific travel-restrictions in Kenya. Kenya lifted the final set of travel restrictions for its metropolitan areas in early July. In Uganda, travel restrictions remained in place for 40 border districts until early August and for a subset of nine districts until late September. During the time of our data collection Kenya experienced two relatively moderate waves of COVID-19 with peaks in the number of confirmed COVID-19 cases in early August and mid-November respectively (Roser et al., 2020).[12] Until early August, Uganda registered only very few COVID-19 cases. The first, smaller wave hit its peak in late September, while the peak of the second wave was registered in mid-December 2020 (Roser et al., 2020).

Figure 5.3: Variation in activity levels across different categories of locations



The graphs compare the variation in activity levels across time for three different Google mobility indices. The data is reported daily at the county level in Kenya and at the district level in Uganda. Each dot represents a county-day (district-day) observation.

Figure 5.3, shows the evolution of changes in activity levels for both Kenya and Uganda between March 2020 and March 2021. The most drastic change for both

[12]https://ourworldindata.org/coronavirus/country/kenya?country=KEN~UGA Data last retrieved 15th April 2023. For a more in-depth discussion of the evolution of the pandemic and related policy measures in Kenya and Uganda see among others Kansiime et al. (2021).

countries is picked up in April when imposed government-restrictions were at their height. While the drop off is rather sudden - in particular in Uganda - the recovery is more gradual and stretches several months. Workplace-related activities, and to some extent also activities at grocery and pharmacy-related localities, recover sooner than those linked to retail and recreation.

Regardless of differential containment strategies, both Kenya and Uganda were affected by disruptions to global supply chains (Chacha et al., 2024) and the emerging global economic crisis (coupled with a locust outbreak in the Horn of Africa). Many households faced substantial income losses (Pape et al., 2020; Kansiime et al., 2021; Egger et al., 2021; Mahmud and Riley, 2021, 2023).

5.4 How did food prices evolve over time?

To get a first idea of how food prices evolved over the course of the COVID-19 pandemic, we employ an event study approach. We regress the log retail price on a series of monthly time dummies, while including item, brand and county/district fixed effects. Given the lack of baseline data prior to the first lockdown phase in March 2020, we use prices in the final weeks of our data collection in February and March 2021 as the comparison group. Figure 5.4 shows that prices were 10% higher during the initial shutdown in Kenya in April 2020 when passenger travel in and out of specific counties was restricted and retail outlets adjusted to complying with COVID-19 safety measures. From June onwards, they are at the same level compared to February 2021. As the last restrictions to the hospitality sector get removed in September, we find that prices remained relatively low for an extended period. As we will discuss below, we cannot rule out a partial role of seasonality. However, the sustained reduction in household incomes due to the pandemic and the associated economic shock is likely a prominent factor (Pape et al., 2020; Kansiime et al., 2021).

The estimates for Uganda suggest a prolonged impact of the COVID-19 shock on retail prices that continues into June and July, the two months in the immediate aftermath of the national lockdown. The results indicate that the initial effects were of a similar magnitude relative to Kenya - with the caveat that we do not observe

prices in April and May in Uganda.[13]

A key limitation of our primary data collection is that it does not allow us to control for seasonality. To partly address this issue, in an alternative specification, we control for historical 5-year monthly average prices relying on the subsample of four items for which FEWS publishes monthly item-level price data for Nairobi. The estimated price effect in April and May now becomes indistinguishable from zero. However, we cannot rule out that this is driven by sample selection as the price effect is zero even in April for the subsample for which FEWS prices are available. In Uganda, the results are unchanged if we control for the national average price in the same month in 2019 reported by UBoS. UBOS prices are available for all items in our sample.

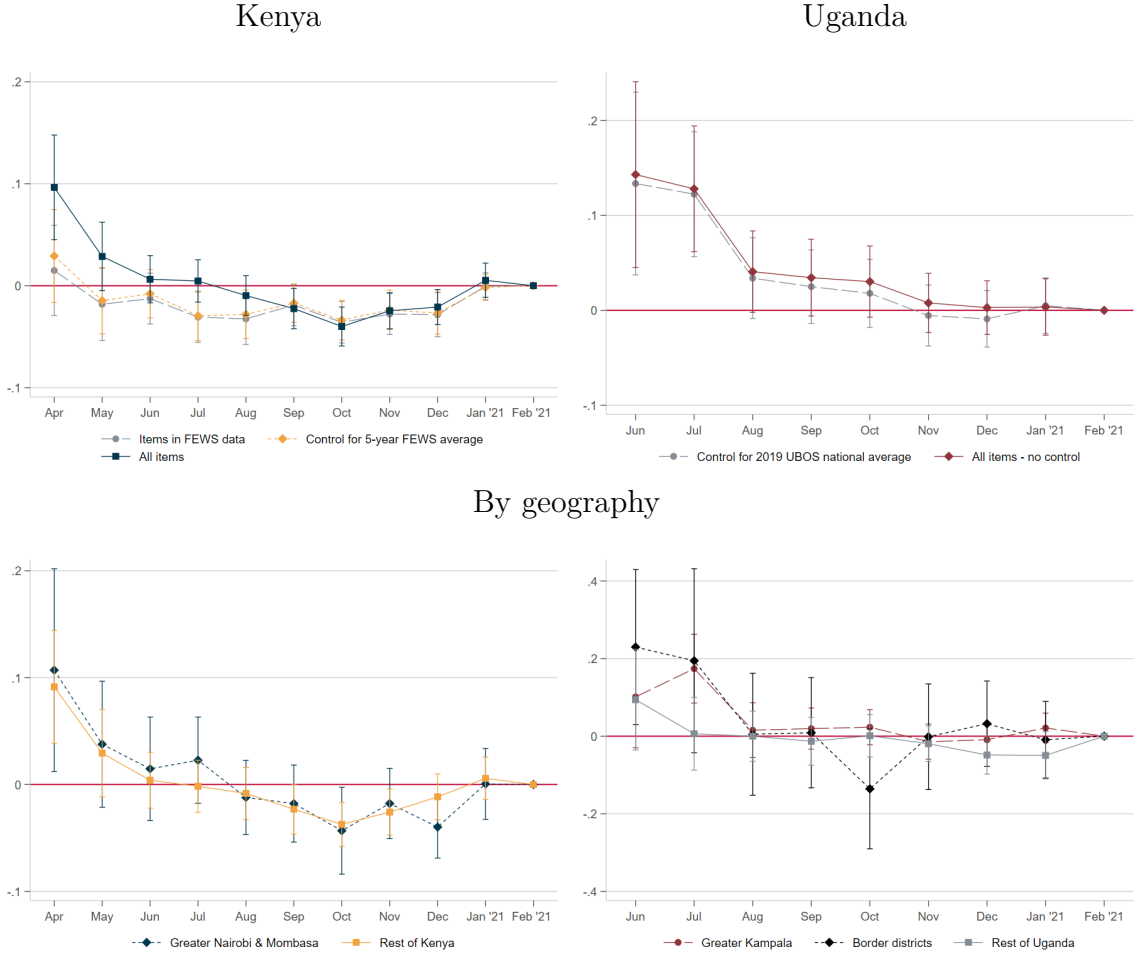
Travel restrictions in and out of the metropolitan areas of Greater Nairobi and Mombasa did not have a differential impact on price dynamics compared to the rest of Kenya (see top-left graph of Figure 5.4). This suggests that the travel restrictions aimed at restricting passenger travel have been successful in its goal to keep disruptions to the transportation of cargo to a minimum. In Uganda, we find evidence that the price level in the 40 border districts which remained in lockdown throughout June 2020 while the rest of the country was opening up, remained higher even for July.[14] Despite the lockdown restrictions being lifted, prices in the metropolitan area of Greater Kampala also remained higher than in the rest of Uganda for the month of July. This could be driven by supply chain disruptions for imported varieties, e.g., of rice, that are more widely consumed in Kampala or supply chain disruptions for fresh items cultivated in rural areas of Uganda.

It is important to highlight that while price levels eventually fell back, even only temporarily, higher prices likely had a substantial impact on food security (Kansiime et al., 2021).

[13] As an alternative specification to using monthly indicator dummies in our event study design, we divide our study period into different phases based on containment policies, following Lowe et al. (2021)'s approach studying food price dynamics in India. The results are covered in Appendix 5.B.

[14] Unfortunately, the collected data for the nine border districts that had restrictions in place until late September is sparse and we are therefore unable to separate them out for the purpose of this exercise.

Figure 5.4: Price dynamics for April 2020 - mid-March 2021



In the above figure we report the results from estimating the following specification: $y_{itd} = \sum_{k=0}^{11} \tau_{-k}$ feb2021$_{t-k} + \phi_i + \eta_{ib} + \delta_c + \varepsilon_{itd}$ where y_{itd} denotes log price of item i on day d, ϕ_i the item FE, η_{ib} the brand FE, and δ_c the county/district FE. The coefficient for February 2021 is normalised to zero. Prices are winsorised at the 1% level. Standard errors are clustered at the item-district/county level. The error bars show the 95% confidence intervals. For Kenya, the data covers the time period between the last week of March 2020 until the first week of March 2021. For Uganda, the data covers the time period between the last week of May 2020 until the second week of March 2021. We further show the results for an alternative specification where we control for historical price trends. In Uganda control for the national average price in the same month in 2019 (grey plot), which is available for all items included in our sample. For Kenya, we show two alternative specifications. First, we restrict the sample to the four items for which historical FEWS data are available (grey plot). Second, we control for the 5-year average price in the corresponding month (orange plot). In the second row, we distinguish between the Rest of Kenya and Greater Nairobi and Mombasa, which have seen travel restrictions imposed between 6th April and 7th July 2020. Correspondingly for Uganda, we distinguish between Greater Kampala, the border districts which were under prolonged lockdowns, and the rest of Uganda.

5.5 Linking mobility patterns and price dynamics

The pandemic altered economic and social interactions through temporary or permanent business closures, changes in consumption due to income losses, but also reduced levels of commuting. Many of these changes persisted beyond the initial lockdown phase (see for example Mahmud and Riley (2021); Egger et al. (2021)). We now turn to the impact of changes in mobility patterns more broadly on food prices.

5.5.1 Empirical strategy

We focus on two alternative baseline specifications, each of which relies on a different sources of variation to estimate the relationship between mobility and price dynamics. The first relies on the variation in mobility patterns across localities within a given month - by including month fixed effects, η_m. This specification allows us to at least partly address the issue of seasonality, which we cannot fully rule out due to lack of baseline price data in Kenya. The second specification relies on within-location changes in mobility over time. By including county/district fixed effects, η_d, this specification controls for the role of location specific characteristics such as fundamental differences in price levels due to local market characteristics. The two specifications including either η_d or η_m are:

$$p_{dibt} = \alpha + \beta * \text{mobility}_{dt} + \gamma_i + \phi_b + \eta_{m/d} + \varepsilon_{dibt} \tag{5.1}$$

where p_{dibt} is the log price of item i of brand b on day t in county/district d.[15] $mobility_{dt}$ captures the deviation from pre-COVID-19 mobility patterns as measured by the daily Google mobility index for county/district d.γ_i is the item FE and ϕ_b the brand FE. In Uganda, we further control for the item-level national average price in the same month of 2019.[16] A limitation is that we are only able to measure changes in mobility patterns at a fairly aggregate level, the county or district level. Kenya has 47 counties and Uganda 136 districts. However, in each case, the Google Mobility Index is not available for all countries and districts. In Table 5.A1 and 5.A2 we report

[15]It is possible that we observe multiple price quotes within an item, brand, day, county/district cell, especially in Kampala or Nairobi. We include each as an individual observation. The results are robust aggregation, with the caveat of less precise estimates.

[16]Results are robust to not including 2019 prices.

additional summary statistics for the subsample of price quotes that we are able to map to the Workplace Mbility Index. In Uganda we can map 63% of the price quotes and 68% in Kenya. Both the average price and standard deviation are very similar to the full sample.

Table 5.3: Price dynamics and mobility

	Kenya		Uganda	
	Month FE	Location FE	Month FE	Location FE
Mobility	-0.032	-0.051**	-0.152***	-0.150**
	(0.029)	(0.023)	(0.053)	(0.071)
$p_{i,2019}$			0.374***	0.395***
			(0.093)	(0.088)
No. observations	16,627	16,627	4,002	4,002
R2 adj.	0.927	0.930	0.928	0.928
Item FE	✓	✓	✓	✓
Brand FE	✓	✓	✓	✓
Month FE	✓	✗	✓	✗
Location FE	✗	✓	✗	✓

The table reports the results from estimating Equation 5.1, where we regress the log price on the Google Mobility Index (workplace) while controlling for item, brand, and month or location fixed effects as specified in the model title of each column. In Uganda, we are further able to control for the nationwide average price recorded by UBoS for the same month in 2019. Standard errors are clustered at the item-district/county level. Prices are winsorised at the 1% level. The Google Mobility Index captures the level of activity at a specified category of locations relative to a baseline period in January and February 2020. *; **; and *** denote significance at the 10; 5; and 1 percent levels, respectively.

5.5.2 Higher mobility levels, lower prices

We find that retail prices decline as activity levels increase (see Table 5.3). In Uganda, a 10-percentage-point difference in the change in mobility leads to a 1.5% decline in prices. In Kenya the estimated effect size is much smaller: prices respond only with a 0.3% to 0.5% decline. Here, we cannot always reject the null hypothesis that the coefficient is indeed zero at the 10% level. This mainly concerns the specification where we rely on variation in changes of mobility patterns across space and within the same month.[17] The coefficients for Uganda are very close to each

[17]Results are largely unchanged if we include week or day fixed effects. Results are available on request.

other in size, irrespective of the type of variation we rely on to estimate it.

Next, we swap out the measure for activity at workplaces for the Grocery and Pharmacy Index and the Retail and Recreation Index for the restricted sample of administrative units for which the two measures are available.[18] To check for the role of sample selection, Figure 5.6 further shows the results for the Workplace Index for the subsample where the other two are available.[19] The estimated effect size increases for both of the alternative mobility measures - with a few qualifications. Relying on the Retail and Recreation Index, a 10-percentage-point decline in mobility is associated with a 1.5% increase in prices in Kenya and a 2.1% increase in Uganda (note the log scale). The results indicate that the retail price index picks up different relevant variations in prices that are not associated with workplace mobility patterns, in particular during the holiday season. However, we cannot reject that the coefficients are indeed the same as for the Workplace Index. Relying on within location variation, the suggested coefficient for Grocery and Pharmacy is a reasonably tightly estimated zero. In both Kenya and Uganda can we rule out that this result is primarily driven by sample selection as the workplace coefficients remain stable for the subsample.

[18] We hold the sample constant with respect to the availability for the Retail and Recreation Index.
[19] The sample size drops by 28% in Uganda and 33% in Kenya.

Figure 5.5: Using alternative measures of mobility



The graph reports the results from estimating Equation 5.1, where we regress the log price on different variations of the Google Mobility Index while controlling for item, brand, and month or location fixed effects. In the column for the Workplace Index, we consider only the subsample of locations for which we also observe the Retail and Recreation Index on the same day. The Google Mobility Index captures the level of activity at a specified category of locations relative to a baseline period in January and February 2020. Standard errors are clustered at the item-county (Kenya) or item-district level (Uganda).

We summarize results from a number of robustness checks in Figure 5.C1 and Figure 5.C2. We first look at the robustness of the results in Kenya to including monthly Nairobi 5-year average prices for the same year drawing on FEWS data for a subset of five items. The second pair of coefficients in Figure 5.C1 shows that the effect is either precisely zero or not significantly different from zero if we only consider cooking oil, maize flour, milk, and bread. The estimates are unchanged if we include the 5-year average prices in the third pair of coefficients.

We then explore the sensitivity of the results with respect to the initial lockdown period (see Figure 5.C2). In Uganda, this extends until August, in Kenya until July for major metropolitan areas. For Uganda, it also coincides with the scale-up of our data collection. We hence call the specification "scale-up" as we cannot rule out a lockdown versus a data collection effect here. Unsurprisingly, the estimated coefficients are now much smaller if we rely on within-location variation in mobility

as mobility patterns are more stable after the lockdown (see Figure 5.3). However, the coefficients still show a negative trend. Further, including outlier prices in the sample does not alter the results. Given the variation in prices across shop types, we lastly include shop type fixed effects. In Kenya the coefficient becomes more clearly indistinguishable from zero, while the coefficient in Uganda is still significant, negative, and sizeable.

By relying on a crowd-sourcing approach, a key concern is that mobility also drives the number and type of observed price quotes. In the following two sections we explore the role of potential composition effects as drivers of the findings.

5.5.3 How relevant are changing shop patterns and product baskets?

Adding shop type fixed effects to the baseline specification does not account for a potential bias that originates in selection. For example, supermarkets are less common in largely rural counties/districts, but these are also areas where mobility was impacted to a lesser extent than in more urbanised counties/districts. The same might hold true across time, in the sense that smaller retail outlets might have been more likely to close down when activity levels were at a minimum. A similar composition effect might hold true for branded versus non-branded products. Non-branded items are often more perishable and tend to be sold in open air markets and smaller shops, which were more likely to close down at the beginning of the pandemic.[20] In addition to being more storable, branded items often follow a national pricing strategy (DellaVigna and Gentzkow, 2019) - our motivation to distinguish between branded versus non-branded items rather than by degree of perish-ability.[21] Figure 5.C3 plots the share of price quotes observed for each product and shop type over time. The broad composition of the price basket is very stable over time with respect to the two outlined categories. Supermarkets and general shops account for 70-80% of the number of price quotes in both countries. Kibandas and stalls account for

[20]These include mostly fresh produce: cabbage, tomatoes, eggs, bananas, avocados, beans, and groundnuts.

[21]Milk is one of the few items that would switch categories based on the alternative sample split. It does not feature in the basket for Uganda. In Kenya, we find milk prices to be among the group of prices that vary the least across locations. Most price quotes are reported for a small set of popular brands that dominate the national market.

20%. Open-air markets account for the residual. In both countries, the majority of items are branded, with branded items being more common in Kenya. To account for the role of sub-national trends, we further regress the county/district-level weekly share of price quotes from each category on the Mobility Index in Figure 5.C4.[22] In Kenya, branded items were recorded more if mobility was lower, while non-branded items account for a larger share in times of higher mobility. The point estimates for Uganda are basically zero. However, in both contexts do we observe a shift from supermarkets to smaller shops, stalls, and open air markets as mobility rises. The effect size is very small, but significant in Kenya (close to significant in Uganda). This finding is in line with the notion that smaller smaller stores and open air markets were more affected by the pandemic. They took longer to adjust to policy measures and have operations that are likely more vulnerable to supply and demand disruptions. The adjustment of the observed basket of prices is both a strength and a weakness of our data collection. Rather than looking at a static set of prices and retail outlets, our price basket adjust with changes in the economic environment. It captures both direct changes in consumer behaviour, e.g., to avoid infections, but also indirect changes via the heterogeneous effects on retail outlets. The U.S. Bureau of Labour Statistics (BLS), for example, highlighted the fast changing consumption patterns as one of the key challenges they faced in their attempt to collect accurate price data and measure effective inflation.[23]

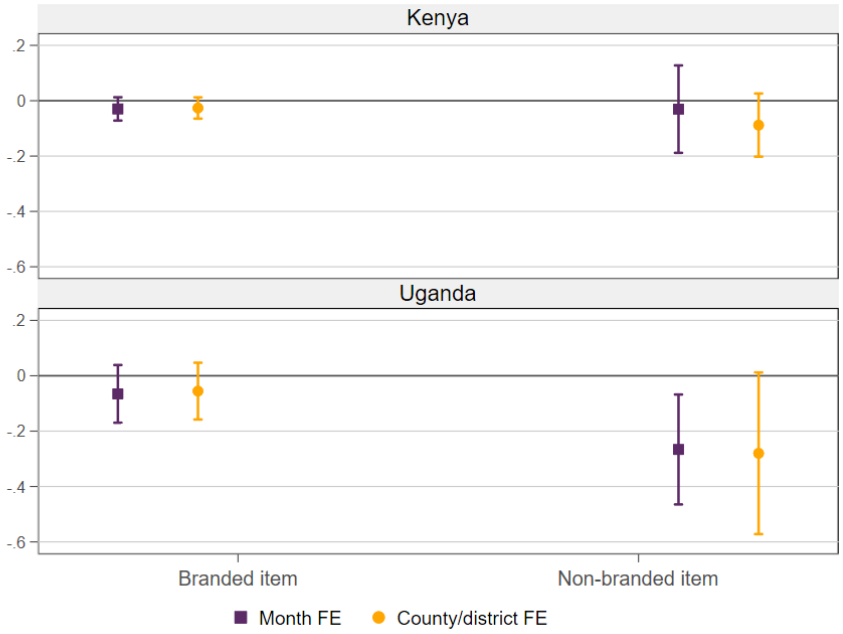
To partly address the potential sample composition effect, we re-run the main specifications separately for each product and shop type in Figure 5.6. In Kenya, the negative association of mobility and food prices is mainly driven by kibandas and small stalls. In each category, the standard errors are too large to reject the null hypothesis that the effect is zero, likely due to insufficient sample sizes. In Uganda, the effect is more detectable across all categories of shops. We do not find much heterogeneity in the results if we consider branded versus non-branded items separately.

[22]When considering the share of different retail outlets we further condition on the item. The unconditional results are very similar.

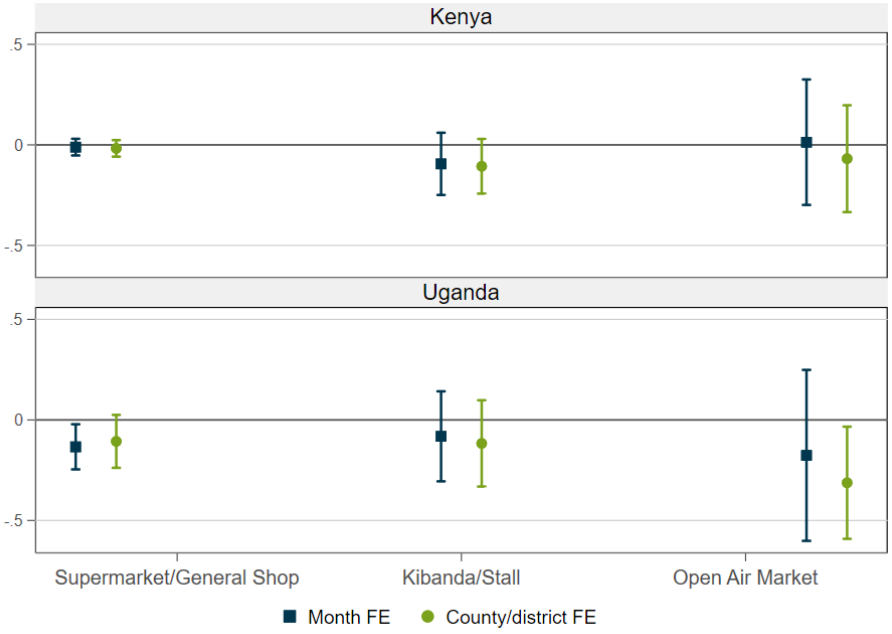
[23]https://www.bls.gov/opub/mlr/2020/article/the-impact-of-the-covid-19-pandemic-on-food-price-indexes-and-data-collection.htm

Figure 5.6: Exploring the role of shop and product types

Branded versus non-branded items



Different types of retail outlets



The table reports the results from estimating Equation 5.1, where we regress the log price on the Google Mobility Index (workplace) while controlling for item, brand, and month or location fixed effects as specified in the legend. The Google Mobility Index captures the level of activity at a specified category of locations relative to a baseline period in January and February 2020. Standard errors are clustered at the item-county (Kenya) or item-district level (Uganda).

5.5.4 Does the data collection vary with mobility?

As a final step, we look at the relevance of mobility for the number of price quotes collected. We will underestimate the effect of mobility on prices, if we observe fewer prices during times when people are more inclined to stay at home and at the same time prices are indeed higher during these times. On the contrary, if more price quotes are recorded during times of a more pronounced reduction in activity, we expect to observe relatively more higher prices quotes and thus potentially overestimate the impact of mobility on prices. This concern might be valid given that we collected more price quotes right after the start of the scale-up in Kenya (May/June 2020) and Uganda (August/September 2020) and slightly fewer in subsequent months due to attrition. In addition, volunteers might have had more time available to fill out the survey during times of the reduced activity.[24]

To test for the endogeneity of our data collection, we check whether the probability to observe a price quote is correlated with our measure of the change in activity levels. We first construct an indicator variable capturing whether or not a price quote has been observed for a given item on a given day in a given county/district. The variable takes a value of 1 if we do not observe at least one price quote in a specific item-day-location cell. We then regress the dummy variable on the Mobility Index controlling for item fixed effects (see Table 5.C1). We find that changes in mobility are indeed correlated with the availability of price quotes. We collected fewer price quotes for days and locations where mobility has reduced less drastically. A 10-percentage-point decline in mobility leads to a 1 percentage point reduction in the probability of observing a price quote in a specific item-day-location cell in Kenya (Uganda). In the first few weeks of the data collection, the number of price quotes we observed was skewed towards Greater Kampala and Greater Nairobi. Both metropolitan areas saw a more drastic decline in mobility than the rest of Kenya and Uganda (see Figure 5.A2 and 5.A1). As expected, this effect is more pronounced in the later periods where we observe the vast majority of the price quotes in locations outside the metropolitan areas. Due to structural differences between locations, e.g., because of differential price levels or differences in localities for which the Mobility Index picks up activities,

[24]While we collected information on product availability through the price survey, the data from this module is rather sparse and barely detects any availability issues.

we control for administrative unit fixed effects (counties/districts) in Table 5.C2. By including location-related fixed effects and solely relying on variation across time, we find a negative relationship between the change in mobility patterns and the probability of missing observations (except for the restricted sample in Kenya). A specification with location fixed effects might thus be biased towards observing lower prices more often, likely due to the roll-out of our data collection. Using month fixed effects instead accounts for this by using the variation in mobility changes within each month and across localities. Applying this specification, the suggested link between changes in mobility and the probability of not observing a price quote is again positive - this time slightly smaller in Kenya, but larger in Uganda (Table 5.C3).
In a nutshell, the likelihood of observing a price quote is indeed correlated with changes in mobility patterns. However, the direction of the bias depends on the type of variation used to estimate the relationship. Given our headline results linking price dynamics and mobility patterns remain unaffected by the type of variation used to estimate the relationship, we are less concerned about this type of bias being the main driver of the results.

5.6 Discussion and conclusion

Given their importance for food security, food price dynamics have been a major concern for policy makers as the COVID-19 pandemic unfolded across the globe. Our results provide insight into the impact of the COVID-19 shock on food prices in Kenya and Uganda, as well as into the general nexus between price dynamics and mobility patterns.

Like prices, activity levels at workplaces, grocery stores and retail locations simultaneously pick up supply and demand side dynamics. On the supply side, the effects could stem from (i) a reduction in working hours, (ii) business closures in the wholesale and retail sector, and (iii) retailers having to increase their margins to cover increased costs. While we cannot determine the relative importance of these channels, related studies suggest that business closures in the retail sector were widespread during the COVID-19 period. This is particularly the case for smaller, household-run businesses that dominate the retail sector in Kenya and Uganda.

While only 4% of the Kenyan household-run enterprises surveyed by Pape et al. (2020) report to have closed down permanently by June 2020, 40% of these operate in the retail and wholesale sector. Mahmud and Riley (2020) find that even by September 2020, 43% of all household-run businesses in Uganda remained closed.
An alternative explanation is that demand for food items was stronger during the initial phase of the lockdown period as households prioritised food purchases over expenditures on durable goods. As the economic impact of the pandemic persisted, stockpiling might have declined and relative consumption levels reduced. Of course, this does not, in turn, rule out a feedback loop of weaker demand into the decision for retailers to remain permanently closed.

Finally, the data collection serves as a proof of concept showing that it is possible to collect high-quality, high-frequency price data at scale through an innovative crowd-sourcing approach. It can be particularly useful to acquire price information with broad spatial coverage in fast changing environments that alter consumption and shopping behaviour. A limitation of our study is the challenge of collecting data in regions with limited Internet connectivity.A more targeted approach to recruiting volunteers in relevant areas might be one way of addressing this issue in future research endeavours. Additionally, for items such as fruits and vegetables, where more crude measures have to be used instead of standardised measures, it is important to keep in mind the limitations of the data for interpretation. We find that price trends for fruits and vegetables, nevertheless, closely align with those collected through traditional channels by the Uganda Bureau of Statistics and the Kenya National Bureau of Statistics.

Acknowledgements for Chapter 5 We are immensely grateful for Mdoe Jackson (Kenyatta University)'s support in the early stages of the project. Elizabeth Wambua and Peter Atwine provided excellent research assistance. We are also grateful for the support of Mary Mang'eli (Kenyatta University and Laikipia University), Priscah Mukii David (Garissa University), David Kamau Karienye (Garissa University), Isack Nyakach (Karatina University), and Urbanus Wambua (Karatina University) in helping us reach out to a fantastic group of student volunteers in Kenya. Our heartfelt thanks to Richard Sebaggala (Uganda Christian University) who helped us reach out to a fantastic group of student volunteers and supported the survey launch in Uganda. We are also grateful for Aisha Jore Ali (University of Oxford), Paul Lakuma (Economic Policy Research Centre), Paul Matovu, and Ameerah Anathallee's (both Vertical and Micro Gardening Uganda) help along the way, especially with reaching a broader pool of volunteers in Uganda. We thank the Uganda Economics Association, the Economic Policy Research Centre (EPRC), TMG research, and Innovations for Poverty Action (IPA) Kenya and Uganda for their support in distributing the link for the online survey. We further thank Mahreen Mahmud (University of Exeter) and Emma Riley (University of Washington) for integrating our price survey questions in an ongoing data collection in Kagadi and Kyenjojo district, and Rose Page and the team at the Centre for the Study of African Economies for invaluable administrative support. We are grateful for Ingo Kleiber (University of Cologne)'s support with the coding of survey components in Java. We thank Lena Anayi (Bank of England), and Mahreen Mahmud for comments on earlier drafts of the paper. We are grateful to the members of the Jeune Street Seminar Series for their invaluable contributions and insightful discussions during the development of this research. We thank Lisa Martin (Oxford) and Raphael Bradenbrink (Oxford) for extensive comments and feedback on this draft. We thank Christopher Woodruff for his guidance and support throughout this endeavour.

The data collection in Kenya was made possible thanks to the outstanding contributions from over 150 volunteers. Special thanks to Faith Mutinda, Felix Ongori Mbaka, Brian Kwanya, Mayaka Jackson, Mary Maina, David Ochieng, Sang Stacey Jebet, Elizabeth Mwikali Wambua, Teresia Wangeci Thiong'o, Mwita Selvenus Gesabo, Lynette Wangeci, Frankline Nisah, Juliet Kemunto Marigwa, Lemiso Patrick Kayioni, Samuel Muiruri Thiongo, Kelvin Mutwiri, Elijah Mbila, Gideon Kipkirui Yegon, James Ngumo Kanji, Billy Osborn Atetwe, Bronson Mwiti, Achaya Everlyne, Maina Albert Chrispin, Jus-

tus Munywoki, Mdoe Samson, Brenda Farida Kemunto Mokua, Mark Lubanga, George Kwanzu, Domnick Oyoo, Victor Okello Sirama, James Odhiambo Ochieng, Janet Mukonyo Kioko, William Musya Kitheka, Cheruiyot Langat, Janet Kipchilis, Brian Mutua, Timothy Kemboi Kipruto, Inviolata Lusweti, Phabian Odhiambo, Sayydd Were, Baserecha Lavender, Antony Ngumbi, Andrea Mwalaga, Hillary Baraka Muuga, Phylis Wangoto, Elizabeth Akinyi Odhiambo Ondego, Tilen Ododa, Desmond Juma, Titi Kibitok Enock, Cosmas Owuor, Alphonce Ochieng, Olivier Masengesho, Gabriel Keith Odunga, Moses Kimani Njoroge, Veronicah Ndungu, William Onura Akwanyi, Denis Kipkoech Ngetich, Precious Kirigo, Grace Wangari Ndungu, Fostina Mirenja, Robert Wafula Odhiambo, Langat Kibet Gideon, Violet Kabuga, Antony Ngumbi, and Augastine Esrom.

The data collection in Uganda was made possible thanks to the outstanding contributions of over 100 volunteers. Special thanks to Peter Bith Manyang Rieth, Onapa Ambrose, Grace Kirabo, Mwebaze Noel, Nalumansi Aisha, Eoju Winnie, Atim Chelsea Okellowange, Ahimbisa Collins, Brian Kali, Kasozi Venansio, Fred Wekesa, Kasozi Venansio, Patricia Joan Nagalwa, Dalia Salim, and Grace Nabitaka.

We gratefully acknowledge funding from the Centre for the Study of African Economies (CSAE) at the University of Oxford and the International Growth Centre (IGC). The project has been reviewed by, and received ethics clearance through, the Research Ethics Committee at the Uganda Christian University and the Departmental Research Ethics Committee for Economics at the University of Oxford (reference # ECON1A20-21-10).

Appendices of Chapter 5

5.A Appendix material for data and sample description

5.A.1 Google mobility data

We use the Google Mobility data to track changes in mobility patterns during our study period. Related studies often rely on fine-grinned high-frequency smartphone data to track changes in consumption patterns and mobility during the COVID-19 pandemic (Couture et al., 2021). However, such data are only available for a limited number of contexts. We therefore rely on more aggregated mobility data published by Google (Aktay et al., 2020). The data has, for example, been used by other researchers to provide cross-country evidence on the link between mobility and the transmission of SARS-CoV-2 (Camehl and Rieth, forthcoming). Similar to other smartphone data, the data is likely bias towards picking up mobility patterns for certain demographics - typically the younger and wealthier in the context of Kenya (Gollin et al., 2020) and Uganda. Despite the skew towards higher income groups and higher educated, Gollin et al. (2020) provide detailed insight into smartphone ownership and mobility patterns in Kenya, showing that especially in urban areas smartphone data gives insights into mobility patterns of a relatively broad subset of the wider population. Nevertheless, we want to highlight that this skew is be particular relevant for the purpose of our study given those groups might also have been better positioned to work from home during the lockdown period and their mobility patterns have been affected more drastically as a result.

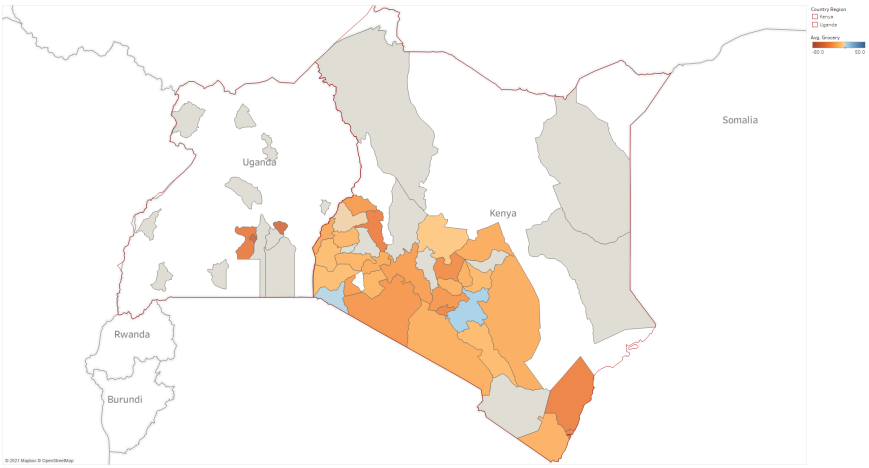
The Google Mobility Index documents how Android smartphone users' movement and time spent at certain locations has been affected by the pandemic relative to a baseline period in January and February 2021. Figure 5.A1 depicts the changes in activity

at grocery and pharmacy-related locations for April and November 2020. The maps are color coded such that dark red colour means a reduction in activity of up to 80% relative to the baseline. Dark blue colour means an increase in activity of up to 50% relative to the baseline. To some considerable extent, activities reduced significantly in most parts of Kenya for which Google has data on. Counties that were targeted by the lockdown measures are seen to reduce activities the most. Data coverage in Uganda is more spares and largely centred around the major urban centres, in particular Greater Kampala. Here the lockdown had the most pronounced impact (dark red region in the top map of Figure 5.A1). In November of 2020, there was improved movement in Kenya even though the coastal region with Mombasa remained constrained.[25] One limitation is that data availability is highly correlated with Internet and mobile phone coverage. Google requires a certain number of data points for it to publish an index for a specific administrative unit. For the blank regions in we do not observe any data throughout the sample period. In Kenya this mostly affects the sparsely populated counties, while large areas of Uganda are not observed in the data. For the grey areas we observe data for at least one of the indices in a given month, but not the one mapped here. Figure 5.A2 highlights the more comprehensive geographic coverage of the Workplace Index. Looking at the Workplace Index in Figure 5.A2, we find even more convincing evidence that workplaces were significantly affected in April of 2020. This was at a time when non-essential workers were advised to work from home. By June of 2020 there was considerable improvements in mobility in some counties/districts. By February 2021 we observe fewer extreme values - darker shades of red or blue - and most values fluctuate around zero. However, the overall tendency still seems to suggest lower activity levels than in January and February 2020. In Uganda, this trend might be partly driven by the aftermath of the January 2021 election.

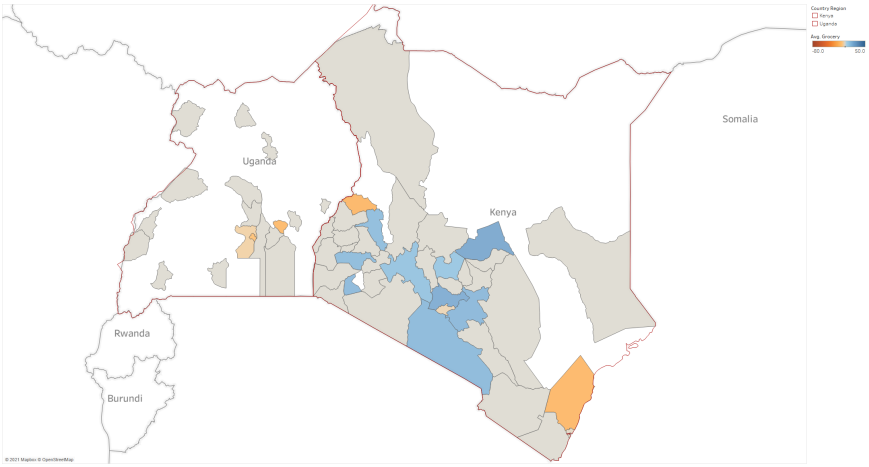
[25]This could be partly driven by the drastic reduction in tourist numbers on the coast.

Figure 5.A1: Mapping mobility: Grocery and Pharmacy Index

April 2020



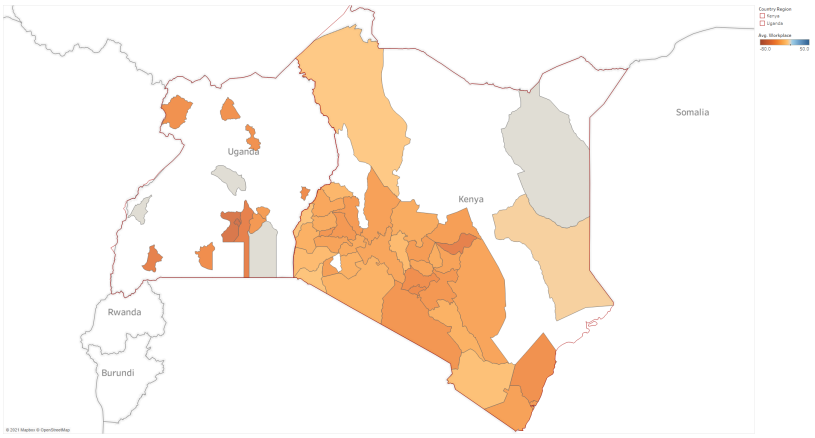
November 2020



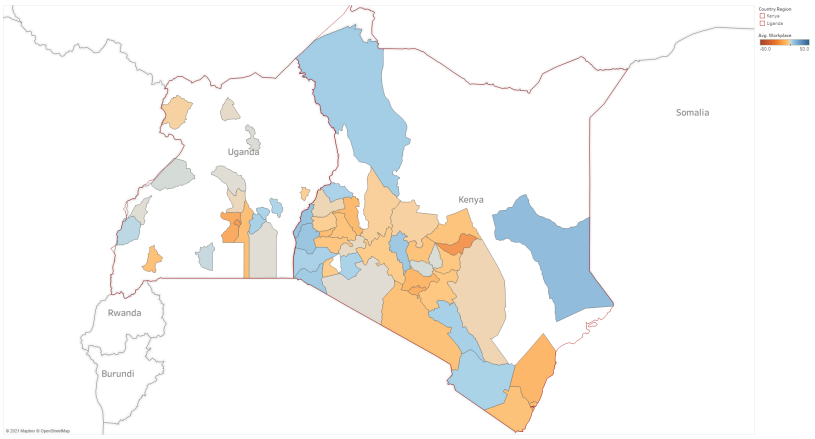
The maps shows the spatial variation in the availability of Google mobility data in Kenya and Uganda for the month of April and November 2020 as well as February 2021. Districts/counties marked in grey are administrative units for which any mobility data is available at some point, however, not in the shown period. Blank areas indicate that mobility data are missing for the time period studied. The colour scheme further indicates the percentage change in activity relative to Google's baseline in January and February 2020. Dark red means a reduction in activity up to 80% relative to the baseline, dark blue means an increase in activity of up to 50% relative to the baseline. The lighter the colour, the closer mobility is to the baseline. The Google mobility data can be downloaded here: https://www.google.com/covid19/mobility/

Figure 5.A2: Mapping mobility: Workplace Index

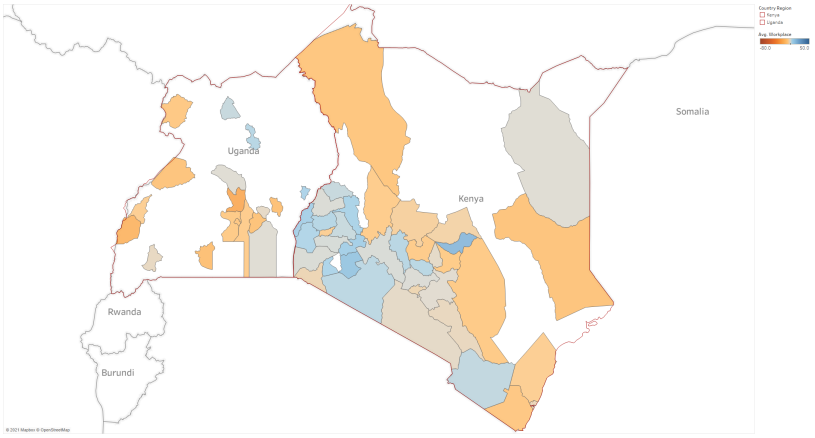
April 2020



June 2020



February 2021

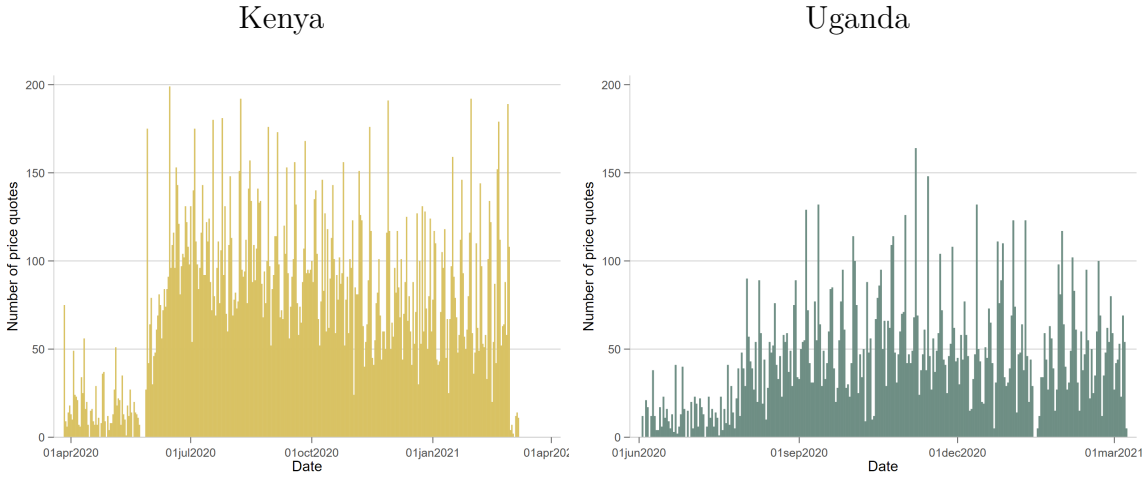


The maps shows the spatial variation in the availability of Google mobility data in Kenya and Uganda for the month of April and November 2020. Districts/counties marked in grey are administrative units for which any mobility data is available at some point, however, not in the shown period. Blank areas indicate that mobility data are missing for the time period studied. The colour scheme further indicates the percentage change in activity relative to Google's baseline in January and February 2020. Dark red means a reduction in activity up to 80% relative to the baseline, dark blue means an increase in activity of up to 50% relative to the baseline. The lighter the colour, the closer mobility is to the baseline.

5.A.2 Price survey

Basic descriptives

Figure 5.A3: Daily number of collected price quotes



The above histogram plots the daily number of collected price quotes. The regular spikes correspond to weekends - in particular Saturdays.

Table 5.A1: Kenya - detailed summary statistics

Item	Quantity	N	N/month	Avg. price	Sd price	#brands	#counties	#counties/month	N mobility	Avg.price mobility	Sd price mobility	#counties mobility	N origin	Avg. distance to origin	Shop type
Rice	2,072	1 kg	156	123	44	11	39	23	1,841	124	44	34	985	168	General shop
Bread (white)	2,061	400 g	153	49	6	15	38	22	1,823	49	6	33	1,713	175	General shop
Sugar	1,556	1 kg	109	112	14	8	37	19	1,367	112	14	33	578	188	General shop
Eggs	1,556	one	163	13	3	1	42	28	1,352	13	3	36			General shop
Wheat flour	1,338	2 kg	96	128	15	13	39	19	1,166	128	15	34	1,192	352	General shop
Maize flour (sifted)	1,323	2 kg	98	121	18	17	39	19	1,182	120	17	34	1,066	175	General shop
Milk (fresh, packaged)	1,246	500 ml	88	51	9	14	39	19	1,135	51	9	33	1,010	167	General shop
Bread (brown)	1,104	400 g	82	49	7	11	40	18	1,001	49	7	36	878	133	General shop
Tomatoes	1,070	one	111	7	3	1	38	23	961	6	3	32			Kibanda
Banana (ripe)	967	one	100	8	4	1	39	24	850	8	4	33			Kibanda
Cooking oil	777	1 litre	52	170	39	10	40	16	687	170	39	34	562	373	General shop
Onions (bulbs)	770	1 kg	80	66	34	1	33	21	684	66	34	30			Kibanda
Cabbage	672	one	73	46	30	1	38	22	586	45	29	31			Open air market
Avocado	657	one	73	19	10	1	35	21	580	20	10	31			Kibanda
Mango	628	one	72	18	10	1	35	20	576	18	10	31			Kibanda
Soda/Soft drink	576	300 ml	64	30	8	1	37	21	506	30	8	33			General shop
Salt	352	1 kg	21	39	15	5	32	10	315	39	15	29	345	515	General shop

Prices are denoted in Kenyan Shilling.

Table 5.A2: Uganda - detailed summary statistics

Item	Quantity	N	N/month	Avg. price	Sd price	#brands	#districts	#districts/month	N mobility	Avg.price mobility	Sd price mobility	#districts mobility	N origin	Avg. distance to origin	Shop type
Rice	1 kg	1,224	264	3,740	1,751	9	58	5	819	3,982	1,991	14	160	65	General shop
Bread (loaf)	400 g	838	134	4,543	719	5	49	5	692	4,553	659	14	621	85	General shop
Beans (dry)	1 kg	771	132	3,587	761	2	48	5	511	3,657	770	14	0		General shop
Sugar	1 kg	761	190	3,357	520	7	51	4	480	3,376	541	13	671	143	General shop
Eggs	one	526	81	413	99	1	42	6	376	399	96	13	0		General shop
Avocado	one	425	70	649	323	1	41	5	306	711	344	15	0		General shop
Groundnuts	1 kg	348	45	5,700	1,190	1	37	5	266	5,849	1,202	14	0		General shop
Maize flour (sifted)	2 kg	333	76	2,020	567	11	37	3	225	2,084	602	11	245	86	General shop
Tomatoes	one	318	54	248	171	1	37	4	245	258	180	11	0		Open air market
Banana (ripe)	one	192	44	337	277	1	39	4	223	376	274	13	0		General shop
Salt	1 kg	182	45	2,128	1,040	6	37	2	129	2,074	1,024	11	3	102	General shop

Prices are denoted in Ugandan Shilling.

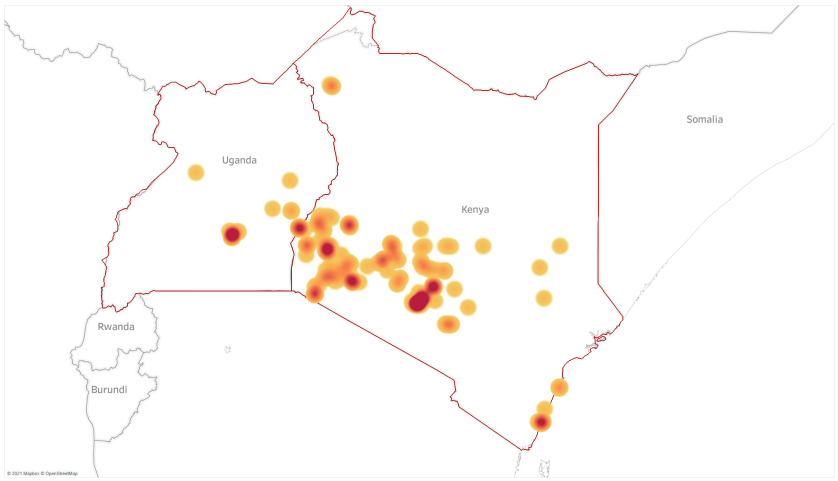
Geographic coverage of the price survey data

The number of collected price quotes highly correlates with population density. In Figure 5.A5 we plot population size for each district (Uganda) and county (Kenya) against the number of price quotes collected (both on a log scale). Each marker represents a district/county and their size is proportional to their population. Larger administrative units reported a higher number of price quotes.

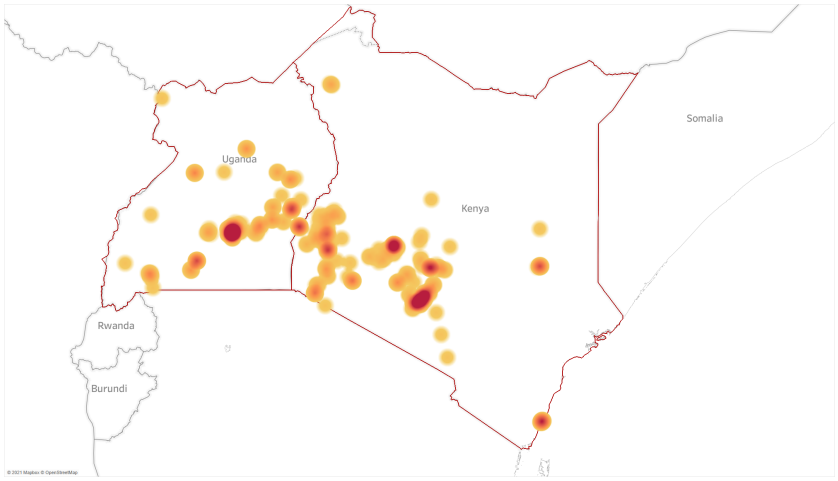
Using additional information from Kenya's latest census (KNBS, 2019a), we further link the number of collected price quotes to population density, mobile phone coverage, and Internet access. In Figure 5.A6, we show the correlation of population density and the number of price quotes per square kilometre, as well as the correlation of the number of price quotes per million people and mobile phone and Internet penetration in Kenya. Each marker again represents a county. The graph on the left panel takes in consideration the number of reported price quotes per square kilometre while in the right panel, the percentage of population with access to Internet/mobile phone is considered. We note that the fitted line is noticeably steeper than a 45° line implying a disproportionate higher concentration of price quotes in densely populated counties.

Figure 5.A4: Geographic dispersion of price quotes for rice over time

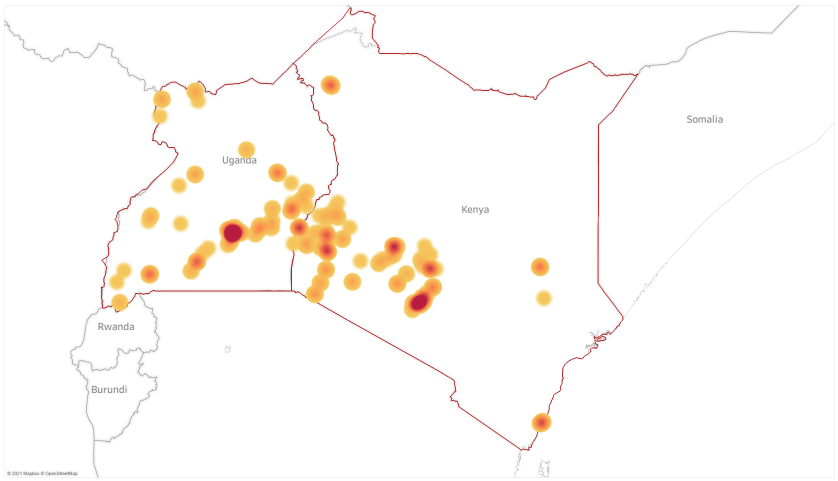
June



October

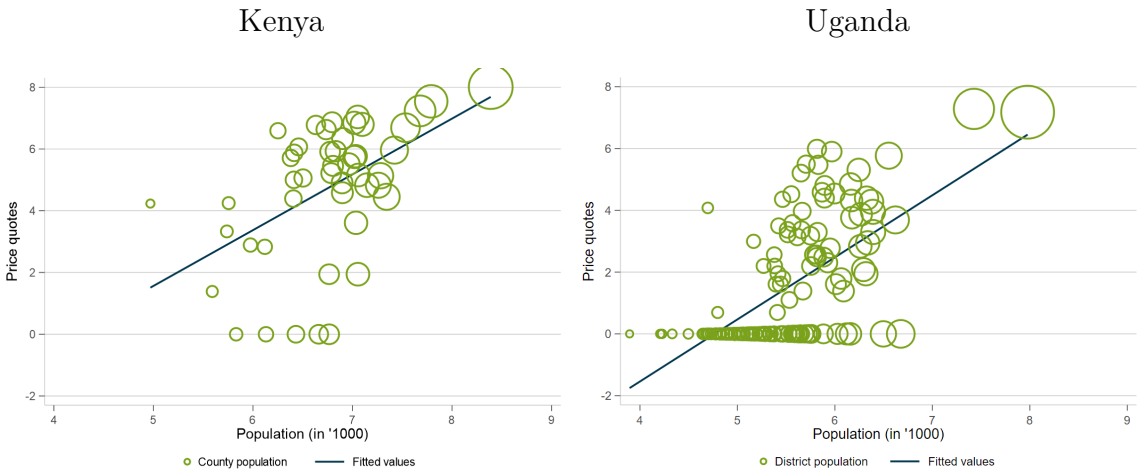


February



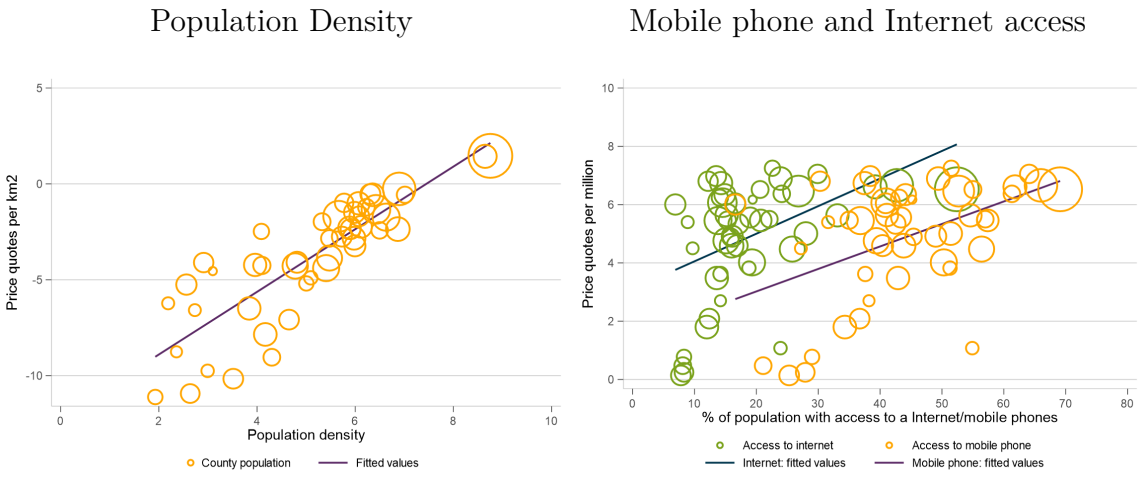
The above maps plot the geographic dispersion of the collected rice prices for the months June 2020, October 2020, and February 2021. Darker colours indicate a greater density of price quotes. Lighter colours indicate fewer price quotes.

Figure 5.A5: Population and the number of collected price quotes across space



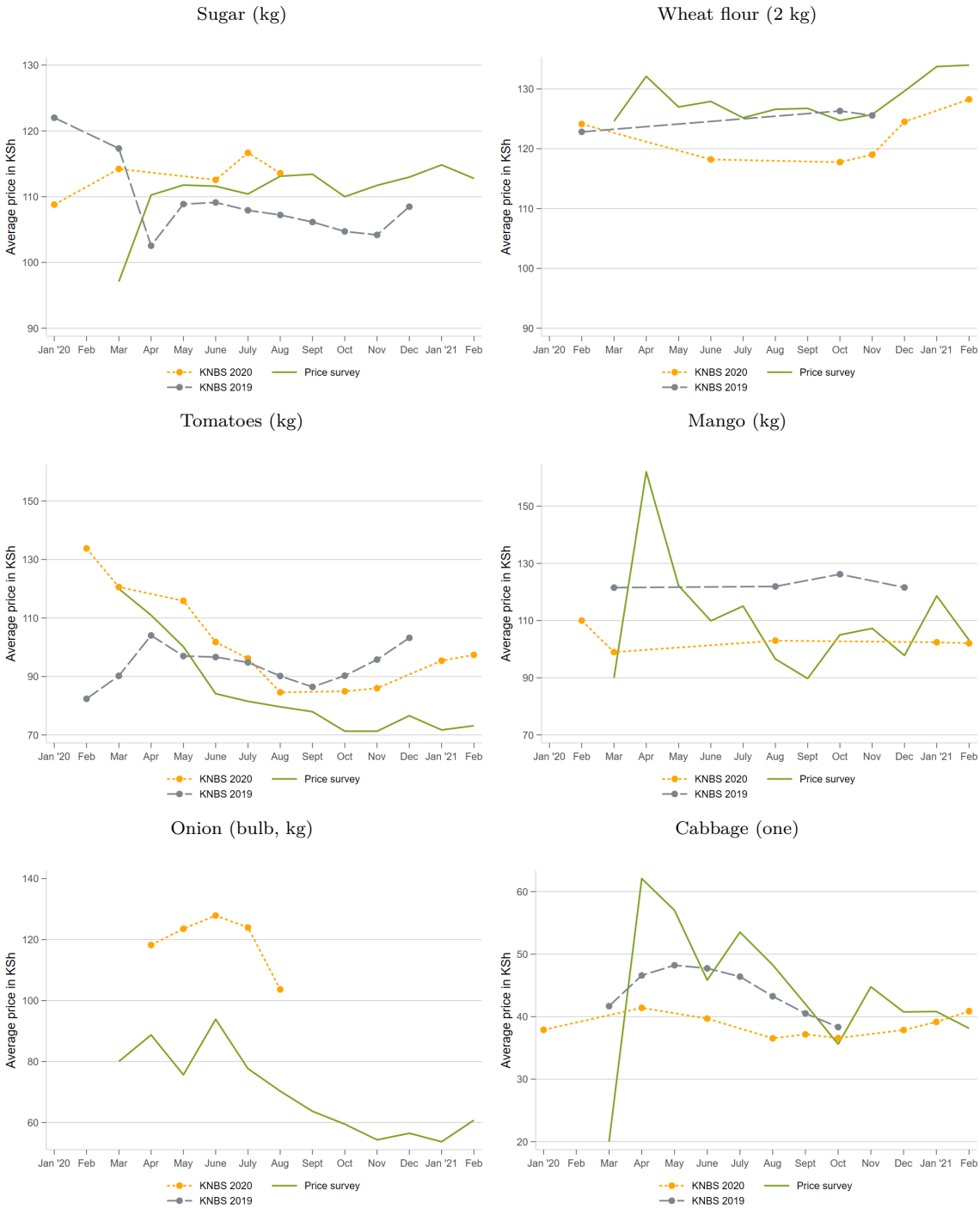
The above graphs plot log population on the x-axis and the log number of price quotes collected in each county (Kenya) and district (Uganda) on the y-axis. Each marker represents one of Kenya's 47 counties or Uganda's 135 districts. The size of each marker is proportional to the county's/district's population. The population data for Kenya are sourced from the 2019 Population and Housing Census (KNBS, 2019a). For Uganda the data are sourced from the Uganda Bureau of Statistics (UBOS, 2020). Population figures for 2020 are projections.

Figure 5.A6: Population density, Internet access, and the number of collected price quotes across space



The above graphs plot the log number of price quotes (per km^2 or million people) collected in each Kenyan county on the y-axis. In the left graph we plot it against population density and in the right graph against the percentage of the population with access to a mobile phone and Internet respectively. Each marker represents one of Kenya's 47 counties. The size of each marker is proportional to the county's population.The slope of the fitted line is slightly steeper than a 45° line, suggesting a dis-proportionally higher concentration of price quotes in densely populated counties. The population data and information on Internet access and mobile phone access are sourced from the 2019 Population and Housing Census (KNBS, 2019a).

Figure 5.A7: Comparison of national average prices with KNBS



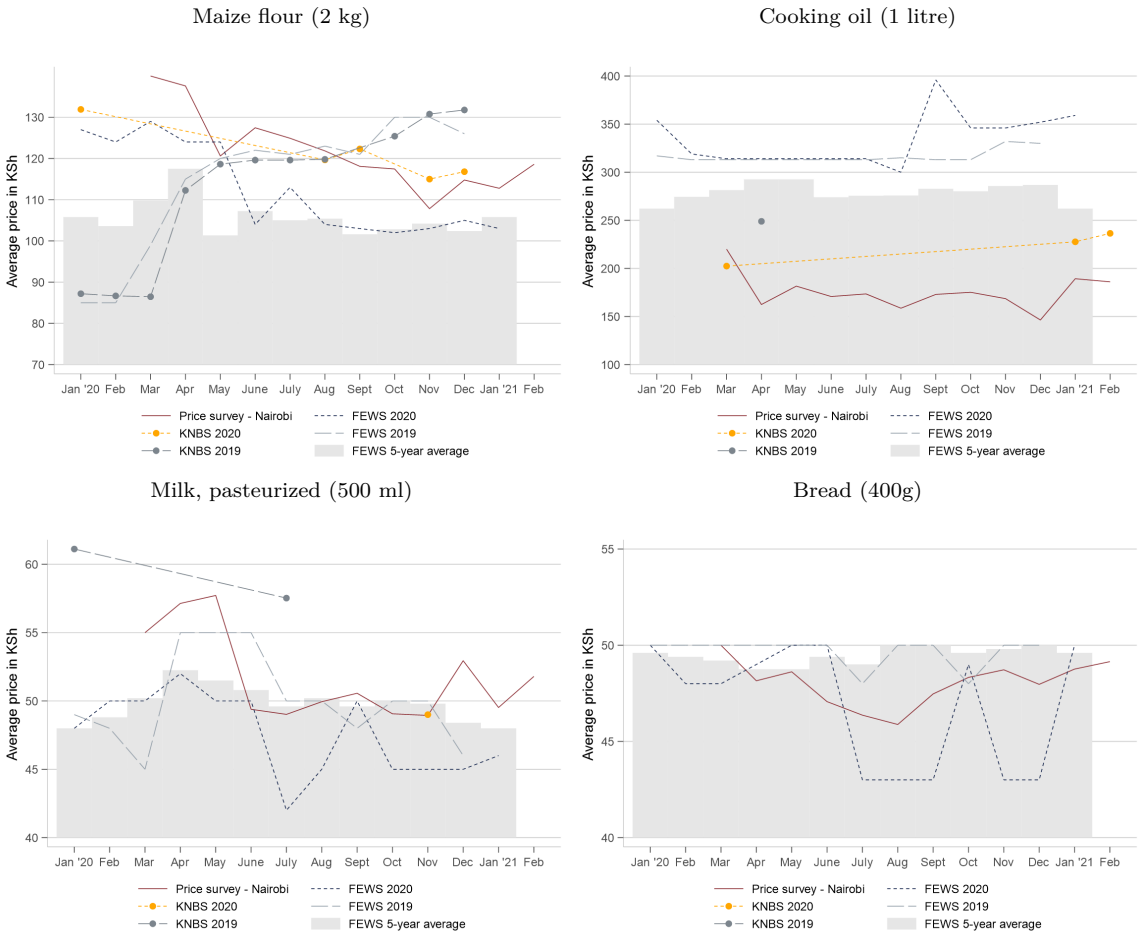
This graph compares the average prices in Nairobi from the price survey to average prices reported by the Kenya National Bureau of Statistics. Note that KNBS only publishes item-level national average prices for a selected number of items each month. In addition, the specific items reported vary from month to month. For the price survey we report the unweighted national average prices. The vast majority of respondents reported prices per piece for mangoes, cabbage, and tomatoes. We therefore standardise all prices accordingly. However, as KNBS report kilo prices only, we re-scale the price for mango and tomatoes for the purpose of this graph. KNBS reports prices for onion bulbs and leeks as one item. In the price survey we specifically collected prices for bulbs.

Figure 5.A8: Comparison of national average prices with FEWS and KNBS



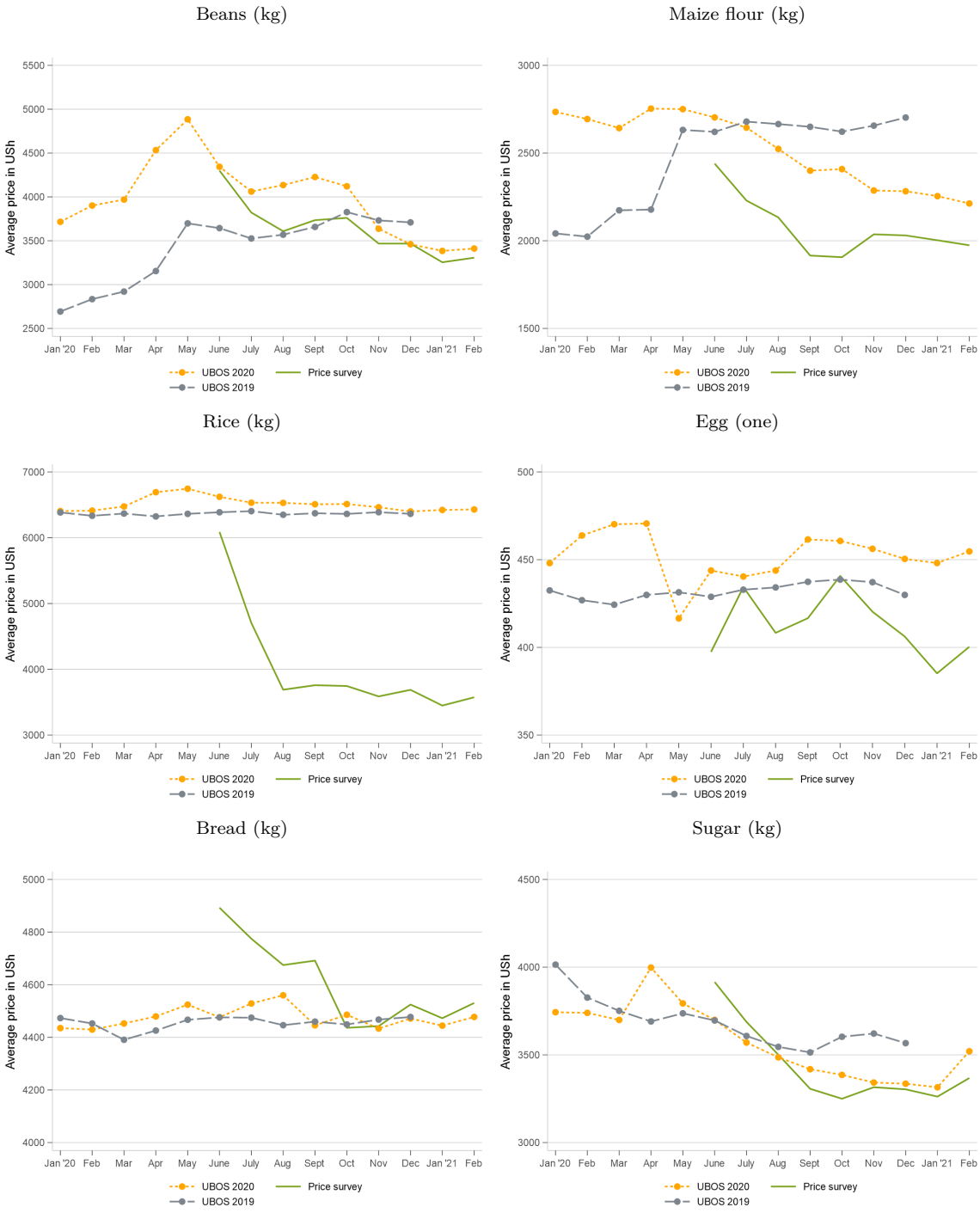
This graph compares the average prices from the price survey to average prices reported by the Kenya National Bureau of Statistics and the Famine Early Warning Systems Network (FEWS). Here we restrict the sample to items for which FEWS reports monthly prices for Nairobi (disregarding gasoline and diesel). Note that KNBS only publishes item-level national average prices for a selected number of items each month. In addition, the specific items reported vary from month to month. For the price survey we report the unweighted national average price.

Figure 5.A9: Comparison of Nairobi average prices with FEWS and KNBS



This graph compares the average prices in Nairobi from the price survey to average prices reported by the Kenya National Bureau of Statistics and the Famine Early Warning Systems Network (FEWS). Here we restrict the sample to items for which FEWS reports monthly prices for Nairobi (disregarding gasoline and diesel). Note that KNBS only publishes item-level national average prices for a selected number of items each month. In addition, the specific items reported vary from month to month. For the price survey we report the unweighted average price for Nairobi.

Figure 5.A10: Comparison of national average prices with UBoS I



The graphs compare national average prices from the price survey prices collected by the Uganda National Bureau of Statistics (UBoS). UBoS prices are collected from 10 centers across the country namely Arua, Fortportal, Masaka, Mbarara, Kampala lower, middle & higher income, Jinja, Mbale, and Gulu. In the UBoS data, prices for salt are reported per 500g while avocado, tomatoes, and bananas are reported per kg. For proper comparison with the price survey, we standardised all UBoS prices by rescalling to price per/kg and price/piece respectively. Bread was also standardised to 1 kg as it was reported in various forms. Nonetheless, the trend of banana prices was not so comparable as UBoS prices comprised of more and various types of ripe bananas compared to the price survey.

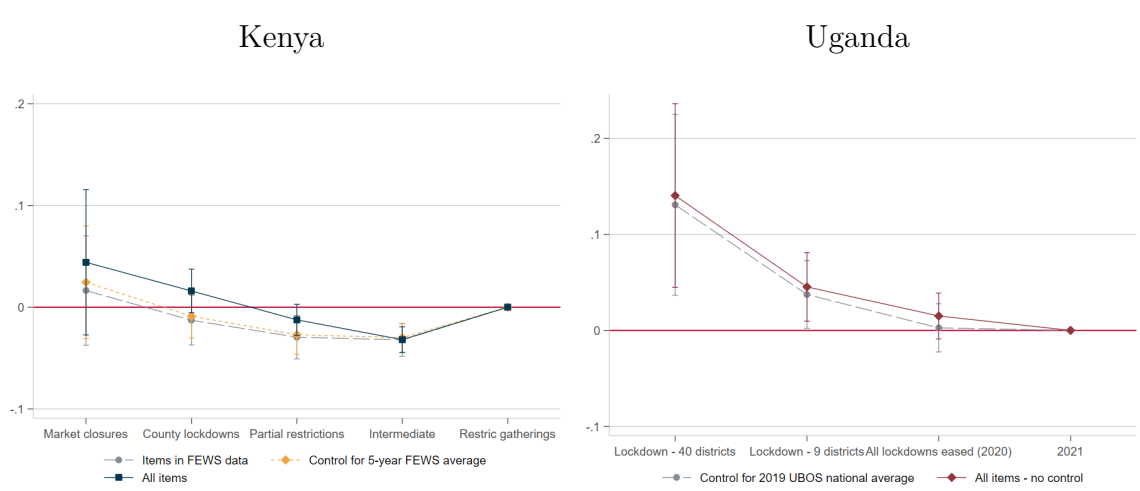
Figure 5.A11: Comparison of national average prices with UBoS II



The graphs compare national average prices from the price survey prices collected by the Uganda National Bureau of Statistics (UBoS). UBoS prices are collected from 10 centers across the country namely Arua, Fortportal, Masaka, Mbarara, Kampala lower, middle & higher income, Jinja, Mbale, and Gulu. In the UBoS data, prices for salt are reported per 500g while avocado, tomatoes, and bananas are reported per kg. For proper comparison with the price survey, we standardised all UBoS prices by rescalling to price per/kg and price/piece respectively. Bread was also standardised to 1 kg as it was reported in various forms. Nonetheless, the trend of banana prices was not so comparable as UBoS prices comprised of more and various types of ripe bananas compared to the price survey.

5.B Additional results for the event study

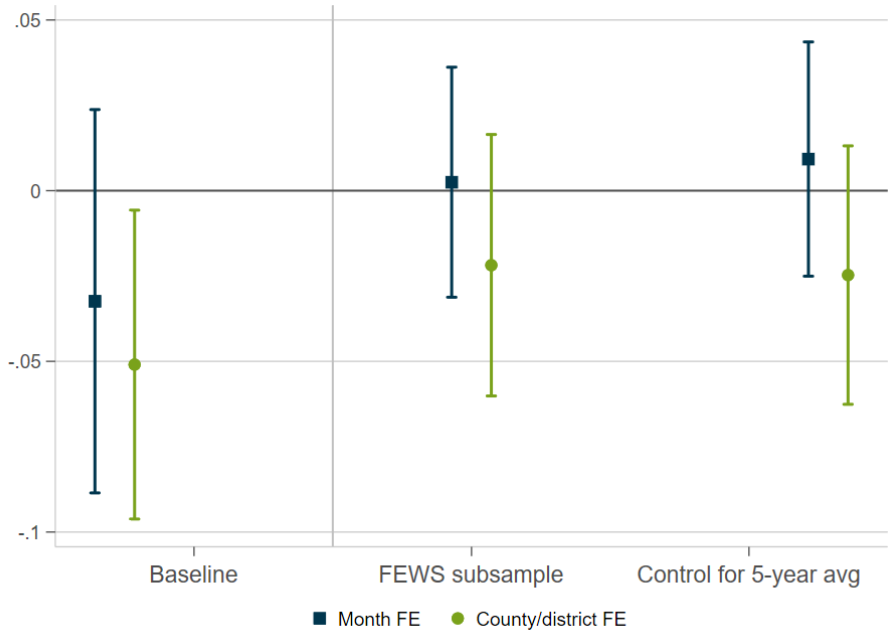
Figure 5.B1: Price dynamics and policy responses to COVID-19



In the above figure we report the results from estimating the following specification: $y_{itd} = \sum_{k=0}^{11} \tau_{-k}$ feb2021$_{t-k} + \phi_i + \eta_{ib} + \delta_c + \varepsilon_{itd}$ where y_{itd} denotes log price of item i on day d, ϕ_i the item FE, η_{ib} the brand FE, and δ_c the county/district FE. The coefficient for February 2021 is normalised to zero. Prices are winsorised at the 1% level. Standard errors are clustered at the item-district/county level. The error bars show the 95% confidence intervals. For Kenya, the data covers the time period between the last week of March 2020 until the first week of March 2021. For Uganda, the data covers the time period between the last week of May 2020 until the second week of March 2021. We further show the results for an alternative specification where we control for historical price trends. In Uganda, we control for the national average price in the same month in 2019 (grey plot), which is available for all items included in our sample. For Kenya, we show two alternative specifications. First, we restrict the sample to the four items for which historical FEWS data are available (grey plot). Second, we control for the 5-year average price in the corresponding month (orange plot). We divide the sample into phases with which government restrictions were lifted. We define the phases as follows: **Kenya:** The first phase (not shown in the plot) spans the period 27th March to 5th April and was marked by the first wave of COVID-19 response measures, most importantly market and restaurant closures. The second phase runs from 6th April to 6th July and marks the time when travel restrictions in and out of Nairobi and Mombasa (plus a few other counties) were in place. The third phase from 7th July until 29th September captures the period when major restrictions on the hospitality sector, domestic and international travel were sequentially lifted. During the intermediate phase only few restrictions, like the nighttime curfew, remained in place. In the final phase, starting 3rd January 2021, public gatherings and events were again prohibited. **Uganda:** The first phase spans the period 26th May to 29th June 2020 and was marked by the easing of COVID-19 response measures across Uganda. However, during this phase 40 border districts remained in lockdown. The second phase runs from 30th June to 23rd September and marks the time when only 9 border districts still had lockdowns imposed. The third phase runs from the date those final lockdown measures were lifted until the end of 2020, i.e. three months after the last lockdown. The final phase (not shown in the plot) combines all data points for 2021. The final split allows us to show the trend for some of the post-lockdown phase.

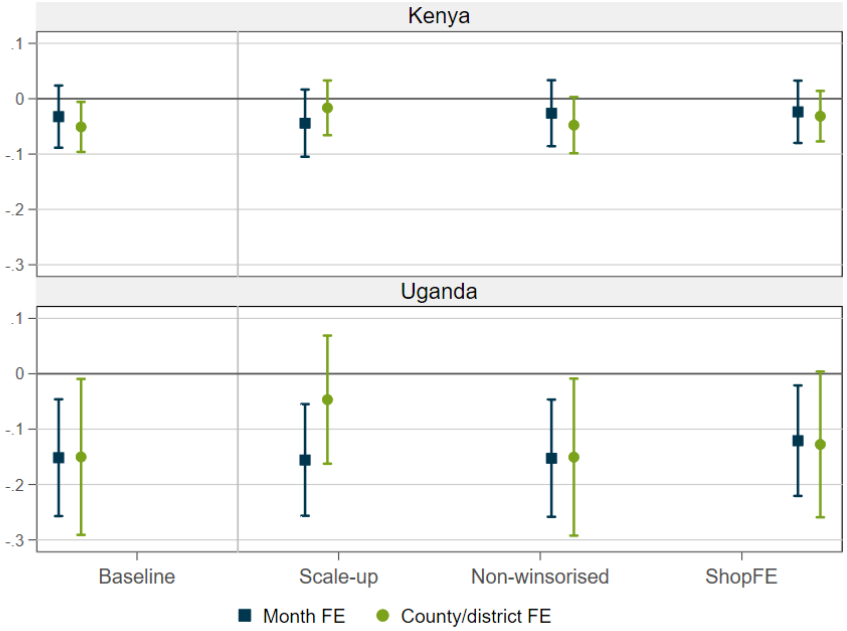
5.C Appendix material on the mobility and food prices nexus

Figure 5.C1: Kenya: Controlling for price trends in previous years



In the above graph we report the results from estimating Equation 5.1 where we regress the log price on the Google Mobility Index (workplace) while controlling for item, and brand fixed effects. Additionally, month or county/district fixed effects are included as specified in the legend. The FEWS data covers monthly average prices in Nairobi for four items: maize flour, cooking oil, milk, and bread (see Figure 5.A9). Standard errors are clustered at the item-county (Kenya). The Google Mobility Index captures the level of activity at a specified category of locations relative to a baseline period in January and February 2020.

Figure 5.C2: Alternative specifications: Price dynamics and mobility

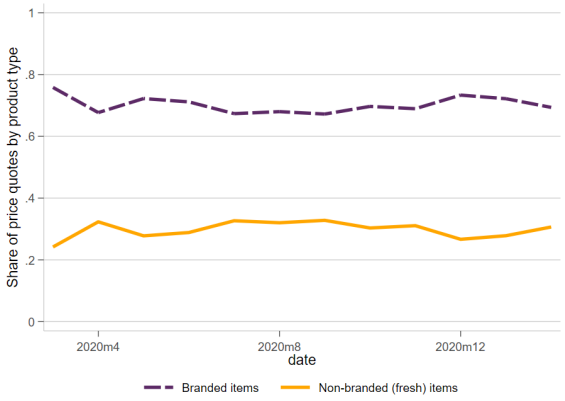


In the above graph we report the results from estimating Equation 5.1 where we regress the log price on the Google Mobility Index (workplace) while controlling for item, and brand fixed effects. Additionally, month or county/district fixed effects are included as specified in the legend. Standard errors are clustered at the item-county (Kenya) or item-district level (Uganda). The Google Mobility Index captures the level of activity at a specified category of locations relative to a baseline period in January and February 2020.

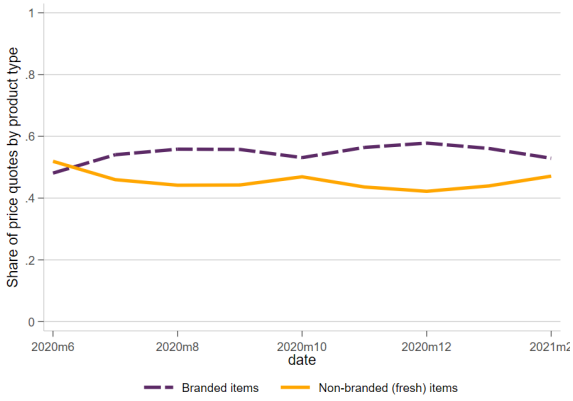
Figure 5.C3: Composition of the price basket over time

Branded versus non-branded items

Kenya

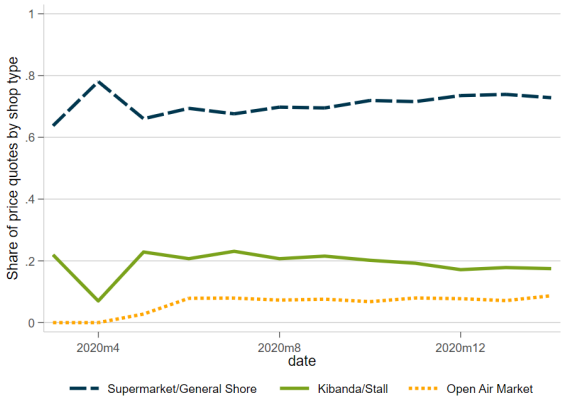


Uganda

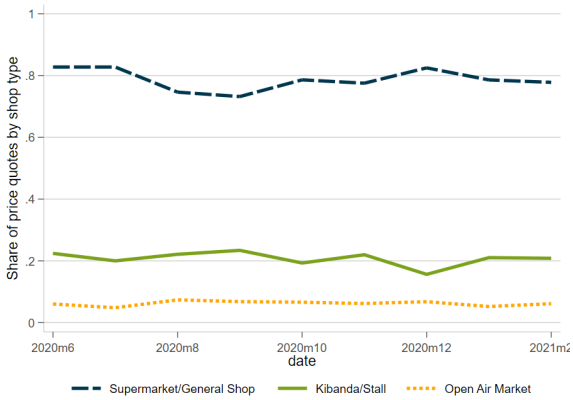


Different types of retail outlets

Kenya



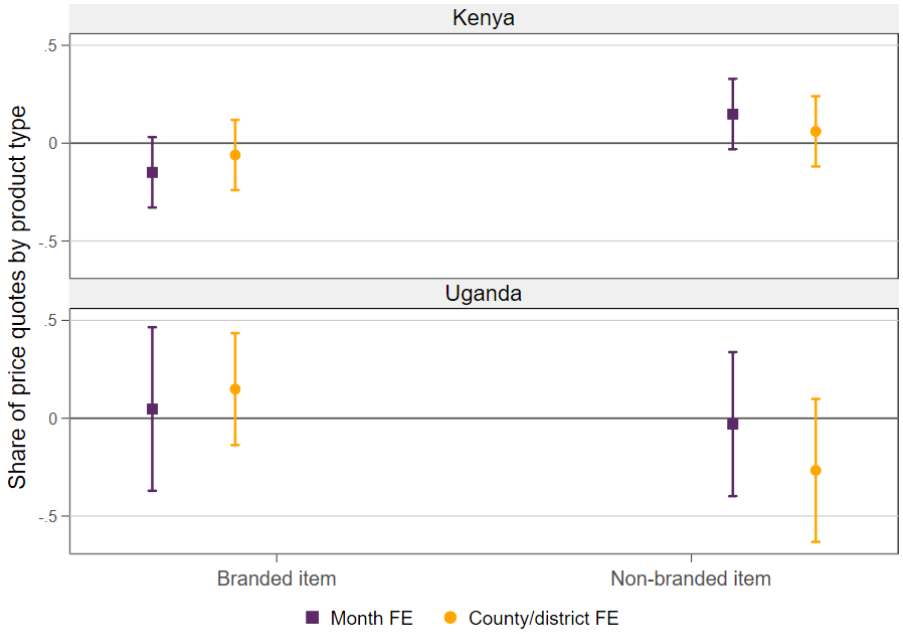
Uganda



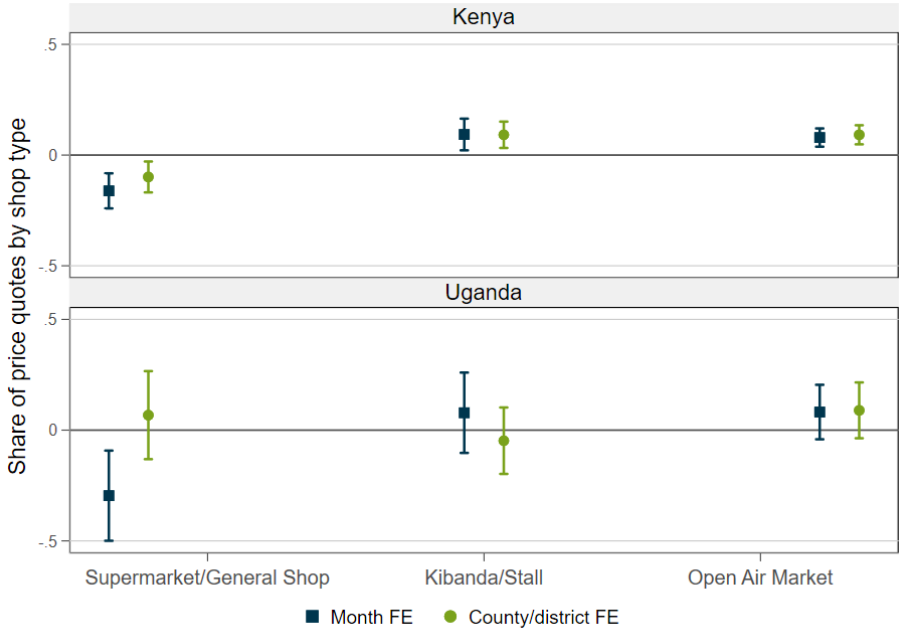
The graphs plot the monthly share of prices collected for each type of retail outlet and type of product.

Figure 5.C4: Mobility and the composition of the observed product basket

Branded versus non-branded items



Different types of retail outlets



In the above graph we regress the weekly share of each product and shop type on the Mobility Index ($s_{iscu} = \alpha \text{mobility}_{cu} + \varepsilon_{isu}$), including either month or county/district FE as specified in the legend. For branded versus non-branded items we consider the overall basket of products. For shop types we consider the share of each shop type amongst the price quotes observed for a certain item within a given county/district (the results are robust to considering within county/district variation only). Standard errors are clustered at the county/district-item level and county/district level respectively.

Table 5.C1: Mobility and data collection - using variation across space and time

	Kenya		Uganda	
	Full sample	Post scale up	Full sample	Post scale up
Mobility	0.055***	0.095***	0.098***	0.114***
	(0.017)	(0.021)	(0.024)	(0.031)
Constant	0.899***	0.898***	0.896***	0.898***
	(0.004)	(0.004)	(0.010)	(0.011)
No. observations	138652	127931	27879	23411
R2 adj.	0.012	0.015	0.024	0.025
R2 adj. within	0.001	0.002	0.002	0.003
Item FE	Yes	Yes	Yes	Yes

In the above table, we estimate the following specification: $x_{idc} = \beta y_{dc} + \delta_i + \varepsilon_{idc}$, where δ_i represents the item FE. The outcome variable x_{idc} captures whether a price quote was recorded for a given day d, item i, and administrative unit c for which Google mobility data were available. The variable takes a value of one if the price was quote was missing. The independent variable y_{dc} captures the change in activity relative to January and February 2020 in percentage points. It takes a value of zero if the observed activity is similar to the baseline period, positive values represent an increase in activity levels, negative values a decline. In Kenya observed values range from -0.7 to 0.4. In Uganda they range from -1 to 0.3. Standard errors are clustered at the item-location level (counties in Kenya, districts in Uganda). *; **; and *** denote significance at the 10; 5; and 1 percent levels respectively.

Table 5.C2: Mobility and data collection - using variation across time

	Kenya		Uganda	
	Full sample	Post scale up	Full sample	Post scale up
Mobility	-0.021*	0.080***	-0.174***	-0.104***
	(0.012)	(0.010)	(0.030)	(0.019)
Constant	0.894***	0.897***	0.874***	0.883***
	(0.002)	(0.002)	(0.006)	(0.006)
No. observations	138652	127931	27879	23411
R2 adj.	0.092	0.107	0.136	0.176
R2 adj. within	0.000	0.001	0.006	0.002
Admin FE	Yes	Yes	Yes	Yes
Item FE	Yes	Yes	Yes	Yes
Day of the week FE	Yes	Yes	Yes	Yes

In the above table, we estimate the following specification: $x_{idc} = \beta y_{dc} + \delta_i + \delta_u + \varepsilon_{idc}$, where δ_i represents the item FE and δ_u the week-day FE. The outcome variable x_{idc} captures whether a price quote was recorded for a given day d, item i, and administrative unit c for which Google mobility data were available. The variable takes a value of one if the price was quote was missing. The independent variable y_{dc} captures the change in activity relative to January and February 2020 in percentage points. It takes a value of zero if the observed activity is similar to the baseline period, positive values represent an increase in activity levels, negative values a decline. In Kenya observed values range from -0.7 to 0.4. In Uganda they range from -1 to 0.3. Standard errors are clustered at the item-location level (counties in Kenya, districts in Uganda). *; **; and *** denote significance at the 10; 5; and 1 percent levels respectively.

Table 5.C3: Mobility and data collection - using variation across space

	Kenya		Uganda	
	Full sample	Post scale up	Full sample	Post scale up
Mobility	0.071***	0.072***	0.155***	0.136***
	(0.020)	(0.021)	(0.032)	(0.033)
Constant	0.900***	0.896***	0.901***	0.899***
	(0.004)	(0.004)	(0.009)	(0.010)
No. observations	138652	127931	27879	23411
R2 adj.	0.024	0.024	0.035	0.033
R2 adj. within	0.001	0.001	0.005	0.004
Item FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes

In the above table, we estimate the following specification: $x_{idc} = \beta y_{dc} + \delta_i + \delta_m + \varepsilon_{idc}$, where δ_i represents the item FE and δ_m the month FE. The outcome variable x_{idc} captures whether a price quote was recorded for a given day d, item i, and administrative unit c for which Google mobility data were available. The variable takes a value of one if the price was quote was missing. The independent variable y_{dc} captures the change in activity relative to January and February 2020 in percentage points. It takes a value of zero if the observed activity is similar to the baseline period, positive values represent an increase in activity levels, negative values a decline. In Kenya observed values range from -0.7 to 0.4. In Uganda they range from -1 to 0.3. Standard errors are clustered at the item-location level (counties in Kenya, districts in Uganda). *; **; and *** denote significance at the 10; 5; and 1 percent levels respectively.

Conclusion

My thesis explores the relevance of economic linkages and inter-connectedness for private sector development and resilience in Kenya. One part of the thesis focuses on documenting the structure of Kenya's formal firm network and its linkages with unobserved informal firms as well as international markets. The second part of my thesis looks at the response of Kenya's private sector to the COVID-19 pandemic and its subsequent recovery.

In Chapter 2, my-coauthors and I explore the spatial configuration of Kenya's domestic firm network. Our findings reveal a high degree of spatial concentration of the network of formal firms around Kenya's metropolitan areas. Using census data and a network formation model we show that solely relying on formal sector network data might underestimate both the connectivity but also vulnerability of smaller regional markets to shocks, especially those passing through hubs like Nairobi.

Chapter 3 examines the potential of Internet connectivity to shape domestic supply chains by reducing information frictions and transaction costs among market participants. I find larger firms to be more likely to purchase inputs from information and communication technology (ICT) providers. Evidence from qualitative interviews suggests that firm size heterogeneity in adopting Internet-based technologies is influenced by firms' capacity to make complementary investments alongside ICT usage.

In Chapter 4, my co-authors and I quantify the extent to which Kenyan private sector firms are exposed to international markets, either through direct imports and exports or through international linkages along their domestic supply chains.

Only nine (five) percent of firms are highly exposed import (export) markets, either directly or indirectly. However, these firms account for the majority of output in the formal economy. In response to international shocks, we find that firms substitute exports for domestic sales. Meanwhile, favourable import shocks do not necessarily crowd out domestic purchases. Examining firm dynamics during the COVID-19 crisis, we find that firms who act as intermediary importers were hit the hardest during the initial downturn due to the collapse in international trade. The recovery of import markets in the aftermath of the initial downturn, however, supported the recovery of Kenya's domestic private sector.

Lastly, Chapter 5 investigates the impact of the COVID-19 pandemic on food prices in Kenya and Uganda using real-time price data collected through a crowd-sourcing approach. My co-authors and I find that reduced mobility during the pandemic led to an increase in food prices in both countries. The negative correlation between mobility and prices persists beyond the initial shutdown phase, an indication that resilience of the retail sector is critical for a post-crisis recovery.

More so than answering questions, this thesis highlighted several avenues for future research. A recurring theme throughout was the centrality of the transport sector (Chapter 1) and its interaction with firms' decision to adopt technologies that have the potential to lower information frictions and transaction cost (Chapter 3). Firms interviewed in Chapter 3 explained how the lack of cost-effective and flexible delivery options prevent them from scaling their business. The lack of entry in the transport sector can partly be explained by economies of scale (Grant and Startz, 2022). However, even firms in Nairobi faced transport-related challenges. While the role of physical transport infrastructure for trade and economic development has been a very active field of research in the past decade (Donaldson, 2015), market structures of the transportation sector and the sector's potential for productivity spillovers to other sectors remains relatively understudied.
Chapter 2 considers spatial firm-to-firm trade in a static framework. Small and mid-sized towns all across Kenya have been growing at a rapid pace in recent years. Exploring how urbanisation has changed the geography of Kenya's supply chains over

time is another promising avenue for future research.

Another natural extension of Chapter 2 would be to collect primary data on the linking probabilities of VAT and non-VAT firms or use linking probabilities of marginally formal firms identified through a natural experiment. Even within the formal sector, matching frictions between small and large firms might exist. In Chapter 1, we compare the correlation coefficients for the degree of trading partners in a simulated network to the real-world data and found a less negative correlation. This suggests that while small firms do indeed link with large ones, they might not do so to extent we would expect them to in the absence of matching frictions. Besides the role transport capacity plays for sourcing decisions of small firms, Chapter 3 further suggest information frictions in markets for trade credit to be relevant.

References

Abala, Daniel Okado, "Determinants of manufactured exports in Kenya: an application of control function approach," *DBA Africa management review*, 2013, *3* (1).

Abay, Kibrom A, Guush Berhane, John Hoddinott, and Kibrom Tafere, "COVID-19 and food security in Ethiopia: do social protection programs protect?," *Economic Development and Cultural Change*, 2023, *71* (2), 373–402.

Abel-Koch, Jennifer, "Who uses intermediaries in international trade? Evidence from firm-level survey data," *The World Economy*, 2013, *36* (8), 1041–1064.

Acemoglu, Daron, Vasco M Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi, "The network origins of aggregate fluctuations," *Econometrica*, 2012, *80* (5), 1977–2016.

Adam, Lishan and Maurice Mubila, "Africa's prospects for infrastructure development and regional integration: information and communications technology sector," in "Infrastructure in Africa," Policy Press, 2017, pp. 333–376.

Adão, Rodrigo, Paul Carrillo, Arnaud Costinot, Dave Donaldson, and Dina Pomeranz, "Imports, exports, and earnings inequality: Measures of exposure and estimates of incidence," *The Quarterly Journal of Economics*, 2022, *137* (3), 1553–1614.

Ades, Alberto F and Edward L Glaeser, "Trade and circuses: explaining urban giants," *The Quarterly Journal of Economics*, 1995, *110* (1), 195–227.

Adewopo, Julius B, Gloria Solano-Hermosilla, Liesbeth Colen, and Fabio Micale, "Using crowd-sourced data for real-time monitoring of food prices during the COVID-

19 pandemic: Insights from a pilot project in northern Nigeria," *Global Food Security*, 2021, *29*, 100523.

Agyei, Samuel Kwaku, Zangina Isshaq, Siaw Frimpong, Anokye Mohammed Adam, Ahmed Bossman, and Oliver Asiamah, "COVID-19 and food prices in sub-Saharan Africa," *African Development Review*, 2021, *33*, S102–S113.

Ahn, JaeBin, Amit K Khandelwal, and Shang-Jin Wei, "The role of intermediaries in facilitating trade," *Journal of International Economics*, 2011, *84* (1), 73–85.

Aker, Jenny C, "Information from markets near and far: Mobile phones and agricultural markets in Niger," *American Economic Journal: Applied Economics*, 2010, *2* (3), 46–59.

Akerman, Anders, Edwin Leuven, and Magne Mogstad, "Information frictions, internet, and the relationship between distance and trade," *American Economic Journal: Applied Economics*, 2022, *14* (1), 133–63.

Aktay, Ahmet, Shailesh Bavadekar, Gwen Cossoul, John Davis, Damien Desfontaines, Alex Fabrikant, Evgeniy Gabrilovich, Krishna Gadepalli, Bryant Gipson, Miguel Guevara et al., "Google COVID-19 community mobility reports: anonymization process description (version 1.1)," *arXiv preprint arXiv:2004.04145*, 2020.

Alfaro-Urena, Alonso, Isabela Manelici, and José P. Vásquez, "The Effects of Multinationals on Workers: Evidence from Costa Rica," 2019.

_ , _ , and Jose P Vasquez, "The effects of joining multinational supply chains: New evidence from firm-to-firm linkages," *The Quarterly Journal of Economics*, 2022, *137* (3), 1495–1552.

_ , Mariany Fuentes Fuentes, Isabela Manelici, and José P Vásquez, "Costa Rican Production Network: Stylized Facts," *Research Paper Banco Central de Costa Rica*, 2018, *2018* (2).

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

Almunia, Miguel, François Gerard, Jonas Hjort, Justine Knebelmann, Dorothy Nakyambadde, Claude Raisaro, and Lin Tian, "An analysis of discrepancies in tax declarations submitted under value-added tax in Uganda," *International Growth Centre Project Report*, 2017.

_ , Jonas Hjort, Justine Knebelmann, and Lin Tian, "Strategic or confused firms? evidence from "missing" transactions in Uganda," *Review of Economics and Statistics*, 2022, pp. 1–35.

Alonso, William, "Urban and regional imbalances in economic development," *Economic development and cultural change*, 1968, *17* (1), 1–14.

Amare, Mulubrhan, Kibrom A Abay, Luca Tiberti, and Jordan Chamberlin, *Impacts of Covid-19 on food security: panel data evidence from Nigeria*, Vol. 1956, Intl Food Policy Res Inst, 2020.

Amendolagine, Vito, Andrea F Presbitero, Roberta Rabellotti, and Marco Sanfilippo, "Local sourcing in developing countries: The role of foreign direct investments and global value chains," *World Development*, 2019, *113*, 73–88.

Amiti, Mary and Jozef Konings, "Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia," *American Economic Review*, 2007, *97* (5), 1611–1638.

Arkolakis, Costas, Federico Huneeus, and Yuhei Miyauchi, "Spatial production networks," Technical Report, National Bureau of Economic Research 2023.

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

Atkin, David and Amit K. Khandelwal, "How Distortions Alter the Impacts of International Trade in Developing Countries," *Annual Review of Economics*, 2020, *12* (1), null.

_ and Dave Donaldson, "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs," Working Paper 21439, National Bureau of Economic Research July 2015.

Augustine, Abraham, "The Next Wave: Africa needs to shed the "mobile-first" garb," *techcabal*, August 2022, *August* (22).

Aum, Sangmin, Sang Yoon Tim Lee, and Yongseok Shin, "COVID-19 doesn't need lockdowns to destroy jobs: The effect of local outbreaks in Korea," *Labour Economics*, 2021, *70*, 101993.

Awori, Kagonya, Melisa Achoko Allela, Stephanie Nyairo, Samuel C Maina, and Jacki O'Neill, "" It's only when somebody says a tool worked for them that I believe it will work for me": Socio-tecture as a lens for Digital Transformation," *Proceedings of the ACM on Human-Computer Interaction*, 2022, *6* (CSCW2), 1–24.

Bailey, Michael, Abhinav Gupta, Sebastian Hillenbrand, Theresa Kuchler, Robert Richmond, and Johannes Stroebel, "International trade and social connectedness," *Journal of International Economics*, 2021, *129*, 103418.

_ , Drew Johnston, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, "Peer effects in product adoption," *American Economic Journal: Applied Economics*, 2022, *14* (3), 488–526.

_ , Patrick Farrell, Theresa Kuchler, and Johannes Stroebel, "Social connectedness in urban areas," *Journal of Urban Economics*, 2020, *118*, 103264.

Baqaee, David and Emmanuel Farhi, "Supply and demand in disaggregated Keynesian economies with an application to the Covid-19 crisis," *American Economic Review*, 2022, *112* (5), 1397–1436.

Baqaee, David Rezza, "Cascading failures in production networks," *Econometrica*, 2018, *86* (5), 1819–1838.

_ and Emmanuel Farhi, "The macroeconomic impact of microeconomic shocks: Beyond Hulten's theorem," *Econometrica*, 2019, *87* (4), 1155–1203.

Barabási, Albert-László, *Network Science: Degree Correlation* CC BY-NC-SA 2.0. 2014.

Barrett, Christopher B, "Urban bias in price risk: The geography of food price distributions in low-income economies," *The Journal of Development Studies*, 1996, *32* (6), 830–849.

_ , "Food security and food assistance programs," *Handbook of agricultural economics*, 2002, *2*, 2103–2190.

_ , Robert Bell, Erin C Lentz, and Daniel G Maxwell, "Market information and food insecurity response analysis," *Food Security*, 2009, *1*, 151–168.

Barrot, Jean-Noël and Julien Sauvagnat, "Input specificity and the propagation of idiosyncratic shocks in production networks," *The Quarterly Journal of Economics*, 2016, *131* (3), 1543–1592.

Barrot, Jean-Noël, Basile Grassi, and Julien Sauvagnat, "Sectoral Effects of Social Distancing," *AEA Papers and Proceedings*, May 2021, *111*, 277–81.

Bassett, Hannah R, Jacqueline Lau, Christopher Giordano, Sharon K Suri, Sahir Advani, and Sonia Sharan, "Preliminary lessons from COVID-19 disruptions of small-scale fishery supply chains," *World Development*, 2021, *143*, 105473.

Bassi, Vittorio, Raffaela Muoio, Tommaso Porzio, Ritwika Sen, and Esau Tugume, "Achieving scale collectively," *Econometrica*, 2022, *90* (6), 2937–2978.

Bergquist, Lauren F, Benjamin Faber, Thibault Fally, Matthias Hoelzlein, Edward Miguel, and Andrés Rodríguez-Clare, "Scaling agricultural policy interventions," Technical Report, National Bureau of Economic Research 2022.

Bergquist, Lauren Falcao and Michael Dinerstein, "Competition and entry in agricultural markets: Experimental evidence from Kenya," *American Economic Review*, 2020, *110* (12), 3705–3747.

_ , Craig McIntosh, and Meredith Startz, "Search cost, intermediation, and trade: Experimental evidence from Ugandan agricultural markets," *eScholarship, University of California*, 2021.

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 Yuan Zi, "Sparse Production Networks," Working Paper 30496, National Bureau of Economic Research September 2022.

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

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

_ , Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes, "The origins of firm heterogeneity: A production network approach," *Journal of Political Economy*, 2022, *130* (7), 1765–1804.

_ , J Bradford Jensen, and Robert Z Lawrence, "Exporters, jobs, and wages in US manufacturing: 1976-1987," *Brookings papers on economic activity. Microeconomics*, 1995, *1995*, 67–119.

_ , Jonathan Eaton, J Bradford Jensen, and Samuel Kortum, "Plants and productivity in international trade," *American economic review*, 2003, *93* (4), 1268–1290.

_ , Marco Grazzi, and Chiara Tomasi, "Intermediaries in international trade: Products and destinations," *Review of Economics and Statistics*, 2015, *97* (4), 916–920.

Berry, Brian JL, "City size distributions and economic development," *Economic development and cultural change*, 1961, *9* (4, Part 1), 573–588.

Berthou, Antoine and Sebastian Stumpner, "Trade Under Lockdown," *Banque de France Working Paper*, 2022, *No. 867.*

Blanchard, Paul, Douglas Gollin, and Martina Kirchberger, "Perpetual Motion: Human Mobility and Spatial Frictions in Three African Countries," *CEPR Discussion Papers*, 2021, *No. 16661.*

Boehm, Christoph E, Aaron Flaaen, and Nitya Pandalai-Nayar, "Input linkages and the transmission of shocks: Firm-level evidence from the 2011 Tōhoku earthquake," *Review of Economics and Statistics*, 2019, *101* (1), 60–75.

Boehm, Johannes and Ezra Oberfield, "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," *The Quarterly Journal of Economics*, 2020, *135* (4), 2007–2058.

_ and Jan Sonntag, "Vertical integration and foreclosure: Evidence from production network data," *Management Science*, 2023, *69* (1), 141–161.

Böhme, Marcus H and Rainer Thiele, "Informal–formal linkages and informal enterprise performance in urban West Africa," *The European Journal of Development Research*, 2014, *26*, 473–489.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

Bramoullé, Yann, Sergio Currarini, Matthew O Jackson, Paolo Pin, and Brian W Rogers, "Homophily and long-run integration in social networks," *Journal of Economic Theory*, 2012, *147* (5), 1754–1786.

Brandt, Nicola, "Informality in Mexico," *OECD Economics Department Working Papers*, 2011, (896).

Bresnahan, Timothy F, Erik Brynjolfsson, and Lorin M Hitt, "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence," *The quarterly journal of economics*, 2002, *117* (1), 339–376.

Brooks, Wyatt, Kevin Donovan, Terence R Johnson, and Jackline Oluoch-Aridi, "Cash transfers as a response to COVID-19: Experimental evidence from Kenya," *Journal of Development Economics*, 2022, *158*, 102929.

Brugués, Felipe, "Take the Goods and Run: Contracting Frictions and Market Power in Supply Chains," *Job Market Paper*, 2023.

Brynjolfsson, Erik and Lorin M Hitt, "Beyond computation: Information technology, organizational transformation and business performance," *Journal of Economic perspectives*, 2000, *14* (4), 23–48.

Bunn, Philip, David Altig, Lena Anayi, Jose Maria Barrero, Nicholas Bloom, Steven J Davis, Brent Meyer, Emil Mihaylov, Paul Mizen, and Greg Thwaites, "COVID-19 Uncertainty: A tale of two tails," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2021, (2021-135).

Butler, Declan, "Crowdsourcing goes mainstream in typhoon response," *Nature News*, 2013.

Böken, Johannes, Lucie Gadenne, Tushar Nandi, and Marta Santamaria, "Community Networks and Trade," *Working Paper*, 2023.

CAK, "Competition in Shipping, Trucking and Haulage Sector Study in East Africa," Technical Report, Competition Authority Kenya 2019.

Camehl, Annika and Malte Rieth, "Disentangling Covid-19, economic mobility, and containment policy shocks," *American Economic Journal: Macroeconomics*, forthcoming.

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

Cardoza, Marvin, Francesco Grigoli, Nicola Pierri, and Cian Ruane, "Worker Mobility in Production Networks," *Working paper*, 2023.

Carvalho, Vasco M, "From micro to macro via production networks," *Journal of Economic Perspectives*, 2014, *28* (4), 23–48.

_ and Alireza Tahbaz-Salehi, "Production networks: A primer," *Annual Review of Economics*, 2019, *11*, 635–663.

_ , Makoto Nirei, Yukiko U Saito, and Alireza Tahbaz-Salehi, "Supply chain disruptions: Evidence from the great east japan earthquake," *The Quarterly Journal of Economics*, 2021, *136* (2), 1255–1321.

Chacha, Peter Wankuru, "For the first time, the relative economic size of Kenya's counties is clear," *World Bank Blogs*, 2019.

_ and Lawrence Edwards, "The growth dynamics of new export entrants in kenya: a survival analysis," *ERSA working paper*, 2017, *712*.

_ and _ , "Imported inputs, Government Support and Performance of manufacturing exporters," *ERSA working paper*, 2018, *740*.

_ and _ , "Exporting to fragile states in Africa: Firm-level evidence," *Review of Development Economics*, 2019, *23* (3), 1177–1201.

_ , Benard Kipyegon Kirui, and Verena Wiedemann, "Mapping Kenya's Production Network Transaction by Transaction," *Working Paper available at SSRN 4315810*, 2022.

_ , _ , and _ , "Supply chains in times of crisis: Evidence from Kenya's production network," *World Development*, 2024, *173*, 106363.

_ , Vatsal Khandelwal, Benard Kipyegon Kirui, and Verena Wiedemann, "Spatial inequality and informality in Kenya's firm network," *Working Paper*, 2023.

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

Chen, Shiyi, Wanlin Liu, and Hong Song, "Broadband internet, firm performance, and worker welfare: Evidence and mechanism," *Economic Inquiry*, 2020, *58* (3), 1146–1166.

Cirera, Xavier, Diego Comin, and Marcio Cruz, *Bridging the technological divide: Technology adoption by firms in developing countries*, World Bank Publications, 2022.

_ , _ , _ , Kyung Min Lee, Pedro Martinez, and Antonio Martins-Neto, "Firm-level Adoption of Technologies in Kenya," 2022.

Consoli, Sarah, Juan José Egas Yerovi, Matteo Machiorlatti, and Cristian Morales Opazo, "Real-time Monitoring of Food Price Policy Interventions during the First Two Years of COVID-19," *Food Policy*, 2023, p. 102405.

Cordaro, Francesco, Marcel Fafchamps, Colin Mayer, Muhammad Meki, Simon Quinn, and Kate Roll, "Microequity and Mutuality: Experimental Evidence on Credit with Performance-Contingent Repayment," Technical Report, National Bureau of Economic Research 2022.

Coşar, A Kerem and Banu Demir, "Shipping inside the box: Containerization and trade," *Journal of International Economics*, 2018, *114*, 331–345.

_ , _ , Devaki Ghose, and Nathaniel Young, "Road capacity, domestic trade and regional outcomes," *Journal of Economic Geography*, 2022, *22* (5), 901–929.

Coulibaly, Romaric, Heddie Moreno, Suwa-Eisenmann Akiko, and Nouhoum Traore, "African Firms in Global Value Chains: What Can We Learn from Firm-Level Data in Cameroon and Côte d'Ivoire?," *AERC Working Paper*, 2022, *GVC-009.*

Couture, Victor, Jonathan I. Dingel, Allison Green, Jessie Handbury, and Kevin R. Williams, "Measuring movement and social contact with smartphone data: a real-time application to COVID-19," *Journal of Urban Economics*, January 2021.

DellaVigna, Stefano and Matthew Gentzkow, "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Demir, B, AC Fieler, DY Xu, and KK Yang, "O-ring production networks," *Journal of Political Economy*, 2023.

Demir, Banu, Beata Javorcik, Tomasz K. Michalski, and Evren Ors, "Financial Constraints and Propagation of Shocks in Production Networks," *The Review of Economics and Statistics*, 01 2022, pp. 1–46.

DeStefano, Timothy, Richard Kneller, and Jonathan Timmis, "The (fuzzy) digital divide: the effect of universal broadband on firm performance," *Journal of Economic Geography*, 2023, *23* (1), 139–177.

Dhyne, Emmanuel and Cédric Duprez, "Has the crisis altered the Belgian economy's DNA?," *Economic Review*, 2015, (ii), 31–43.

_ , Ayumu Ken Kikkawa, Magne Mogstad, and Felix Tintelnot, "Trade and domestic production networks," *The Review of Economic Studies*, 2021, *88* (2), 643–668.

_ , Glenn Magerman, and Stela Rubínova, "The Belgian production network 2002-2012," Technical Report, National Bank of Belgium 2015.

Dix-Carneiro, Rafael, Pinelopi K Goldberg, Costas Meghir, and Gabriel Ulyssea, "Trade and informality in the presence of labor market frictions and regulations," Technical Report, National Bureau of Economic Research 2021.

Donaldson, Dave, "The Gains from Market Integration," *Annual Review of Economics*, 2015, *7* (1), 619–647.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, "An anatomy of international trade: Evidence from French firms," *Econometrica*, 2011, *79* (5), 1453–1498.

_ , Samuel S Kortum, and Francis Kramarz, "Firm-to-Firm Trade: Imports, exports, and the labor market," Technical Report, National Bureau of Economic Research 2022.

Egger, Dennis, Edward Miguel, Shana S Warren, Ashish Shenoy, Elliott Collins, Dean Karlan, Doug Parkerson, A Mushfiq Mobarak, Günther Fink, Christopher Udry et al., "Falling living standards during the COVID-19 crisis: Quantitative evidence from nine developing countries," *Science advances*, 2021, *7* (6), eabe0997.

Elgin, Ceyhun, M Ayhan Kose, Franziska Ohnsorge, and Shu Yu, "Understanding informality," *CERP Discussion Paper*, 2021, *16497*.

Erol, Selman and Rakesh Vohra, "Network formation and systemic risk," *European Economic Review*, 2022, *148*, 104213.

Fafchamps, Marcel, *Market institutions in sub-Saharan Africa: Theory and evidence*, MIT press, 2003.

Fort, Teresa C, "Technology and production fragmentation: Domestic versus foreign sourcing," *The Review of Economic Studies*, 2017, *84* (2), 650–687.

Foster, Christopher, Mark Graham, Laura Mann, Timothy Waema, and Nicolas Friederici, "Digital control in value chains: Challenges of connectivity for East African firms," *Economic Geography*, 2018, *94* (1), 68–86.

Freund, Caroline, Aaditya Mattoo, Alen Mulabdic, and Michele Ruta, "Natural disasters and the reshaping of global value chains," *IMF Economic Review*, 2022, *70* (3), 590–623.

Fujiy, Brian Cevallos, Devaki Ghose, and Gaurav Khanna, "Production Networks and Firm-level Elasticities of Substitution," *STEG Working Paper Series*, 2022, *WP027*.

Gabaix, Xavier, "Power laws in economics and finance," *Annu. Rev. Econ.*, 2009, *1* (1), 255–294.

_ , "The granular origins of aggregate fluctuations," *Econometrica*, 2011, *79* (3), 733–772.

_ and Yannis M Ioannides, "The evolution of city size distributions," in "Handbook of regional and urban economics," Vol. 4, Elsevier, 2004, pp. 2341–2378.

Gadenne, Lucie, Samuel Norris, Monica Singhal, and Sandip Sukhtankar, "In-kind transfers as insurance," Technical Report, National Bureau of Economic Research 2021.

_ , Tushar K Nandi, and Roland Rathelot, "Taxation and Supplier Networks: Evidence from India," *Working Paper*, 2022.

Gebrewolde, Tewodros M, Michael Koelle, Pramila Krishnan, and Andualem T Mengistu, "Currency Shocks and Firm Behaviour in Ethiopia and Uganda," *Journal of African Economies*, 2022, *31*, i59–i82.

Gerard, François, Clément Imbert, and Kate Orkin, "Social protection response to the COVID-19 crisis: options for developing countries," *Oxford Review of Economic Policy*, 2020, *36* (Supplement_1), S281–S296.

Gereffi, Gary, John Humphrey, and Timothy Sturgeon, "The governance of global value chains," *Review of international political economy*, 2005, *12* (1), 78–104.

Ghassibe, Mishel, "Endogenous Production Networks and Non-Linear Monetary Transmission," in "Working Paper" 2021.

Goldberg, Pinelopi K and Tristan Reed, "Demand-side constraints in development: The role of market size, trade, and (in)equality," *Econometrica*, 2023.

Goldenberg, Jacob and Moshe Levy, "Distance is not dead: Social interaction and geographical distance in the internet era," *arXiv preprint arXiv:0906.3202*, 2009.

Gollin, D, P Blanchard, and M Kirchberger, "Perpetual motion: human mobility and spatial frictions in three African countries," *CSAE Working Paper Series*, 2020, *2020-18.*

Gollin, Douglas, "Nobody's business but my own: Self-employment and small enterprise in economic development," *Journal of Monetary Economics*, 2008, *55* (2), 219–233.

Goolsbee, Austan and Chad Syverson, "Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020," *Journal of public economics*, 2021, *193*, 104311.

Granér, Mats and Anders Isaksson, "Firm efficiency and the destination of exports: evidence from Kenyan plant-level data," *The Developing Economies*, 2009, *47* (3), 279–306.

Grant, Matthew and Meredith Startz, "Cutting out the middleman: The structure of chains of intermediation," Technical Report, National Bureau of Economic Research 2022.

Grassi, Basile et al., "Io in io: Size, industrial organization, and the input-output network make a firm structurally important," *Work. Pap., Bocconi Univ., Milan, Italy*, 2017.

Greenaway, David and Richard Kneller, "Firm heterogeneity, exporting and foreign direct investment," *The Economic Journal*, 2007, *117* (517), F134–F161.

Grigoli, Francesco, Emiliano Luttini, and Damiano Sandri, "Idiosyncratic shocks and aggregate fluctuations in an emerging market," *Journal of Development Economics*, 2023, *160*, 102949.

Hansen, Ulrich Elmer, Ivan Nygaard, Mike Morris, and Glen Robbins, "Servicification of manufacturing in global value chains: Upgrading of local suppliers of embedded services in the South African market for wind turbines," *The Journal of Development Studies*, 2022, *58* (4), 787–808.

Hassan, Mai and Friedrich Schneider, "Size and development of the shadow economies of 157 countries worldwide: Updated and new measures from 1999 to 2013," *IZA Discussion Paper*, 2019, *No. 10281.*

Head, Keith and Thierry Mayer, "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of international economics," Vol. 4, Elsevier, 2014, pp. 131–195.

Henderson, Vernon, "Urban primacy, external costs, and quality of life," *Resource and Energy Economics*, 2002, *24* (1-2), 95–106.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi, "New evidence on sectoral labor productivity: Implications for industrialization and development," Technical Report, National Bureau of Economic Research 2022.

Hjort, Jonas and Lin Tian, "The economic impact of internet connectivity in developing countries," *PEDL Synthesis Series*, 2021, *No. 6.*

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang, "The wage effects of offshoring: Evidence from Danish matched worker-firm data," *American Economic Review*, 2014, *104* (6), 1597–1629.

Huneeus, Federico, "Production Network Dynamics and the Propagation of Shocks," *Job Market Paper*, 2018.

Huneeus, Frederico, Kory Kroft, Kevin Lim, and David J. Price, "Earnings Inequality in Production Networks," 2020.

Iacovone, Leonardo, Mariana Pereira-López, and Marc Schiffbauer, "ICT use, competitive pressures, and firm performance in Mexico," *The World Bank Economic Review*, 2017, *30* (Supplement_1), S109–S118.

Ishikawa, Jota and Nori Tarui, "Backfiring with backhaul problems: Trade and industrial policies with endogenous transport costs," *Journal of International Economics*, 2018, *111*, 81–98.

Iyoha, Ebehi, "Estimating productivity in the presence of spillovers: Firm-level evidence from the US production network," *Working Paper*, 2022.

Jackson, Matthew O and Brian W Rogers, "Meeting strangers and friends of friends: How random are social networks?," *American Economic Review*, 2007, *97* (3), 890–915.

Jaravel, Xavier and Martin O'Connell, "Real-time price indices: Inflation spike and falling product variety during the Great Lockdown," *Journal of Public Economics*, 2020, *191*, 104270.

Jefferson, Mark, "The law of the primate city," *Geographical Review*, 1939, *29* (2), 226–232.

_ , "Why geography? The law of the primate city," *Geographical Review*, 1989, *79* (2), 226–232.

Jensen, Robert, "The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector," *The quarterly journal of economics*, 2007, *122* (3), 879–924.

Jiang, Xian, "Efficiency Gains from Information and Communication Technology: A Spatial Analysis of Firm Geographic Expansion," *Working Paper*, 2022.

Johnson, Simon, John McMillan, and Christopher Woodruff, "Property rights and finance," *American Economic Review*, 2002, *92* (5), 1335–1356.

Jorgenson, Dale W, Mun S Ho, and Kevin J Stiroh, "A retrospective look at the US productivity growth resurgence," *Journal of Economic perspectives*, 2008, *22* (1), 3–24.

Juhász, Réka and Claudia Steinwender, "Spinning the web: Codifiability, information frictions and trade," *NBER working paper*, 2018.

Kaberia, Salome K and Stephen Muathe, "Effect of Covid-19 pandemic on performance of women owned micro, small and medium enterprises in Kenya," *Int'l J. Soc. Sci. Stud.*, 2021, *9*, 7.

Kansiime, Monica K, Justice A Tambo, Idah Mugambi, Mary Bundi, Augustine Kara, and Charles Owuor, "COVID-19 implications on household income and food security in Kenya and Uganda: Findings from a rapid assessment," *World development*, 2021, *137*, 105199.

Kashiwagi, Yuzuka, Yasuyuki Todo, and Petr Matous, "Propagation of economic shocks through global supply chains—Evidence from Hurricane Sandy," *Review of International Economics*, 2021, *29* (5), 1186–1220.

Kassem, Dana, "Does Electrification Cause Industrial Development? Grid Expansion and Firm Turnover in Indonesia," *Working Paper*, 2021.

Kawakubo, Takafumi and Takafumi Suzuki, "Do Supply Chain Disruptions Harm Firm Performance? Evidence from Japan," *Job Market Paper*, 2022.

Khera, Reetika, "Cash vs. in-kind transfers: Indian data meets theory," *Food Policy*, 2014, *46*, 116–128.

King, Maia, Bassel Tarbush, and Alexander Teytelboym, "Targeted carbon tax reforms," *European Economic Review*, 2019, *119*, 526–547.

Kinyanjui, George, Doreen Rubatsimbira, and Verena Wiedemann, "Changing mobility patterns and food price dynamics during the COVID-19 pandemic in Kenya and Uganda," *Working Paper*, 2023.

Klenow, Peter J and Andres Rodriguez-Clare, "The neoclassical revival in growth economics: Has it gone too far?," *NBER macroeconomics annual*, 1997, *12*, 73–103.

KNBS, "Basic Report on the 2010 Census of Industrial Production," 2010.

_ , "Micro, Small and Medium Establishment (MSME) Survey: Basic Report2016," Technical Report, Kenya National Bureau of Statistics 2016.

_ , "Report on the 2017 Kenya Census of Establishments (CoE)," Technical Report, Kenya National Bureau of Statistics 2017.

_ , *2019 Kenya Population and Housing Census* number Volume I, Kenya National Bureau of Statistics, November 2019.

_ , "2019 Kenya Population and Housing Census: Volume I," Technical Report, Kenya National Bureau of Statistics 2019.

_ , "Economic Survey 2021," Technical Report, Kenya National Bureau of Statistics 2021.

_ , "Gross County Product 2021 (GCP)," Technical Report, KNBS 2022.

Korovkin, Vasily and Alexey Makarin, "Production Networks and War: Evidence from Ukraine," *Available at SSRN 3969161*, 2022.

Kose, M. Ayhan, Franziska Ohnsorge, Shu Yu, Mohammad Amin, Sinem Kilic Celik, Gene Joseph Kindberg Hanlon, Ergys Islamaj, Sergiy Kasyanenko, Cedric Okou, Naotaka Sugawara, Temel Taskin, and Wheeler. Collette, *Global economic prospects: A World Bank Flagship Report*, World Bank,

Koutroumpis, Pantelis, Farshad Ravasan, and Taheya Tarannum, "The arrival of fiber and the digital dividend gap: Evidence from the asymmetric homebuyers' responses," *Available at SSRN 4109470*, 2022.

Kremer, Michael, "The O-Ring Theory of Economic Development," *The Quarterly Journal of Economics*, 1993, *108* (3), 551–575.

Kumar, Krishna, Raghuram Rajan, and Luigi Zingales, "What determines firm size?," Technical Report, NBER Working Paper 1999.

Laeven, Luc and Christopher Woodruff, "The quality of the legal system, firm ownership, and firm size," *The Review of Economics and Statistics*, 2007, *89* (4), 601–614.

Lafrogne-Joussier, Raphael, Julien Martin, and Isabelle Mejean, "Supply shocks in supply chains: Evidence from the early lockdown in China," *IMF economic review*, 2023, *71* (1), 170–215.

Lane, Nathan, "Manufacturing Revolutions - Industrial Policy and Industrialization in South Korea," *Quarterly Journal of Economics*, 2023.

Lee, Hau L, Venkata Padmanabhan, and Seungjin Whang, "Information distortion in a supply chain: The bullwhip effect," *Management science*, 1997, *43* (4), 546–558.

Lim, Kevin, "Endogenous Production Networks and the Business Cycle," *Job Market Paper*, 2018.

Liu, Ernest, "Industrial policies in production networks," *The Quarterly Journal of Economics*, 2019, *134* (4), 1883–1948.

Lowe, Matt, GV Nadhanael, and Benjamin N Roth, "India's food supply chain during the pandemic," *Food Policy*, 2021, *105*, 102162.

Macchiavello, Rocco and Ameet Morjaria, "The value of relationships: evidence from a supply shock to Kenyan rose exports," *The American Economic Review*, 2015, *105* (9), 2911–2945.

Mahajan, Kanika and Shekhar Tomar, "COVID-19 and Supply Chain Disruption: Evidence from Food Markets in India," *American Journal of Agricultural Economics*, 2021, *103* (1), 35–52.

Mahmud, Mahreen and Emma Riley, "Impact of Covid-19 on Economic Outcomes and Well-being of Rural Communities in Western Uganda," *Policy brief*, 2020, *No. 3*.

_ and _ , "Household response to an extreme shock: Evidence on the immediate impact of the Covid-19 lockdown on economic outcomes and well-being in rural Uganda," *World Development*, 2021, *140*, 105318.

_ and _ , "Adapting to an aggregate shock: The impact of the Covid-19 crisis on rural households," *Review of Economics of the Household*, 2023, *21* (1), 19–36.

Majune, Socrates Kraido, "The Effect of Lockdown Policies on International Trade Flows from Developing Countries: Event Study Evidence from Kenya," 2020.

_ , Eliud Moyi, and Kamau Gathiaka, "Explaining export duration in Kenya," *South African Journal of Economics*, 2020, *88* (2), 204–224.

Malgouyres, Clément, Thierry Mayer, and Clément Mazet-Sonilhac, "Technology-induced trade shocks? Evidence from broadband expansion in France," *Journal of International Economics*, 2021, *133*, 103520.

Marshall, A, "Principles of Economics. Palgrave Macmillan," 1890, *2013 edition.*

McCaig, Brian and Nina Pavcnik, "Informal employment in a growing and globalizing low-income country," *American Economic Review*, 2015, *105* (5), 545–550.

McMillan, John and Christopher Woodruff, "Interfirm relationships and informal credit in Vietnam," *The Quarterly Journal of Economics*, 1999, *114* (4), 1285–1320.

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

Memon, Pyar Ali, "Urban primacy in Kenya," *IDS Working Paper Series, University of Nairobi*, 1976, *282.*

Metters, Richard, "Quantifying the bullwhip effect in supply chains," *Journal of operations management*, 1997, *15* (2), 89–100.

Ministry of ICT, Innovation and Youth Affairs, *The Kenya National Digital Master Plan 2022-2032* 2022.

Minten, Bart, Anneleen Vandeplas, and Johan Swinnen, "Regulations, brokers, and interlinkages: The institutional organization of wholesale markets in India," *Journal of Development Studies*, 2012, *48* (7), 864–886.

Miyauchi, Yuhei, "Matching and Agglomeration: Theory and Evidence from Japanese Firm-to-Firm Trade," *Working Paper*, 2023.

MoF, "Survey of Distribution (1960)," Technical Report, Republic of Kenya, Ministry of Finance, Republic of Kenya 1963.

_ , "Urban Population Projections During 1969-2000, Within the Context of Urban Development Strategy," Technical Report, Republic of Kenya, Ministry of Finance and Planning 1974.

Mok, Diana, Barry Wellman, and Juan Carrasco, "Does distance matter in the age of the internet?," *Urban Studies*, 2010, *47* (13), 2747–2783.

Mukherjee, Sanghamitra Warrier, "Digital Technology Adoption: Subsidizing Learning Costs for Firms in India," *mimeo*, 2023.

Murthy, S V Ramana, "Measuring the Informal Economy in India," 2019.

Narayanan, Sudha and Shree Saha, "Urban food markets and the COVID-19 lockdown in India," *Global Food Security*, 2021, *29*, 100515.

Naritomi, Joana, "Consumers as tax auditors," *American Economic Review*, 2019, *109* (9), 3031–72.

Navas-Alemán, Lizbeth, "The impact of operating in multiple value chains for upgrading: the case of the Brazilian furniture and footwear industries," *World development*, 2011, *39* (8), 1386–1397.

Ndubuisi, Gideon and Solomon Owusu, "Wage effects of global value chains participation and position: An industry-level analysis1," *The Journal of International Trade & Economic Development*, 2022, *31* (7), 1086–1107.

Nechifor, Victor, Maria Priscila Ramos, Emanuele Ferrari, Joshua Laichena, Evelyne Kihiu, Daniel Omanyo, Rodgers Musamali, and Benson Kiriga, "Food security and welfare changes under COVID-19 in Sub-Saharan Africa: Impacts and responses in Kenya," *Global Food Security*, 2021, *28*, 100514.

Newman, Mark, *Networks*, Oxford university press, 2018.

Nordhagen, Stella, Uduak Igbeka, Hannah Rowlands, Ritta Sabbas Shine, Emily Heneghan, and Jonathan Tench, "COVID-19 and small enterprises in the food supply chain: Early impacts and implications for longer-term food system resilience in low-and middle-income countries," *World Development*, 2021, *141*, 105405.

Obudho, Robert A, "Nairobi: National capital and regional hub," *The urban challenge in Africa: Growth and management of its large cities*, 1997, pp. 292–334.

O'Connell, Martin, Áureo De Paula, and Kate Smith, "Preparing for a pandemic: Spending dynamics and panic buying during the COVID-19 first wave," *Fiscal Studies*, 2021, *42* (2), 249–264.

Otiso, Kefa M, "Kenya's secondary cities growth strategy at a crossroads: Which way forward?," *GeoJournal*, 2005, *62*, 117–128.

Panigrahi, Piyush, "Endogenous Spatial Production Networks: Quantitative Implications for Trade and Productivity," *Working Paper*, 2022.

Pape, Utz Johann, Antonia Delius, Ritika Khandelwal, and Rhea Gupta, "Socio-Economic Impacts of COVID-19 in Kenya : Results Update," Technical Report, World Bank Group, Washington, D.C. 2020.

Paula, Aureo De and Jose A Scheinkman, "Value-added taxes, chain effects, and informality," *American Economic Journal: Macroeconomics*, 2010, *2* (4), 195–221.

Pavcnik, Nina, "Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants," *The Review of Economic Studies*, 2002, *69* (1), 245–276.

Porta, Rafael La and Andrei Shleifer, "Informality and development," *Journal of Economic Perspectives*, 2014, *28* (3), 109–26.

Pred, Allan, *Urban growth and city systems in the United States, 1840-1860*, Vol. 11, Harvard University Press, 1980.

Prete, Davide Del, Giorgia Giovannetti, and Enrico Marvasi, "Global value chains participation and productivity gains for North African firms," *Review of World Economics*, 2017, *153* (4), 675–701.

Rankin, Neil, Måns Söderbom, and Francis Teal, "Exporting from manufacturing firms in sub-Saharan Africa," *Journal of African Economies*, 2006, *15* (4), 671–687.

Reardon, Thomas, Lenis Saweda O Liverpool-Tasie, and Bart Minten, "Quiet Revolution by SMEs in the midstream of value chains in developing regions: wholesale markets, wholesalers, logistics, and processing," *Food Security*, 2021, *13*, 1577–1594.

Roser, Max, Hannah Ritchie, Esteban Ortiz-Ospina, and Joe Hasell, "Coronavirus Pandemic (COVID-19)," *Our World in Data*, 2020. https://ourworldindata.org/coronavirus.

Ruan, Jianqing, Qingwen Cai, and Songqing Jin, "Impact of COVID-19 and nationwide lockdowns on vegetable prices: evidence from wholesale markets in China," *American Journal of Agricultural Economics*, 2021, *103* (5), 1574–1594.

Rudin-Rush, Lorin, Jeffrey D Michler, Anna Josephson, and Jeffrey R Bloem, "Food insecurity during the first year of the COVID-19 pandemic in four African countries," *Food Policy*, 2022, *111*, 102306.

Sargent, Thomas J and John Stachurski, "Economic Networks: Theory and Computation," *arXiv preprint arXiv:2203.11972*, 2022.

Schneider, Friedrich and Dominik H Enste, "Shadow economies: Size, causes, and consequences," *Journal of Economic Literature*, 2000, *38* (1), 77–114.

Sheffi, Yossi, *The power of resilience: How the best companies manage the unexpected*, mit Press, 2015.

_ and James B Rice Jr, "A supply chain view of the resilient enterprise," *MIT Sloan Management Review*, 2005.

Shupler, Matthew, James Mwitari, Arthur Gohole, Rachel Anderson de Cuevas, Elisa Puzzolo, Iva Čukić, Emily Nix, and Dan Pope, "COVID-19 Lockdown in a Kenyan Informal Settlement: Impacts on Household Energy and Food Security," *medRxiv*, 2020.

Soo, Kwok Tong, "Zipf's Law for cities: a cross-country investigation," *Regional Science and Urban Economics*, 2005, *35* (3), 239–263.

Spray, John, "Search Externalities in Firm-to-Firm Trade," Job market paper 2019.

_ and Sebastian Wolf, "Industries without smokestacks in Uganda and Rwanda," *Industries without Smokestacks: Industrialization in Africa Reconsidered*, 2018, pp. 341–363.

Startz, Meredith, "The value of face-to-face: Search and contracting problems in Nigerian trade," *Working Paper*, 2021.

Storeygard, Adam, "Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa," *The Review of Economic Studies*, 2016, *83* (3), 1263–1295.

Tabe-Ojong, Martin Paul Jr., Emmanuel Nshakira-Rukundo, and Bisrat Haile Gebrekidan, "COVID-19 and food insecurity in Africa: A review of the emerging empirical evidence," *European Review of Agricultural Economics*, July 2023, *50* (3), Pages 853–878.

Tian, Lin, "Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil," *CEPR Discussion Paper*, 2021, (DP16590).

Türkcan, Kemal, Socrates Majune Kraido, and Eliud Moyi, "Export margins and survival: A firm-level analysis using Kenyan data," *South African Journal of Economics*, 2022, *90* (2), 149–174.

UBOS, *Subnational Population Statistics*, OCHA Regional Office for Southern and Eastern Africa, 2020.

Ulyssea, Gabriel, "Firms, informality, and development: Theory and evidence from Brazil," *American Economic Review*, 2018, *108* (8), 2015–47.

Venables, Anthony J, "Equilibrium locations of vertically linked industries," *International Economic Review*, 1996, pp. 341–359.

Wagner, Joachim, "Exports and productivity: A survey of the evidence from firm-level data," *World Economy*, 2007, *30* (1), 60–82.

Wong, Woan Foong, "The round trip effect: Endogenous transport costs and international trade," *American Economic Journal: Applied Economics*, 2022, *14* (4), 127–66.

Yao, Yuliang, Yongrui Duan, and Jiazhen Huo, "On empirically estimating bullwhip effects: measurement, aggregation, and impact," *Journal of Operations Management*, 2021, *67* (1), 5–30.

Zárate, Román D, *Spatial misallocation, informality, and transit improvements: Evidence from mexico city*, The World Bank, 2022.

Zhou, Yue, "The Value Added Tax, Cascading Sales Tax, and Informality," in Maurizio Bussolo and Siddharth Sharma, eds., *Hidden Potential: Rethinking Informality in South Asia*, World Bank Publications, 2022, chapter The Value Added Tax, Cascading Sales Tax, and Informality, pp. p. 61–90.