GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Essays on Technological Change and Trade in Development Economics

Dissertation in order to acquire the degree of "Doctor rerum politicarum" from the Faculty of Economic Sciences at the Georg-August-Universität Göttingen

Submitted by
Henry Walter Scott Stemmler
born in Freiburg im Breisgau, Germany

Göttingen, 2022

First examiner:

Professor Dr. Holger Strulik

Professor of Economics
Chair of Macroeconomics and Development
Department of Economics
Georg-August-Universität Göttingen

Second examiner:

Professor Dr. Krisztina Kis-Katos

Professor of Economics
Chair of International Economic Policy
Department of Economics
Georg-August-Universität Göttingen

Third examiner:

Professor Dr. Axel Dreher

Professor of Economics
Chair of International and Development Policy
Alfred-Weber-Institute for Economics
Ruprechts-Karls-Universität Heidelberg

Acknowledgements

I would like to take this page to express my gratitude and appreciation towards the many people, without whom this thesis would not have been possible. I am greatly indebted to those that gave me guidance, encouragement and kept the spirits high throughout this journey.

First, I thank my principal supervisor, Holger Strulik, who always provided swift and excellent feedback, and the freedom to pursue the various topics that I found interesting. I especially appreciate discussing various academic and non-academic issues over long forest-walks in times of social distancing. I am very grateful to my second supervisor, Krisztina Kis-Katos, for the continuous support and countless talks with academic and professional advice. Her dedication to academic research but also to the well-being of the PhD students of the GlaD program have been truly inspiring. I thank Axel Dreher, the third member of my thesis committee, for his greatly valued feedback on research projects of different stages and academic insights. I would also like to express my gratitude to my co-authors Richard Haarburger, Krisztina Kis-Katos, Tobias Korn, Friederike Lenel, Jan Priebe, Ute Rink, Florian Unger and Christoph Weisser with whom countless hours were spent pondering over issues to identification, setbacks during fieldwork or the sudden appearance of similar new papers.

I have been very lucky to have been part of the fantastic group of people in the third and final cohort of the GlaD program, funded by the German Research Foundation (DFG). Laura Barros, Andrea Cinque, Anna Gasten, Richard Haarburger, Tobias Korn, Johannes Matzat, Tatiana Orozco García, Anna Reuter, Lisa Rogge, Claudia Schupp, Lukas Wellner and Yuanwei Xu: Thank you all for the many good memories, I look forward to many more in the future! A special shout-out to Richard, for the good vibes in the office (pre-Covid), but also the ”magic” times with various card and board games. Also a big thanks to our PostDoc Friederike, who opened many doors and was a huge help with academic and career questions. I'm grateful to Antje Juraschek, Helga Hemmersmeier, Sarah Khan, Celia Mewes, Christiane Fuhlrott and Melanie Grosse for the administrative support and having a clear view in the confusing world that is University regulations.

Beyond the academic sphere, the company of my friends and family have given me the motivation to pursue this endeavor, but also much needed distractions and times to reset. I'm especially grateful to my parents, Judith and Gerhard, and my siblings, Nelly and Harvey, for always supporting and encouraging me on my path, and my nieces, Liv and Cleo, for bringing so much joy to this world. Above all, I thank Anne, for standing by my side, believing in me and giving me the strength to go through with this thing.

Contents

List of Figures

List of Tables

Introduction

I.1 Thesis Outline

The recent COVID-19 pandemic has demonstrated how interconnected our world today is. The closure of borders in many countries, restriction of travel and the disruption of supply chains have reminded many people of the luxury that free travel is and the costs that an interruption of the flow of goods entail. At the same time, the global economy held its breath over a vessel, which, stuck in the Suez Canal, single-handedly caused massive delays in international shipments. Anthony Fauci, the Chief Medical Advisor to the president of the United States, stated that "The world is a place that is so interconnected that what happens in another part of the world will impact us.". The pandemic has showcased the accuracy of this statement - not only in relation to health, but also to our economic system as a whole.[1]

Global interactions, interrelations and inter-dependencies, or globalization, in its many facets, has long been at the center of economic research and public debate (Banerjee and Duflo, 2019). The effects of economic integration and the removal of trade barriers have been studied for decades (Ricardo, 1817; Amiti and Konings, 2007; Goldberg and Pavcnik, 2007; Topalova and Khandelwal, 2011). Naturally, globalization is not merely related to the flow of goods (Dreher, 2006). It is impossible to imagine a world today without financial flows, for instance in the shape of investments, remittances or aid (Giuliano and Ruiz-Arranz, 2009; Bruno and Shin, 2015; Rey, 2015; Dreher and Langlotz, 2020). Migration, as one of the most visible aspects of globalization, significantly shapes political discourses (Alesina et al., 2018; Dustmann et al., 2019).[2]

Not only the spread of people has increased substantially, but also of ideas and social norms (Kis-Katos et al., 2018; Zhuravskaya et al., 2020). In that sense, when thinking about our intertwined global society, for most people, most likely new technologies come to mind before trade. The spread of fast internet and the emergence of social

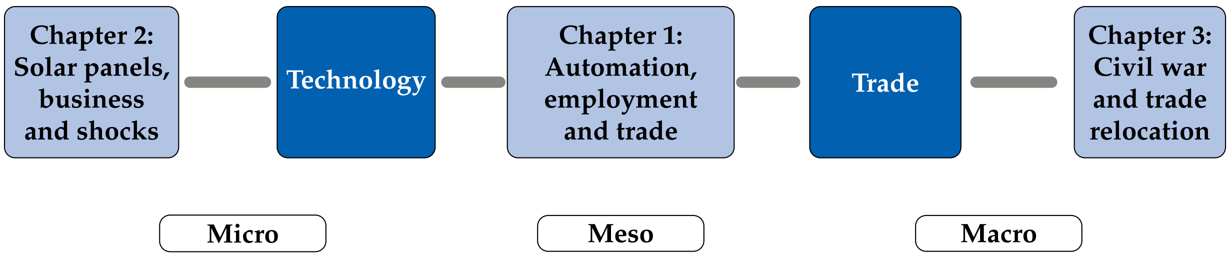
[1] How intertwined these two are is best exemplified by the closure of major cargo ports in Southern China related to COVID-19 outbreaks, which caused large-scale congestions and delays in deliveries.

[2] Global migration flows have not increased between between 1960 and 2000, but migration is concentrated on fewer destination countries (Czaika and De Haas, 2014). However, in general, there is a misconception about the share of migrants and their characteristics (Alesina et al., 2018).

media have made the world a ”global village” (McLuhan et al., 1968). Technological advancements are happening at increasingly faster rates; as the famous ”Moore’s Law” predicted, the number of transistors on integrated circuits increases exponentially, roughly doubling every two years (Moore et al., 1965). A vast literature has explored how new technologies have shaped societies, the international economies and labor markets, for instance in relation to electrification (Dinkelman, 2011), the spread of internet (Hjort and Poulsen, 2019) or the rise of automation technologies (Graetz and Michaels, 2018; Acemoglu et al., 2020).

This dissertation investigates how indirect and subtle effects related to international trade and technological change affect international economic development and the well-being of individuals in developing countries. The world is interconnected to a degree that changes in policies or drastic events happening in one or between multiple countries can have far-reaching effects on other, not directly targeted or affected countries. Such indirect effects in relation to globalization are most often studied in relation to trade agreements or production networks. Dai et al. (2014) show that countries signing trade-agreements trade more with each other, and less with other countries. Therefore, trade flows relocate away from countries not directly involved in the agreement. Furthermore, idiosyncratic shocks to one firm, location or sector can have negative consequences for other firms and customers through upstream and downstream supply chains (Acemoglu et al., 2016; Barrot and Sauvagnat, 2016). Since production networks today span across borders, such shocks can cause ripple effects through the international economy. For example, quite literally, Boehm et al. (2019) show that an earthquake in Japan disrupted production of affiliate firms in the USA, as the supply of inputs fell substantially. Also new technologies, which for instance alter production systems, may induce transformations beyond the location of their operation (Graetz and Michaels, 2018; Krenz et al., 2021), and those which introduce new applications may create new opportunities beyond their primary field of use (Aker and Mbiti, 2010; Suri, 2017). Figure I.1 illustrates the outline of the thesis. The two overarching themes, technology and trade in the context of development, are studied together or separately in each of the three chapters. Moreover, each study marks a different level of analysis, from the micro-level in chapter 1, to the macro-level in chapter 3.

Figure I.1: Thesis outline



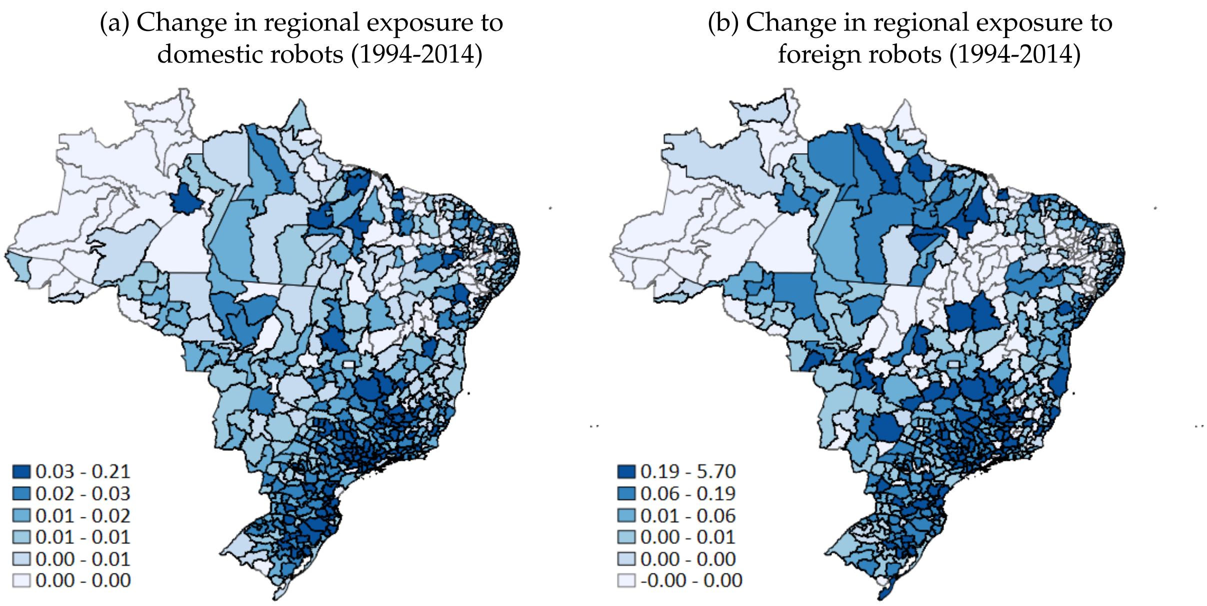
The first study combines the two themes at the meso-level: It shows that the uptake of automation technologies around the globe leads to lower employment rates in the manufacturing sector in Brazil and higher employment rates in extractive industries. The main channel through which automation in other countries affects local labor markets in Brazil is via input-output linkages and a shift in global trade patterns. Chapter two shows how a new technology, namely solar panel home systems, is used by farmers in rural areas to generate income after experiencing agricultural shocks, by providing energy to run small-scale businesses. Chapter three returns to the topic of trade. It analyzes how a civil war in one country alters trade-flows between two other, not directly affected countries.

Methodologically, this thesis uses a variety of estimation techniques, types of data-sets and levels of observation. Estimation strategies include the shift-share approach, instrumental variables, high-dimensional fixed effects models, gravity-estimations and general equilibrium analyses. On the micro-level, chapter 2 combines individual-level and high-frequency data with machine learning techniques to identify certain types of behavior from electricity consumption patterns. Furthermore, geo-referenced climatic indicators are used to define local agricultural shocks in Tanzania. For chapter 1, I combine a rich administrative employee-level data-set, regional trade-flows and global data on sector-level robots to calculate the exposure of local labor markets in Brazil to domestic and foreign automation. Chapter 3 takes a macro-level perspective, analyzing trade flows of 180 countries. In the study, we develop a method which allows to estimate how an isolated shock, here in the form of civil wars, affects two other countries within a Structural Gravity framework (Yotov et al., 2016b). In the following, each of the chapters is summarized in more detail.

I.2 Summary of Chapters

Chapter 1 In the first paper of this thesis, I investigate how the growing robotization of the global economy affects an emerging economy, namely Brazil. Developing and emerging economies may be especially vulnerable to automation technologies, due to the large share of lower-skilled workers engaging in routine-manual tasks (Maloney and Molina, 2016). However, up-to-date, research on the effects of automation has focused primarily on industrialized economies. Even though developing and emerging economies themselves have adopted relatively few robots, robots in other countries may already indirectly affect their labor markets (Krenz et al., 2021; Faber, 2020). The paper first develops a theory of how robot adoption in domestic and foreign industries may have differential effects on local labor market employment. I then empirically study these effects for the case of Brazil. The identification strategy is based on a shift-share approach and instrumental variables. The exposure of local labor markets to domestic automation is constructed by combining sectoral stocks of robots with the initial industry-employment distribution within local labor markets and inter-sectoral linkages within Brazil. Input-output linkages between local labor markets and foreign industries are used to construct an index of exposure to foreign automation. Figure I.2 shows the regional variation in the change in the exposure to domestic (panel (a)) and foreign (panel (b)) robots, with darker colors representing a larger exposure. While the regions in the industrial areas in the Southeast of Brazil are especially exposed to domestic and foreign robots, there is quite some regional heterogeneity in the two measures.

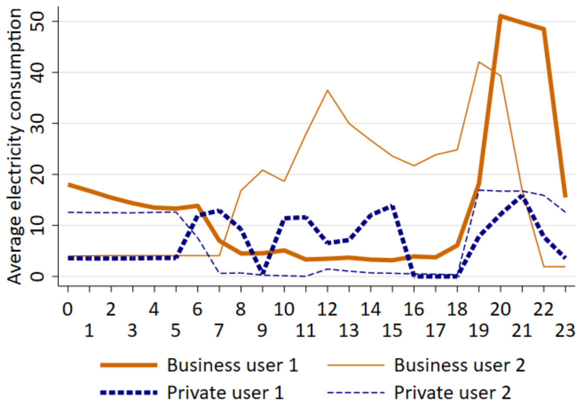
Figure I.2: Regional exposure to domestic and foreign robots



The empirical results demonstrate that automation in export destination countries leads to declining employment in the manufacturing sector in Brazil and an increase in employment in the extractive, or mining, sector. These shifts are driven by changes in the demand for export goods: Regions more exposed to foreign automation face on average declines in exports of manufacturing goods, while exports of raw materials increase. These findings indicate that automation in industrialized economies increases their production in the manufacturing sector, thereby lowering the demand for imports of such goods. At the same time, larger production requires more inputs, mirroring rising exports of raw materials. Domestic automation in turn is found to benefit higher skilled and female workers. The results are robust to a number of robustness tests and alternative specifications. The findings demonstrate how technological change in industrialized countries may cause premature deindustrialization and developmental setbacks in developing and emerging economies.

Chapter 2 The second chapter of my thesis, co-authored with Krisztina Kis-Katos, Friederike Lenel and Christoph Weisser, studies whether a relatively new technology, namely small-scale solar panel home systems, can help farmers in rural areas of Tanzania to mitigate income losses induced by agricultural shocks. Social safety nets and off-farm income generation possibilities are often not available in rural areas, leaving farmers vulnerable to such shocks (Dercon and Krishnan, 1996; Barrett et al., 2001). To explore whether solar panels enable farmers to generate additional income, we make use of high-frequency data on loan repayments and electricity usage that we obtain from a solar panel company which operates in East Africa. Combining electricity consumption patterns and a small survey that includes questions on the types of application of the solar panel, we predict the likelihood of customers using the solar panel for business purposes with supervised machine learning methods. Figure I.3 displays different electricity consumption patterns of 2 labeled private and 2 labeled business users over the course of an exemplary day. In general, business users have a larger consumption of energy, especially in the early and late evening hours. But the types of businesses differ; business user 2 requires more energy during the day, while business user 1 powers appliances only in the evening. Such patterns allow us to identify days on which customers are likely to use their system for business purposes outside of the labeled survey data.

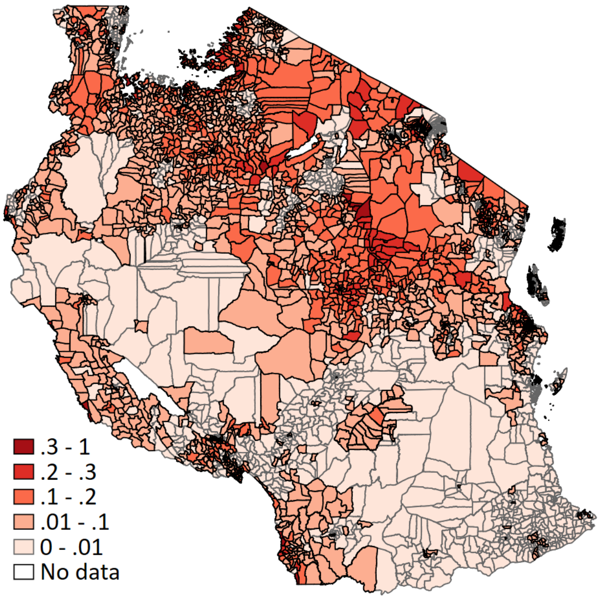
Figure I.3: Electricity use patterns of selected customers



In addition, we use deviations from long-term geo-referenced vegetation indices to identify agricultural shocks. The regional heterogeneity of such shocks is depicted in Figure I.4, where darker red colors indicate a stronger deviation in plant health in 2017 as compared to the long-term average.

Three main results follow from our analysis: First, we find that agricultural shocks lead to income losses for farmers. Second, after such shocks, farmers are more likely to use their solar panel home system for business purposes. Third, using the system for income generation reduces the magnitude of the negative income shock. These findings

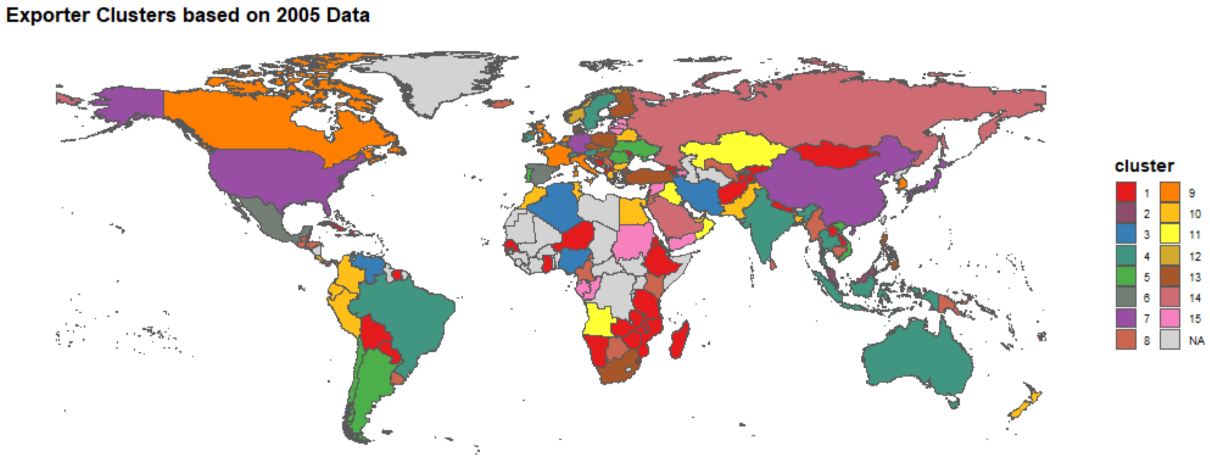
Figure I.4: Agricultural shocks in 2017



suggest that solar panels can help farmers to cushion income losses from agricultural shocks. Furthermore, we find that farmers with fewer resources and who live in more remote areas are more likely to adjust the way they use their solar panel, and that this adjustment is only a short-term strategy. With increasingly frequent and severe climatic stress events, strategies to mitigate income shocks will become increasingly important. Low-cost and flexible solar panels are a promising way to provide means for income generation in areas where access to electricity and off-farm employment opportunities are scarce.

Chapter 3 In the third study of my thesis, which is joint work together with Tobias Korn, we develop a novel method which allows to directly estimate how a shock in one country may affect other countries, that the shock does not immediately concern. It is derived from the structural gravity model of international trade and translates the triadic relationship between a conflict country and an exporter-importer pair into an estimable dyadic relationship. More specifically, we construct triadic relationships based on the conflict country having been an important exporter for a given importer, and another exporter having a similar production structure as the conflict country. The logic behind this relationship is that an importer, for which an important exporter can no longer provide goods due to the onset of a civil conflict, will look for other exporters producing similar goods. We allocate countries with a similar production structure into clusters, which is depicted in Figure I.5 for the year 2005. The map shows which countries are defined as similar, for instance Canada and several European countries, or Brazil, India and Australia.

Figure I.5: Export similarity clusters



We thereby extend the literature on trade relocation effects. Such effects have so far been estimated either in a dyadic, partial equilibrium framework, for instance with only one country joining a trade-agreement (Dai et al., 2014; Mattoo et al., 2017), or in a general equilibrium, case-study type of analysis (Anderson et al., 2018a; Felbermayr et al., 2019a). We take this method to data in a structural gravity framework. The estimation results show that importers divert trade flows from conflict countries to alternative export partners in the agricultural, manufacturing and minerals sectors, but not in the fuels sector. This trade diversion effect persists up to 9 years after civil conflicts end. Complementing this finding, trade diversion fosters market integration via trade-agreements among affected trade-partners. The findings indicate a double penalty of civil conflict. In addition to the immediate costs and dire consequences of conflicts on the society and economy, economic recovery of conflict countries is hindered by the fact that supply chains are manifested between other countries. Our method is variable and applicable to many other research topics. For example, it possible to study how climatic shocks affect trade relations between other countries. Our method can be applied and extended to other outcomes than trade, such as FDI or migration flows, or other one-sided shocks, such as droughts or natural disasters.

The findings of the individual chapters highlight how important it is to take into account the indirect and unintended consequences that small- and large-scale events and actions have in the global economy. The rise of automation technologies already today affects developing and emerging economies. A civil war in one country leads to, under certain conditions, increasing trade-flows and economic integration between two other countries, resulting in a double penalty for nations in conflict. On a more positive note, small-scale and low-cost solar panels do not only contribute to electrification in rural and poor regions, but can also provide the means for farmers in rural areas to cope with income shocks.

Chapter 1

Automated Deindustrialization: How Global Robotization affects Emerging Economies - Evidence from Brazil

Abstract This paper investigates how domestic and foreign automation impact an emerging economy. The empirical analysis builds on a shift-share approach, exploiting differences in regional industrial compositions, inter-sectoral input-output connections and differential linkages of local labor markets to foreign industries. Instrumental variables account for endogeneity in robot adoption. Larger exposure to foreign automation is found to decrease the manufacturing employment ratio and increase in the mining sector employment ratio. These shifts are driven by changes in the demand for export goods from local labor markets. Domestic automation has lesser effects, but benefits higher skilled and female workers.

JEL classification codes: O10, O14, O19, O33, J23
Keywords: automation, development, trade, employment

1.1 Introduction

The emergence of robots and automation technologies has been one of most impactful global developments in recent years. While the discussion about the labor-replacing effects of robotics is centered around advanced economies, also less technologically-advanced countries may already be directly or indirectly affected by the rise of automation. Declining prices of robotics reduce production costs especially in industrialized countries, increasing competition in exporting markets for developing and emerging economies. As a consequence, production and employment in these countries could shift to light manufacturing, raw-material extraction and the service sector (Rodrik, 2016).

This paper analyzes whether automation taking place domestically and in more industrialized countries has induced such trends in Brazil, a resource-rich emerging economy. The declining manufacturing share in GDP and changing export composition illustrate that there has been a deindustrialization trend in Brazil (Jenkins, 2015).[1] In the early 2000's, Brazil had a relatively diverse export structure, both in terms of export destinations and producing sectors. It has however become increasingly reliant on primary goods.[2] The recession that hit Brazil in 2014 as a consequence of the collapse of international commodity prices can at least to some part be attributed to the concentration on mineral extraction as compared to other industries and revealed the dangers of such an economic development (Spilimbergo and Srinivasan, 2019).

The empirical analysis of this paper is based on a Ricardian model of trade with production of intermediate and final goods, in which domestic robots replace labor in certain industry specific tasks. The model shows that foreign automation affects local labor markets (LLMs) through changing expenditures on intermediate and final goods in the global economy. Lower production costs increase exports from automating industries and at the same time increase demand for intermediate goods and raw materials. Thus, automation in advanced economies indirectly affects less technologically-advanced countries. These direct and indirect mechanisms are empirically tested by exploiting differential exposures of LLMs to domestic and foreign automation. Sectoral stocks of robots, the initial industry-employment distribution within LLMs and intersectoral linkages within Brazil compose the domestic robot exposure. Input-output linkages between LLMs and foreign industries are used to construct an index of exposure to foreign automation. Instrumental variables (IVs) for domestic and foreign robot exposure, which build on the incentives and feasibility of industries to automate, ac-

[1]See Figure 1.A.1 in the Appendix.

[2]In 2000, after Brazil's trade liberalization, the largest exporting sectors were transportation and machine manufacturing, while mineral products made up only 8% of exports (Simoes and Hidalgo, 2011). In 2014, exports of mineral products rose to 22%, while transportation and machine manufacturing together made up less than 15%.

count for endogeneity in the decision to use robots.

The results demonstrate that increased usage of robots in advanced economies leads to employment shifting from higher value-added manufacturing industries to lower value-added raw-material extraction activities in Brazil. Foreign automation moreover causes the same pattern for sectoral exports from LLMs, suggesting that shifts in demand for export goods are causing employment changes. Conversely, employment in foreign owned companies does not change. Changes in employment thus are not caused by the direct replacement of workers through robots in multinational companies. Regions with an average change in the exposure to foreign robots experience a 0.043 percentage point slower growth in their manufacturing employment ratio and a 0.042 percentage points larger growth of the employment ration in the raw materials sector. Exposure to domestic automation has lesser effects. However, female and higher-skilled workers gain in terms of employment ratios, suggesting skill complementary with the new technology.

The literature about automation has mostly been focused on developed countries, while there is little evidence as to how emerging economies are affected by robotization.[3] Evidence about the latter is almost exclusively centered around the notion of "reshoring".[4] As robots become cheaper, it may become more profitable for firms in industrialized countries to shift back production facilities from developing countries. Carbonero et al. (2018) document that advanced economies decrease their offshoring activities, which has a negative employment effect for emerging economies. Similarly, Krenz et al. (2021) find that robot adoption leads to reshoring, benefiting high-skilled workers in advanced economies. Using data on LLMs in the USA, Bonfiglioli et al. (2021) find that automation has a weaker labor displacement effect in commuting zones that are more exposed to offshoring. The authors conclude that automation induces reshoring and displaces labor in non-offshorable occupations. Faber (2020) demonstrates that robot adoption in the United States leads to reshoring of production entities from Mexico. As a consequence, employment in LLMs in Mexico declines.[5] Also studying the impact of automation in the United States on labor markets in Mexico, Artuc et al. (2019b) document lower exports of consumption and intermediate goods from Mexico, consistent with reshoring patterns. The authors however don't find that

[3]Studies have for instance shown that automation in industrialized countries has a labor-replacing effect (Graetz and Michaels, 2018; Bessen et al., 2019; Acemoglu and Restrepo, 2020) and increases output and profits (Koch et al., 2019; Acemoglu et al., 2020; Aghion et al., 2020).

[4]There are several studies which look at the broader theme of occupational shifts, which are at least partly attributed to skill-biased technological change, in developing and emerging countries (Maloney and Molina, 2016; Apella and Zunino, 2017; Brambilla and Tortarolo, 2018; Brambilla et al., 2021; Gasparini et al., 2021; Messina and Silva, 2021).

[5]Cortes and Morris (2020) neither find evidence for offshoring of middle-skill routine manual tasks within industries from the USA to Mexico, nor that workers with such tasks are replaced through the use of Mexican intermediate inputs. This is taken as suggestive evidence for new technologies substituting labor in tasks, but could also be driven by offshoring to third countries, such as China.

these changes in exports have consistently led to employment declines in Mexico.

Less attention has been paid to linkages between developing and developed countries aside from specific offshoring activities, but rather through global trade flows. That a substantial share of exports from developing countries can potentially be produced by robots in advanced economies is documented by Artuc et al. (2019a). The authors show that automation increases net-exports in the Global North, but does not harm exports from commodity exporters, as there is a larger demand for raw material inputs and intermediate goods.[6] While higher foreign demand for raw materials might be beneficial for emerging economies with natural resources in the short-run, lower value-added production can hinder growth in the long-run and increase the likelihood of deindustrialization and getting caught in a resource trap. Up to date, these effects have however not been empirically shown.[7] This paper is the first to provide evidence of how automation in advanced economies affects an emerging economy through input-output linkages. The findings are in line with the notion of global automation inducing deindustrialization and a focus on raw-material extraction in emerging economies (Rodrik, 2016).

The rest of the paper is structured as follows. In Section 1.2, the theoretical model and empirical strategy is developed. Section 1.3 gives a short overview of the different data sets used in the analysis. Thereafter, Section 1.4 estimates the effects of domestic and foreign automation on LLMs in Brazil and runs a number of robustness tests, such as tests for the validity of the shift-share estimation (Adão et al., 2019; Goldsmith-Pinkham et al., 2020; Borusyak et al., 2018), a pre-trend analysis and alternative empirical specifications. Section 1.5 concludes.

1.2 Theory and Empirical Strategy

1.2.1 Theoretical model

This section develops a model which identifies how domestic and foreign automation affect local labor markets. The model builds on Artuc et al. (2019a), Caliendo and Parro (2015) and Acemoglu and Restrepo (2020). Regions in Brazil are denoted by r, foreign economies by j, with there being J regions and countries overall.[8] There are K

[6] Similarly, Aghion et al. (2020) find that robot-adoption increases profits and production of exporting firms in France, which can reduce costs and serve a larger international market. Cilekoglu et al. (2021) provide evidence that firms which adopt robots indeed increase imports of intermediate goods.

[7] Kunst (2020) descriptively shows that changes in the occupational composition in developing countries do not point to deindustrialization. However, the author does not look at specific shocks, but rather at broad changes within economies.

[8] As in Autor et al. (2013) and Acemoglu and Restrepo (2020), workers are immobile between regions and countries, but can move between industries.

industries, represented by i (own industry) and k (partner industry). The production stage s of intermediate goods is defined as $s = 1$ and of final goods as $s = 2$. Households in region r maximize utility by consuming final goods, given by a Cobb-Douglas utility function

$$U_r = \prod_{i}^{K} (Q_{ri}^2)^{\gamma_{ri}}, \tag{1.1}$$

where $\sum_i \gamma_{ri} = 1$ and Q_{ri}^2 is the demand of final goods of industry i in region r. Households receive income by supplying labor L_{ri} at wage w_{ri}^L. Firms in region r produce varieties $\omega \ \epsilon \ [0,1]$ as intermediate goods or final goods. Firms use three inputs for production: Capital F, a task input T and intermediate inputs Q^1, including raw materials.[9] The Cobb-Douglas production function of variety ω is

$$q_{ri}(\omega_i) = z_{ri}(\omega_i) F_{ri}(\omega_i)^{\alpha_{ri}^F} \prod_{k}^{K} Q_{rik}^1(\omega_i)^{\alpha_{rik}^M} T_{ri}(\omega_i)^{\alpha_{ri}^T}. \tag{1.2}$$

The parameter α_{rik}^M denotes the share of inputs used from industry k for the production of variety ω_i, where $\sum_k \alpha_{rik}^M = 1 - \alpha_{ri}^F - \alpha_{ri}^T$ (Caliendo and Parro, 2015). Intermediates from other industries can be used for the production in industry i, permitting to examine important Input-Output (I-O) patterns. The efficiency of production in region r industry i is drawn from the *Fréchet* distribution as in Eaton and Kortum (2002) and given by z_{ri}.[10] Due to the probabilistic distribution of technology, countries and sectors have different levels of productivity. As in Caliendo and Parro (2015), λ_{ri} denotes the location parameter varying by country and sector, and θ_i the shape parameter, which is the industry specific variation of efficiency in production. Varieties ω_i can be sourced from international suppliers.[11] The expenditure share of country j on goods from region r and industry i can thus be written as

$$\pi_{jri} = \frac{\lambda_{ri} \left[(c_{ri} \tau_{rji}) \right]^{-\theta_i}}{\sum_h \lambda_{hi} \left[(c_{hi} \tau_{hji}) \right]^{-\theta_i}}, \tag{1.3}$$

where τ_{rji} are iceberg-type costs per unit shipped between region r and country j and c_{ri} is the cost function of producing ω_i.[12]

Tasks T_{ri} in production are allocated between human labor and robot capital in the

[9] As in Artuc et al. (2019a), the intermediate good Q^1 is used only as a production input, while Q^2 is consumed by households.

[10] See equation 1.A.5 in the Appendix.

[11] The production technology is an aggregator as in Caliendo and Parro (2015), see equations 1.A.2 and 1.A.4 in Appendix 1.A.2 for further derivations.

[12] See equation 1.A.1 in the Appendix.

continuum $b \epsilon [0,1]$. There is an automation frontier, which is given by C_i. This cut-off defines the advancement in robot technology in industry i, which, due to technological diffusion, is the same across the world. Let labor be defined as L and robots as R with the productivity of labor being $\gamma_L(b)$ and of robots $\gamma_R(b)$.[13] With $\gamma_L(b)/\gamma_R(b)$ increasing, labor has a comparative advantage in tasks which are closer to 1 (Acemoglu and Restrepo, 2020). Tasks from 0 to C_i can be performed by robots or human workers, where producers will choose the, productivity-adjusted, cheaper input.[14] Unit labor costs are given by w^L_{ri} and robot rental price per unit by w^R. If $\gamma_R(b)w^L_{ri}/\gamma_L(b)w^R > 1$, a robot will be used to complete task b (Artuc et al., 2019a). Therefore, tasks for which robots are used differ among regions and countries. Less developed economies are less likely to employ robots than industrialized countries, as wages are much lower. As the adoption of robots changes the share of workers, production costs change according to

$$c_{ri} = \psi_{ri} {f_{ri}}^{\alpha^F_{ri}} \prod_k^K {p^1_{rik}}^{\alpha^M_{rik}} \left(C'_{ri} \frac{w^R}{\gamma_R} + (1 - C'_{ri}) \frac{w^L_{ri}}{\gamma_L} \right)^{\alpha^T_{ri}}, \tag{1.4}$$

where f_{ri} is the rental rate for capital and P^1_{rik} the price of intermediate goods. Automation thus affects international sourcing and trade through (1) a shift in trade patterns (as can be seen by plugging in equation 1.4 into equation 1.3) and (2) by increased productivity and higher demand (demonstrated when plugging in equation 1.4 into equation 1.2).[15] Labor market clearing implies that labor income equals the labor's share of region r's and industry i's share of exports and domestic sales (Eaton and Kortum, 2002), such that, analogous to Acemoglu and Restrepo (2020),

$$w^L_{ri} L_{ri} = \alpha^T_{ri} s^L_{ri} P_{ri} \sum_j E_{jri}, \tag{1.5}$$

where $\sum_j E_{jri}$ is the expenditure for goods of industry i and region r from all countries and regions, s^L_{ri} is the share of labor in production and P_{ri} is the unit price of the composite good.[16] As goods can be sourced internationally, demand for goods of industry i in region r is made up of demand for final and intermediate goods from all countries. Plugging the share of labor in production tasks, expenditure shares from equation 1.3

[13] For simplicity, I here abstain from modelling labor of different skill levels. As in Acemoglu and Restrepo (2018), high-skilled labor could be modelled to be complementary in robot technology and low-skilled labor to be substitutable.

[14] See Appendix 1.A.2 for further derivations.

[15] Aghion et al. (2020) for instance show that automation induced productivity and output gains in France are driven by exporting firms. Artuc et al. (2019a) present a formal derivation of these two channels.

[16] See equations 1.A.7 and 1.A.10 in the Appendix.

and cost funtion 1.4, taking logs and differentiating equation 1.5 yields[17]

$$
\begin{aligned}
d\ln L_{ri} = \frac{-dC'_{ri}}{(1-C'_{ri})} - \left(\frac{1}{\alpha^T_{ri}} - 1\right) d\ln P_{ri} + d\ln \sum_j E_{ji} \pi_{rji}^{\frac{\theta_i \alpha^T_{ri}}{1+\theta_i \alpha^T_{ri}}} \frac{1}{1+\tau_{rji}} \\
+ \frac{1}{\alpha^T_{ri}} d\ln \prod_k p^1_{rik}{}^{\alpha^M_{rik}}.
\end{aligned} \tag{1.6}
$$

Equation 1.6 shows the differential effects of automation on changes in employment in region r and industry i. The first term, $-dC'_{ri}/(1-C'_{ri})$ captures how changes in domestic automation increase the threshold C' and thereby decrease labor in some tasks.[18] The second term is what Acemoglu and Restrepo (2020) call the 'composition effect', industries expanding at the expense of others. The summation term reflects changing expenditures for goods from region r from all countries and regions induced by automation, which also includes gains in productivity through automation. As automation reduces costs and thereby prices in certain industries, other countries shift their expenditures towards the cheapest supplier, which is captured by changes in π_{rji}. At the same time, automation increases productivity and therefore overall expenditure, which leads to an increase in E_{ji}. The last expression reflects the complementarity between the expenditure on inputs and required labor in production. Consider an industry i in region r automating. On the one hand labor is displaced by the new technology. At the same time, as production costs decrease, demand increases in the global economy. The higher production in turn increases labor demand (Acemoglu et al., 2020). Now consider a foreign country automating. A decrease in the robot rental rate w^R decreases production costs especially in developed countries, where wages are higher a priori. As production becomes cheaper, manufacturing of certain goods, intuitively more complex goods, shifts to these countries. To produce these goods, more raw materials and intermediate goods are required. Since it is neither cheaper nor feasible to produce everything, some goods will still be sourced from less developed economies, where wages are lower and (other) natural resources available. Thus, even if a country doesn't automate itself, it will still be affected by foreign automation.

[17] Fixed capital is exogenously given as in Acemoglu and Restrepo (2020). Equation 1.A.11 in the Appendix displays the equation without taking logs and differentiating.

[18] I here abstain from the possibility that automation may create new tasks, which is described in Acemoglu and Restrepo (2019b).

1.2.2 Empirical Specification

Equation 1.6 is the basis for the empirical specifications capturing how domestic and foreign automation affect employment in Brazil. It has been well established in the literature that workers are imperfectly mobile across space, but move between industries (Autor et al., 2013; Dix-Carneiro and Kovak, 2015; Kovak, 2013). Therefore, LLMs are used as the unit of analysis instead of industries.[19] In order to estimate the effects of robot adoption on labor markets, a shift-share approach translates the industry-level increases in robots to regions in Brazil.[20]

To map industry-level increases in robot stocks to the regional level, the change in these stocks are weighted with the initial share of each industry's employment in the region. Exposure to domestic automation is defined as

$$RE_r^{dom} = \sum_i \frac{L_{ri,1994}}{L_{r,1994}} \frac{dAR_{Bra,i}}{L_{i,1994}}, \tag{1.7}$$

where the left fraction on the right-hand side denotes the "share" of the shift-share operator, namely the initial share of employment in industry i in micro-region r.[21] The "shift" is the change in the adjusted sectoral stock of robots $dAR_{BRA,i}$ per 1000 workers. The stock of robots is adjusted with an additional weight to account for input-output linkages, which is explained in more detail in Appendix 1.A.3. The adjustment takes into account that robots in one sector will affect other sectors within the economy. As an alternative specification, robot exposure without input-output adjustments, i.e. simply inserting the sectoral stock of robots $dR_{BRA,i}$ instead of adjusted robots $dAR_{BRA,i}$, will be used in the empirical analysis. Comparing these two measures, coupled with an analysis of regional GDP growth, allows to tentatively disentangle the replacement and productivity effects of domestic automation, described in equation 1.6. The adjustment moreover accounts for the composition effect described above. The main specification used throughout the paper are stacked-differences between 1994, 2004 and 2014. Other specifications, such as long-differences between 1994 and 2014, are used to test robustness of the main results.

The decision to use robots in production is most likely a choice of cost-savings and quality of production (Acemoglu and Restrepo, 2019a). Since these factors clearly depend on local labor market conditions, the above defined exposures to robots are

[19] In addition, more variation between regions than between industries can be exploited in the data, which is described below.

[20] Shift-share specifications have been implemented in similar contexts for instance by Acemoglu and Restrepo (2020), Faber (2020) and Dauth et al. (2017).

[21] 1994 is the first available year in the data that I can use, due to different region and industry classifications before. Borusyak et al. (2018) discuss the shift-share design using panel data, specifically whether to use yearly industry shares or an initial share. In the case of serial correlation in the data, which in the automation data is present, the authors advocate using initial shares.

susceptible to endogeneity concerns. An IV following the instrument introduced by Graetz and Michaels (2018) is used to mitigate these concerns. The IV is constructed by substituting domestic robots $\mathrm{d}R_i$ in equation 1.7 with the term

$$dAR_i^{IV_{dom}} = dR\, a_{i,1994}, \tag{1.8}$$

where $\mathrm{d}R$ is the change in the global stock of robots, a proxy for the price decline in robots (Artuc et al., 2019a). The global decline in robot prices is exogenous to single countries and industries and is interacted with the share of "automizable" workers in industry i in 1994 (incentive to automate), $a_{i,1994}$.[22] A worker is defined to be "automizable", if the description of the occupation's task matches the description of robot applications (International Federation of Robotics, 2018). The instrument makes use of exogenous time-variation, which is interacted with a cross-sectional variable to predict in which industry robots are likely to be installed. In addition, an alternative instrument is used, which exploits the average number of robots in other emerging economies in the data set as an exogenous source for technological progress, as it is very unlikely that robot adoption in these countries is driven by changes in the Brazilian labor market.[23]

Automation abroad can have differential effects on employment in Brazil, through changing expenditures on goods from different industries. Naturally, not all regions are equally exposed to foreign automation. Therefore, changes in foreign robots are weighted by a measure of how important exports of each LLM to a foreign industry are for the region's GDP.[24] Regional exposure to foreign automation is defined as

$$RE_r^{for} = \sum_j \sum_i \frac{X_{rji,2000}}{G_{r,2000}} dAR_{ji}, \tag{1.9}$$

where $X_{rji,2000}$ are exports of industry i in region r to country j in the year 2000 and $G_{r,2000}$ is the region's GDP for this year. The "share" thus now is not the initial share of industry employment in micro-region r, but measures how exposed a micro-region is to a sector in a foreign economy. The term dAR_{ji} is the change in the adjusted foreign robots per workers, as explained in Appendix 1.A.3. Analogous to the measure of domestic robots, the adjustment takes into account that exports from one sector can be used for production in different industries abroad. This constitutes a novel way of

[22] To again account for input-output linkages, $a_{i,1994}$ again includes the inter-sectoral adjustment term laid out in Appendix 1.A.3.

[23] These other emerging economies are India, Indonesia, Turkey, and Mexico. Similar instruments have been used by Acemoglu and Restrepo (2020), Faber (2020) and Micco (2019).

[24] Hakobyan and McLaren (2016) introduce a measure capturing the vulnerability of each local labor market in the United States to NAFTA, by using geographic variation in the exposure to imports from Mexico. A similar approach is used here.

measuring exposure to foreign automation. Putting equation 1.9 into words, a region in Brazil will be more affected by foreign automation if its goods are exported to a higher automating industry and if these exports make up a large share of region r's GDP.[25]

The variable RE_r^{for} reflects how foreign automation affects more exposed regions in Brazil through changing expenditures.[26] Artuc et al. (2019b) use a similar shift-share measure for foreign robot exposure, which accounts for the share of exports from regions in Mexico to the United States. The measure developed here furthermore accounts for input-output linkages between sectors and multiple trade-partners. Exposure to foreign robots may also be endogenous if local labor market conditions directly affect foreign firms' automation decisions. While improbable, high prices of export goods from Brazil could induce foreign firms to use robots, to lower costs. It is unlikely that a local labor market in Brazil will have such impact on global markets, however, to rule out endogeneity concerns, an IV for foreign robot exposure is used as a robustness test in the empirical analysis. The instrument is explained in more detail in section 1.A.3 in the Appendix.

The main estimation equation is a stacked-difference regression defined as

$$dY_{rt} = \beta_0 + \beta_1 dRE_{rt}^{dom} + \beta_2 dRE_{rt}^{for} + \chi_r + \gamma_{mt} + \varepsilon_r, \tag{E.1}$$

which is estimated via OLS and a two-stage least squares procedure, with the above defined instrument for dRE_r^{dom} being estimated in the first-stage. dY_r denotes the outcome variable of interest. The main outcomes are changes in employment, for instance by sector, between 1994, 2004 and 2014 at the micro-region level in Brazil. The vector χ_r includes a set of initial micro-region characteristics which could also affect labor demand. Moreover, state-year fixed effects γ_{mt} are used to control for unobserved confounders over time on the state level. For instance, policies and local spending decisions are administered in states in Brazil. Table 1.A.1 in the Appendix presents summary statistics of the main outcome variables and covariates. All variables' means are shown over all regions, as well as in regions which are below and above median exposed to domestic and foreign automation. As could be expected, regions differ for instance in terms of employment and GDP.

[25] One caveat with the available data is that the year 2000 is the first point in time in which regional exports, regional GDP, input-output linkages between two industries and sectoral labor can be consistently matched. However, as robot uptake only took off in the 2000s, it is unlikely that using data from 2000 as an initial value biases the results. Moreover, as Brazil's trade liberalization took place in the mid 1990s, using trade-flows from 2000 may prevent confounding factors (Pavcnik et al., 2004).

[26] Similar to Autor et al. (2013), this paper mainly focuses on one way that a foreign shock affects Brazilian labor markets, that is through the channel of exports. Foreign adoption of robots changes expenditure from on goods from Brazil, which as denoted by changes in $E_{ji}\pi_{rji}$ in equation 1.6.

1.3 Data

Data on the stock of country, industry and year specific robots are obtained from the International Federation of Robotics (2018).[27] Panel A of Figure 1.A.2 in the Appendix illustrates the data for the six industries with the most robots in Brazil. The automotive industry (right axis) has by far largest amount and growth rate of robots. Panel B displays the six industries that were mostly exposed to export weighted foreign robots. The graphs show that there is variation between the sectors, their growth in the stock of robots and between domestic and to foreign robot adoption.

Regional export and import data comes from SECEX (Secretaria de Comércio Exterior - Foreign Trade Secretariat), which contains information about trade volumes by product and destination market for each municipality in Brazil. Products are classified by the international Harmonized System, which are translated to the industry codes used in the analysis.[28] Initial sector level employment, the wage bill and output of foreign economies are taken from WIOD-SEA (Timmer et al., 2015). The WIOD-SEA 2016 version has the largest correspondence with the IFR data, but employment is available only from 2000 on. The sample consists of 43 countries.

Local labor market data from Brazil comes from the Relação Anual de Informações Sociais (RAIS) database of the Brazilian Ministry of Labor.[29] The RAIS is an annual administrative census, covering 99% of the formal enterprises, which report employee-specific data on a yearly basis.[30] Among others, data in the RAIS includes the employees' age, gender, educational attainment, wage and social benefits. Micro-regions in Brazil are used as the unit of local labor markets.[31] In total, there are 558 micro-regions, 137 meso-regions and 27 states in Brazil. The sample of workers is limited to individuals aged 16–65, to only observe the working-age population. Individuals working in public administration are excluded, as the public sector operates differently than other sectors (Dix-Carneiro and Kovak, 2017). One caveat of the data is that

[27] A Robot is defined by ISO 8373:2012 as an automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes, which can be either fixed in place or mobile for use in industrial automation applications (International Federation of Robotics, 2018).

[28] Goods that can fall into multiple industry categories, are allocated to industries based on the share of municipality exports of each industry. Since municipalities are such a small geographic unit, it is unlikely that many different unclassified industries exist. If there are no exports from the industries into which a product could fall, they are weighted by the share of micro-region, meso-region or state level export shares instead.

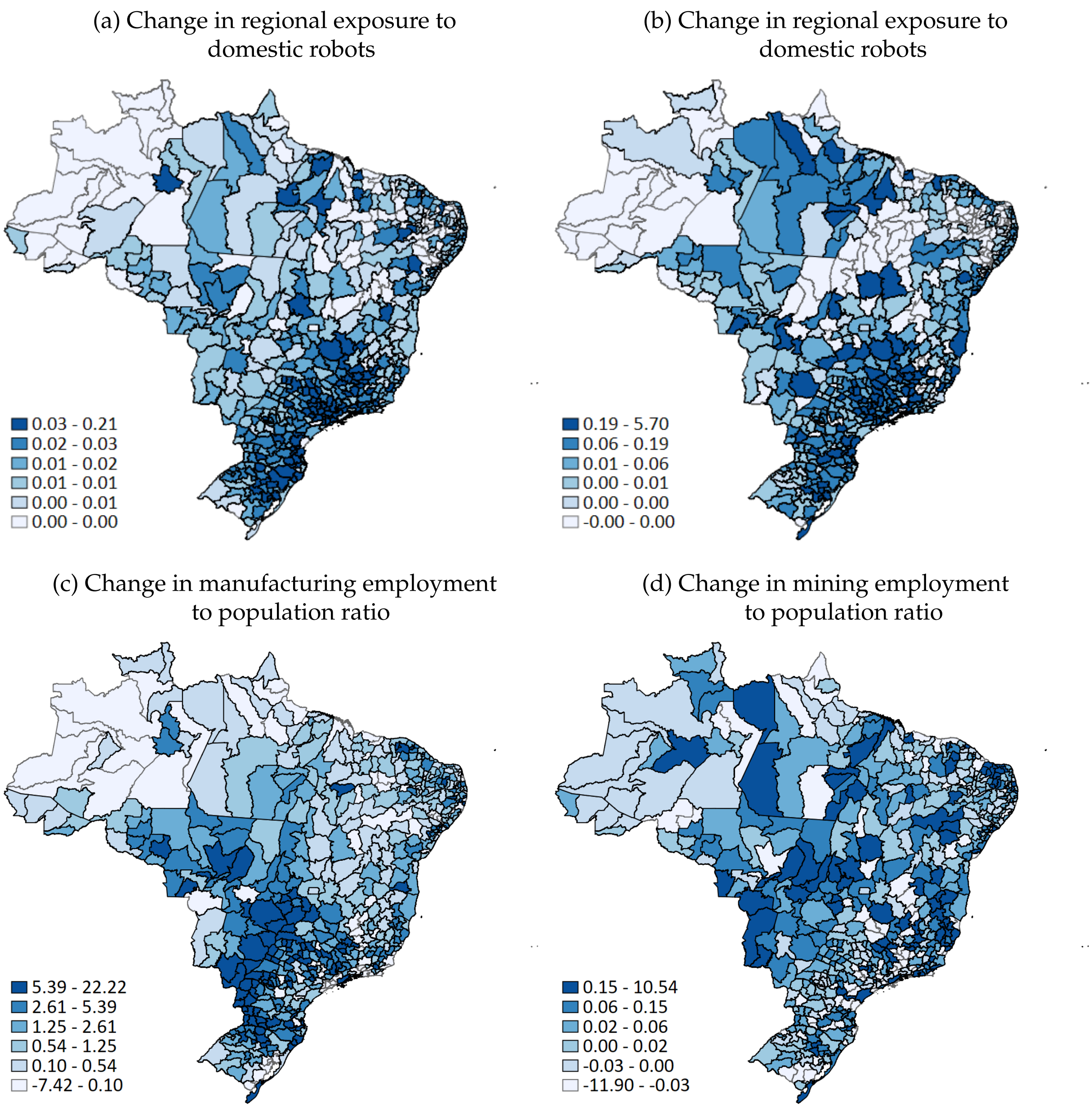
[29] For further information on the dataset, see also Helpman et al. (2017), Dix-Carneiro and Kovak (2015), Dix-Carneiro and Kovak (2017) or Alvarez et al. (2018).

[30] Firms are classified into sectors on a five-digit or six-digit level according to the CNAE code by the Instituto Brasileiro de Geografia e Estatistica (IBGE), which is similar to ISIC, and jobs are classified into the CBO code by the Ministry of Labor, similar to ILO's ISCO codes, also on a five-digit level.

[31] There is a strand of literature that investigates local labor market outcomes using micro-regions in Brazil, majorly of trade liberalization (Kovak, 2013; Costa et al., 2016; Dix-Carneiro and Kovak, 2015; Hirata and Soares, 2016).

it does not cover informal firms. Informal firms are unlikely to use robots in production or export though, thus it is unlikely that the omission biases the estimation. Nevertheless, individuals moving from formal to informal employment as a consequence of robot competition cannot be observed. Figure 1.1 displays the regional variation in exposure to foreign and domestic robots, as well as regional variation in changes in the manufacturing and mining employment to population ratio between 1994 and 2014. The figure gives a first indication that there is substantial variation both within and between regions in terms of exposure to domestic and foreign automation.

Figure 1.1: Changes in regional exposure to robots and employment (1994-2014)



Notes: The figure displays regional variation in the differences in domestic automation, foreign automation, the manufacturing employment ratio and the mining employment ratio between 1994 and 2014.

1.4 Estimation Results

In this section, the results of estimating the effects of exposure to domestic and foreign robots on labor market outcomes, as well as alternative specifications and several robustness tests are presented. Throughout, the coefficients of the domestic and foreign robot exposure differences are standardized to a standard deviation of 1.

1.4.1 The Effect of Automation on LLM Employment

The empirical investigation begins with OLS regressions in Table 1.1, which introduces the main setup and control variables. In all panels the outcome variable is a measure of the respective employment to population ratio, in stacked-differences. Column 1 includes state-year fixed effects, controlling for all developments on the state-level over time, as well as regional baseline characteristics. These are the initial micro-region population and GDP (in logs), the initial share of female workers, the average high school and university graduate rates, the routine task intensity index and the share of foreign owned enterprises.[32] The routine task intensity index is included to capture the regional susceptibility to general technological advancements, such as computerization (Faber, 2020). The share of foreign owned enterprises measures offshoring behavior of foreign companies. As industrial robots are most often used in manufacturing sectors, in column 2, the initial employment shares in manufacturing and light-manufacturing are added.[33] Column 3 controls for changes in overall regional imports and Chinese imports specifically, as it has been shown that Chinese imports have had a strong impact on local labor market outcomes in Brazil (Benguria and Ederington, 2017; Jenkins, 2015).[34] Column 4 uses initial population weights, to adjust for differences in the overall size of regions. Panel A of Table 1.1 displays the regression results with the total employment to population ratio as the outcome, while in the subsequent panels the dependent variable is the employment to population ratio of the manufacturing, mining and service sectors, respectively. The coefficient of foreign robot exposure is negative and statistically significant with the manufacturing employment ratio as the outcome, and positive and statistically significant with the mining employment ratio as the outcome in all specifications. This pattern indicates that a higher exposure to foreign automation leads to declines in manufacturing employment, but increases in mining employment. Domestic automation exposure is only statistically significant at

[32] The first year GDP is available is 1997, and the share of foreign firms cannot be calculated before 1995, because of different classifications of the firms' judicial status. All other variables are from 1994.

[33] Light manufacturing industries are the textile and the paper and publishing industries, following Acemoglu and Restrepo (2019b).

[34] Changes in imports are measured as the difference between the inverse hyperbolic sine of each region's imports between 1994-2004, and 2004-2014.

the 10% level in some specifications for total and manufacturing employment.

Table 1.1: OLS estimations: Employment ratios, overall and by sector (1994-2004-2014)

	Stacked differences			
	(1)	(2)	(3)	(4)
Panel A: Changes in the total employment ratio				
Domestic Robot Exp.	0.26	0.30*	0.28*	0.57*
	(0.18)	(0.18)	(0.17)	(0.32)
Foreign Robot Exp.	-0.11	-0.12	-0.11	-0.40**
	(0.19)	(0.19)	(0.19)	(0.18)
Panel B: Changes in the manufacturing employment ratio				
Domestic Robot Exp.	0.15*	0.13	0.13	0.19*
	(0.08)	(0.09)	(0.09)	(0.10)
Foreign Robot Exp.	-0.13***	-0.12***	-0.12***	-0.15***
	(0.04)	(0.04)	(0.04)	(0.04)
Panel C: Changes in the mining employment ratio				
Domestic Robot Exp.	0.00	0.01	0.00	-0.00
	(0.03)	(0.03)	(0.03)	(0.01)
Foreign Robot Exp.	0.11***	0.11***	0.11***	0.05***
	(0.04)	(0.04)	(0.04)	(0.02)
Panel D: Changes in the service employment ratio				
Domestic Robot Exp.	0.06	0.05	0.05	0.11
	(0.10)	(0.11)	(0.10)	(0.18)
Foreign Robot Exp.	0.00	0.00	0.01	-0.12
	(0.07)	(0.07)	(0.07)	(0.10)
Observations	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓
Regional char.	✓	✓	✓	✓
Industry controls		✓	✓	✓
Regional import growth			✓	✓
Population weighted				✓

Notes: Standard errors, in parentheses, are clustered on the meso-region level. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.2 introduces the IV estimations, which provide causal evidence, as automation exposure may be endogenous to labor market characteristics. The first two columns use the instrument for domestic automation, columns 3 and 4 for foreign automation and in columns 5 and 6 both instruments are employed. The full set of controls are used in all specifications. The same pattern as with the OLS regressions emerges and the coefficients remain similar in size. Throughout, foreign automation exposure is statistically significant in panels B and C. Foreign automation thus indeed leads to a shift from manufacturing industries to raw materials extraction. Taking the results of columns 1 and 2, a one standard-deviation increase in the exposure to foreign automation decreases the manufacturing employment ratio by 0.11 percentage points in the unweighted and by 0.14 percentage points in the weighted case. Put differently, regions with an average change in exposure to foreign automation (which is 0.1) have roughly a 0.04 or 0.05 percentage point slower rise in the manufacturing employment ratio.[35] In turn, a one standard-deviation increase in the exposure to foreign automation increases the ratio of employment in the raw materials sector by 0.11 percentage points in the unweighted and by 0.05 in the weighted case. Exposure to domestic automation in turn does not have a consistent effect on manufacturing or total employment. One explanation could be that automation is not as wide spread in emerging economies as in developed countries (see also Faber (2020)). Furthermore, it could be that direct employment losses due to installments of robots are offset by productivity gains, leading to the creation of other jobs, which will be tested for later on. The service sector appears to be neither affected by domestic nor foreign automation (Panel D). Hence, workers losing jobs in the manufacturing sector do not seem to shift to the service sector, as has been observed in other countries.[36] The Kleibergen-Paap F-Statistic (which is presented in the bottom of the table) is above the usual thresholds in all specifications.[37] Table 1.A.2 in the Appendix presents the first-stage results of estimation equation E.1. In Panel A the outcome domestic robot exposure and in Panel B foreign robot exposure. All variables are measured in stacked-differences, and all previous control variables are included. Panel A shows that the IV constructed in equation 1.8 is a strong predictor of regional exposure to domestic robots. A concern with regard to the identification strategy would be that regional exposure to domestic robots would fully explain foreign robot exposure or vice versa, which however does not seem to be the case. The first-stage results for foreign robot exposure are similar (see Panel B). The corresponding IV is a strong predictor, while domestic robot exposure and foreign exposure through imports are statistically insignificant.

A number of alternative specifications underpin the results. Table 1.A.3 in the Appendix repeats the analysis with an alternative specifications. To start with, in the first two columns the instrument for domestic robot exposure switches to the alternative instrument. As outlined in section 1.2, it is constructed by using the average number of robots in other emerging economies as an exogenous source for robot adoption (similar to the instruments used by Acemoglu and Restrepo (2020), Faber (2020) and Micco

[35] This is roughly half the magnitude which has been found by Faber (2020) for the case of Mexican labor markets and automation in the US. This reflects that the links between these two economies are stronger than in the case of Brazil.

[36] Dauth et al. (2017) for instance find that employment declines through automation were fully offset in business related services in Germany.

[37] Note that the thresholds for two instruments are lower, where a F-statistic above 7 indicates identification.

Table 1.2: IV estimations: Employment ratios, overall and by sector (1994-2004-2014)

	Domestic Robot Exposure IV		Foreign Robot Exposure IV		Dom. & For. Robot Exposure IV	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Changes in the total employment ratio						
Domestic Robot Exp.	0.23	0.58*	0.22	0.50	0.16	0.49
	(0.17)	(0.33)	(0.17)	(0.33)	(0.17)	(0.33)
Foreign Robot Exp.	-0.10	-0.41**	0.07	-0.18	0.08	-0.18
	(0.19)	(0.18)	(0.21)	(0.22)	(0.21)	(0.22)
Panel B: Changes in the manufacturing employment ratio						
Domestic Robot Exp.	0.07	0.16	0.17*	0.20*	0.11	0.17
	(0.09)	(0.10)	(0.10)	(0.11)	(0.10)	(0.10)
Foreign Robot Exp.	-0.11***	-0.14***	-0.23***	-0.17**	-0.22***	-0.16**
	(0.04)	(0.04)	(0.06)	(0.08)	(0.06)	(0.08)
Panel C: Changes in the mining employment ratio						
Domestic Robot Exp.	0.01	-0.00	-0.00	-0.01	0.00	-0.01
	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)
Foreign Robot Exp.	0.11***	0.05***	0.13**	0.07**	0.13**	0.07**
	(0.04)	(0.02)	(0.05)	(0.03)	(0.05)	(0.03)
Panel D: Changes in the service employment ratio						
Domestic Robot Exp.	0.05	0.10	0.01	0.08	0.00	0.07
	(0.09)	(0.17)	(0.10)	(0.19)	(0.09)	(0.18)
Foreign Robot Exp.	0.01	-0.12	0.13	-0.03	0.13	-0.02
	(0.07)	(0.10)	(0.09)	(0.14)	(0.09)	(0.14)
Observations	1114	1114	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓	✓	✓
Regional char.	✓	✓	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓	✓	✓
Regional import growth	✓	✓	✓	✓	✓	✓
Population weighted		✓		✓		✓
KP F-Statistic	1800	1281	46.4	54.5	24.6	28.6

Notes: Standard errors, in parentheses, are clustered on the meso-region level. Columns 1 and 2 instrument for domestic robot exposure, columns 3 and 4 for foreign robot exposure and columns 5 and 6 for both. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

(2019)). The results are almost identical and the coefficient of domestic robot exposure with the total and manufacturing employment ratios as outcomes turns statistically significant. In columns 3 and 4, long-differences between 1994 and 2014 replace the previous stacked-difference specification, with everything else being equal. The coefficients increase in size for foreign automation. For domestic automation none of the coefficients is statistically distinguishable from zero. In the last two columns, the dependent and independent variables are not measured in differences, but microregion-fixed effects are employed. Thus, all time-invariant characteristics at the microregion level are controlled for. The previous baseline control variables here have yearly variation. The results remain, with foreign automation leading to a decline in manufacturing, but an increase in mining employment.

The fact that domestic automation does not have a strong and robust effect on employment ratios could, in addition to the number of robots in Brazil not being large enough to see effects, potentially be driven by the displacement and the productivity effects canceling each other out, or only single industries being affected by domestic automation.[38] Table 1.A.4 in the Appendix displays the regression results without adjusting domestic robots with within-country input-output linkages, which does not substantially change the coefficients. The specification of domestic robot exposure with input-output linkages is likely to capture any productivity effects, as robot installments in one sector would increase the demand for inputs of other sectors. Within-sector productivity effects cannot be accounted for with the given data. That both measures yield similar results gives some indication that the two effects are not canceling each other out. Further, in Table 1.A.6 the effect of exposure to domestic and foreign automation on changes in regional GDP, as a proxy for regional productivity gains from automation, and population is estimated. Neither domestic nor foreign automation exposure appear to have an effect on changes in regional GDP (Panel A) nor on population (Panel B), mitigating concerns that the previously found effects could be driven by automation induced in- or out-migration. The fact that there is also no effect for GDP is further evidence against domestic automation largely increasing productivity in Brazil up to 2014.

Overall, the above evidence is in line with the main hypotheses of automation in advanced economies causing deindustrialization in Brazil. Exposure to foreign automation leads to a decrease in the manufacturing employment ratio.[39]

As a next step, Figure 1.2 disaggregates employment by industry. The light-blue colored bars display exposure to domestic robots and the dark-blue colored bars show exposure to foreign robots. The same specification as in column 1 of Table 1.2 is used. That is, variables are measured in stacked-differences, domestic robot exposure is instrumented for and all controls are included.[40] Domestic automation has a statistically

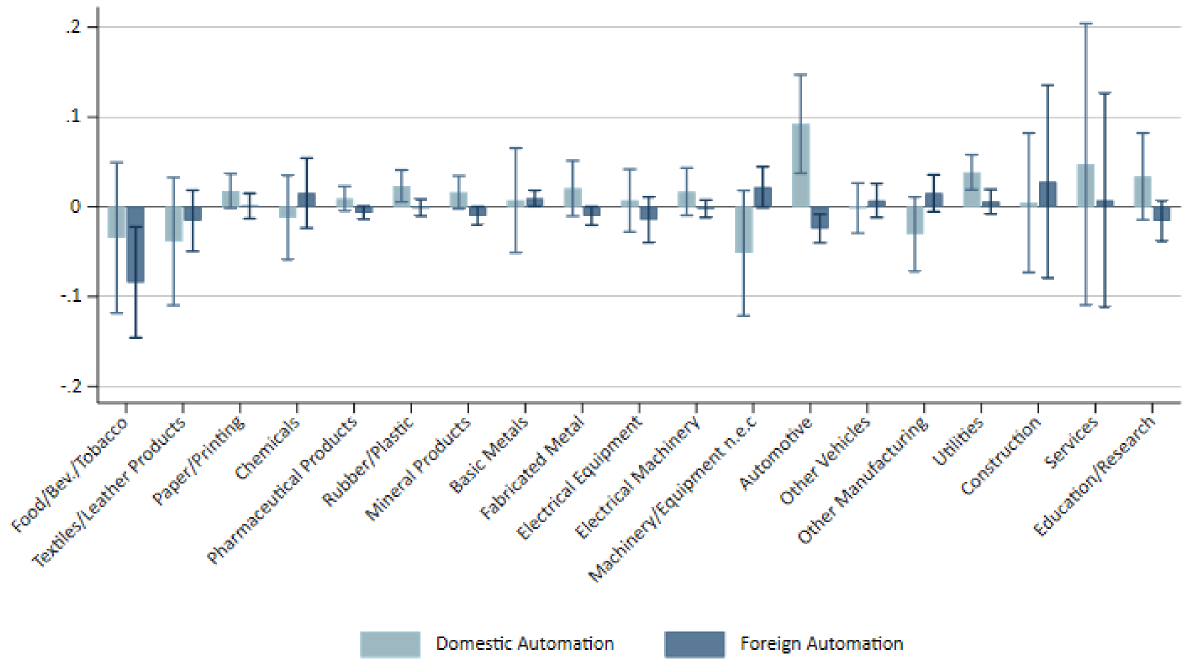
[38] Koch et al. (2019) and Aghion et al. (2020) for instance find productivity gains in automating firms, which lead to net-employment gains.

[39] Table 1.A.5 in the Appendix shows that the effects of automation are rather on the extensive than the intensive margin. If anything, domestic automation seems to slightly depress overall wages but increase manufacturing wages.

[40] Only the manufacturing industries are displayed for visibility and due to large confidence intervals in the service sectors. The only coefficient in service sectors which is statistically different from zero is that

significant and positive effect on employment in the automotive, the paper and printing, the pharmaceutical, the rubber and plastic and the minerals industries. A negative effect of foreign automation on employment is found mainly for the food, beverages and tobacco industry (which also is the largest exporting industry), but also for the pharmaceutical, the rubber and plastic, fabricated minerals and automotive industries. Foreign automation also has a labor-enhancing effect in two industries: The basic metals industry, which is complementary to the mining industry, and the machinery and equipment industry.

Figure 1.2: Robot Exposure and Industry Employment Ratio (1994-2004-2014)



Notes: The figure displays the coefficients of regressing industry level employment ratios on domestic robot exposure (light-blue) and foreign robot exposure (dark-blue). Standard errors, in parentheses, are clustered on the meso-region level. All specifications include state-year fixed effects and control for the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises, log microregion GDP, the share of manufacturing workers, the share of light-manufacturing workers and changes in overall and Chinese imports.

Table 1.A.7 in the Appendix shows that female and skilled workers are the ones that benefit from domestic automation, while there is no effect of foreign robot exposure on employment by gender or skill levels.[41] The installment of robots in Brazil thus appears to have a labor enhancing effect for workers with higher skill levels, which has also been found for industrialized countries (Aghion et al., 2020). However, the pattern

of domestic robot exposure in the utilities sector. Furthermore, statistically significant coefficients are found for agricultural employment.

[41] Skill-levels are defined according to the standard 1-digit ISCO classification: Groups 1 and 2 are skilled, 4-6 are medium skilled and 7-9 are low skilled.

of unskilled workers being substituted is not apparent here. Occupations which are female-dominated seem to be more complementary to automation technology than male-dominated occupations.

1.4.2 Exports

The main hypothesis of the paper is that foreign automation affects regional employment through changes in export demand. On the one hand, automation decreases the cost of production especially in advanced economies and thus decreases demand for certain manufactured goods from other countries, but at the same time increases the demand for raw materials and inputs needed for production.

So far the analysis has run under the assumption that regional exports flows adjust to foreign automation, which was captured by using initial export shares to weight foreign robots, leading to regional employment ratio changes. Figure 1.A.4 displays correlations between changes in regional export to GDP ratios and changes in exposures to foreign robots. Panel a) shows a raw negative correlation between exposure to robots and manufacturing exports and a positive correlation in mining exports in panel b). To test whether foreign automation in fact changes expenditures on goods from Brazil, the effects of exposure to foreign automation on changes in exports are examined in Table 1.3. Changes in regional exports are measured as stacked-differences in the inverse hyperbolic sine (asinh) of exports between 1997, 2006 and 2014.

Table 1.3: Changes in *asinh* regional exports (1997-2006-2004)

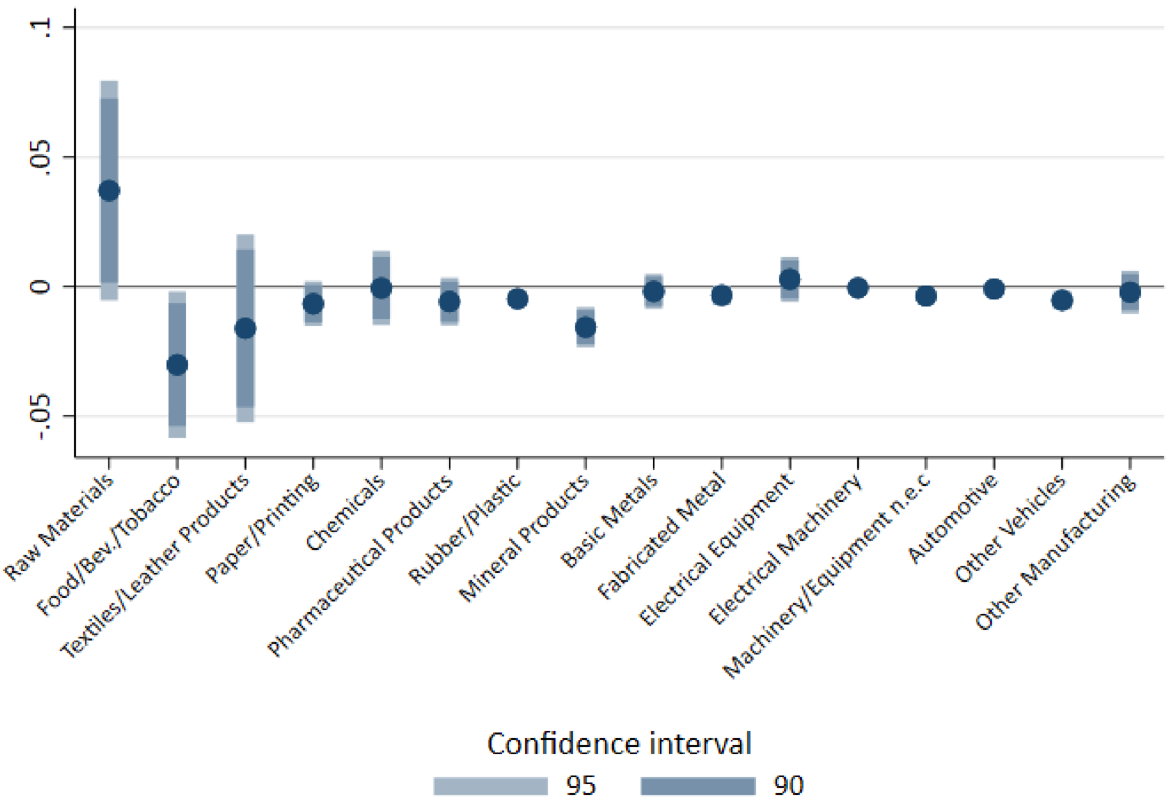
	Manufacturing		Mining	
	(1)	(2)	(3)	(4)
Domestic Robot Exp.	-0.04	-0.01	0.00	-0.01
	(0.03)	(0.02)	(0.05)	(0.06)
Foreign Robot Exp.	-0.04*	-0.04***	0.05	0.03
	(0.02)	(0.01)	(0.03)	(0.04)
Observations	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
Population weighted		✓		✓
KP F-Statistic	1863	1055	1863	1055

Notes: Standard errors, in parentheses, are clustered on the meso-region level. All specifications include state-year fixed effects. Controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises, log microregion GDP, the share of manufacturing workers and light-manufacturing workers. Outcomes are stacked-differences of log exports (1997-2006-2014). * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

The first two columns display changes in manufacturing exports. The coefficient of domestic automation exposure is again statistically indifferent from zero. This in line with the modest adjustments of employment to exposure to domestic robots found above. Foreign automation has a negative effect on changes of manufacturing exports, mirroring the decrease in manufacturing employment. For mining exports, shown in columns 3 and 4, the coefficient of foreign robot exposure is positive but not statistically significant.

In Figure 1.3 the effect of foreign robots on dyadic exports between microregions and more developed foreign countries are estimated directly and not over exposures. Exports of each sector in a micro-region to each foreign country are regressed on the foreign sector's stock of robots per worker, adjusted to account for input-output linkages. Section 1.A.3 in the Appendix explains the underlying estimation strategy in more detail. All possible time varying trends and time invariant characteristics are controlled for. Foreign automation leads to lower exports of most manufacturing industries, which largely correspond to those in which large declines in the employment ratios were found in Figure 1.2.

Figure 1.3: Dyadic exports



Notes: The figure displays the coefficients of regressing log exports of sector i in micro-region r to foreign country j on that country's adjusted sectoral stock of robots per worker (see section 1.A.3 in the Appendix for more details. Standard errors are clustered on the micro-region-country level. Each regression includes micro-region × sector × year fixed effects, micro-region × country × year fixed effects, and micro-region × country × sector fixed effects.

The only industry in which exports increase is the raw materials sector. The results are thus in line with the previously found changes in employment. Conversely, raw material imports are not affected by foreign automation, as is shown in Figure 1.A.5 in the Appendix. However, automation in partner-countries increases imports from several manufacturing sectors. This pattern reconciles with automation reducing production and thereby output costs (see equation 1.3), which increases exports of robot-adopting countries in the global economy.

An alternative explanation for the found patterns would be multinational companies directly substituting employment in manufacturing factories in Brazil with robots in their home economy. Evidence for such movements has been demonstrated by Faber (2020) and Artuc et al. (2019b) for the case of special manufacturing plants in Mexico (Maquiladoras). Column 3 of table 1.A.7 in the Appendix shows that exposure to neither domestic, nor to foreign automation has an effect on employment in foreign owned enterprises in Brazil. Therefore, decreases in manufacturing employment appear not to be driven by multinationals directly "reshoring" production activity back to their home economy, but rather by a diversion in demand for certain goods in the global economy.

1.4.3 Robustness Tests

Table 1.4 presents a number of robustness tests, which validate the previous findings of section 1.4.1. The table presents unweighted regression results, while Table 1.A.8 in the Appendix shows the same regressions with initial population weights. In the first column, instead of state-year fixed effects, meso-region-year fixed effects are included. Column 2 tests whether clustering standard errors on a different level, namely on the state instead of the meso-region level, affects the results (Borusyak et al., 2018). In column 3, the five regions most exposed to foreign automation are excluded, to test whether these are solely driving the results. While differences in imports are already controlled for, column 4 additionally includes foreign robot exposure through imports, to test whether a similar or opposing pattern as compared to exposure through exports is observable. In the next columns, additional sector controls are included: Initial employment shares in the automotive sector, as it has the largest numbers of robots both domestically and abroad, and the initial share of agricultural sector employment, since regions heavily specialized in agriculture are likely to evolve differently than others.[42] As regions more exposed to foreign automation are by construction those that were initially more engaged in trade, the last columns control for additional trade-openness characteristics. To dismiss the possibility that the results are driven by more open

[42] Bustos et al. (2020) for instance demonstrate that the rise in soy-bean production lead to specialization and capital outflows to more industrialized areas.

regions evolving differently because they are more integrated into global markets, initial exports (in logs) is added. Furthermore, the initial shares of exports of the food, beverages and tobacco and the agricultural sectors are controlled for, being the most important exporting industries in Brazil.

The main results remain statistically significant in almost all specifications. The only exception is the slightly smaller coefficient of foreign robot exposure with mining as an outcome, in the unweighted setting (compare Table 1.A.8 in the Appendix). In all other

Table 1.4: Robustness tests (unweighted) - Stacked differences (1994-2004-2014)

	Unweighted					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Changes in the manufacturing employment ratio						
Domestic Robot Exp.	0.17*	0.07	0.10	0.09	0.15	0.23
	(0.10)	(0.09)	(0.10)	(0.09)	(0.14)	(0.14)
Foreign Robot Exp. (through exports)	-0.07**	-0.11***	-0.11**	-0.11***	-0.12***	-0.11***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Foreign Robot Exp. (through imports)				-0.04	-0.04	-0.03
				(0.05)	(0.05)	(0.05)
Panel B: Changes in the mining employment ratio						
Domestic Robot Exp.	-0.01	0.01	0.03	0.01	0.04	0.04
	(0.04)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
Foreign Robot Exp. (through exports)	0.07	0.11***	0.08***	0.11***	0.10***	0.10**
	(0.05)	(0.03)	(0.02)	(0.04)	(0.04)	(0.04)
Foreign Robot Exp. (through imports)				-0.00	-0.00	-0.00
				(0.01)	(0.01)	(0.01)
Observations	1114	1114	1104	1114	1114	1114
Controls	✓	✓	✓	✓	✓	✓
Mesoregion Year FE	✓					
State Year FE		✓	✓	✓	✓	✓
State-level std. error clustering		✓				
Excluding top 5 exposed			✓			
Additional industry controls					✓	✓
Additional export controls						✓
KP F-Statistic	994	1431	1388	2718	572	562

Notes: Regressions are unweighted. Standard errors, in parentheses, are clustered on the meso-region level apart from column 2. In columns 1, meso-region-year fixed effects, while in all other columns state-year fixed effects are used. All specifications control for the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises, log microregion GDP, the share of manufacturing workers and light-manufacturing workers. In column 3 the five regions most exposed to foreign automation are excluded. Column 4 adds foreign robot exposure through imports. Additional industry controls are the initial shares of workers in the agricultural and automotive sectors. Additional export controls include the initial exports (in logs) and the initial shares of the agricultural and food, beverages and tobacco sectors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

estimations, foreign automation exposure remains statistically significant and positive for mining employment and negative for manufacturing employment. With state-level clustering, the standard errors slightly reduce in size, as compared to the baseline results. Excluding the top-5 exposed micro-regions slightly reduces the coefficient size in the weighted setting, however the results do not appear to be driven solely by the most exposed regions. Both additional industry and export controls do not change the results.

Since residuals might not be only correlated between regions of close vicinity but also between regions with a similar industry structure, in Table 1.A.9 shift-share adjusted standard errors are used, following Adão et al. (2019).[43] In the table, Eicker-Huber-White standard errors are reported in parantheses, and the shift-share adjusted (AKM) standard-errors are presented at the bottom of Panel A for foreign robot exposure, and at the bottom of Panel B for domestic robot exposure. For foreign robot exposure, the AKM standard errors are slightly smaller except for column 3, where the outcome is mining employment and the regressions are unweighted. Since the shift-share standard errors yield tighter confidence intervals as regional clustering (except for column 3), there seems to be little cross-sectoral correlation in the residuals driven by the shift-share structure.[44] The coefficient is still statistically significant though. For the domestic robot exposure, AKM standard-errors are substantially larger for manufacturing employment and slightly smaller for mining employment.

An alternative way to establish causality with the shift-share estimator is exogeneity of initial industry shares. The identification strategy so far has exploited exogenous variation in shocks through instruments. The initial share assumption is tested following Goldsmith-Pinkham et al. (2020) to ensure validity of the shift-share estimator. The authors demonstrate that a sufficient condition for identification in a shift-share estimation is that initial local industry shares are exogenous to employment changes (conditional on the included controls).[45] First, the existence of pre-trends is tested in the first two columns of Table 1.5, by estimating whether micro-regions that were more exposed to domestic or foreign automation already had similar employment trends before automation took off. Should similar employment trends be found, it is likely that these are driven by other trends than exposure to domestic or foreign automation or by the region's underlying industry structure (Goldsmith-Pinkham et al., 2020).[46] All included variables are constructed as in the long-difference setting of Table 1.A.3. Robot exposures are measured between 1994 and 2014, and baseline control variables

[43] I am grateful for helpful comments and the authors' estimation code.

[44] This may be due to a small sample bias of having relatively few industries. However, over-rejection is more severe in the case of a small number of sectors and clustering on the regional level.

[45] With a large number of regions and a small number of industries, exogeneity in initial shares would suffice to establish exogeneity in the shift-share estimator (Borusyak et al., 2018).

[46] Pre-trends are analyzed between 1985 and 1994, where the usage of robots was not yet spread. Brazil for instance had the first recorded robots in 1999.

in 1994. The results do not suggest the presence of confounding pre-trends. The only existing pre-trend is a decline in manufacturing employment in regions which were more exposed to domestic automation between 1994-2014. For the main period of analysis however, if at all, a positive effect from domestic automation for manufacturing workers was found. In columns 3 and 4, the specification switches back to the stacked-difference setting between 1994, 2004 and 2014, but the pre-trend period change in the respective outcome variable is controlled for. The coefficients of foreign robot exposure are almost identical to the results in the main specification of Table 1.2. The coefficients of domestic robot exposure increase in size for manufacturing employment.

Table 1.5: Pre- and long-term effects

	Pre-trend (1985-1994)		Stacked-differences (1994-2004-2014)	
	(1)	(2)	(3)	(4)
Panel A: Changes in the manufacturing employment ratio				
Domestic Robot Exp.	-0.89***	-0.84***	0.17*	0.27**
	(0.29)	(0.25)	(0.10)	(0.12)
Foreign Robot Exp.	0.06	-0.09	-0.12***	-0.13***
	(0.20)	(0.19)	(0.04)	(0.05)
Panel B: Changes in the mining employment ratio				
Domestic Robot Exp.	-0.05	-0.00	0.00	-0.00
	(0.05)	(0.03)	(0.03)	(0.01)
Foreign Robot Exp.	-0.01	-0.04	0.11***	0.05**
	(0.03)	(0.04)	(0.04)	(0.02)
Observations	557	557	1114	1114
Regional controls	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓
Import growth	✓	✓	✓	✓
State FE	✓	✓		
State Year FE			✓	✓
Initial population weighted		✓		✓
Pre-Trend Controls			✓	✓
KP F-Statistic	1372	1173	1853	1297

Notes: Standard errors, in parentheses, are clustered on the meso-region level. In columns 1-2 the outcome is the difference of manufacturing or mining employment in the pre-period. The explanatory variables are measured as long-differences between 1994 and 2014. Columns 3 and 4 switch to the stacked-difference setting as in Table 1.2 and include the respective pre-trend period outcome as a control. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

As a next step, Rotemberg weights are calculated to test whether industries with the largest weights are driving the results (Goldsmith-Pinkham et al., 2020). Table 1.A.10 displays the weights of the eight industries with the largest exposures. There are two clear outlier sectors. For domestic automation, this is, unsurprisingly, the automotive sector, which has by far the most robots. For foreign automation, the sector with the largest weight is the basic metals industry. Though it is not among the industries with the most robots globally, it is a supplier for other automating industries and important for the mining industry in Brazil. In order to see whether one of these sectors is driving the observed changes in employment, exposures to domestic and foreign automation are calculated without the two industries in Table 1.6. In columns 1-3, the automotive industry is excluded. The employment differences both in the manufacturing sector, as well as the raw materials sector are still statistically significant. The F-Statistic decreases, remains above the usual thresholds though. Without the basic metals sector the standard errors increase and the coefficients reduce slightly in size with raw material employment as an outcome. This indicates the importance of the basic metals sector for the downstream supply chain of the mining sector. The results however appear not to be driven solely by outlier industries.

Table 1.6: Excluding highest Rotemberg-weight sectors

	Excl. automotive sector		Excl. basic metals sector	
	(1)	(2)	(3)	(4)
Panel A: Changes in the manufacturing employment ratio				
Domestic Robot Exp.	0.17	0.29**	0.23**	0.25**
	(0.14)	(0.12)	(0.09)	(0.10)
Foreign Robot Exp.	-0.13***	-0.10**	-0.13*	-0.17***
	(0.04)	(0.05)	(0.07)	(0.05)
Panel B: Changes in the mining employment ratio				
Domestic Robot Exp.	0.04	0.00	-0.03	0.00
	(0.04)	(0.02)	(0.03)	(0.01)
Foreign Robot Exp.	0.10**	0.08***	0.08*	0.04**
	(0.04)	(0.03)	(0.05)	(0.02)
Observations	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓
Regional controls	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓
Import growth	✓	✓	✓	✓
Initial population weighted		✓		✓
KP F-Statistic	361	527	935	1135

Notes: Standard errors, in parentheses, are clustered on the meso-region level. In columns 1-3 exposures are calculated without the automotive sector, and in columns 4-6 without the basic metals sector. All specifications include state-year fixed effects. Regional controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. Exposures to the excluded sectors are included as separate controls. Variables are measured in stacked-differences between 1994-2004-2014. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Lastly, following Borusyak et al. (2018), the region-level dataset is reweighted to the industry level, controlling for regional characteristics, in Table 1.A.12. Due to the small number of sectors in the robot data, the sample size drops extensively. The IV results correspond to the basic shift-share estimations, with a statistically significant and negative coefficient of foreign robot exposure for the manufacturing employment outcome, and a positive coefficient for the mining employment outcome. The domestic robot exposure variable is only statistically significant without including broad sector trends. Even though it is difficult to argue that sectoral shocks are independent from each other in the robot data (Caselli et al., 2019), it is reassuring to find significant results on the sectoral level, when controlling for broader sector trends.

1.5 Conclusion

This paper sheds light on the question whether automation-technology advancements lead to deindustrialization in a resource-rich emerging economy. Building on a Ricardian model of trade with a two-stage production technology, a novel approach is used to capture the effects of foreign automation has on LLMs in Brazil, which accounts for inter-sectoral input-output linkages. Automation in export destination countries is found to decrease employment in the manufacturing sector but to increase employment in the raw materials sector in Brazil. These shifts are mirrored by changes in demand for export goods from Brazil. Foreign automation leads to smaller exports of manufacturing goods and rising exports of raw materials. Employment is less affected by exposure to domestic automation. Higher skilled and female workers as well as workers in the automotive sector benefit most in terms of employment from domestic automation. A number of robustness tests strengthen the results. These include testing for standard errors not being correctly specified in the shift-share setting (Adão et al., 2019), validity of the shift-share estimator (Goldsmith-Pinkham et al., 2020; Borusyak et al., 2018), pre-trends, and excluding outliers.

The findings may cause further concerns of technological change in developed economies causing premature deindustrialization in emerging economies (Rodrik, 2016, 2018). Emerging economies may not only be affected by developed countries reversing their offshoring activities, but also by shifting demand-patterns in international trade. Thereby, as was shown for the case of Brazil, especially resource-rich countries face incentives to shift employment to raw material extraction, leaving whole economies more vulnerable to external price shocks and less active in higher value-added production.

Traditional strategies to foster manufacturing sector growth in developing and emerging economies may therefore no longer be viable. This calls for more research on how technological change in advanced economies affects industry compositions and labor demand in developing and emerging economies. For instance, investigating linkages between automating firms and input-providing firms in global supply chains could be a promising avenue for future research. New strategies, which enable resource-rich developing countries to cope with or even to benefit from global technological changes, are needed in order to prevent long-term setbacks in developmental trajectories.

1.A Appendix

1.A.1 Acknowledgements

I am grateful to Holger Strulik, Krisztina Kis-Katos, Ana Abeliansky, Lukas Wellner, Richard Haarburger, Laura Barros, the participants of the Globalization and Development (GlaD) seminar, and participants of the SMYE 2021, GlaD 2021 and SBE 2019 conferences for valuable feedback and comments and Kathrin Rupieper for excellent research assistance. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project RTG 1723 in the framework of the research training group on "Globalization and Development".

1.A.2 Theory Appendix

From the production function in equation 1.2 follows the cost function of producing ω_i

$$c_{ri}(\omega_i) = \psi_{ri} f_{ri}^{\alpha^F_{ri}} \prod_k^K p^1_{rik}{}^{\alpha^M_{rik}} w_{ri}{}^{\alpha^T_{ri}}, \tag{1.A.1}$$

with f_{ri} as the rental rate for capital, P^1_{rik} the price of intermediate goods and w_{ri} the task input price. Producers of the composite good of stage s supply Q^s_{ri} by minimizing the costs of the intermediate variety ω_i from international suppliers. The production technology is an aggregator as in Caliendo and Parro (2015), given by

$$Q^s_{ri} = \left[\int (m_{ri}(\omega_i))^{\sigma_i - \frac{1}{\sigma_i}} \right]^{\frac{\sigma_i}{\sigma_i - 1}}, \tag{1.A.2}$$

where σ_i is the elasticity of substitution across intermediate goods. Region r and industry i's demand of intermediate good ω_i from the lowest cost supplier across all countries is denoted as $m_{ri}(\omega_i)$.[47] Intermediate inputs and final goods can be sourced internationally. The price of an intermediate or final good for region r and sector i is

$$p_{ri}(\omega_i) = \min_j \left[\frac{c_{ji}\tau_{rji}}{z_{ji}(\omega_i)} \right], \tag{1.A.3}$$

where c_{ji} is defined as in equation 1.A.1. The price of the composite good can be expressed with the properties of the *Fréchet* distribution (Eaton and Kortum, 2002). Due to the probabilistic distribution of technology, countries and sectors have different levels of productivity. As in Caliendo and Parro (2015), λ_{ri} denotes the location parameter varying by country and sector, representing absolute advantage of a region in industry i. The shape parameter θ_i is the industry specific variation of efficiency in production, thereby representing comparative advantage. The price of the composite good is then

$$P_{ri} = A_i \left[\sum_j \lambda_{ji} (c_{ji}\tau_{rji})^{-\theta_i} \right]^{-\frac{1}{\theta_i}}, \tag{1.A.4}$$

where A_i is a constant. The efficiency of production of each country/region in industry i is the realization of the random variable z_{ri} from the probability distribution $F_{ri} = Pr\left[Z_i \leq z\right]$, which with the properties from the Fréchet distribution becomes

$$F_{ri} = e^{-T_{rji}^{-\theta_i}}, \tag{1.A.5}$$

[47]Demand for the intermediate is given by $m^s_{ri}(\omega_i) = \left(\frac{p_{ri}(\omega_i)}{P_{ri}}\right)^{-\sigma_i} Q^s_{ri}$, where P_{ri} is the unit price of the composite good.

with $T_{rji} = \lambda_{ri}(c_{ri}\tau_{rji})^{-\theta_i}$. Caliendo et al. (2015) show that the lowest price of a good Ω_i in region r also has a Fréchet distribution

$$Pr\left[p_{ri} \leq p\right] = 1 - e^{-\psi_{ri} p^{-\theta_i}}, \tag{1.A.6}$$

where $\psi_{ri} = \lambda_{ri}(c_{ri}\tau_{rji})^{-\theta_i}$, which is a statistic of technologies, input costs, geographic barriers and tariff policies (Eaton and Kortum, 2002; Caliendo and Parro, 2015). The authors further show that due to the properties of probability density functions the price index is

$$P_{ri} = A_i \psi_{ri}^{\frac{-1}{\theta_i}}$$

, which is equivalent to equation 1.A.4. The expenditure of country j on goods from region r and industry i can be denoted in probability terms as

$$E_{jri} = Pr\left[\frac{c_{ri}\tau_{rji}}{z_{ri}(\Omega_i)} \leq \min_{h \neq r} \frac{c_{hi}\tau_{hji}}{z_{hi}(\Omega_i)}\right] E_{ji}, \tag{1.A.7}$$

where overall expenditure of country j industry i E_{ji} is multiplied by the probability of region r having the lowest price. To derive equation 1.3, equations 1.A.5 and 1.A.6 are plugged into 1.A.7. Tasks are allocated to labor and robots according to

$$T_{ri}(b) = \begin{cases} \gamma_L(b)L_{ri}(b) + \gamma_R(b)R_{ri} & \text{if b} \leq C_i \\ \gamma_L(b)L_{ri}(b) & \text{if b} > C_i. \end{cases} \tag{1.A.8}$$

Following Acemoglu and Restrepo (2018), a region-specific threshold C'_r is defined, below which producers in region r use robots instead of human capital. The threshold C'_{ri} is given by $w^L_{ri}/\gamma_L(C'_{ri}) = w^R/\gamma_R(C'_{ri})$. The threshold implies a reduction in task related costs, according to

$$\Omega_{ri} = \frac{(1 - C'_{ri})w^L_{ri} + C'_{ri}w^R}{w_{ri}}. \tag{1.A.9}$$

The share of labor in production tasks then is

$$s^L_{ri} = \frac{(1 - C'_{ri})\frac{w^L_{ri}}{\gamma_L}}{C'_{ri}\frac{w^R}{\gamma_R} + (1 - C'_{ri})\frac{w^L_{ri}}{\gamma_L}}, \tag{1.A.10}$$

which is plugged into equation 1.A.1 to obtain equation 1.4. Equation 1.5 can be rewritten as

$$w_{ri}^L L_{ri} = \alpha_{ri}^T (1 - C'_{ri}) \sum_j \pi_{jri} E_{ji},$$

since expenditure on goods of region r and industry i is $Y_{ri} = \sum_j \frac{\pi_{jri} E_{ji}}{1+\tau_{jri}}$. Using the share of labor in production tasks (see equation 1.A.10), expenditure shares from equation 1.3 and cost funtion 1.4, equation 1.5 can be rewritten as:

$$w_{ri}^L L_{ri} = \frac{\alpha_{ri}^T (1 - C'_{ri.}) \frac{w_{ri}^L}{\gamma_L} P_{ri} \sum_j \pi_{jri} E_{ji} \frac{1}{1+\tau_{jri}}}{\left(\lambda^{1/\theta_{ri}^i} P_{ri} \tau_{rji}^{-1} \pi_{jri}^{-1/\theta_i}\right)^{1/\alpha_{ri}^T} f^{-\alpha_{ri}^F/\alpha_{ri}^T} \prod_k p_{rik}^{1} {}^{-\alpha_{rik}^M/\alpha_{ri}^T}}. \tag{1.A.11}$$

Taking logs and differentiating finally yields equation 1.6.

1.A.3 Additional empirical strategies

Input-output adjusted robots Intranational production and international trade are composed of intra-sectoral, but also inter-sectoral flows of goods. To capture the fact that exports from Brazil can be used for production in different industries abroad (see equation 1.2), I construct an adjusted robots per worker $AR_{ji,t}$, which accounts for input-output linkages of sectors i in Brazil with foreign economies j. Similarly, for domestic automation exposure, the measure captures how production in one sector is composed of inputs from other sectors:

$$AR_{ji,t} = \sum_k \frac{X_{ijk,2000}}{X_{jk,2000}} \frac{R_{jk,t}}{L_{jk,2000}}. \tag{1.A.12}$$

The term $X_{ijk,2000}/X_{jk,2000}$ captures sectoral linkages between Brazil and the foreign country. It represents the share of exports from industry i in Brazil to industry k in country j from overall Brazilian exports to country j.[48] For domestic adjusted robots, country j simply represents Brazil. The measure ensures that exports from sector i of micro-region r are matched correctly to robots of country j. For instance, exports from the raw materials sector (e.g. sector i) are likely not only affected by robots in foreign the raw materials sector, but also sectors which use such materials for inputs as the automotive sector (e.g. sector k). $R_{jk,t}$ is the stock of robots in country j, sector k and year t, which is divided by $L_{jk,2000}$, employment in 1000s in that sector.

Foreign robot exposure IV To mitigate endogeneity concerns of foreign robot adoption, the following instrument is used:

[48] Data of regional exports from Brazil does not contain information about the destination industry, which is why regional exports are weighted by sectoral linkage data on the national level.

$$dR_{ji}^{IV_{for}} = dR\, w_{ji,2000}\, o_{ji,2000}, \tag{1.A.13}$$

where dR is the change in global robots (as in the IV for domestic robot exposure), w_{ji} is the wage bill and o_{ji} output per worker of industry i in country j.[49] The underlying idea is that an industry with a larger wage bill will have a larger incentive to install robots, in order to cut production costs. Industries with larger output in turn are more likely to have the capacity to purchase robots. Accordingly, the instrument for foreign automation is generated by replacing dAR_{ji} in equation 1.9 with $dR_{ji}^{IV_{for}}$. Both IVs are based on of the fact that the global decline in robot prices is exogenous to single countries and industries, as well as to cross-sectional predictors that drive automation.

Micro-region exports This section describes the estimation of dyadic exports between micro-regions in Brazil and foreign economies on their stock of robots in Section 1.4.2. The estimation equation underlying Figure 1.3 is

$$X_{rji,t} = \beta_0 + \beta_1 AR_{ji,t} \times \iota_i + \xi_{ri,t} + \psi_{rj,t} + \gamma_{rji} + \epsilon_{rji,t},$$

where $X_{rji,t}$ are exports from micro-region r and sector i to country j in year t measured in logs and $AR_{ji,t}$ is the adjusted stock of robots in country j and sector i, as defined in equation 1.A.12. The adjusted robot term is interacted with a dummy for sector i to obtain sector specific coefficients, which are displayed in Figure 1.3. To control for all possible time varying trends and time invariant characteristics, and following the structural gravity literature, the regression controls for all possible fixed effects combinations (Yotov et al., 2016b).

1.A.4 Data Appendix

Routine Task Intensity A measure of routine task intensity of each micro-region is included to control for other technological progresses affecting employment of certain workers, most importantly computerization Frey and Osborne (2017). Data of routine task intensity comes from the O*NET 2000 release, which associates the importance of certain tasks to 800 occupations in the United States. Following Almeida et al. (2017), occupations from the US are matched with the Brazilian occupation system CBO[50].

[49] As for the adjusted foreign robot measure, wages and output in the foreign sector are transformed between foreign sector k and sector i: $w_{ji} = \sum_k \frac{X_{ijk,2000}}{X_{jk,2000}} w_{jk}$ and $o_{ji} = \sum_k \frac{X_{ijk,2000}}{X_{jk,2000}} o_{jk}$. As detailed task descriptions about workers' occupations are not available for each trade-partner country, not the same IV for domestic and foreign robots could be used.

[50] I use the crosswalks developed by Hardy et al. (2018) and Muendler et al. (2004) to match O*NET SOC occupations to CBO Brazil occupations.

Robot Data Adjustments Two measures for robots are provided in the data, annual deliveries (sales) and a yearly robot stock. The IFR acknowledges annual shipments to be more accurate. I therefore follow Graetz and Michaels (2018) and construct a robot stock based on annual deliveries and a depreciation rate of ten percent. As the number of unclassified robots is high for certain countries and years, these are proportionally distributed to those that are classified into industries, following Acemoglu and Restrepo (2020).[51]

Additional Data Sources Micro-region population and GDP are obtained from the Brazilian Institute of Geography and Statistics (IBGE). Regional GDP is available from 1997 and population from 1994. Value added data for the mining and industry sectors, used to construct the instrumental variable, also come from the IBGE. Value added in other sectors, namely agriculture and service sectors are are taken from The World Bank (2020). Exports and imports are deflated using the Personal Consumption Expenditure deflator. To harmonize wage bills and output data across countries, exchange rates from The World Bank (2020) are used.

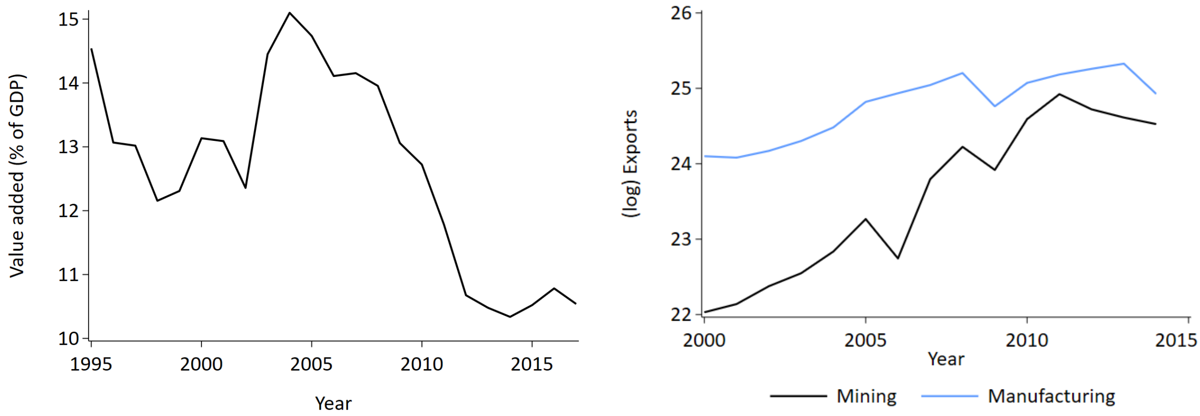
[51]For countries, in which robots are unclassified before a certain year, robots are distributed according to the sectoral share of robotsin the first year in which not all robots are unclassified. Also, before 2011, robots for Mexico, Canada and the United States are reported collectively as North America. To obtain country and industry-specific data for the whole period, I follow Acemoglu and Restrepo (2020) and construct the country-specific industry shares before 2011 based on the yearly sectoral share of each country relative to North America as a whole.

1.A.5 Additional Figures

Figure 1.A.1: Manufacturing and mining sector trends in Brazil

(a) Share of manufacturing in value added

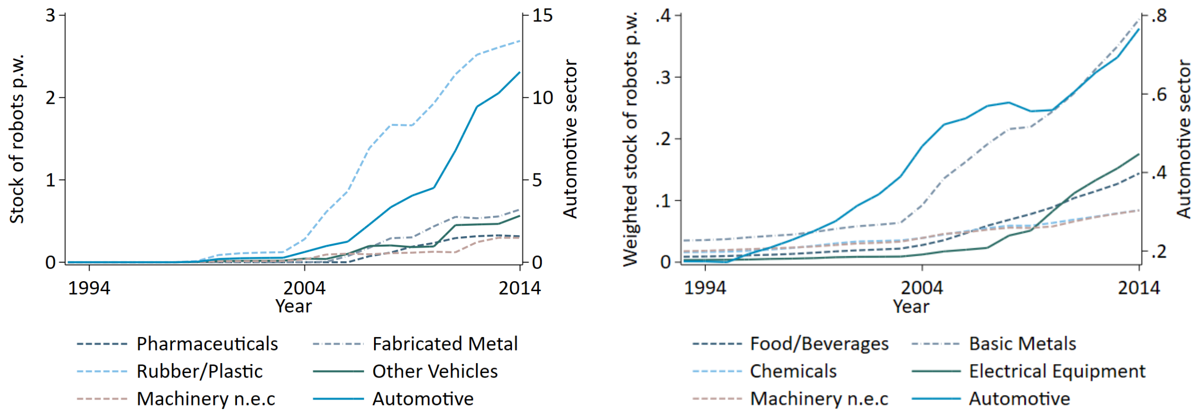
(b) Sectoral exports



Notes: The figure in Panel A displays the share of manufacturing in value added in Brazil between 2000 and 2017. Data comes from World Bank National Accounts data and OECD National Accounts data files. The figure in Panel B displays the log exports of the Manufacturing sector and the Mining sector of Brazil between 2000 and 2014. Data comes from the Comtrade Database.

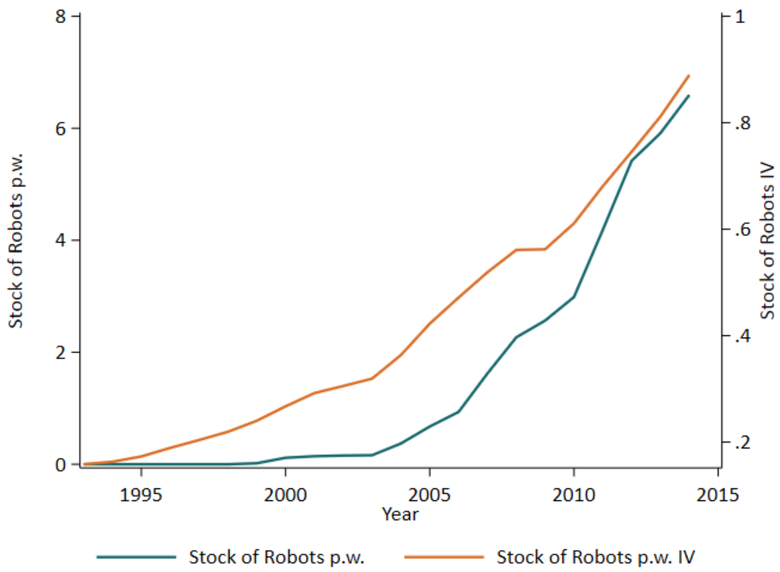
Figure 1.A.2: Stock of Robots in and abroad by Industry

(a) Stock of robots in Brazil (b) Weighted foreign robots



Notes: The figure in Panel A displays the stock of robots in Brazil between 2000 and 2014 in the 6 industries with the most robots. In Panel B, the stock of robots in a partner country's industry is weighted by the share of each sector's exports from Brazil to this industry. The right axis refers to the stock of robots of the Automotive industry in both panels.

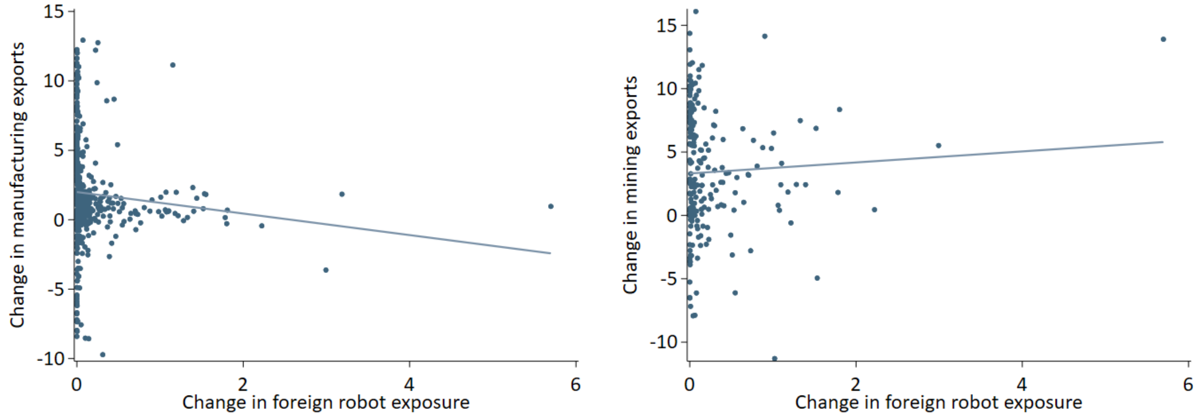
Figure 1.A.3: Aggregate stock of robots per worker in Brazil and its instrumental variable



Notes: The figure displays total robots per worker in Brazil and the instrument for domestic robots as outlined in equation 1.8.

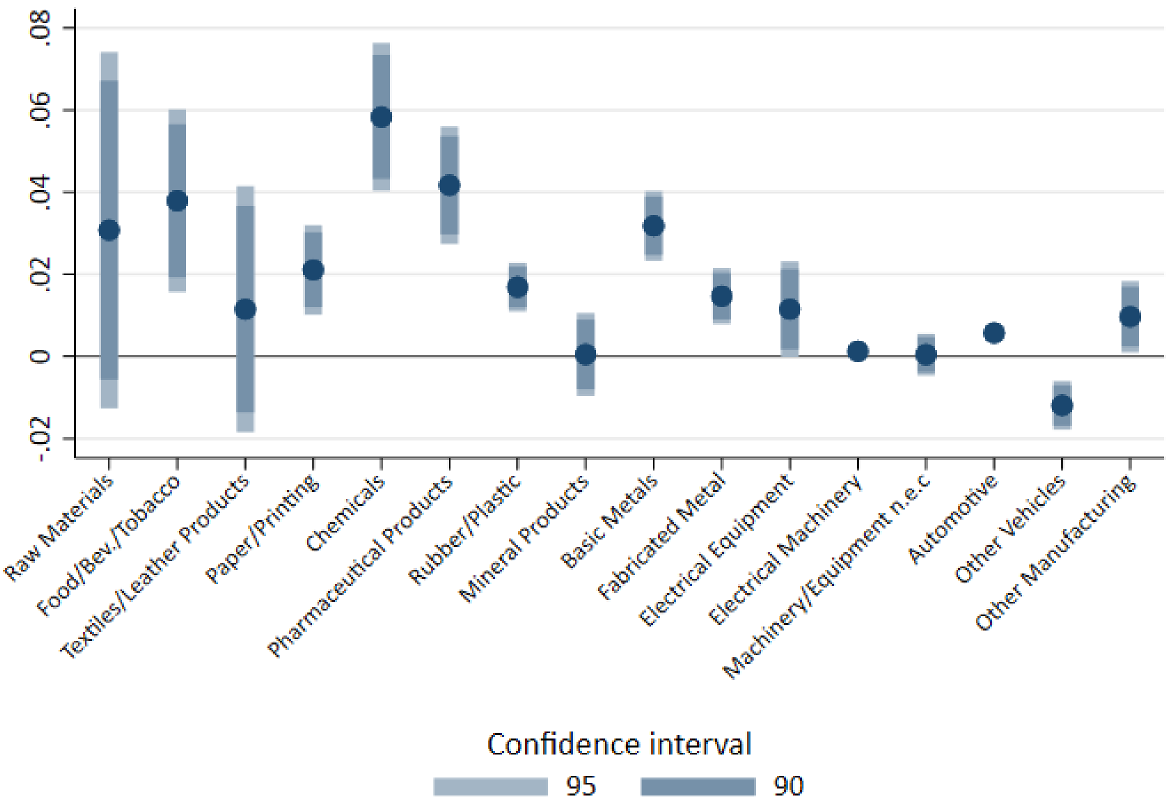
Figure 1.A.4: Exposure to foreign robots and micro-region export ratios

(a) Changes in regional manufacturing exports (log) (b) Changes in regional mining exports (log)



Notes: The figure in Panel A displays changes of micro-region manufacturing exports (in logs) and changes in foreign robot exposure between 1997 and 2014. The figure in Panel B displays changes of micro-region mining exports (in logs) and changes in foreign robot exposure between 1997 and 2014.

Figure 1.A.5: Dyadic imports



Notes: The figure displays the coefficients of regressing log imports of sector i in micro-region r from foreign country j on that country's adjusted sectoral stock of robots per worker (see section 1.A.3 in the Appendix for more details. Standard errors are clustered on the micro-region-country level. Each regression includes micro-region $\times$ sector $\times$ year fixed effects, micro-region $\times$ country $\times$ year fixed effects, and country $\times$ sector $\times$ year fixed effects.

1.A.6 Additional Tables

Table 1.A.1: Descriptive statistics - Differences between 1994 and 2014 or initial values

		Exposure to Domestic Robots		Exposure to Foreign Robots	
Variables	All Regions	<Median	>Median	<Median	>Median
	Panel A: Outcomes				
Tot. Employment Ratio	13.85	10.02	17.68	10.90	16.80
	[9.53]	[8.98]	[8.47]	[9.27]	[8.86]
Manuf. Employment Ratio	2.42	1.75	3.09	2.19	2.64
	[3.44]	[2.74]	[3.91]	[3.21]	[3.65]
Mining Employment Ratio	0.11	0.07	0.14	0.05	0.17
	[0.88]	[0.97]	[0.79]	[0.83]	[0.93]
Log Export	0.08	2.28	1.66	2.57	1.37
	[0.84]	[4.43]	[2.61]	[4.46]	[2.45]
Log Manuf. Export	0.06	1.51	1.59	1.83	1.26
	[0.79]	[4.17]	[2.84]	[4.15]	[2.84]
Log Mining Export	0.05	0.92	1.62	0.89	1.65
	[0.73]	[3.13]	[3.53]	[3.04]	[3.61]
	Panel B: Main Covariates				
Domestic Robot Exp.	0.02	0.01	0.03	0.01	0.03
	[0.03]	[0.01]	[0.03]	[0.01]	[0.03]
Foreign Robot Exp.	0.15	0.11	0.19	0.00	0.30
	[0.42]	[0.38]	[0.46]	[0.00]	[0.56]
HS Graduate Rate	0.23	0.25	0.21	0.25	0.20
	[0.09]	[0.10]	[0.08]	[0.10]	[0.07]
Higher Education Rate	0.04	0.04	0.04	0.04	0.04
	[0.02]	[0.03]	[0.02]	[0.03]	[0.02]
(log) Population	11.94	11.60	12.27	11.57	12.30
	[0.92]	[0.82]	[0.90]	[0.79]	[0.91]
(log) GDP	12.89	12.26	13.53	12.23	13.56
	[1.30]	[1.06]	[1.20]	[0.99]	[1.23]
Share of For. owned Firms (*100)	0.04	0.02	0.05	0.02	0.06
	[0.28]	[0.10]	[0.39]	[0.08]	[0.39]
Share of Female Workers	0.28	0.28	0.28	0.29	0.27
	[0.10]	[0.12]	[0.07]	[0.12]	[0.07]
Avg. Routine Task Intensity	3.16	3.14	3.18	3.14	3.18
	[0.10]	[0.11]	[0.08]	[0.10]	[0.09]
Manuf. Empl. Share	0.21	0.12	0.30	0.16	0.27
	[0.17]	[0.12]	[0.16]	[0.15]	[0.17]
Light Manuf. Empl. Share	0.06	0.04	0.07	0.04	0.07
	[0.09]	[0.09]	[0.09]	[0.08]	[0.09]
Log Imports	0.04	0.86	1.08	0.81	1.13
	[0.42]	[2.05]	[1.60]	[2.08]	[1.56]
Log Chinese Imports	0.17	3.01	5.02	3.09	4.94
	[1.08]	[3.64]	[3.17]	[3.72]	[3.13]

Notes: This table provides descriptive statistics, where regions are split into above and below median exposure to domestic and foreign automation. Employment ratios, exports and imports, and the exposure to robots are given as differences between 1994 and 2014, log micro-region GDP is from 1997 and otherwise values are from 1994. Sample means and standard deviation (in brackets) are displayed. Light Manufacturing is composed of workers in the Textile and the Paper & Printing industries.

Table 1.A.2: First stage - Stacked differences (1994-2004-2014)

	(1)	(2)	(3)	(4)	(5)
Panel A: Changes in domestic robot exposure					
Domestic Robot Exp. IV	0.90***	0.90***	0.91***	0.93***	0.94***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Foreign Robot Exp. (through exports)		0.00	0.01	0.02	0.01
		(0.01)	(0.01)	(0.01)	(0.03)
Foreign Robot Exp. IV			-0.01	-0.02	0.01
			(0.01)	(0.01)	(0.02)
Foreign Robot Exp. (through imports)				-0.08***	-0.07***
				(0.02)	(0.02)
Panel B: Changes in foreign robot exposure					
Foreign Robot Exp. IV	0.74***	0.72***	0.72***	0.72***	0.64***
	(0.11)	(0.11)	(0.11)	(0.11)	(0.09)
Domestic Robot Exp.		0.08	0.05	0.10	0.08
		(0.06)	(0.06)	(0.07)	(0.16)
Domestic Robot Exp. IV			0.03	-0.03	-0.00
			(0.07)	(0.09)	(0.16)
Foreign Robot Exp. (through imports)				0.05	0.06
				(0.04)	(0.04)
Observations	1114	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓	✓
Regional char.	✓	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓	✓
Regional import growth	✓	✓	✓	✓	✓
Population weighted					✓

Notes: Standard errors, in parentheses, are clustered on the meso-region level. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.3: IV estimations: alternative specifications

	Alternative Domestic Robot Exposure IV		Long Difference (1994-2014)		Microregion FEs (1994-2004-2014)	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Total employment						
Domestic Robot Exp.	0.38**	0.63*	-0.33	-0.20	0.21	0.41**
	(0.17)	(0.32)	(0.36)	(0.32)	(0.16)	(0.16)
Foreign Robot Exp.	-0.13	-0.42**	-0.37	-0.88**	0.21	0.01
	(0.19)	(0.18)	(0.42)	(0.42)	(0.22)	(0.21)
Panel B: Manufacturing employment						
Domestic Robot Exp.	0.19**	0.24**	-0.22	-0.30	0.02	-0.05
	(0.09)	(0.11)	(0.21)	(0.20)	(0.08)	(0.10)
Foreign Robot Exp.	-0.13***	-0.16***	-0.33***	-0.31**	-0.13***	-0.16***
	(0.04)	(0.04)	(0.12)	(0.13)	(0.05)	(0.06)
Panel C: Mining employment						
Domestic Robot Exp.	0.00	-0.01	-0.01	-0.02	-0.01	-0.01
	(0.02)	(0.01)	(0.05)	(0.02)	(0.02)	(0.01)
Foreign Robot Exp.	0.11***	0.05***	0.20**	0.13**	0.10**	0.06**
	(0.04)	(0.02)	(0.09)	(0.05)	(0.04)	(0.02)
Observations	1114	1114	557	557	1671	1671
State Year FE	✓	✓	✓	✓	✓	✓
Regional char.	✓	✓	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓	✓	✓
Regional import growth	✓	✓	✓	✓	✓	✓
Population weighted		✓		✓		✓
Microregion FE					✓	✓
KP F-Statistic	1983	1592	1438	1208	1444	899

Notes: Standard errors, in parentheses, are clustered on the meso-region level. In columns 1 and 2, the alternative instrument for domestic robot exposure is used, while in columns 3-6 the main instrument is used. In columns 1 and 2, the specification follows column 2 of Table 1.2. In columns 3 and 4 the outcome and independent variables are defined in long-difference (1994-2014), while in columns 5 and 6 the whole panel is used. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 1.A.4: Domestic robots without I-O adjustments (1994-2004-2014)

	OLS		IV	
	(1)	(2)	(3)	(4)
Panel A: Manufacturing employment ratio				
Domestic Robot Exp.	0.14**	0.15	0.14*	0.18*
	(0.07)	(0.09)	(0.08)	(0.11)
Foreign Robot Exp.	-0.11***	-0.13***	-0.11***	-0.14***
	(0.04)	(0.04)	(0.04)	(0.04)
Panel B: Mining employment ratio				
Domestic Robot Exp.	-0.03	-0.01	-0.03	-0.01
	(0.02)	(0.01)	(0.02)	(0.01)
Foreign Robot Exp.	0.11***	0.05***	0.11***	0.05***
	(0.04)	(0.02)	(0.04)	(0.02)
Observations	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓
Regional char.	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓
Regional import growth	✓	✓	✓	✓
Population weighted		✓		✓
KP F-Statistic	.	.	345	465

Notes: Standard errors, in parentheses, are clustered on the meso-region level. Domestic robots are not weighted by sectoral input-output linkages. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 1.A.5: IV estimations: Wages (in logs) - Stacked differences (1994-2004-2014)

	All		Manufacturing)		Mining	
	(1)	(2)	(3)	(4)	(5)	(6)
Domestic Robot Exp.	-0.06*	0.01	0.03*	0.02	0.00	-0.02
	(0.03)	(0.06)	(0.02)	(0.01)	(0.02)	(0.02)
Foreign Robot Exp.	-0.01	-0.04	0.01	0.01	-0.02	-0.02
	(0.03)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	1114	1114	1114	1114	844	844
State Year FE	✓	✓	✓	✓	✓	✓
Regional char.	✓	✓	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓	✓	✓
Regional import growth	✓	✓	✓	✓	✓	✓
Population weighted		✓		✓		✓
KP F-Statistic	1800	1281	1800	1281	1783	1544

Notes: Standard errors, in parentheses, are clustered on the meso-region level. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 1.A.6: Changes in GDP and population - Stacked differences

	Population (log)		Regional GDP (log)	
	(1)	(2)	(3)	(4)
Domestic Robot Exp.	-0.000	-0.004	0.012	0.007
	(0.003)	(0.004)	(0.012)	(0.013)
Foreign Robot Exp.	-0.005	0.003	-0.015	-0.011
	(0.004)	(0.004)	(0.011)	(0.008)
Observations	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓
Regional char.	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓
Regional import growth	✓	✓	✓	✓
Population weighted		✓		✓
KP F-Statistic	1800	1281	1800	1281

Notes: Stacked-differences are measured between 1994, 2004 and 2014 for population (Panel and) and between 1997, 2004 and 2014 for GDP (Panel B). Standard errors, in parentheses, are clustered on the meso-region level. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 1.A.7: Employment by gender, skill and foreign owned enterprises - Stacked differences (1994-2004-2014)

Employment ratios	Female	Male	Skilled	Medium skilled	Low skilled	Foreign owned firms
	(1)	(2)	(3)	(4)	(5)	(6)
Domestic Robot Exp.	0.151**	0.076	0.101***	0.221**	-0.096	0.004
	(0.075)	(0.117)	(0.026)	(0.102)	(0.081)	(0.003)
Foreign Robot Exp.	0.015	-0.118	-0.013	-0.084	-0.007	0.001
	(0.050)	(0.150)	(0.017)	(0.100)	(0.097)	(0.001)
Observations	1114	1114	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓	✓	✓
Regional char.	✓	✓	✓	✓	✓	✓
Industry controls	✓	✓	✓	✓	✓	✓
Regional import growth	✓	✓	✓	✓	✓	✓
KP F-Statistic	1800	1800	1800	1800	1800	1800

Notes: Standard errors, in parentheses, are clustered on the meso-region level. All specifications include state-year fixed effects. Regional characteristic controls are the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises and log microregion GDP. Industry controls are the share of manufacturing workers and light-manufacturing workers. Import growth refers to changes in overall and Chinese imports. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.A.8: Robustness tests (weighted) - Stacked differences (1994-2004-2014)

	Initial population weighted					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Changes in the manufacturing employment ratio						
Domestic Robot Exp.	0.19**	0.16**	0.21**	0.17*	0.42***	0.50***
	(0.09)	(0.07)	(0.10)	(0.09)	(0.12)	(0.13)
Foreign Robot Exp.	-0.11***	-0.14**	-0.09*	-0.14***	-0.15***	-0.14***
(through exports)	(0.03)	(0.06)	(0.05)	(0.04)	(0.04)	(0.04)
Foreign Robot Exp.			0.01	-0.02	-0.02	-0.03
(through imports)			(0.04)	(0.05)	(0.04)	(0.04)
Panel B: Changes in the mining employment ratio						
Domestic Robot Exp.	-0.03	-0.00	-0.00	0.00	0.02	0.02
	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Foreign Robot Exp.	0.08***	0.05**	0.06**	0.05***	0.05***	0.05***
(through exports)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Foreign Robot Exp.			-0.00	-0.00	-0.01	-0.00
(through imports)			(0.01)	(0.01)	(0.01)	(0.01)
Observations	1114	1114	1104	1114	1114	1114
Controls	✓	✓	✓	✓	✓	✓
Mesoregion Year FE	✓					
State Year FE		✓	✓	✓	✓	✓
State-level std. error clustering		✓				
Excluding top 5 exposed			✓			
Additional industry controls					✓	✓
Additional export controls						✓
KP F-Statistic	688	728	1818	2222	297	323

Notes: Regressions are weighted by initial population. Standard errors, in parentheses, are clustered on the meso-region level apart from column 2. In columns 1, meso-region-year fixed effects, while in all other columns state-year fixed effects are used. All specifications control for the log microregion population, the share of female workers, the routine task intensity index, the average high school graduate rate, the average university graduate rate, the share of foreign owned enterprises, log microregion GDP, the share of manufacturing workers and light-manufacturing workers. In column 3 the five regions most exposed to foreign automation are excluded. Column 4 adds foreign robot exposure through imports. Additional industry controls are the initial shares of workers in the agricultural and automotive sectors. Additional export controls include the initial exports (in logs) and the initial shares of the agricultural and food, beverages and tobacco sectors. * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 1.A.9: Shift-share adjusted (AKM) standard errors

	Manufacturing Employment		Mining Employment	
	(1)	(2)	(3)	(4)
Panel A: OLS regression				
Foreign Robot Exp.	-0.121	-0.158	0.113	0.043
	(0.047)	(0.053)	(0.037)	(0.036)
AKM Standard Errors (foreign robot exposure)	.0403	.0538	.00579	.00296
Panel B: IV regression				
Domestic Robot Exp.	0.100	0.311	0.012	0.009
	(0.089)	(0.095)	(0.030)	(0.014)
AKM Standard Errors (domestic robot exposure)	.0966	.159	.026	.0106
Observations	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
Population weighted		✓		✓

Notes: EHW Standard errors in parentheses. AKM standard errors for the instrumented variable are reported at the bottom of the table. All specifications include meso-region dummies. Regional characteristic controls are the log micro-region population, the share of female workers, the routine task intensity index, the average highschool graduate rate, the average university graduate rate in 1994, the share of foreign owned enterprises in 1995 and log micro-region GDP in 1997. Industry controls are the share of manufacturing workers and light-manufacturing workers in 1994. Export controls include the initial share of exports of the food and beverage, basic metals, other manufacturing, chemicals and agricultural sectors. Employment outcomes are the changes in the ratio of the respective employment to population ratio between 1994 and 2014.

Table 1.A.10: Rotemberg weights

	Domestic Automation			Foreign Automation		
Sectors	(1)	(2)	(3)	(4)	(5)	(6)
Agriculture	-0.010	-0.009	0.000	-0.252	0.192	0.028
Food, Beverages, Tobacco	-0.001	-0.007	-0.030	-0.203	-0.342	-0.014
Paper and Printing	0.004	0.003	0.003	0.093	0.059	0.050
Rubber and Plastic	0.069	0.060	0.052	-0.002	-0.003	-0.002
Mineral Products	0.067	0.070	0.051	-0.056	-0.031	-0.023
Basic Metals	0.211	0.224	0.218	1.549	1.252	1.054
Fabricated Metal Products	0.045	0.042	0.036	0.008	0.006	0.005
Motor Vehicles	0.557	0.563	0.635	-0.064	-0.050	-0.047
Other Manufacturing	0.016	0.017	0.010	-0.050	-0.035	-0.024
Regional char.		✓	✓		✓	✓
Industry controls			✓			✓

Notes: The table presents Rotemberg weights for the eight industries with the largest weights, following Goldsmith-Pinkham et al. (2020). Columns 1-3 show weights for domestic automation exposure and columns 4-6 weights for foreign automation exposure. Regional characteristics are the log micro-region population, the share of female workers, the routine task intensity index, the average highschool graduate rate, the average university graduate rate in 1994, the share of foreign owned enterprises in 1995 and log micro-region GDP in 1997. Industry controls are the share of manufacturing workers and light-manufacturing workers in 1994.

Table 1.A.11: Regional imports - Stacked differences (1997-2006-2014)

	Difference in asinh(imports)			
	Manufacturing		Mining	
Domestic Robot Exp.	0.07**	0.04*	-0.04	-0.03
	(0.03)	(0.02)	(0.03)	(0.03)
Foreign Robot Exp.	-0.01	-0.01	-0.02*	-0.02
(through imports)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	1114	1114	1114	1114
State Year FE	✓	✓	✓	✓
Baseline controls	✓	✓	✓	✓
Population weighted		✓		✓
KP F-Statistic	2742	945	2742	945

Notes: Standard errors, in parentheses, are clustered on the meso-region level. All specifications include meso-region trends. Regional characteristic controls are the log micro-region population, the share of female workers, the routine task intensity index, the average highschool graduate rate, the average university graduate rate in 1994, the share of foreign owned enterprises in 1995 and log micro-region GDP in 1997. Industry controls are the share of manufacturing workers and light-manufacturing workers in 1994. Export controls include the initial share of exports of the food and beverage, basic metals, other manufacturing, chemicals and agricultural sectors. Outcomes are the changes in imports between 1997 and 2014.

Table 1.A.12: Sector level analysis - Stack differences (1994-2004-2014)

	OLS		IV	
	(1)	(2)	(3)	(4)
Panel B: Changes in the manufacturing employment ratio				
Domestic Robot Exposure	0.25***	0.15	0.26***	0.18
	(0.08)	(0.13)	(0.08)	(0.12)
Foreign Robot Exposure	-0.39	-0.35	-0.40*	-0.37*
	(0.24)	(0.25)	(0.23)	(0.22)
Panel C: Changes in the mining employment ratio				
Domestic Robot Exposure	-0.03	-0.01	-0.04	-0.03
	(0.03)	(0.03)	(0.03)	(0.03)
Foreign Robot Exposure	0.23**	0.23**	0.23**	0.25***
	(0.10)	(0.10)	(0.09)	(0.09)
Observations	42	42	42	42
Year FE	✓		✓	
Broad sector $\times$ year FE		✓		✓
KP F-Statistic	.	.	45.6	25.1

Notes: Heteroscedasticity robust standard errors are in parentheses. Broad sector fixed effects include indicators for the agricultural, light manufacturing, manufacturing and service sectors. Employment outcomes are the changes in the ratio of the respective employment ratio between 1994-2004-2014.

Chapter 2

Dealing with Agricultural Shocks: Income Source Diversification through Solar Panel Home Systems

joint work with Krisztina Kis-Katos, Friederike Lenel, Christoph Weisser

Abstract Solar panel home systems can help farmers in rural areas to mitigate income losses when they experience agricultural shocks. To study this, we exploit a unique dataset, containing information on hourly electricity usage and daily loan repayment of 20,000 customers of a solar panel company in Tanzania. Customer survey data combined with supervised machine learning allows us to classify the customers' daily electricity usage patterns and predict the likelihood of electricity usage for business purposes. We measure agricultural shocks based on local variation in plant health across growing seasons, based on remotely sensed data. Our results show that farmers make use of the solar panel system for income generation when facing harvest loss. This is in particular driven by customers with fewer resources and those living in more remote areas where little alternative employment opportunities exist. Furthermore, customers who are more likely to use their solar panel for business purposes face less loan repayment difficulties after an agricultural shock. We find the adjustment to be of temporary nature: farmers do not shift their resources into the new business activity permanently but rather use the panel as a tool for short-term relief to stabilize their income situation.

Keywords: Agricultural shocks, farming, solar panels, loan repayment

JEL Classification: O13, Q15, Q42, C55

2.1 Introduction

Climate change is increasingly leading to unpredictable weather events around the world (Lobell et al., 2011; Serdeczny et al., 2017). With more than half of its population employed in the agricultural sector (OECD & FAO, 2016) and 95% of cultivated land being rain-fed (Nash et al., 2013), Sub-Saharan Africa (SSA) is particularly exposed to changing climatic conditions. Shocks to crops and agricultural land, in the form of droughts, floods and pests have increased in frequency, duration and magnitude over the past half century and will continue to do so in the future (IPCC, 2021). Agricultural production in SSA is predicted to decrease by between 8% and 22%, depending on the type of crop, by 2050 (Schlenker and Lobell, 2010). As agricultural yields are becoming more volatile, it will be vital to increase the resilience of farmers to such shocks while at the same time reducing their exclusive reliance on farming activities.

Farmers have been shown to cope with vegetation shocks by adapting their farming practices (Bryan et al., 2009; Roncoli et al., 2001; Thomas et al., 2007), by diversifying their income sources through off-farm engagements (Branco and Féres, 2021) or, in more extreme cases, through migration (Gray and Mueller, 2012; Cattaneo and Peri, 2016). The set of alternative coping strategies depends on the available resources and local circumstances, often resulting in sub-optimal adaptation decisions with severe long-term consequences.[1] Diversification is usually costly, difficult to maintain in the long-run and successful only when markets are well-functioning and assets are available (Dercon and Krishnan, 1996; Barrett et al., 2001; Call et al., 2019).

In this paper, we investigate the potential of solar panel systems, a technology that by now is widely available in SSA, to diversify income sources and thereby mitigate income losses after vegetation shocks. In the context of rural Tanzania, we show how using solar panels for business purposes can mitigate credit repayment difficulties in times of harvest loss, by enabling individuals to generate additional income in the local service economy. We exploit a unique set of high-frequency electricity usage and daily repayment data, provided to us by a clean energy company that sells solar panel home systems in East Africa. We focus on roughly 20,000 farmers in Tanzania, who purchased a solar panel system on loan between 2015 to 2018. Each system comes with a number of different appliances (such as lights, TV, mobile phone charger, radio, etc.) that can be used both for private consumption and for small-scale business purposes. The repayments for the loan can be flexibly made within three years. Each repayment charges the solar system automatically, similar to pay-as-you-go products; the system

[1] In a review of rural livelihood diversification strategies, Alobo Loison (2015) emphasizes that only wealthier smallholders are successful at diversifying their income sources. Asset constraints and limited access to credit and new technologies hinder a large proportion of smallholders to benefit from diversification (Barrios et al., 2008). As a result, agricultural shocks have a stronger negative impact on workers living close to subsistence (Jayachandran, 2006).

shuts off automatically when payments are insufficient. In the long run, a customer risks losing the system once the system has been shut down for more than 30.5 days per year. More frequent system shut-downs after agricultural shocks can therefore be taken as an indicator of a customer's increased financial distress.

Our analysis consists of three steps. We first identify locality-specific vegetation shocks in Tanzania for the 1,919 wards in our sample by calculating deviations from a long-term average vegetation index over the growing season. Relying on daily vegetation estimates based on satellite data (NOAA STAR, 2018), we are able to identify the exact timing of local agricultural shocks. As agricultural seasons vary widely across Tanzania (Kaminski et al., 2016), we define location-specific growing and harvesting seasons in order to assess when the relevant negative vegetation shocks occur. In a second step, we use individual hourly electricity usage data to identify days during which customers are more likely to have used their solar system for the purposes of small-scale business. We rely on information from a detailed customer survey that asked randomly selected households whether they have used their solar system for business purposes during the last month. We combine this information with daily electricity usage data to train a supervised machine learning classifier to predict for each customer-day observation the likelihood that the solar panel was used for business purposes. Based on these results we make out-of-sample predictions for all remaining customers and time-spans.

Third, we combine the localized vegetation shocks with predicted individual business usage and loan repayment data in a regression framework to analyse how farmers cope with agricultural loss. Having the means to generate additional income on the side is especially valuable for farmers who are increasingly exposed to extreme weather events. We therefore hypothesize, that (1) farmers are more likely to make use of their solar panel system for income generating activities when facing a harvest loss, and that (2) this in turn helps them to mitigate the negative income effects. As localized crop failure at an aggregate scale can be considered exogenous to each farmer's individual behavior, this approach helps to causally identify whether individuals in rural areas leverage electricity access through solar panels to diversify their income and mitigate income losses. We include a rich set of fixed effects in our estimations, controlling for all unobserved customer characteristics and district-level developments over time, and address further concerns to causal identification in a number of robustness tests. In addition, we study which factors determine the feasibility of this strategy. Guided by the literature on income source diversification, we focus on wealth, alternative employment opportunities and farming practices. As solar panel-based small scale businesses require little additional investment, we expect the panel to provide valuable off-farm business opportunities in particular for households with limited resources. We expect solar panels to be also more useful for income generation in more remote areas where few off-farm employment alternatives exist and access to finance is limited.

Three results follow from our empirical analysis. First, local vegetation shocks increase the probability of subsequent loan default. A one standard deviation decrease in vegetation health during the growing season increases the number of system shut-off days due to insufficient payments in the harvest season by about 6.7 percent. Second, farmers tend to rely more on their solar panel system for business purposes in a harvest season that follows a weaker-than-usual growing season, although the average increase only amounts to about 2 percent of a standard deviation. Third, farmers who use their solar panel system for business encounter fewer cash-flow problems when facing harvest loss: they accumulate a lower number of system shut-off days and are able to make on average larger repayments. Our results are robust to a number of different specifications, using different vegetation shock and business classification measures and controlling for other factors that might induce or prevent farmers from using their system for business purposes. We furthermore find that especially poorer farmers and farmers who live in more remote areas are more likely to use their solar panel for business in the aftermath of an agricultural shock, probably as they lack access to alternative coping strategies. Finally, we show that farmers do not shift their resources into the new business activity permanently, but only use the system as a tool for temporary relief in order to stabilize their income situation.

This paper makes three main contributions to the literature. First, it highlights the potential of solar panel home systems as a tool for income diversification. Our results indicate that the additional income not only helps repaying the large investment but also provides means to reduce the vulnerability to extreme weather events. To the best of our knowledge, this potential has so far received limited attention. Indeed, the literature on uptake and impacts of low-cost solar panels is still in its infancy. In terms of income generation, Wassie and Adaramola (2021) find that micro-businesses in South Ethiopia that use solar panel systems tend to be more productive, while in a survey of the existing literature, Lemaire (2018) documents only a modest correlation of solar panel usage with income and poverty indicators.[2] All existing studies on solar electricity usage behavior are exclusively based on survey data, which inhibits analyzing

[2]The broader evidence on the socioeconomic impacts of electrification in Africa, mostly measured as household connection to the electricity grid, is mixed. In contrast to positive findings in Asia and Latin America especially in terms of manufacturing employment, labor supply and educational attainment (Lipscomb et al., 2013; Thomas et al., 2020; van de Walle et al., 2017), there is no clear evidence about electrification causing economic development and alleviating poverty in Africa (Bayer et al., 2020; Lee et al., 2020b; Lenz et al., 2017; Peters and Sievert, 2016). Even though households in Africa are willing to spend large parts of their income on electricity, a major constraint is the level of costs of access to electricity, which hamper private investment in electricity grids (Grimm et al., 2020). Households that are willing to spend more on electricity are those that will gain most from electrification (Lee et al., 2020a). Grimm et al. (2020) conclude that small-scale and low-cost solar panels are a preferred measure to reach electrification in rural areas.

changes in usage behavior over time.[3]

Second, high-frequency observational data on daily repayment and electricity consumption allows us to study the direct economic impact of agricultural shocks on small-holder farmers as well as their adaptation dynamics. Combined with rich survey and geo-referenced data, we are able to study the potential drivers of adaptation strategies as well as their success. This contributes to the literature on farmers' adaptation strategies to agricultural shocks, which typically relies on survey data: Farmers respond to rainfall shocks by taking-up off-farm employment (Branco and Féres, 2021; Chuang, 2019). Yet, such income diversification depends on local conditions, such as the availability of off-farm employment (Jayachandran, 2006; Blakeslee et al., 2020) or flexible labor regulations (Colmer, 2021) and available income (Macours et al., 2012).[4] Furthermore, such diversification strategies are not always beneficial but can also result in harmful outcomes: off-farm employment is often associated with low wages and poor working conditions and can lead to on-farm labor shortages (Antwi-Agyei et al., 2018). Indeed, Fink et al. (2020) show that alleviating temporary liquidity constraints of farmers through season-specific loans can reduce the necessity to search for low-wage piece work. We contribute to this literature by identifying economic effects and shifts in labor supply based on observational data and by studying behavioral dynamics across seasons.

Third, we identify localized vegetation shocks based on satellite data that directly reflect plant health and capture what farmers observe on the ground. We therefore do not focus only on a specific type of shock that affects crops, such as droughts or floods, but include all types of shocks that are relevant for farmers in our setting. This adds to the literature that studies the impact of localized weather shocks on the household or individual level. Rainfall deviations and drought are usually found to be associated with lower health outcomes, educational attainment, socio-economic status and consumption, and increasing poverty (Mueller et al., 2014; Kjellstrom et al., 2016; Jessoe et al., 2018; Hyland and Russ, 2019; Joshi, 2019). Also, loan repayment is shown to suffer in the aftermath of extreme climatic events (Castro Iragorri and Garcia, 2014; Pelka et al., 2015; de Roux, 2021).[5] Even though farmers in rural areas report pests

[3]Relying on similar data from the same company, Weisser et al. (2021) introduce the classification approach of electricity usage for business purposes and show that business usage is correlated with a lower likelihood of default. To the best of our knowledge, this is the only paper that uses high-frequency observational data to study solar panel usage behavior.

[4]Declining trade costs and the resulting market integration decrease the magnitude of income losses through rainfall shocks and induce farmers to plant crops with less volatile yields (Burgess and Donaldson, 2010; Allen and Atkin, 2016).

[5]Collier et al. (2011) find a stark increase in restructured loans in response to the 1997–1998 El Nino in Peru. Linking precipitation data to loan repayment over a longer time period, Pelka et al. (2015) show an increase in delinquency rates of farmers in Madagascar in the aftermath of excessive rains; similarly, de Roux (2021) shows that excessive rains increase the likelihood of late payments of coffee farmers in Colombia, though he finds no negative impact in the long run. Studying default risk models on a more

and infestations to be among the most frequently experienced causes for crop failure (Abid et al., 2020; Salazar and Rand, 2020), only a small number of papers considers these types of shocks. These studies typically rely on survey data where pests are one of several categories for causes of crop-loss (Beegle et al., 2006; Dercon and Krishnan, 2007; Porter, 2012). Our measure of vegetation stress accounts for such shocks beyond drought and excessive rain. At the same time, we show in further robustness checks that our mitigation results can be reproduced relying on the arguably more exogenous rainfall deviation data as well.

The rest of the paper is structured as follows. In the following section, we present the setting of our analysis in more detail and derive our hypotheses. Section 2.2 presents the different data sources, whereas section 2.3 outlines the empirical approach by deriving the measures for vegetation shocks, and presenting the machine learning approach to classify business usage as well as the empirical models. Section 2.4 presents the empirical results and addresses robustness issues. Section 2.4.3 discusses possible channels and alternative explanations, whereas Section 2.5 concludes.

2.2 Setting and Data

2.2.1 Setting

Our analysis relies on proprietary customer data, shared with us by a clean energy company that operates in several countries in East Africa. We focus on Tanzania, where the company has been selling solar panel home systems to households on credit since 2011. The data span until the end of 2018. By then, the company's customer base included about 100,000 households.

The solar panel home system comes with a number of appliances (TV, radio, mobile phone charger, lamps, etc.). The company sells its products mainly through its own outlets that are located in towns throughout the country. The typical customers are low-income households living in rural areas, where solar panel home systems serve as an alternative energy source due to very low rates of electrification. In 2016, less than 10% of the population living in rural areas in Tanzania had access to electricity (World Bank, 2015). The company offers three different system types to match the varying electricity needs and payment abilities of the different customer groups. The systems cost between 600 USD and 1,300 USD, requiring a substantial financial investment as compared to the average income in Tanzania.[6] The types differ in the panels' power

aggregate level, Castro Iragorri and Garcia (2014) shows that rainfall variability is an important factor in predicting credit default.

[6]The annual GNI per capita in current terms in Tanzania for the years 2014 to 2017 was 970 USD https:

(80Wp, 120Wp or 200Wp) and the additional appliances that come with each system. All solar home systems come with a four year warranty and close customer support.

Almost all customers purchase their solar home systems on credit; by the end of 2018, only 5% of all systems were paid in full at the time of their purchase. A detailed phone interview assesses eligibility for the loan. After approval, borrowers are required to pay 5% of the total system costs immediately, receive the system, and have then three years to repay the remaining amount. There is an incentive to repay faster: borrowers receive a 10% discount on the system price if they repay within 2 years and a 20% discount if they repay within one year. The company facilitates loan repayment by operating a highly flexible repayment system that allows customers to choose the timing and amount of their payments. Customers use mobile money to make their payments (usually through MPesa) either through their own account or through a mobile money agent. As long as the loan has not been fully repaid, the system operates like any pay-as-you go device where each loan payment also charges the system to function for as many days as the payment translates to. The system shuts down as soon as the full charge has been used up (that is when a next payment is due) and will only become functional again once a new payment has been made. The off-periods of the system are recorded and can be used to assess the timing and length of repayment difficulties that each customer encounters. Customers are allowed to have their system off for an accumulated period of 30.5 days per year. If this period is exceeded, customers are attended to by a loan field officer. If a customer is not able or willing to pay, the system is de-installed. Using the same data, Grohmann et al. (2020) show that customers flexibly adjust loan repayments to their individual income streams.[7]

The system can be used both for private consumption and to generate income. Survey data suggests that around 20–30% of the borrowers use their system to also generate income. Of those, some use the panel exclusively for business purposes, in particular, for shops, bars and restaurants. Lights allow for longer opening hours, while the radio and the TV make the venues more attractive. Yet, most use the system primarily for private purposes and generate income with it only on the side. They then supplement their main income e.g., by charging the phones of other villagers against a small fee or by operating home cinemas for a certain period of time. These types of businesses do not require large investments and can be set up relatively easily. In particular, when facing income losses from their main occupation, customers might make use of this possibility.

//data.worldbank.org/indicator/NY.GNP.PCAP.CD?view=chart.

[7] In general, the literature on flexible loan contracts shows that such contracts can enhance entrepreneurial risk taking and increase business profits (Battaglia et al., 2018). They also help to mitigate income shocks but this may come at the cost of increased default rates (Czura, 2015). In our setting, however, Grohmann et al. (2020) show that customers who repay in a less regular manner are not more likely to default.

2.2.2 Data

For our empirical analysis, we combine customer-specific information provided to us by the clean energy company—namely, customer characteristics, repayment behavior, and electricity consumption—with geo-referenced indicators of local fluctuations in plant health. We obtained full customer data on all borrowers from July 2015 to November 2018. Although the company started its operations in Tanzania in 2011, we only use data starting with July 2015 as by this time the company started to follow standardized and well-established procedures and had a substantial enough customer base.

To increase the comparability across customers, we focus on customers that bought the more common 80W or 120W systems and exclude the most expensive system type of 200W. Our data only includes borrowers who bought the system on credit and who took at least one year to repay the system (excluding about 5% of all customers who pay the system by cash on receipt and 7% who take less than one year to repay the full loan). Customers can purchase multiple systems and additional appliances from the company, for each of which a new credit line is then opened. For the 10% of customers that have more than one, we focus on the first credit line. We exclude systems that are very rarely used (with on average less than 10 Watt hours of electricity consumption per day).

Since we are interested in whether solar panels can help farmers to mitigate agricultural income shocks that are linked to crop failure, we focus on rural customers who reported at the time of the purchase farming as a source of income. We classify areas as rural according to the definition of National Bureau of Statistics of Tanzania and identify farmers based on the types of occupation reported in an initial loan eligibility interview. This leaves us with a total of 19,939 borrowers and 13,566,412 borrower-day observations for a time period from July 2015 to November 2018. Figure 2.A.2 in the Appendix displays how our sample of customers is spread out across Tanzania.

Customer characteristics We derive customer characteristics from socio-economic information collected by the company during the loan-eligibility interviews. As the questionnaire of the loan-elegibility interview changed over time, certain characteristics are not available for all customers in our sample. We combine this with location specific information on financial access points using data from the Financial Sector Deepening Trust (FSDT).[8] Panel A of Table 2.A.1 in the appendix shows the socio-economic profile for our sample of customers. 16% of borrowers are women and they live in relatively large households with 4.6 people per household on average. Farmers produce on aver-

[8] https://financialaccessmaptz.com.

age 46 bags of maize in a season (90% of farmers reported to grow maize and for 75% of them, maize is the main crop). In our sample of rural areas, the average distance to the next town (with more than 20,000 inhabitants) is 50 kilometers. Nonetheless, most of the customers live near to some kind of financial service provider: Within 5 kilometers from their home, 63% have a mobile money agent and about 38% have a bank agent.[9] Half of the farmers in our sample run a business on the side (already at the time of the loan interview) and 10% are employed. Farmers produce on average 46 bags of maize in a season (90% of farmers reported to grow maize and for 75% of them, maize is the main crop). 40% of the farmers state that they usually spend money on farming equipment and 22% irrigate their farm-land. Compared to the Tanzanian population, farmers who purchase a solar panel on credit are slightly older and have larger household sizes; they are furthermore more likely to run their own business but less likely to be wage employed, while, compared to the population of farmers, they more often irrigate their lands.[10]

Loan repayment Each payment a borrower makes is recorded by the mobile money operator at the time of the payment. The data is instantly shared with the solar panel company so that the system can be charged accordingly. From this data we can infer the timing, frequency and amount of payments as well as the timing, frequency and duration of system shutdowns. We aggregate the borrowers' payments on a daily level. For each day in the life of a borrower, we know the amount of payments a borrower made and how much charge-time such a payment translates to. We use this information to further calculate whether and for how long a system was switched off if no payment was made. The repayment data shows that customers indeed repay flexibly. The median customer charges the system for 11 days, but there is large variation (see Panel B of Table 2.A.1). In our sample of farmers in rural areas, two-thirds of customers have had their system turned-off for at least one day, i.e. they were late in their payments at least once. Yet, customers tend to repay fast. The average period of consecutive off-days does not exceed two days for half of the customers. Among the remaining half, the average number of consecutive off-days is fifteen. A prolonged system shut-down is indicative of more substantial cash-flow problems and can finally result in default.

[9] We link financial access points and distance to the next city to customers via their GPS coordinates. For customers for whom this information is missing, we calculate the ward-level average distance, and assume that they have access to the respective financial service if more than half of the customers in our sample living within the same ward do.

[10] According to the 2014-2015 wave of the Living Standards Measurement Survey, 29% of Tanzanian households were headed by a women. The average age of the population is 22, while the average age of a household head is 44. The average number of household members is 3.7. Half of the household heads report farming as their main activity, whereas about 17% run their own business and 13.5% are wage employed. From the sample of households with farming activities and 3% irrigate their land (https://microdata.worldbank.org/index.php/catalog/lsms).

Roughly 8% of our customers defaulted on their solar-panel system loan altogether.[11]

Electricity usage High-frequency electricity usage data allows us to observe the electricity consumption behavior of all customers throughout time. Each solar panel home system records electricity generation and outflow at two different loads (small and big load) every ten minutes. We aggregate this data by hour and load to describe the electricity usage patterns for each customer-day observation. On a sunny day, an 80W (120W) system can produce up to 400 (600) Watt per day; energy production reduces by 30–40% on cloudy days and by 60–80% on completely rainy days. The solar panel system is set up in a way that it only allows plugging in appliances purchased from the solar panel company. Hence, electricity usage depends on the type of appliances purchased with the system and their intensity of usage throughout the day. On average, customers consume 7.7 Watt per hour, with the system being used primarily in the late evening, and least in the early morning (see Panel C of Table 2.A.1). 95% of the customers use less than 300 Watt hours of electricity per day, and hence the daily productive capacity of the solar panel even on a cloudy day is larger than what most of the users can consume.[12]

Vegetation data To define vegetation shocks, we use variations from the Normalized Difference Vegetation Index (NDVI) by looking at deviations of region-specific seasonal NDVI from its long-term average. The NDVI is calculated with remotely sensed radiance values in both the visible and near-infrared channels. It measures the presence of the chlorophyll pigment, indicating the health of vegetation (Tucker, 1979; Tucker et al., 2005). Data comes from the Center for Satellite Applications and Research of the National Oceanic and Atmospheric Administration (NOAA STAR, 2018). It contains weekly NDVI estimates on a 4km raster level. Although in economic studies the NDVI has received limited attention, it is widely used for drought and early famine detection, index-based insurance, crop yield forecasts and measuring environmental change (Roerink et al., 2003; Peters et al., 2002; Mkhabela et al., 2005; Pettorelli et al., 2005; Karnieli et al., 2010; Chantarat et al., 2013; IPCC, 2019; Meroni et al., 2019).[13] Panel

[11] As an illustration, Figure 2.A.3 in the Appendix shows repayment of two random borrowers who received the system at the end of April 2017. Repayment behavior differs strongly both in amounts and regularity.

[12] In all analyses, we control for cloud coverage to account for lower electricity production. Yet, since our main outcomes are measured outside of the rainy season, where days are typically sunny, sometimes cloudy but rarely rainy, for most customers fluctuations in electricity usage reflect changes in the demand for electricity, whereas supply factors play a relatively minor role.

[13] In economic studies, the NDVI has been used for instance by Emerick (2018), who estimates trading frictions in new crop varieties, using the index to observe plant health after floods. Beg (2021) analyzes the economic effects of a property rights reform in Pakistan, with the NDVI as measure of agricultural output. Pape and Wollburg (2019) use the NDVI to measure the impact of drought on poverty, consumption, and hunger. Lybbert and McPeak (2012) study uncertainty and risk preferences among pastoralists

D of Table 2.A.1 shows that, on average, the local NDVI during our sample period is similar to its long-term average.

Further data sources We rely on daily precipitation data on the 0.5 degree grid-cell level to establish the local timing of growing seasons. Data comes from the Climate Hazards Group Infrared Precipitation with Station database (CHIRPS) (Funk et al., 2015). We calculate regional and seasonal cloud cover with data from MODIS Terra 8-day surface reflectance products (Vermote, 2015). Ward level population statistics and rural/urban definitions come from the 2012 Census of Tanzania. Further, we use data from a survey conducted with a small subset of customers, in order to identify those which use their system for business purposes (more details are provided in section 2.3.2).

2.3 Empirical Approach

In order to estimate whether the usage of solar panel home systems contributes to mitigating income shocks in the aftermath of crop loss, we first specify the local agricultural seasons and define vegetation shock indicators. In a second step, we derive a time-variant measure for the likelihood of business activity based on electricity consumption patterns. Our subsequent empirical analysis links customers' repayment behavior in the harvest season to the agricultural shocks experienced in the preceding growing season and investigates whether business usage of solar panels can have a moderating effect.

2.3.1 Measuring Seasons and Vegetation Shocks

Defining Growing Seasons Agricultural shocks arise primarily during the growing season and hence, their precise measurement requires information on the timing of seasons. As Tanzania covers a large number of climatic zones, with large spatial differences in yearly rainfall patterns and both unimodal and bimodal yearly rainfall regimes, the measurement of growing seasons needs to be location-specific. We build on Dunning et al. (2016) and Liebmann et al. (2012) to define location-specific growing seasons for each of the 1,919 wards (ADM3 regions) included in our sample. We first determine the local length and approximate timing of growing seasons based on historical rainfall patterns and then establish the onset of each annual growing season based on actual rainfall.

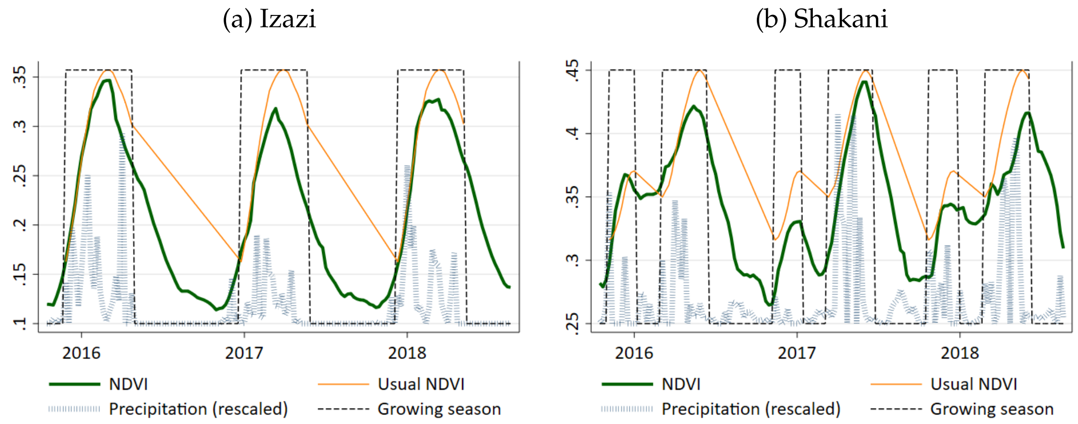
in Kenya, estimating pastoral land quality through the NDVI.

In the first step, we identify for each ward the 'usual' onset and cessation of a growing season following the procedure suggested by Dunning et al. (2016). We aggregate rainfall to the ward level by taking the average over each 0.05 degree grid-cell within the ward. We first calculate the cumulative daily precipitation anomaly, which is the cumulative sum of the difference between the daily precipitation and its historical average, smoothed over one month. The local minimum and the subsequent maximum of this curve (calculated over nine days each) define the start and end point of each growing season. A season must last for at least one month to be considered and there must be at least 30 days between two growing seasons. This ensures that there is a maximum of two growing seasons per ward and year. We calculate historical averages based on 20 years of data from 1995 to 2014 in order to avoid concerns about long-run changes in rainfall patterns (Kotir, 2011), and to ensure that our definition of seasons is not confounded by weather shocks that occur during our study period.

The second step accounts for the fact that rainy seasons start at different points in time across years. We observe the minimum daily cumulative rainfall deviation within 50 days (20 days) before and after the 'usual' season onset and end to define the year-specific season onset for unimodal (bimodal) rainfall regimes. By doing so, we hold the local length of each growing season constant (in contrast to Dunning et al., 2016) but flexibly adjust its starting point. As drought shocks mechanically lead to shorter growing seasons, fixing the seasonal length allows us to construct agricultural shock variables that are comparable over time. Moreover, since the procedure associates little overall rainfall with a late onset of the growing season, we restrict the yearly rainy season to start at most 5 weeks before or after the onset of the usual growing season for unimodal and 3 weeks for bimodal rainfall regimes. Figure 2.1 displays weekly precipitation in two exemplary wards and overlays it with our definition of local growing seasons and further variables. The graphs show a substantial overlap between rainfall patterns and our definition of local growing seasons, but at the same time they demonstrate a substantial variation in rainfall over time. In 84% of all wards the growing seasons are unimodal like in Izazi (panel a). When growing seasons are bimodal like in Shakani (panel b), in our empirical analysis we focus on the growing season that is followed by the longer harvest season.

Measuring agricultural shocks Our agricultural shock variables rely on modelled local NDVI data to assess deviations in plant health within each growing season from its long-term local average. Typically the NDVI will increase throughout the growing season and will start to decrease before or right after its end (Mkhabela et al., 2005). Figure 2.1 shows that increases in the weekly NDVI in our two selected wards follow increases in rainfall with a lag and usually peak close to the end of the growing season,

Figure 2.1: Season definition, precipitation and NDVI



Notes: This figure presents weekly precipitation, season definitions, weekly NDVI and the usual historical NDVI values for two ADM3 regions in Tanzania. Precipitation is re-scaled to fit the graph. Within each region the length of each season is kept stable over time, while the season start varies. Historical NDVI values are calculated as the weekly mean between 1995-2014.

reflecting vegetation growth before the harvest. To assess the strength of agricultural shocks within the season, we compare the values of actual NDVI to a measure of 'usual' NDVI that is computed based on weekly data from 1995 to 2014 during each growing season. Our main agricultural shock variable captures the relative gap between the seasonal average of current and 'usual' NDVI in form of a percentage deviation. We set the shock variable to zero when average NDVI and hence plant health is above its long-run average. We expect shocks to be more harmful, the longer the vegetation stress lasts and the larger the NDVI deviation is from its historical value. Both dimensions are reflected in our measure of relative deviation. Alternatively, we use the deviation of the season's median NDVI from the long-run median for robustness checks. To ease interpretation, in regressions we standardize the shock variable to have a mean zero and standard deviation of one. As displayed in Figure 2.1, in Izazi, the largest relative negative deviation as compared to the usual NDVI occurred in 2017, with no negative average deviation in 2016 and a more moderate one in 2018. In Shakani, NDVI losses during the second growing season have been relatively more substantial in 2016 and 2018 and virtually absent in 2017. In a similar vein, Figure 2.A.4 in the Appendix shows substantial spatial variation in the average magnitude of negative vegetation shocks in the year 2017.

Relying on the NDVI to measure vegetation shocks has several advantages in our context. First, as the index reflects plant health, it provides a good indication of what farmers observe on the ground and thus can react to. At the same time, due to its relatively coarse spatial resolution, the index only reflects aggregate outcomes at the ward level and is unlikely to be affected by the behavior of individual farmers. Second, defining shocks based on the NDVI data reduces the potential of mis-measurement driven by single days with large deviations in rainfall or temperature. As rainfall is characterized by much larger fluctuations between different days, not all deviations from norm will directly impact plant health. Because of this, NDVI is also less sensitive to mis-measurement in the start or end-dates of seasons. Third, NDVI fluctuations also capture other sources of plant health stress beyond weather shocks and thus can help to identify a wider range of shocks affecting agricultural productivity. For instance, local plant health may decline not only due to droughts or excessive rain, but also due to other natural catastrophes like a pest infestation that destroys the crops, or a pandemic that causes a shortage in agricultural labor.

2.3.2 Classifying Business Users by Machine Learning

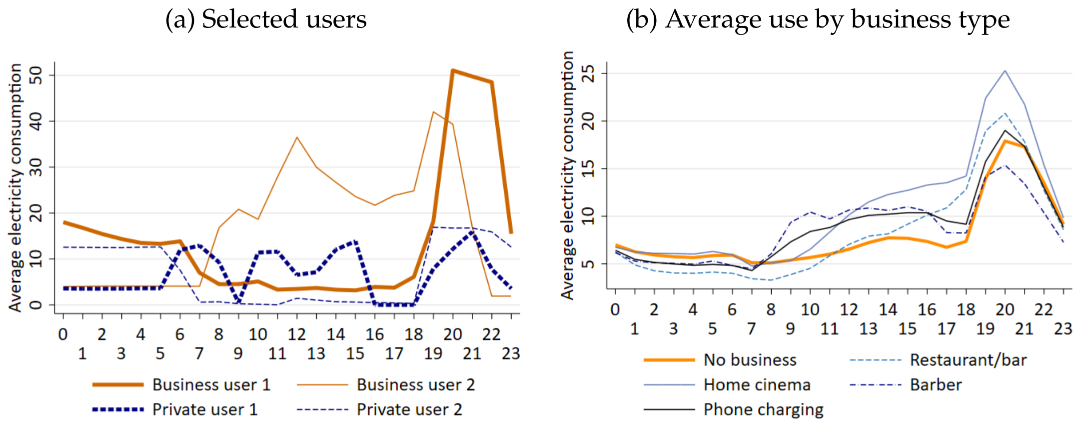
In order to understand whether solar panels are also used as a shock-coping strategy, we need comprehensive information on business activities for each customer. Although information on income generating activities could theoretically be also collected through repeated customer surveys, generating comprehensive data on business usage is costly and may suffer from reporting and recollection biases. This is especially problematic as we expect households to use solar panels for income generation only periodically and hence small-scale business usage is likely to fluctuate substantially over time. High-frequency electricity consumption data serves as a promising alternative data source for inferring whether customers are operating a small-scale business at any point in time based on the patterns of their electricity consumption. Anecdotal evidence from the field together with insights from a small-scale survey suggest that customers who operate small-scale businesses use electricity in a different way than those who only consume electricity privately.

Insights from survey data We identify differences in electricity usage patterns across customers by relying on customer survey data that records detailed information on the purpose of electricity usage. The survey was conducted in November 2018 with 315 customers and asked about whether they have been using electricity from their solar panel system for business purposes during the past month.[14] Of the surveyed cus-

[14] The customers targeted by the survey were randomly selected from all customers that had the system for at least 6 months and had had at least one system shut-off day in the past. They are thus not necessarily representative of our sample. Table 2.A.3 compares the main characteristics of the small survey sample and the full sample. They are quite similar in terms of electricity consumption (Panel C), but differ in terms of repayment. Customers in this survey sample repay less on average, but also have fewer off-days, compared to the full sample. In terms of customer characteristics, respondents of the small survey are more often female, employed and have smaller household sizes, but have similar maize yields and farming practices. Lastly, the regions of the two samples do not differ in terms of the average vegetation index (Panel D).

tomers, 22% indicated to have used the system also for business purposes whereas the other customers were using electricity exclusively for private consumption. Panel (a) of Figure 2.2 illustrates electricity consumption patterns of four customers on a characteristic day in the month before the survey, two of whom reported to have also been using the system for business. Business user 1 and private user 2 both use electricity for overnight lighting, but the private user does not power other appliances during the day whereas the electricity consumption of the business user increases sharply in the evening. By contrast, business user 2 relies on large electricity loads throughout the whole day and shuts down all appliances overnight. Finally, private user 1 consumes larger amounts of electricity in the morning and the evening, but also moderate levels of electricity during the daytime.

Figure 2.2: Daily electricity usage profiles of selected users and on average



Notes: Panel (a) presents the electricity usage profile of four customers on one selected day (two business and two private users). Panel (b) presents average daily usage profiles by the type of business.

When we average these patterns across all customers during the four weeks preceding the survey in panel (b) of Figure 2.2, we see a somewhat higher electricity usage among business customers also on average (see also Figure 2.A.5 in the Appendix), although the differences are less striking. While private and business users consume on average similar levels of electricity during the morning, the difference becomes more pronounced during the day and early evening, before converging again at night. Customers who use their solar panel system to run a home cinema have the largest electricity consumption from noon until the night. The second highest energy flows throughout the day stem from phone charging businesses that can charge up to 10 mobile phones in parallel. Restaurant and bar operators in turn require a lot of electricity only in the afternoon until early evening. Purely private customers use their system primarily in the evening. This could be driven by returning home from work, and powering lights as well as entertainment appliances.

Generating labels and features Based on this contextual information and the insight that daily energy usage patterns are likely to change with business usage, we use a machine learning approach to determine the likelihood that each customer-day observation reflects energy use for business purposes. To train the machine learning classifier, we use the electricity usage data of the surveyed customers over three months preceding the survey. The survey data provides us with so-called *labels*, which classify customer-day observations either as business or private consumption days on the basis of the customer survey. From the included total 22,068 customer-day observations 4,588 are labelled as business user days and 17,480 as non-business user days. Business customers are remarkably similar to pure private users (as displayed in Table 2.A.2 in the Appendix), only differing in using more electricity per day. As a predictor of the likelihood of business usage, we only rely on daily electricity usage profiles of the survey participants and do not use any other time-invariant customer characteristics collected in the survey. This will allow us to extrapolate business usage to the full customer base of the firm by performing out-of-sample predictions relying on the trained model. In order to increase the interpretability but also efficiency of our approach, we define a set of *features* that distill the information on daily electricity usage dynamics in the form of 84 explanatory variables. These variables distinguish between usage recorded for small and big load appliances as well as their average and the difference between the two types of loads. The features capture the intensity of usage at different times of the day as well as its variability and changes in usage intensity across hours and time periods. Appendix 2.A.2 provides a definition of all utilized features.

The XGBoost algorithm We use the Extreme Gradient Boosting (XGBoost) algorithm (Chen and Guestrin, 2016), which belongs to the family of decision tree-based methods. Tree-based methods are especially powerful to detect non-linear relationships and complex interactions across predictive features (see Molnar, 2022, for an overview).[15] Similarly to a random forest, gradient boosting is an ensemble learning method that aims at sequentially dividing the training sample of customer-day observations based on randomly selected cut-offs of each feature in such a way that by the end, business and private customers become as dissimilar to each other in their usage characteristics as possible. As a result of the sequential partitioning of the data, we end up with a 'tree', with structurally similar customer groups ending up on the same 'leaf' of the tree. In order to avoid over-fitting, the XGBoost algorithm recursively combines a large number of strongly simplified (so-called 'shallow') trees, all of which are weak predictors on their own but achieve a strong predictor in their combination. For a more

[15]Within the field of economics, the use of supervised machine learning classifiers for prediction or data generation purposes has been steadily gaining ground (see e.g., Oster, 2018; Albanesi and Vamossy, 2019; Davis et al., 2021).

technical description of the XGBoost model see Appendix 2.A.2.

Classification results We divide all customer-days in our sample into a training and test dataset consisting of 80% and 20% of the customer-day observations respectively. The sampling and hyperparameter tuning decisions are listed in Appendix 2.A.2. Figure 2.3 displays the results of the classification exercise within our test sample, plotting the distribution of the predicted business probabilities of survey participants during the three months prior to the survey. The probability distribution shows that private users almost never end up with high predicted business probabilities. Business users however are assigned low business probabilities on some days.[16] This shows that the classification procedure faces a pertinent issue of measurement error. For a customer who reports having operated a business recently, we label each day in the past three months as a business day. In reality however, on many of these days most likely no business activities have taken place, for instance due to holidays and off-times, sickness, festivities, or travel, and hence these days are mislabelled in our data. Table 2.A.2 shows for the full sample that business users are assigned much higher business probabilities on average as compared to private users, and have a somewhat larger electricity consumption. Business users do not indicate substantially larger maximum affordable payment amounts.

For our subsequent empirical analysis, we will consider a day to be a business day if the predicted probability of business usage on this day surpasses the relatively high cut-off of 75%. This choice is informed by the measurement error described above. The cut-off of 0.75 nearly coincides with the average daily predicted business probability among business users in our sample, which lies at 0.76 (see Table 2.A.2 in the Appendix).

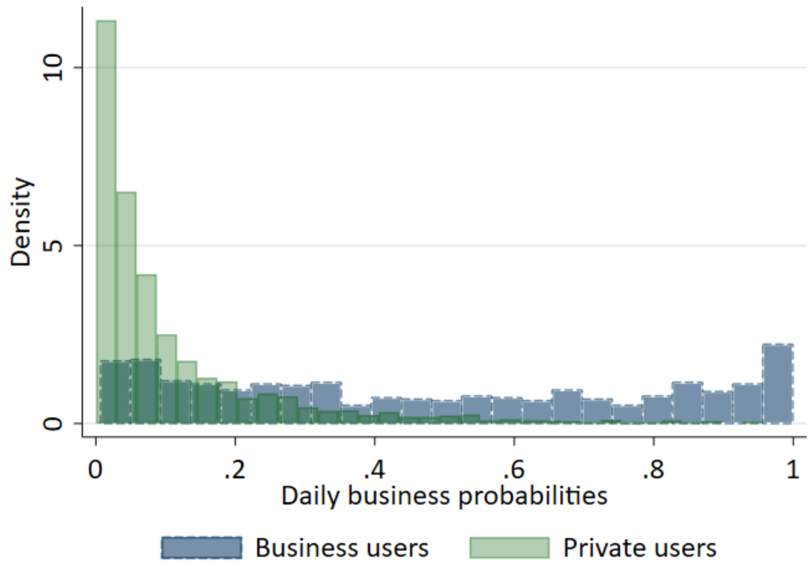
Choosing this cut-off results in a very high precision rate for our model and hence only a very low number of false positives. Less than 9% of the private customer days are wrongly classified as business days in our test sample at this cut-off (type one error). This comes at the expense of increasing the number of false negatives (type two errors). With a recall of 0.391, only about 40% of business customer days are classified as days on which the business has been operated with a large likelihood.[17] In further robustness checks we will assess the sensitivity of our results to the cut-off choice.

Figure 2.4 lists the 20 most important features (those with the largest predictive power) that result from this classification algorithm where feature importance is based on the so-called gain metric. The gain metric reflects improvements in the precision of the

[16] We achieve an accuracy (AUC, area under the curve) metric of 0.883, which gives us the area share under the receiver operating curve (ROC) that plots the false and true positive rates against each other.

[17] Overall this provides us with an F1-score of 0.547, which is calculated as the harmonic mean of precision and recall and measures the overall fit of the model.

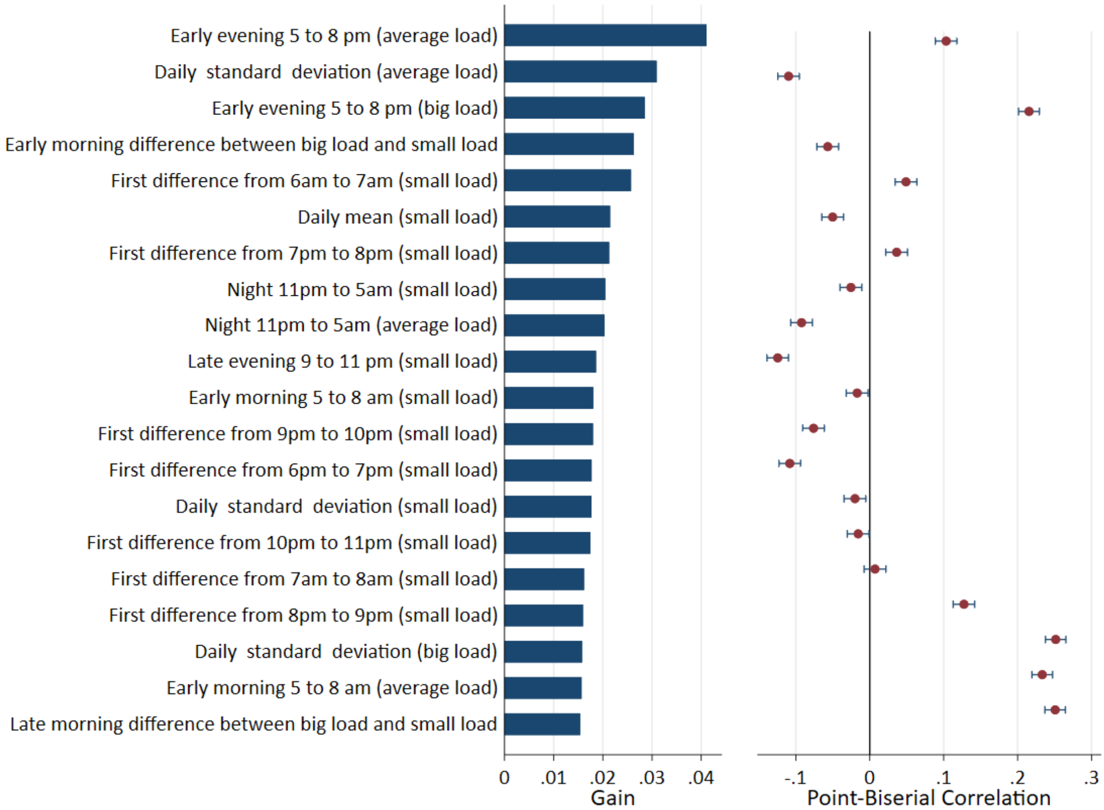
Figure 2.3: Predicted daily business probabilities of survey respondents in the test sample



classification by capturing how often each variable is selected for splitting the tree and by how much the predictive power of the model improves due to each split (Chen and Guestrin, 2016; Chen et al., 2020). The right panel of the graph displays the simple correlation between each of the variables and the business usage indicator. The strongest predictor of business usage is the intensity of early evening use as business users tend to use more electricity especially from 5 to 8pm. Their electricity usage is also less volatile throughout the day, resulting in a lower daily standard deviation. Private customers tend to use more electricity (especially small load) at night, whereas business and private customers tend to power up and down their system at different times in the day. These simple correlations provide a first intuition on the possible role of these factors, but the boosted tree approach relies on a more complex non-linear combination of a series of all 84 features.

Out-of-sample predictions of business use We use the trained model to make out-of-sample predictions of the likelihood of business usage for all customer-days in the full dataset, for 37,485,575 customer-day observations in total. Figure 2.A.6 in the Appendix shows the distribution of the predicted likelihood of business days in the full sample. Although the major bulk of our customer-day observations shows relatively low predicted business probabilities, we also observe a right tail with very high probabilities of predicted business. At our preferred probability cut-off of 0.75, about 5% of customer-day observations are considered as business days in our sample and about 4% of customer-days during the relevant harvest periods are labelled as business days. Figure 2.A.7 in the Appendix shows that, similarly to the energy usage patterns of the

Figure 2.4: Most important features of electricity usage in the XGBoost classification



Notes: The left panel in the figure displays the importance of the 20 features with the largest predictive power in the XGBoost classification procedure according to the gain metric. The right panel presents the correlation of each of these variables with business usage for illustrative purposes.

labelled customers displayed in Figure 2.2, average electricity consumption is larger during daytime on days that are labelled as business days than as non-business days but differs less during other periods.

2.3.3 Estimation Strategy

Empirical models Our empirical analysis proceeds in three steps. We first estimate the impact of agricultural shocks in the growing season on loan repayment in the following harvest season to capture the effects of the income shock that occurred due to harvest loss. In a second step, we study whether farmers make use of their solar panel to buffer the experienced shock; in particular, we estimate whether agricultural shocks increase the likelihood that the systems are used for business purposes. Finally, we combine these two analyses and estimate whether farmers who use the system for business purposes can mitigate the negative impact of agricultural shocks on repayment. For our analyses, we rely on the idiosyncratic spatio-temporal variation in agricultural shock occurrence. Including customer as well as district-year fixed effects allows us to identify the effects of localized weather shocks within customers while controlling for yearly variations at the level of 133 ADM2 districts.

The effects of agricultural shocks on repayment and business usage For the first two analyses, our estimations follow the specification

$$Y_{i\omega rt} = \alpha_{i\omega r} + \beta_1 S_{\omega rt} + \beta_2 C_{i\omega rt} + \delta_{rt} + \epsilon_{i\omega rt}, \tag{2.3.1}$$

where $Y_{i\omega rt}$ captures either customer i's loan repayment behavior or the extent to which a customer uses the system for business purposes. Our main measure for repayment behavior is the number of days that the solar panel system was turned off due to non-repayment (transformed as inverse hyperbolic sine to control for outliers). To capture business usage behaviour, we use the proportion of business days during the harvest season; as derived in section 2.3.2, business days are defined as days with an electricity usage that results in a predicted business probability of above 75%. Both outcomes are measured for the harvest season, when income shocks due to harvest loss are expected to materialize.

The coefficient of interest is β_1, which indicates the effect of a localized agricultural shock, $S_{\omega rt}$, experienced in the growing season that preceded the main harvest season in ward ω (ADM3 region) in district r (ADM2 region) in year t. The shock is measured by the percentage deviation of the seasonal average NDVI from its long-run local average as derived in section 2.3.1. We standardize the shock variable to have a standard deviation of one to ease interpretation. Customer fixed effects $\alpha_{i\omega r}$ control for all time-invariant household characteristics that could affect electricity usage and repayment behavior, whereas district-year fixed effects δ_{rt} control for a whole range of economic and policy shocks that occur at the level of larger administrative regions (ADM2). For seasons that span over two years, the fixed effect accounts for the year with the larger part of the season.

The vector of controls $C_{i\omega rt}$ contains the length of the harvest season measured in weeks, the number of weeks the customer has the system during the harvest season and the average cloud cover in region r during the harvest season. The length of the harvest season varies across years due to an earlier or later start of the growing season. Moreover, some customers purchase their system in the course of the specific harvest season and can operate the panel only for a shorter time. A shorter time horizon mechanically reduces the scope for experiencing repayment difficulties but also increases the relative importance of short periods of business usage. Finally, we expect the likelihood of business usage to increase over time and be relatively lower among the most recent customers. We control for average cloud cover to account for any electricity supply effects. Vegetation shocks in the growing season are unlikely to affect the capacity of solar panels to produce electricity in the harvest season. Nevertheless, unusual cloud cover during the harvest season could bias our results if not controlled for. See Table 2.1 for summary statistics for all variables included in our main analyses at the level of customer-season cells.

Table 2.1: Descriptives on main variables

Variable	Median	Mean	SD	Min	Max
System-off days in harvest season	2.37	7.56	11.20	0.00	88.26
Consecutive system-off days in harvest season	0.00	2.59	6.45	0.00	30.00
Charged days per payment in harvest season	9.11	16.72	25.38	0.10	941.88
Proportion of business days (75% cut-off)	0.01	0.04	0.09	0.00	0.97
Proportion of business days (50% cut-off)	0.05	0.10	0.14	0.00	0.99
Avrg. hourly electricity use in harvest season (in Watt)	7.52	7.87	3.04	0.38	23.02
Vegetation shock in growing season (mean deviation)	0.04	0.06	0.07	0.00	0.96
Vegetation shock in growing season (median deviation)	0.05	0.07	0.08	0.00	0.99
Dev. between season and usual sum NDVI	0.95	0.95	0.06	0.00	1.00
No. weeks of harvest season	25.00	24.52	5.58	5.00	46.00
No. weeks owning system in harvest season	22.00	20.86	7.07	1.00	42.00
Average regional cloud cover in harvest season	0.01	0.03	0.04	0.00	0.27

Notes: Descriptive statistics vary on the customer $\times$ harvest season level, with a total number of 43,207 observations. Business proportion cut-offs refer to the threshold after which a given day is counted as a business day.

Business usage as a moderating factor To study the potentially mitigating effect of business usage on repayment in the aftermath of an agricultural shock we interact our shock measure $S_{\omega rt}$ with business usage $B_{i\omega rt}$ and estimate

$$Y_{i\omega rt} = \alpha_{i\omega r} + \beta_1 S_{\omega rt} + \beta_2 B_{i\omega rt} + \beta_3 S_{\omega rt} \times B_{i\omega rt} + \beta_4 C_{i\omega rt} + \delta_{rt} + \epsilon_{i\omega rt}, \quad (2.3.2)$$

where the dependent variable $Y_{i\omega rt}$ captures a customer's loan repayment behavior. All remaining controls are specified as above. In this setting, the coefficient of interest is β_3, which measures the impact of an agricultural shock in the growing season on repayment in dependence on the degree of business usage.

Issues of identification As compared to the classically used rainfall variables (Shah and Steinberg, 2017; Tiwari et al., 2017; Chuang, 2019), our NDVI-based shock measure captures a much wider variety of reasons for crop loss. This raises the concern that aggregate crop losses could also be influenced by shocks to agricultural technology, urbanization, health or the local labor market and hence are less exogenous than rainfall itself. Conditional on customer and district-year fixed effects and season length and cloudiness controls, our measure of localized agricultural shocks in the growing season captures idiosyncratic variation in crop losses within each ward. Focusing on rural wards only and including district-year fixed effects mitigates such concerns, as it is unlikely that other, labor-market related shocks are distinctive to individual wards only. From the perspective of individual customers, we interpret the estimated coefficients in a causal way as measuring the differential effects of aggregate agricultural losses (due to climatic but also other reasons) on customer behavior. Furthermore, as a robustness check, we show that our NDVI-based shock measure is strongly correlated with negative rainfall shocks (see also Anyamba et al. (2002)). Our repayment and moderation results persist also when we use the arguably more exogenous variation in rainfall to define shocks, although we do not see the same adjustment in business usage itself.

Among further robustness checks, we demonstrate that our results are robust to measurement issues. We specifically show that the business usage proxy, which we derived from electricity consumption with the help of machine learning, captures a separate phenomenon from average electricity consumption. Although these two variables are positively correlated, they respond differently to agricultural shocks and our results are robust to the inclusion of electricity consumption as a further control. We also show that our results are robust to how we define the agricultural shock, how we measure loan repayment, or which cutoffs we choose to define business usage days.

The analysis of the moderating role of business usage yields a more descriptive result. The choice to use a solar panel for business purposes at any point in time is endogenous as it depends on customer and location characteristics and also changes in response to the shock. Customer fixed effects control for time invariant, but not for time varying factors driving this choice. By interacting business usage with an aggregate vegetation shock variable, we are able to at least partly draw on variation that is based on changes over time and is triggered by region-level fluctuations in crop-loss. However, we still have to make sure that the moderating effect of business usage does not simply proxy for some other omitted factor. We show that the effect persists when interactions between the shock and a series of further possible moderating factors are also controlled for, including the intensity of electricity usage and a list of further customer and location characteristics that all could also drive the individual shock response. Finally, analyses of temporal dynamics show only responses to current but neither to past, nor to future shocks, emphasizing the temporary (but also quasi-exogenous) nature of aggregate vegetation shocks.

2.4 Results

2.4.1 Baseline Results

The effects of agricultural shocks on repayment and business usage To assess the average effects of agricultural shocks, we first investigate whether a local vegetation loss occurring in the growing season affects customers' income in the following harvest season substantially enough to reduce their ability to repay the solar panel system. We estimate equation (2.3.1) with the dependent variable measuring the number of days that the solar panel system was turned off due to non-payment during the harvest season. The results are presented in Table 2.2 in columns 1-3. In the first column, we control for customer and year fixed effects only, while in column 2 we add interacted district-year fixed effects. In our preferred specification in column 3, we include the length of the harvest season, the number of weeks the customer has the system during the harvest season and the average cloud cover during the harvest season as further controls. The results confirm that agricultural shocks during the growing season lead to more severe cash constraints in the harvest season, when crop losses are expected to translate to income shocks. In our fully specified model (column 3), a one standard-deviation larger vegetation shock increases the number of system-off days due to non-repayment by 3 percent. For an average customer, this effect adds about 0.2 days to the average of 7.5 shut-off days per harvest season.

Columns 4 to 6 of Table 2.2 test whether farmers respond to the agricultural shock by using their solar panel system for business purposes. The proportion of business days during the harvest season increases significantly with the strength of the vegetation shock throughout all specifications. The results of the fully specified model in column 6 indicate that at least some farmers start to rely more on their system for income generation in the aftermath of an agricultural shock. A one standard-deviation larger vegetation shock during the growing season leads to an average increase in the proportion of business usage days by 0.2 percentage points during the harvest season. Although this effect is of a relatively modest magnitude, it moves the proportion of business usage days by about 5% relative to its average of 0.04.

The moderating role of business usage Can the use of the solar panel system for income generating activities help to mitigate the income losses experienced in the aftermath of agricultural shocks? We study this by estimating specification 2.3.2 with repayment as the dependent, the vegetation shock the main explanatory and business usage the moderating variable, all defined as before. Table 2.3 presents the results. Without a vegetation shock, customers with a larger proportion of business days experience on average fewer system shut-off days (columns 1 and 2). Conditional on

Table 2.2: The effect of agricultural shocks on loan repayment and business usage

Outcome:	*asinh* System-off days in harvest season			Prop. business usage days in harvest season		
	(1)	(2)	(3)	(4)	(5)	(6)
Vegetation shock in growing season	0.095***	0.051***	0.030**	0.001*	0.002**	0.002**
	(0.022)	(0.018)	(0.014)	(0.001)	(0.001)	(0.001)
asinh No. weeks of harvest season			0.107			-0.005
			(0.108)			(0.005)
asinh No. weeks owning system in harvest season			0.637***			0.007***
			(0.022)			(0.001)
Average seasonal cloud cover			0.453			-0.058**
			(0.613)			(0.025)
Dependent mean	1.729	1.729	1.729	0.041	0.041	0.041
Observations	43207	43207	43207	43207	43207	43207
Customer FE	✓	✓	✓	✓	✓	✓
Year FE	✓			✓		
ADM2 × year FE		✓	✓		✓	✓

Notes: The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. ADM2-year fixed effects are set at the district level. Business days are defined as days with an electricity usage that results in a predicted business probability above 75%. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

customer fixed effects, this correlation suggests that customers who use the electricity for income generation in the current harvest season are better able to prioritize loan repayment due to improvements in their cash-flow. When the proportion of business usage days increases by about 10 percentage points, this reduces the number of system-off days by about 19%. Columns 3 and 4 introduce the interaction between the agricultural shock and the business measure. Its coefficient is negative and statistically significant, even when including district-year fixed effects (column 4), indicating that customers using their solar panel system for business activities respond less strongly to the income loss triggered by an agricultural shock. Customers experiencing a one standard deviation larger agricultural shock increase the number of system-off days by about 4.4%. This negative shock can be offset by slightly more than half (2.8%) if at the same time customers also increase the number of business usage days by 10 percentage points. These results indicate that rural farmers who use their solar panels for business purposes after an agricultural shock are better able to counteract the negative income shock.

These results provide evidence that having access to solar panels, coupled with flexible loan repayment schedules, can be crucial to deal with more frequently occurring vegetation shocks. Farmers that experience such shocks can forego credit default by adjusting their repayment schedule and by using clean electricity generating devices

Table 2.3: Agricultural shocks and repayment: The role of business usage

Outcome:	*asinh* System-off days in harvest season			
	(1)	(2)	(3)	(4)
Vegetation shock in growing season	0.059***	0.034**	0.070***	0.044***
	(0.020)	(0.014)	(0.020)	(0.014)
Business usage in harvest season	-1.989***	-1.929***	-2.002***	-1.944***
	(0.126)	(0.128)	(0.120)	(0.123)
Business usage × vegetation shock			-0.272***	-0.279***
			(0.074)	(0.075)
Observations	43207	43207	43207	43207
Customer FE	✓	✓	✓	✓
Year FE	✓		✓	
ADM2 × year FE		✓		✓
Controls	✓	✓	✓	✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. The proportion of business days and the number of system-off days are measured during the first harvest season that follows the growing season. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to generate off-farm income.

2.4.2 Robustness Checks

We conduct various robustness checks to ensure that our results are not driven by confounding factors and test the sensitivity of our results to using alternative measures for the agricultural shock, repayment behavior and business usage. Finally, we confirm our underlying story by studying the effects for farmers growing crops with multiple harvests per year as well as for non-farmers—for both groups we should not expect to find strong effects.

Electricity usage as a confounding factor Customers using the system for business purposes also consume on average more electricity. Thus, our results might just mask an increase in electricity consumption in the aftermath of shocks. For instance, customers could start using their system more inside their home as the opportunity costs of leisure time decline. Alternatively, a bad harvest could also induce customers to search for extra work outside of the household, reducing their home electricity use. To assess the importance of such confounding dynamics, we regress the average seasonal electricity consumption over all days on which the system was not shut off on the shock and test whether our main estimation results persist when controlling for electricity consumption and its interaction with the vegetation shock variable. Tables 2.A.4 and 2.A.5 in the Appendix display the estimation results. We find that agricultural shocks actually decrease electricity consumption in the following harvest season (columns 1 and 2 in Table 2.A.4), possibly reflecting that some farmers pursue other income generating activities outside of the household. Yet, the estimated effect of shocks on business usage remains stable once controlling for electricity consumption (column 3). This suggests that the increase in business days is not merely an artifact of changing usage intensity in the aftermath of agricultural shocks. When focusing on repayment in Table 2.A.5 instead, electricity consumption is negatively correlated with the number of system-off days, indicating that customers who use their system more intensively are less likely to run into delinquency in general. The inclusion of the electricity consumption variable as a further control leaves the coefficients of the vegetation shock and its interaction with business usage largely unaffected (columns 1 and 2). In column 3, we additionally include an interaction between electricity usage and the vegetation shock to control for changes in the intensity of solar panel usage behavior after vegetation shocks. Our main coefficient of interest, the interaction between business usage and vegetation shocks, remains statistically significant. This makes it more likely that we are indeed measuring changes in business behavior, rather than general adjustments of electricity consumption after vegetation shock events.

Customer and location characteristics as confounding factors Whether customers decide to use their system for business purposes is likely affected by a number of customer and location-specific characteristics. This being an endogenous choice renders a causal interpretation of the mitigation effect difficult. Including customer fixed effects allows us to control for time-invariant customer characteristics, yet not for time-varying factors driving this decision. To ensure that the shock-mitigating effects of business usage do not simply capture the role of underlying factors that make farmers generally more resilient to income shocks, we first interact our shock variable with a number of customer characteristics measured at the time of the purchase (see Table 2.A.6 in the Appendix). Controlling for interactions between the vegetation shock and the customer's gender, household size, an indicator for animal farmers, those with a side-business, and wage-employed, as well as the value of farm output (captured by reported yields of maize, the most commonly produced crop) does not change the shock-mitigating effects of business usage. As a second check, we control for interactions between the vegetation shock and additional locational characteristics (see Table 2.A.7 in the Appendix). A better access to infrastructure and credit affects the likelihood of starting a business, yet also the extent to which an income shock can be mitigated. Controlling for interactions between our shock measure and an indicator for the presence of a mobile money agent, a bank, or a lending group within a 5 kilometer radius around the customer's location, or with the distance to the next town does not change our baseline results. Therefore, our findings of income-loss mitigation do not merely reflect that business users may have a better access to financial services in times of vegetation stress.

Past business usage as a confounding factor Past business usage might make current business usage also more likely, while at the same time, customers who are more experienced in running a business might also be able to better cope with shocks. We test if the mitigation effect is solely driven by past business experience in Table 2.A.15 in the Appendix. Indeed, we find that past business usage, measured as the proportion of business days in the last season, is associated with a larger proportion of business days in the current season (column 1). Customers already experienced in using the solar panel system in the past thus are more likely to use the system for business purposes today. Past experience however neither alters the effect of agricultural shocks on current business usage (column 2) nor does it affect the repayment ability in the current season, independently of the occurrence of a shock (columns 3 and 4). The insignificant interaction coefficient and the fact that the interaction between current business usage and the vegetation shock remains statistically significant, indicates that it is not business experience mitigating the vegetation shock. Rather, farmers who adjust their usage behavior in the current season are the ones who can reduce the negative income shock.

Measurement and identification: Is loan repayment indicative of cash constraints? So far, we have interpreted the negative interaction coefficient between vegetation shocks and business usage, when explaining system off-days, as an indicator for additional income being generated when the system is used for business purposes. However, such a result might also reflect differences in system valuation. When using the system for business purposes, customers might arguably be more dependent on the system and more willing to prioritize loan repayment to avoid system-off days. While the valuation of the system is likely a function of the expected cash flow, it can also have a psychological component, making the number of system-off days an imperfect proxy for actual cash constraints. We therefore also study two alternative measures for repayment behavior that are presumably less affected by system valuation but reflect cash availability more directly—namely, the average number of charged days during the harvest season and the average payment amount translated to charged days. These two variables capture the frequency and size of the loan repayments. We expect cash constraints to result in generally smaller payment sums, which translate to lower number of charged days and smaller average payments. Finally, we also use a more stringent measure for system-off days that places a larger emphasis on longer shut-down-periods: the maximum period length during which the system is turned off (in contrast to the total number of days) during the harvest season. Results are shown in Figure 2.A.8 in the Appendix. Vegetation shocks lead to a significant decline in the average number of paid days, and an increase in the number of days during which the system is turned off; farmers using their system for business purposes, however, make on average larger and less frequent payments and experience less system-off days in the aftermath of a shock. These findings suggest that the better repayment behavior when a system is used for business purposes in the aftermath of a shock can indeed to a large extent be explained by cash availability.

Measurement and identification: Using rainfall shocks We test the sensitivity of our results to how we measure vegetation shocks by focusing on deviations in rainfall instead of deviations in the NDVI. As rainfall shocks reflect climatic variability only, they are arguably more exogenous than the ward-level vegetation shocks. The reduced form relationship between NDVI and rainfall shocks at the level of wards shows that NDVI responds to rainfall shocks (see Table 2.A.10 in the Appendix). As rainfall varies substantially less across regions than the NDVI, these estimations only include ward and year but not district-year fixed effects. In the first column, rainfall shocks are defined as the percentage deviation of the pre-period average seasonal rainfall and the current season's average rainfall. In column 2, we disaggregate the rainfall deviation into negative and positive rainfall shocks. For both variables, a larger value indicates a larger absolute deviation from the average. The negative rainfall shock is highly correlated with our vegetation shock, which indicates that drought periods severely affected plant health during our period of observation. Positive rainfall shocks in turn do not lead to vegetation loss, indicating that floods were less detrimental.[18] Table 2.A.11 in the Appendix replicates our previous analyses using negative rainfall shocks (i.e. droughts) instead of the NDVI based vegetation shocks. While, we do not find that rainfall shocks lead to an average increase in business usage (column 1), the effects on repayment (column 2) and in the mitigation analysis (column 3) are qualitatively very similar to our previous findings. Less-than-usual rainfall in a growing season increases the number of days on which the customer's system is shut off, but this effect is successfully mitigated by increased business usage.

Measurement: Sensitivity to the seasonal calendar To identify the relevant vegetation shocks, in our analysis we focus on the main growing season in a year. However,

[18]We cannot build a viable instrumental variable strategy predicting NDVI variation by negative rainfall shocks as in our fixed effects specifications, rainfall turns out to be a relatively weak instrument for NDVI in the first stage.

some of the crops planted follow different growing calendars, for which our results may not hold.[19] We use information from the loan-eligibility survey about the number of seasons that farmers plant their crops per year to verify our results. In Table 2.A.12 in the Appendix, we split our sample between farmers whose main crop has only one harvest season per year and farmers whose main crop is harvested several times per year. The table shows that the effects of vegetation shocks are only statistically significant for crops that grow in one season (columns 1 and 2) and not for crops that grow in more than one season (columns 3 and 4). This confirms that our methodology truly captures the intended seasonality. Moreover, we make sure that our analysis is not sensitive to the specification of bimodal seasons, for which we up to now relied only on the main (longer) growing season. In the first three columns of Table 2.A.13 in the appendix, we exclude regions with bimodal rainfall regimes altogether. The business and mitigation outcome results are robust, while for the repayment outcome (column 2) the vegetation shock coefficient is less precisely measured and does not reach statistical significance. In the last three columns of the table, for all regions that have bimodal seasons, we aggregate the vegetation shock measure over the short and the long growing season. The business outcome and repayment results are robust to the modification.

Measurement: Sensitivity to defining business usage In order to verify that the business outcome results are not driven by how we select the cut-off to define a day with business-like usage, we rerun our preferred specification using a series of alternative thresholds to identify days of business usage (see Table 2.A.8 in the Appendix). The shock variable remains statistically significant and similar in size for similar cut-offs, and turns insignificant for values that are either too low (50%) or too high (90%) to distinguish business-days, resulting either in too many or too few business days. Next, we use the same alternative business day cut-offs to verify that our mitigation results are not sensitive to the choice of probability cut-off that we use to classify business days in Table 2.A.9. The results remain stable throughout, while higher cut-offs increase the coefficient-size of the interaction term.

Adjustment by non-farmers We expect that localized vegetation shocks lead to direct income losses mainly among the local farmers. Although crop losses could also affect the real income of rural consumers through their effects on prices, lower overall demand and available income (Acevedo et al., 2020), trade in agricultural products across wards reduces the likelihood of highly localized price hikes. Table 2.A.16 in the Appendix supports this expectation by showing no statistically significant responses

[19] In our sample, 90% of farmers indicate that they grow maize and for 75% maize is the main crop.

to vegetation shocks among non-farmers, confirming that non-farmers are less vulnerable to localized crop loss, but also showing that our empirical strategy does not capture second-round general equilibrium effects.[20] Therefore, our measure of agriculture shocks indeed appear to capture shocks that are primarily relevant for farmers.

2.4.3 Dynamics and Heterogeneities

Our results demonstrate that vegetation shocks impair the ability of farmers to repay their loans. This is in line with the literature that documents increasing in the incidence of poverty due to weather shocks (Hallegatte et al., 2015; Pape and Wollburg, 2019; Fink et al., 2020). In addition, we show that farmers make use of their solar panel system for income diversification in the aftermath of agricultural shocks and that this can help to mitigate the experienced income losses. To better understand the dynamics and potential underlying mechanisms, we conduct a number of additional analyses. First, we investigate whether farmers adjust their behavior already before the shock materializes, i.e. in anticipation of a harvest loss, and whether adjustments to shocks and business usage in the past are transitory or if adjustments persist over time. We then study to which extent our results depend on particular customer and location specific characteristics; namely, whether only customers with sufficient financial resources can make use of their system's income generating possibilities in the aftermath of a shock, whether it depends on the availability of alternative off-farm job opportunities and the customer's farming practices.

Dynamics: Anticipation So far, we have assumed that vegetation shocks that occur during the growing season affect farmers when they generate income, that is during the harvest season. However, if farmers anticipate a bad harvest during the growing season, they could immediately reduce loan payments for their solar panel systems (in order to save for more uncertain times in the future) and make use of alternative income sources (e.g., with the help of their system). Table 2.A.14 in the appendix tests whether vegetation shocks affect customers' usage and repayment behavior already during the growing season. On average, we see no changes in system usage during the growing season while the shock occurs (column 1). Farmers thus seem not to adjust their usage behavior directly, but only in the next harvest season, when the income shock materializes. This may be indicative of farmers continuing to work on the field in an effort to save the harvest, rather than taking up alternative income generating activities. That the income shock does not materialize before the harvest season is shown in column 2; there is no statistically significant effect of vegetation shocks on the num-

[20] The same holds true if we focus on self-employed non-farmers only, who are arguably more affected by demand shocks.

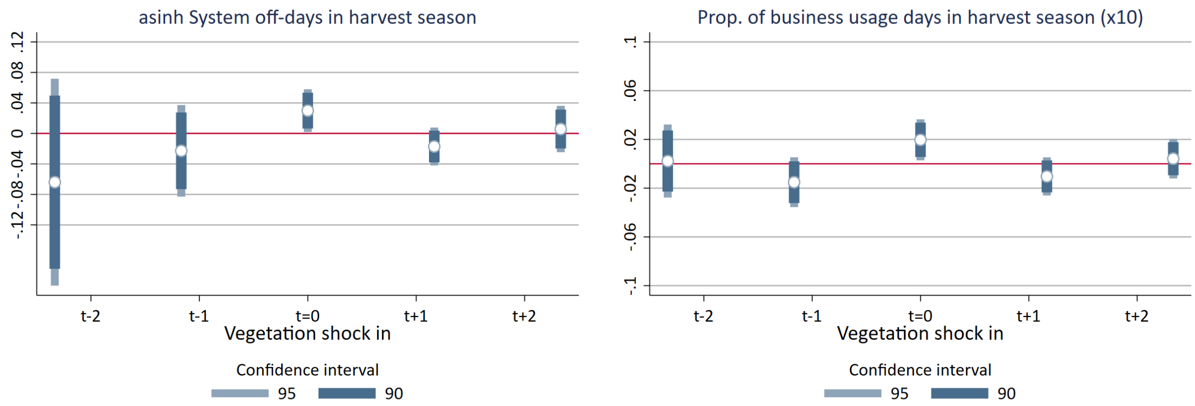
ber of days the system is off in the growing season. This also indicates that the farmers in our sample do not engage in precautionary savings. Further, the interaction between the vegetation shock and business usage during the growing season is also not statistically significant (column 3). Finally, we find no indication that business usage during the growing season can mitigate repayment difficulties in the harvest season in the aftermath of a shock (column 4). In conclusion, even though farmers' expectations of their future income streams may be adjusted downwards when experiencing a vegetation shock, farmers only start using the system for small-scale business purposes when the income shock is manifested (i.e. in the following harvest season).[21]

Dynamics: Transitory versus persistent adjustments and future shocks Do farmers make use of the system for income-generation on a more permanent basis or do they only use it as a transitory remedy until their income has stabilized? Figure 2.5 shows that past shocks have no lasting impact on behavior in the current season. Farmers neither exhibit different repayment amounts (left panel) nor business usage (right panel) after shocks in the last year (t-1) or two years ago (t-2). Therefore, small-scale business activity relying on solar panels in the aftermath of shocks appears to be transitory, used to stabilize customers' income in the short run. In the long run, the opportunity costs of no longer engaging in farming activities could be too large to permanently shift away from agriculture and into running small-scale businesses. Finally, including additionally shock coefficients for the next year (t+1) and in two years' time (t+2) provides a further placebo check, documenting no effects of future shocks on current repayment or business usage.

Heterogeneities: The role of wealth In order to understand which customers succeed to rely on solar panels for income shock mitigation, we assess how the effects of the vegetation shock vary with customer characteristics. Should only wealthier customers be able to diversify their income, business use would not necessarily be driven by necessity, as wealthier customers could both be able to start businesses and have a lower risk of running into repayment trouble. By contrast, diversification after income shocks can be vital for poor households if off-farm income generating opportunities exist (Asfaw et al., 2019). We use two measures to proxy for a farmer's wealth or income. In panel (a) of Figure 2.6 we proxy the farmers' income with the last realized maize yield as indicated in the eligibility interview. The findings in the left graph of panel (a) suggest that farmers with below median yields face longer system shutdowns after income shocks. This could be indicative of these farmers having fewer reserves and thus being more exposed to agricultural shocks. However, we do not

[21] We cannot rule out the possibility that farmers already started setting up the business in the growing season, yet were only able to generate sufficient demand in the following harvest season.

Figure 2.5: Adjustments to past and future shocks

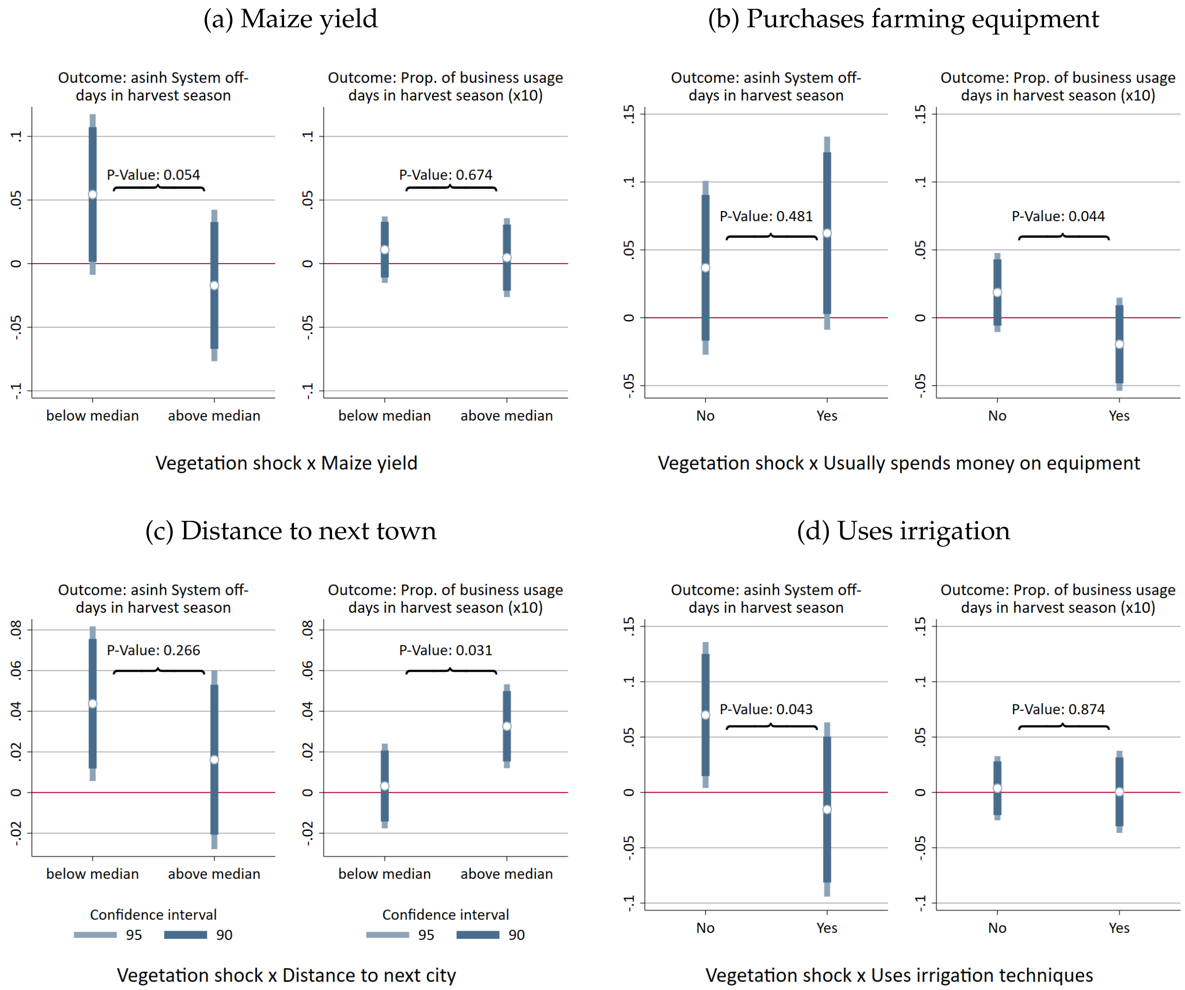


Notes: This figure displays the coefficients of vegetation shocks in the growing season of the same year, the past year, two years ago, one year ahead and two years ahead. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. All controls from columns 3 and 6 of Table 2.2 are included. Standard errors are clustered on the level of districts (ADM2). * p< 0.1, ** p< 0.05, *** p< 0.01

find differential changes in business usage by farmers' income. Conversely, panel (b) shows that farmers who report to spend money on farming equipment are similarly affected by the vegetation shock, but are less likely to respond to it by using their panel for business purposes.Purchasing farming equipment, such as tractors, entails large investment costs and is indicative of higher wealth. This finding could thus be in line with more resource-constrained customers having a stronger need to generate additional income in the aftermath of agricultural shocks (Jayachandran, 2006). Alternatively, it could indicate that farmers that already invested heavily in their farms, are less willing to diversify into other businesses.

Heterogeneities: Remoteness and professionalization of farming Income diversification is more likely to be successful in areas where there are more off-farm income generation opportunities (Chuang, 2019), whereas in remote rural areas it is more likely to be used as a means for survival (Etea et al., 2019). In the absence of other off-farm income-generating opportunities, solar electricity could provide a valuable source for income diversification by powering small businesses especially in rural areas. While we find no statistically significant differences in repayment behavior by remoteness, customers are more likely to increase their business use if they live in more remote regions (Panel (c) of Figure 2.6). This highlights one of the unique features of solar panels: They allow individuals to offer services which are not available in the most remote areas, and thereby weaken dependency on off-farm employment opportunities. Panel (d) of Figure 2.6 studies whether effects differ by the degree of professionalization and resilience to agricultural shocks in the farming business, by differentiating farmers by

Figure 2.6: Differences by individual and regional characteristics



Notes: This figure displays coefficients of the vegetation shock interacted with indicator variables as indicated in each sub-title. In each sub-graph, in the left panels, the outcome is the *asinh* number of system-off days in the harvest days, and in the right panel the proportion of business days with an electricity usage that results in a predicted business probability above 75% multiplied by factor 10. All controls from columns 3 and 6 of Table 2.2 are included. Standard errors are clustered on the level of districts (ADM2).

whether they employ irrigation techniques. As could be expected, vegetation shocks have a larger effect on repayment of farmers who do not irrigate. However, we see no differences in business usage.

Heterogeneities: The role of market saturation Lastly, one would expect that using the system for business purposes is more profitable when solar panel saturation is low and there is thus higher demand for electricity related services. Figure 2.A.9 in the appendix investigates whether customers behave differently depending on how many other individuals in the ward purchased a solar panel system. The results suggest that farmers only increase business usage after vegetation shocks if they live in areas in which there are few other solar panel users of the same firm. However, as the share of customers per ward is below 1% on average and we cannot observe solar panel penetration from other companies, these results should be treated with caution.

These heterogeneity results highlight the unique features that make solar panels an attractive asset in order to mitigate income losses. First, they enable off-farm diversification in remote areas, where typically off-farm income generation is less widely available (Chuang, 2019). Second, while asset constraints hinder off-farm income generation (Alobo Loison, 2015; Barrios et al., 2008), farmers who are not able to purchase productive farming assets (equipment) are more likely to operate a solar-panel-run business after agricultural shocks. In the absence of other assets, solar-panels can be flexibly used to generate income with little additional investment needed. Nevertheless, even though the solar-panel company targets low-income households, not all households will be able to afford an asset such as the solar panel. Third, and related to the previous point, solar panels are a flexible asset that allows farmers to temporarily diversify their income away from agriculture. We show that farmers adjust their business behavior after shocks only in the short-run as a mean for temporary relief. There is thus not the risk of farmers permanently moving out of agriculture and on-farm labor shortages (Antwi-Agyei et al., 2018).

2.5 Conclusion

More than 600 million individuals in SSA still lack access to electricity (IEA, 2019). Small-scale solar panels have the potential to supply clean energy especially to rural areas, where connections to the electricity grid remain scarce and expansion is still costly. This paper shows that farmers in rural areas can make use of such panels to diversify their income and thereby mitigate the negative effects of vegetation shocks. We leverage a detailed dataset with daily electricity usage behavior to predict the likelihood of business activities using supervised machine-learning methods. The paper shows the advantages and opportunities that such methods entail to conduct rigorous economic analyses especially on rural areas of low-income countries, in the absence of survey data. Three findings emerge from our analysis: First, we show that vegetation shocks reduce farmers' income. Second, farmers use their solar panel system for business purposes in the aftermath of vegetation shocks. Third, using the system for income generation enables farmers to cope with the negative consequences of such shocks. Especially farmers who are less wealthy and live in more remote areas are more likely to adjust the way they use their solar panel. Thus, solar panels offer those farmers a new shock-coping option who have fewer other alternatives for diversifying their income.

With climatic shocks occurring more frequently and with larger magnitude, more easily accessible means for adaptation, both on- and off-farm, will be pertinent to secure the livelihoods of farmers. Our findings suggest that solar panels should be further promoted not only as a mean to energy access but also as a tool that can help farmers to temporarily mitigate negative income losses due to agricultural shocks, in particular in remote areas where little alternative off-farm employment opportunities exist. Complementary measures such as business training or safety nets are essential to enable farmers to take full advantage of the potential of solar panel technology. Notwithstanding, once an expansion of the electricity grid or of solar panel systems makes electricity more readily available to a large proportion of the population, providing electricity related services becomes less profitable.[22] The potential of solar panel systems for income diversification is thus by nature rather temporary. The technology can nevertheless be an important bridge for farmers to cope with income shocks, before other developments open alternative opportunities. Moreover, they could help farmers to develop business-related skills, which may benefit them in the long-run.

[22] In a similar vein, Burke et al. (2019) show that low rates of loan accessibility benefit farmers who get a loan, but a higher saturation of loans at least partly eliminates these effects by diminishing the arbitrage opportunities due to lower price fluctuations.

2.A Appendix

2.A.1 Acknowledgements

We are grateful to an unnamed company for providing access to their proprietary customer data. We thank Yao Lu for excellent research assistance and Antonia Grohmann, Thomas Kneib, Rahel Laudien, Jann Lay, Reimund Rötter and participants at research seminars in Göttingen, Mainz, Paris, and Regensburg for valuable comments and discussions.

2.A.2 Classifying Business Usage

The Extreme Gradient Boosting (XGBoost) algorithm

XGBoost is one of the most powerful machine learning classifiers for structured data that utilizes a random forest algorithm for prediction. It optimizes a regularized objective function:

$$L(f(\boldsymbol{x})) = \sum_{i=1}^{n} L(y_i, f(\boldsymbol{x}_i)) + \Omega(f(\boldsymbol{x})).$$

where $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$ denote observations on a response variable y and feature vector $\boldsymbol{x}$. $L(y_i, f(\boldsymbol{x}_i))$ is a loss function for prediction $f(\boldsymbol{x}_i)$ and response y_i. The regularisation term is given by:

$$\Omega(f(\boldsymbol{x})) = \gamma T + \frac{1}{2}\lambda \|\boldsymbol{w}\|^2.$$

The number of terminal leaves is denoted with T. The vector $\boldsymbol{w}$ denotes the leaf weights. $\gamma > 0$ and $\lambda > 0$ are the corresponding regularization parameters. $\hat{f}(\boldsymbol{x})$ is iteratively updated in iteration t to minimise the objective function via gradient-based boosting:

$$\hat{f}^{(t)}(\boldsymbol{x}) = \hat{f}^{(t-1)}(\boldsymbol{x}) + \hat{g}^{(t)}(\boldsymbol{x}).$$

The random forest update $\hat{g}^{(t)}(\boldsymbol{x})$ is greedily determined in the t-th iteration of the boosting procedure. A second order Taylor expansion to the loss function is used to improve the speed of the optimization (Chen and Guestrin, 2016). We use the XGBoost implementation in R by Bischl et al. (2016).

Hyperparameters for XGBoost

We oversample the minority class of business user days to train XGBoost on a balanced data set of business and non-business user days. We train XGBoost on this oversampled training data and apply extensive hyperparameter tuning to optimize the performance of XGBoostm, relying the following parameterization. The parameters that are selected based on the hyperparameter tuning are displayed in parentheses. XGBoost uses stochastic boosting, so that a sample of the data is used to construct a tree. We set the range for the sub-sample, which refers to the share of the observations to be stochastically drown in each iteration, to 0.5 to 1 (0.701). Additionally, XGBoost samples the variables (features) that are used to construct a tree for which we specify the range of 0.5 to 1 (0.94). Thus, the optimized model relies almost on all available features. We select a range of 100 to 500 number of iterations for the number of boosting iterations (232). We set the the learning rate η to 0.1. We specify a range of 3 to 12 for the maximum depth of a tree (12), which refers to the number of splits the tree contains. We specify a range of 1 to 10 for the minimum number of observations in the terminal node (3.24). For the k-fold cross-validation we set k to 5.

Feature generation

We select 84 explanatory features to implement the XGBoost classifier. The generation of features reduces the dimensionality of the high-frequency usage data and enables us to evaluate which features help to discriminate between private and business usage customer-day observation. Note that similar features are also generated in Weisser et al. (2021). All features are calculated for for small and big load separately and some features also add average load:

1. *Usage features*:
   daily mean and daily standard deviation of electricity usage (for average load, small load and big load) [6 in total];

2. *Counting features*:
   number of hours with low usage (below the 25th percentile), number of hours with intensive usage (above the 75th percentile), number of hours with zero use (for average load, small load and big load) [9 in total];

3. *Within-day features:*:
   usage calculated at 7 time intervals of the day (early morning 5–8 am, late morning 8–11 am, noon 11 am–2pm, afternoon 2–5 pm, early evening 5–8 pm, late evening 8–11 pm, night 11 pm–5 am) (for average load, small load and big load) [21 in total];

4. *Features for usage changes over time*:

   a) Delta between the average usage of every two neighboring hours (only for small load and big load) (excluding the hours from 0 am to 4 am) [38 in total];

   b) The difference between big load and small load calculated at the 7 time intervals of the day that we outline above [7 in total];

   c) Delta of cumulative usage at prime time (8pm–11 pm) and non-prime time (11pm –8am) (for average load, small load and big load) [3 in total].

2.A.3 Additional Figures

Figure 2.A.1: Process for generating and using daily business usage probabilities.

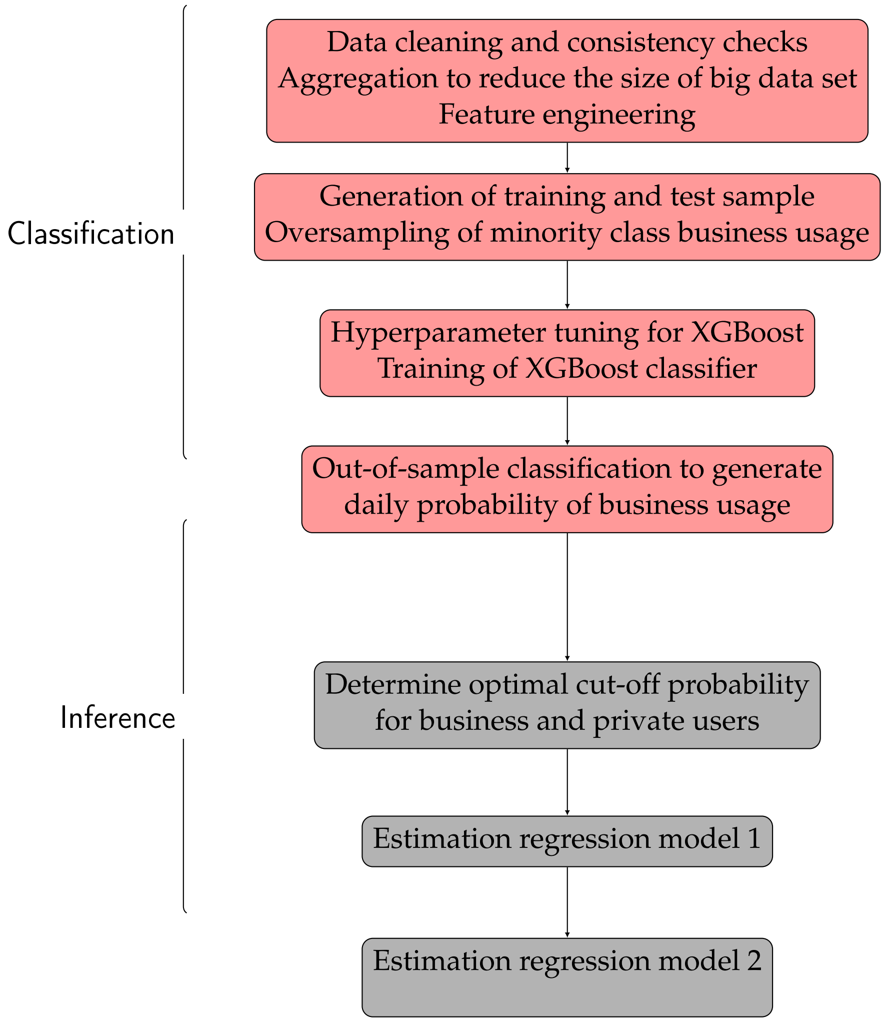
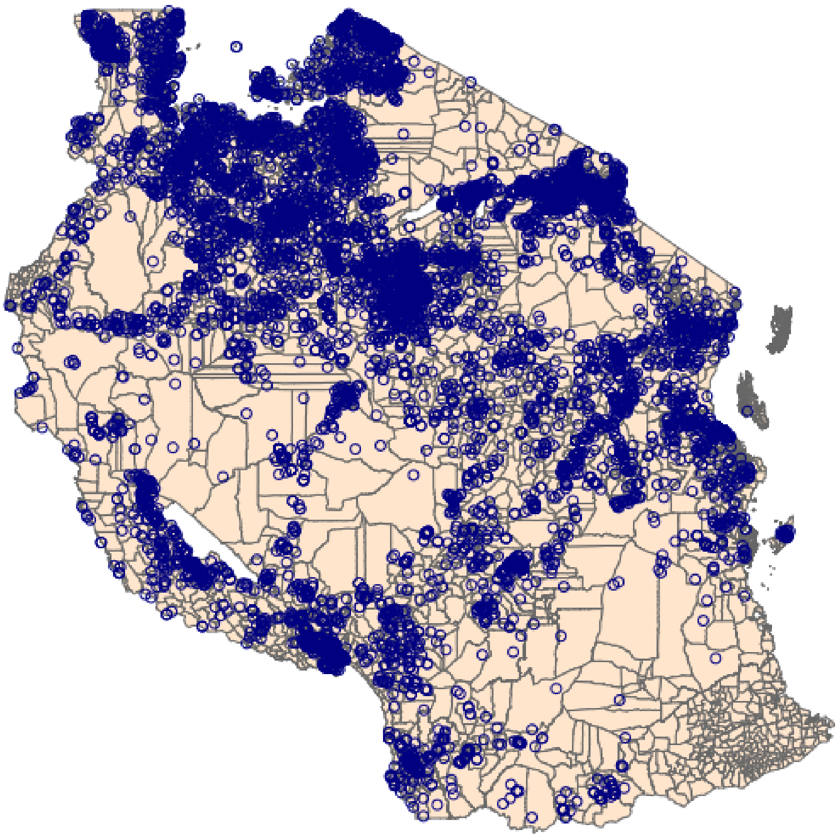
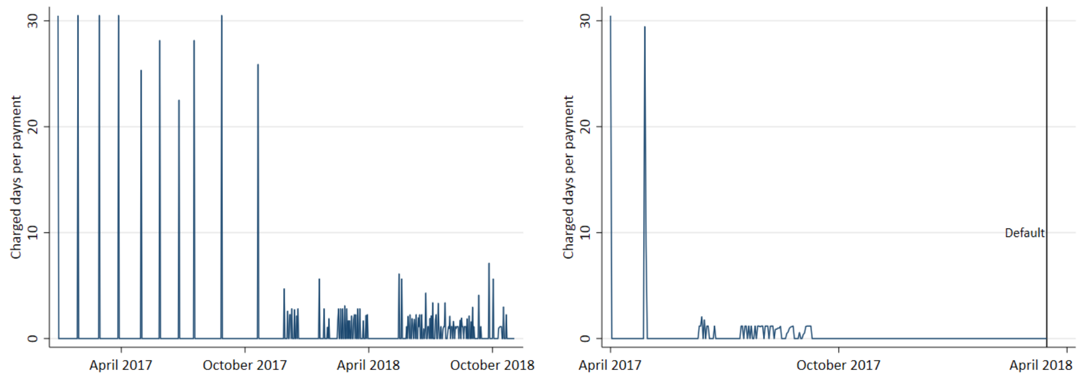


Figure 2.A.2: Customer locations



Notes: This figure displays the locations of all customers in the sample.

Figure 2.A.3: Customer repayment behavior



Notes: This figure displays the repayment schedule of two exemplary customers. Repayment differs by the paid amount per charge (measured by the number of charged days the payment is buying) and payment frequency.

Figure 2.A.4: NDVI deviation of usual NDVI in 2017

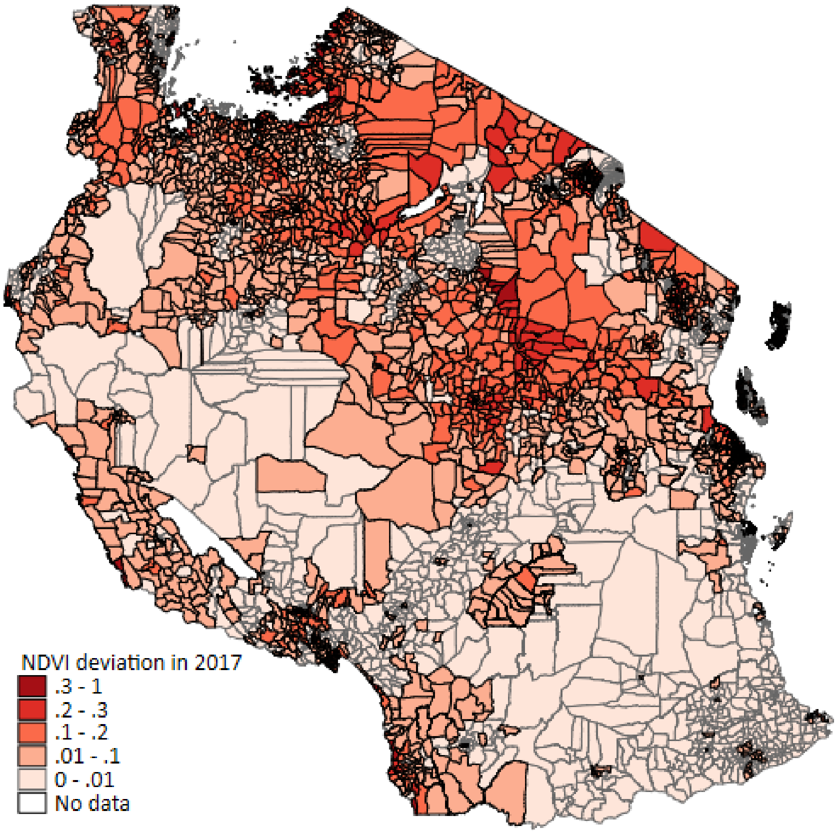
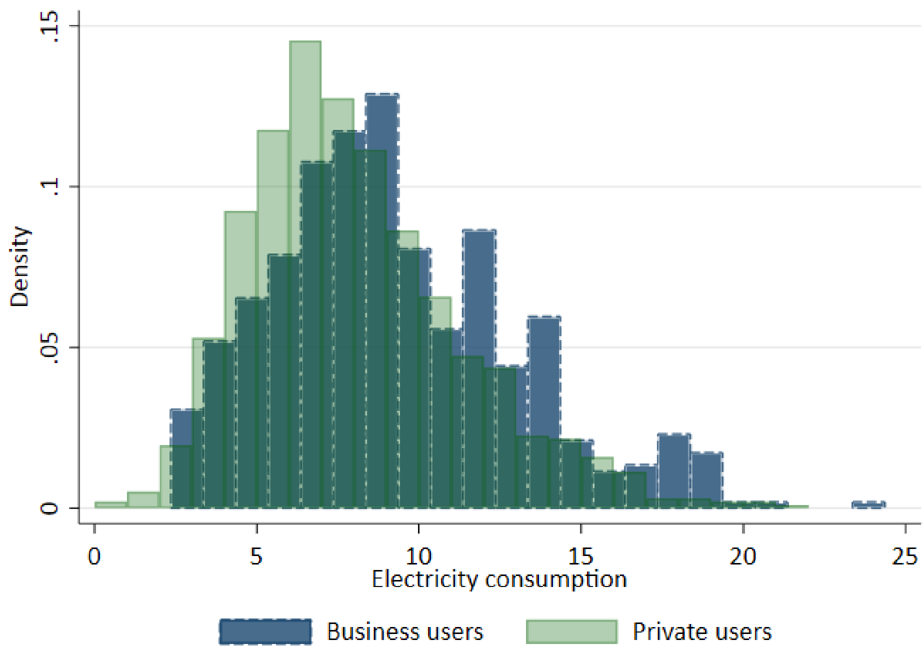
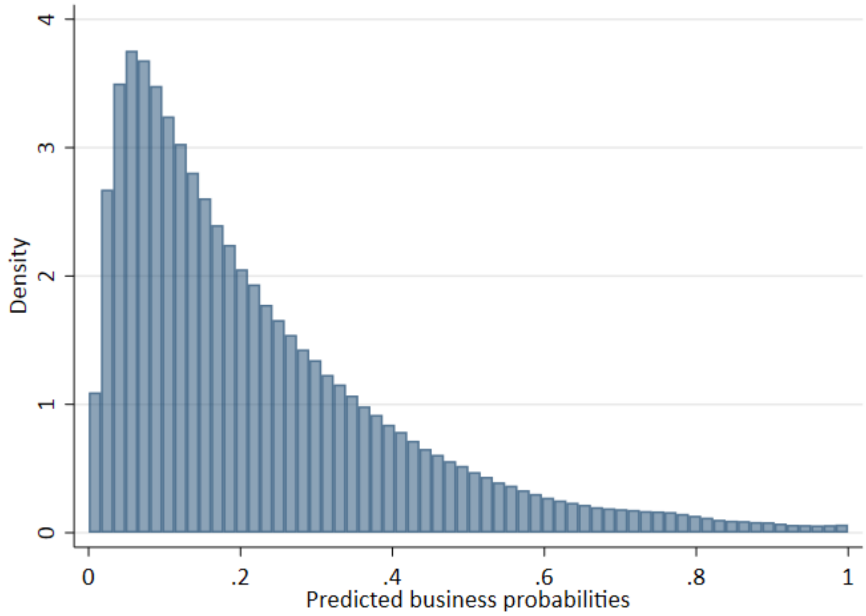


Figure 2.A.5: Electricity consumption by business and private users during the survey periods



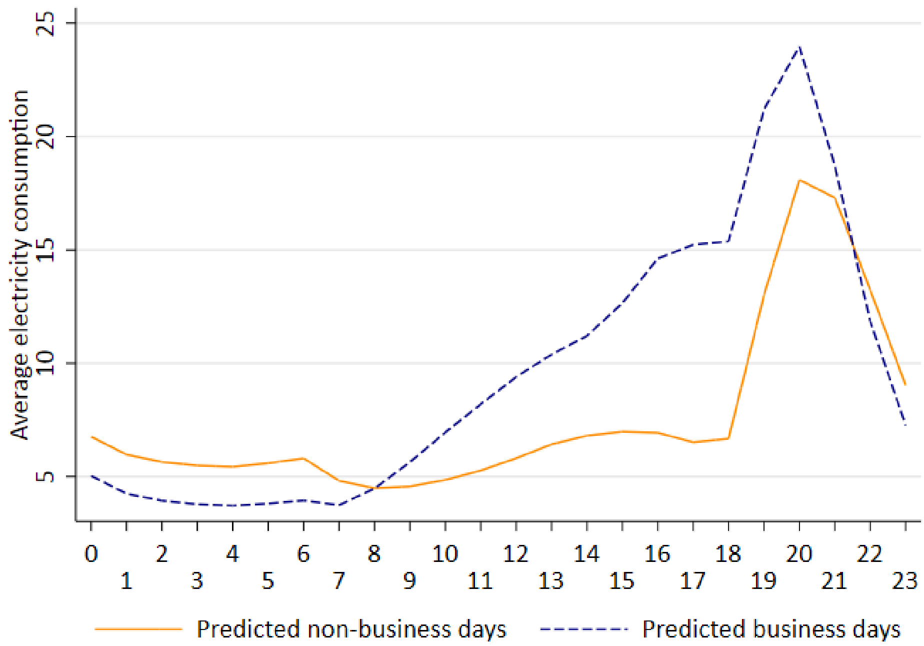
Notes: This figure presents the distribution of electricity consumption by business and private users.

Figure 2.A.6: Out-of-sample prediction of business probabilities



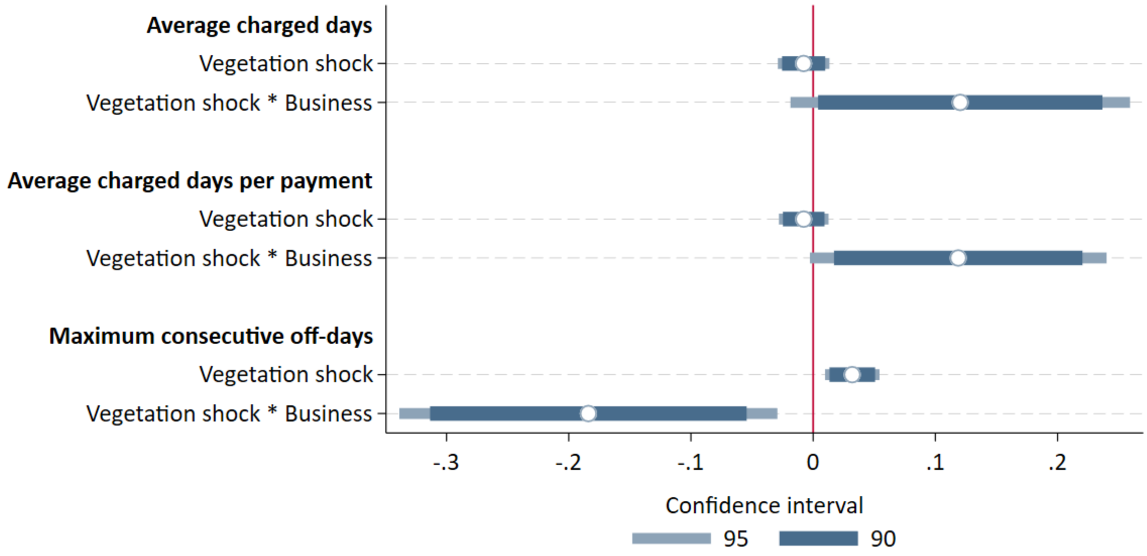
Notes: This figure presents the distribution of predicted business probabilities of the sample of 19,939 customers.

Figure 2.A.7: Electricity consumption by predicted business-days



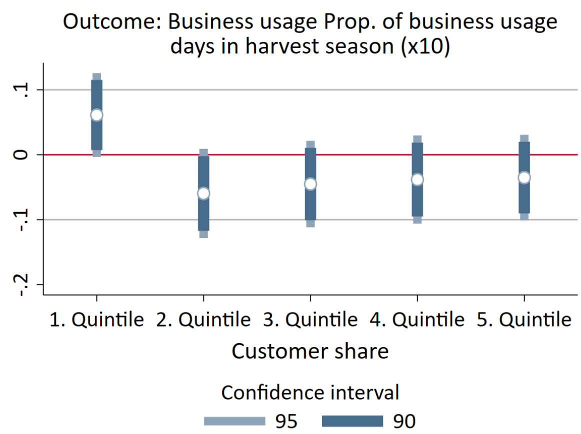
Notes: This figure displays the average electricity consumption during predicted business and non-business days for the sample of 19,939 customers.

Figure 2.A.8: Alternative repayment and default measures



Notes: This figure displays the coefficients of vegetation shocks in the growing season and the interaction between vegetation shocks in the growing season and business usage in the harvest season. The regressions are specified as in Table 2.3, with alternating dependent variables (indicated in bold on the y-axis). The dependents are the *asinh* average number of charged days, the *asinh* average charged days per payment, and the*asinh* maximum consecutive system-off days, all during the harvest season. Business usage is defined as the proportion of business days with an electricity usage that results in a predicted business probability above 75% multiplied by factor 10. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. All controls from columns 3 and 6 of Table 2.2 are included. Standard errors are clustered on the level of districts (ADM2). * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Figure 2.A.9: Differences by customer saturation



Notes: This figure displays coefficients of the vegetation shock interacted with quintiles of the distribution of the regional customer share. The outcome variable is the proportion of business days with an electricity usage that results in a predicted business probability above 75% multiplied by factor 10. All controls from columns 3 and 6 of Table 2.2 are included. Standard errors are clustered on the level of districts (ADM2).

2.A.4 Additional Tables

Table 2.A.1: Descriptives on borrower characteristics

Variable	Median	Mean	SD	Min	Max	Obs
Panel A: Sample description						
Female	0.00	0.16	0.36	0.00	1.00	15277
Household size	4.00	4.41	1.95	1.00	30.00	20021
Distance to next town (km)	46.31	49.59	36.10	0.13	192.85	15153
Mobile money agent within 5km	1.00	0.63	0.48	0.00	1.00	15153
Bank agent within 5km	0.00	0.39	0.49	0.00	1.00	15153
Has side-business	1.00	0.56	0.50	0.00	1.00	15277
Is employed	0.00	0.10	0.30	0.00	1.00	15277
Maize yield (in bags)	35.00	44.96	52.58	0.00	2500.00	10474
Spends money on farming equipment	0.00	0.39	0.49	0.00	1.00	6809
Uses irrigation	0.00	0.22	0.41	0.00	1.00	6809
Panel B: Repayment data						
Average system-off days	7.06	9.75	10.30	0.00	88.26	20040
Average charged days per payment	10.80	16.73	26.69	0.18	941.88	19858
System off at least one full day	1.00	0.64	0.48	0.00	1.00	20040
Max. consecutive days the system is turned off	2.30	7.89	10.65	0.00	30.00	20040
Customer defaulted	0.00	0.08	0.27	0.00	1.00	20040
Panel C: Usage data						
Avrg. hourly electricity consumption (in Watt)	7.42	7.70	2.69	0.38	23.02	20040
in the morning	4.19	4.78	3.02	0.00	30.75	20040
in the afternoon	6.50	7.42	4.66	0.00	40.36	20040
in the late evening	15.43	16.07	5.92	0.00	62.32	20040
in the night	6.01	6.22	2.76	0.00	26.02	20040
Panel D: Vegetation data						
Average pre-period NDVI (1995-2014)	0.31	0.31	0.05	0.07	0.51	19947
Average NDVI (2015-2018)	0.30	0.30	0.06	0.01	0.53	19947

Notes: Descriptive statistics are based on the sample of 20,040 farmers in rural areas who bought a system between 01-2015 and 11-2018 and who took part in an eligibility interview. Variables are on the customer level. Data are provided by the clean energy company.

Table 2.A.2: Descriptives on survey respondent characteristics

	Private users		Business users		Difference
Variable	Mean	SD	Mean	SD	t-Stat.
Average business probability	0.13	0.08	0.43	0.19	-16.25
Avrg. hourly electricity consumption (in Watt)	7.32	2.74	8.54	2.83	-2.65
Customer is female (based on NAM)	0.20	0.40	0.21	0.41	-0.11
customer's household size	3.45	1.40	3.50	1.34	-0.22
Maize yield	42.04	34.39	41.79	35.65	0.03
Distance to next town	57.41	42.47	46.29	34.38	1.03
Any mobile money agent within km	0.62	0.49	0.50	0.51	0.89
Any bank within km	0.27	0.45	0.22	0.43	0.44
Is an employee	0.21	0.41	0.15	0.36	0.94
Any animal farming	0.38	0.49	0.33	0.48	0.60
Has a business	0.59	0.49	0.65	0.48	-0.74
Is an employee	0.21	0.41	0.15	0.36	0.94
Usually spends money on equipment	0.36	0.48	0.27	0.47	0.54
Farmer irrigates	0.25	0.43	0.09	0.30	1.12
System off at least one full day	0.47	0.50	0.50	0.51	-0.40

Notes: This table presents differences of customer characteristics between 232 labelled private and business users of the small survey.

Table 2.A.3: Descriptives on small survey sample vs. main sample

	Whole sample		Small sample		Difference
Variable	Mean	SD	Mean	SD	t-Stat.
Panel A: Sample description					
Female	0.16	0.36	0.20	0.40	-1.93
Household size	4.42	1.96	3.46	1.38	7.41
Distance to next town (km)	49.57	36.07	55.21	41.07	-1.49
Mobile money agent within 5km	0.63	0.48	0.59	0.49	0.75
Bank agent within 5km	0.39	0.49	0.26	0.44	2.40
Has side-business	0.56	0.50	0.60	0.49	-1.06
Is employed	0.10	0.30	0.19	0.40	-4.49
Maize yield (in bags)	44.95	52.71	41.98	34.51	0.58
Spends money on farming equipment	0.40	0.49	0.34	0.48	0.84
Uses irrigation	0.22	0.41	0.22	0.42	-0.00
Panel B: Repayment data:					
Average system-off days	9.78	10.33	5.51	5.21	6.28
Average charged days per payment	16.78	26.74	10.82	9.96	3.40
System off at least one full day	0.64	0.48	0.47	0.50	5.28
Max. consecutive days the system is turned off	7.93	10.68	2.61	5.01	7.58
Customer defaulted	0.08	0.27	0.00	0.00	4.46
Panel C: Usage data:					
Average business probability	0.22	0.14	0.19	0.17	2.65
Avrg. hourly electricity consumption (in Watt)	7.69	2.70	7.58	2.80	0.57
in the morning	4.78	3.02	4.66	3.04	0.59
in the afternoon	7.42	4.67	7.28	4.99	0.44
in the late evening	16.06	5.94	15.98	6.11	0.20
in the night	6.21	2.77	6.19	2.76	0.09
Panel D: Vegetation data:					
Average pre-period NDVI (1995-2014)	0.31	0.05	0.31	0.05	-0.15
Average NDVI (2015-2018)	0.30	0.06	0.30	0.05	-0.75

Notes: This table presents differences of the main variables between 232 small-survey and 19,900 main sample customers (excluding those in the small-survey sample). Data are provided by the clean energy company.

Table 2.A.4: Agricultural shocks and electricity consumption

Outcome	*asinh* Average electricity consumption in harvest season		Proportion of business usage days in harvest season
	(1)	(2)	(3)
Vegetation shock in growing season	-0.006	-0.005	0.002**
	(0.003)	(0.003)	(0.001)
Average electricity consumption			-0.001
			(0.007)
Observations	43207	43207	43207
Customer FE	✓	✓	✓
ADM2 × year FE	✓	✓	✓
Controls		✓	✓

Notes: Business days are defined as days with an electricity usage that results in a predicted business probability above 75%. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. ADM2-year fixed effects are set at the district level. Standard errors, in parentheses, are clustered on the level of districts (ADM2). Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 2.A.5: The role of electricity use for repayment

Outcome:	*asinh* System-off days in harvest season		
	(1)	(2)	(3)
Vegetation shock in growing season	0.029**	0.043***	-0.027
	(0.014)	(0.015)	(0.055)
Business usage		-1.946***	-1.950***
		(0.115)	(0.115)
Average electricity consumption	-0.187***	-0.187***	-0.187***
	(0.036)	(0.033)	(0.033)
Business usage × vegetation shock		-0.271***	-0.289***
		(0.076)	(0.076)
Electricity consumption × vegetation shock			0.027
			(0.020)
Observations	43207	43207	43207
Customer FE	✓	✓	✓
ADM2 × year FE	✓	✓	✓
Controls	✓	✓	✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. Average electricity usage is the asinh of the mean electricity usage during the season. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. The proportion of business and high usage days, and the number of system-off days are measured during the first harvest season that follows the growing season. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 2.A.6: Agricultural shocks and repayment: Differences by individual characteristics

Outcome:	*asinh* System-off days in harvest season					
	(1)	(2)	(3)	(4)	(5)	(6)
Vegetation shock in growing season	0.023	0.051*	0.029	0.025	0.034*	0.049*
	(0.019)	(0.029)	(0.020)	(0.020)	(0.019)	(0.028)
Business usage	-1.678***	-1.986***	-1.677***	-1.678***	-1.677***	-1.831***
	(0.152)	(0.126)	(0.152)	(0.152)	(0.152)	(0.181)
Business usage × vegetation shock	-0.347***	-0.261***	-0.350***	-0.348***	-0.351***	-0.432***
	(0.103)	(0.076)	(0.104)	(0.105)	(0.104)	(0.128)
Observations	30059	43166	30059	30059	30059	18053
Customer FE	✓	✓	✓	✓	✓	✓
ADM2 × year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓
Shock × customer female	✓					
Shock × household size		✓				
Shock × animal-farmer			✓			
Shock × side-business				✓		
Shock × wage-employed					✓	
Shock × maize yield						✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. The proportion of business days and the number of system-off days are measured during the first harvest season that follows the growing season. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.A.7: Agricultural shocks, business usage and repayment: Locational characteristics

Outcome:	*asinh* System off-days in harvest season			
	(1)	(2)	(3)	(4)
Vegetation shock in growing season	0.013 (0.026)	0.037 (0.024)	0.045 (0.028)	0.069 (0.043)
Business usage	-2.092*** (0.148)	-2.092*** (0.148)	-2.092*** (0.148)	-2.092*** (0.148)
Business usage × vegetation shock	-0.294*** (0.079)	-0.297*** (0.079)	-0.296*** (0.080)	-0.296*** (0.080)
Observations	34036	34036	34036	34036
Customer FE	✓	✓	✓	✓
ADM2 × year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Shock × mobile money agent 5km	✓			
Shock × bank 5km		✓		
Shock × lending group 5km			✓	
Shock × *asinh* distance next town				✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. The proportion of business days and the number of system-off days are measured during the first harvest season that follows the growing season. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * p< 0.1, ** p< 0.05, *** p< 0.01

Table 2.A.8: Robustness: Using alternative cut-offs to identify business user-days

Outcome:	Proportion of business usage days in harvest season				
Cut-off:	50%	60%	70%	80%	90%
Vegetation shock in growing season	0.002 (0.001)	0.002* (0.001)	0.002** (0.001)	0.002** (0.001)	0.001 (0.000)
Observations	43207	43207	43207	43207	43207
Customer FE	✓	✓	✓	✓	✓
ADM2 × year FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 50%, 65%, 70%, 80% and 90%, respectively. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. The proportion of business days are measured during the first harvest season that follows the growing season. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * p< 0.1, ** p< 0.05, *** p< 0.01

Table 2.A.9: Robustness: Using alternative cut-offs to identify business users

Outcome:	*asinh* System off-days in harvest season				
Business cut-off	50 percent	60 percent	70 percent	80 percent	90 percent
Vegetation shock in growing season	0.051***	0.048***	0.046***	0.041***	0.035**
	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
Business usage	-1.636***	-1.659***	-1.793***	-2.152***	-2.485***
	(0.087)	(0.102)	(0.116)	(0.134)	(0.216)
Business usage × vegetation shock	-0.170***	-0.197***	-0.258***	-0.260***	-0.310**
	(0.048)	(0.057)	(0.068)	(0.093)	(0.143)
Observations	43207	43207	43207	43207	43207
Customer FE	✓	✓	✓	✓	✓
ADM2 × year FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 50%, 60%, 70%, 80% and 90%, respectively. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. The proportion of business days and the number of system-off days are measured during the first harvest season that follows the growing season. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * p< 0.1, ** p< 0.05, *** p< 0.01

Table 2.A.10: Rainfall and Vegetation shocks shocks

Outcome:	Vegetation shocks (NDVI)	
	(1)	(2)
Rainfall shock	0.042	
	(0.044)	
Negative rainfall shock		0.071**
		(0.036)
Positive rainfall shock		0.008
		(0.031)
Average growing season cloud cover	-8.300***	-8.084***
	(1.153)	(1.149)
Observations	9493	9493
ADM3 FE	✓	✓
Year FE	✓	✓

Notes: The vegetation and rainfall shocks are standardized measures of the long-term mean deviation of the average seasonal NDVI/ rainfall from their. Larger values of the rainfall shock indicate less rain in the observed season compared to the long-term mean. Estimations are run at the ward-season level. Controls include the average ward-level cloud cover during the growing season, ward (ADM3) fixed effects and year fixed effects. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * p< 0.1, ** p< 0.05, *** p< 0.01

Table 2.A.11: The effect of rainfall shocks on repayment and business usage

Outcome	Proportion of business usage days in harvest season	*asinh* System-off days in harvest season	
	(1)	(2)	(3)
Rainfall shock in growing season	-0.001 (0.001)	0.065*** (0.019)	0.069*** (0.019)
Business usage in harvest season			-1.971*** (0.126)
Business usage × rainfall shock			-0.130** (0.057)
Observations	43207	43207	43207
Customer FE	✓	✓	✓
Year FE	✓	✓	✓
Controls	✓	✓	✓

Notes: Business days are defined as days with an electricity usage that results in a predicted business probability above 75%. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. ADM2-year fixed effects are set at the district level. Standard errors, in parentheses, are clustered on the level of districts (ADM2). Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 2.A.12: Agricultural shocks, business usage and repayment: Number of seasons

Outcome:	*asinh* System off-days in harvest season			
	One season		More than one season	
	(1)	(2)	(3)	(4)
Vegetation shock in growing season	0.019 (0.023)	0.040* (0.023)	0.017 (0.036)	0.017 (0.038)
Business usage		-1.674*** (0.153)		-1.922*** (0.344)
Business usage × vegetation shock		-0.465*** (0.111)		0.063 (0.267)
Observations	20036	20036	5488	5488
Customer FE	✓	✓	✓	✓
ADM2 × year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 2.A.13: Agricultural shocks, business usage and repayment: Varying bimodal seasons

	Excluding regions with bimodal seasons			Vegetation shocks over short and long bimodal seasons		
Outcome:	Business	Repayment		Business	Repayment	
	(1)	(2)	(3)	(4)	(5)	(6)
Vegetation shock in growing season	0.002* (0.001)	0.024 (0.015)	0.035** (0.016)	0.002* (0.001)	0.032** (0.015)	0.043** (0.017)
Business usage		-1.824*** (0.130)	-1.840*** (0.125)		-1.928*** (0.128)	-1.944*** (0.123)
Business usage × vegetation shock			-0.272*** (0.078)			-0.286*** (0.074)
Observations	37207	37207	37207	43207	43207	43207
Customer FE	✓	✓	✓	✓	✓	✓
ADM2 × year FE	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. Repayment is the number of *asinh* System-off days in the harvest season. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 2.A.14: Repayment and business usage in growing season and anticipation effects

Outcome:	Proportion of business days in growing season	*asinh* System-off days in growing season		*asinh* System-off days in harvest season
	(1)	(2)	(3)	(4)
Vegetation shock in growing season	0.001 (0.001)	0.027 (0.020)	0.020 (0.020)	0.037** (0.016)
Business usage in growing season			-1.613*** (0.194)	-0.207 (0.168)
Business usage × vegetation shock			0.194** (0.089)	-0.109 (0.115)
Observations	33663	33663	33663	32245
Customer FE	✓	✓	✓	✓
ADM2 × year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system and the average ward-level cloud cover during the respective season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

Table 2.A.15: Agricultural shocks, business usage and repayment: The role of past business usage

Outcome	Proportion of business usage days in harvest season		*asinh* System off-days in harvest season	
	(1)	(2)	(3)	(4)
Vegetation shock in growing season	0.002**	0.002**	0.046***	0.046***
	(0.001)	(0.001)	(0.014)	(0.014)
Business usage			-1.971***	-1.971***
			(0.125)	(0.125)
Business usage × vegetation shock			-0.327***	-0.321***
			(0.083)	(0.096)
Past business usage	0.049***	0.048***	0.491***	0.491***
	(0.016)	(0.016)	(0.117)	(0.117)
Past business usage × vegetation shock		0.003		-0.015
		(0.015)		(0.101)
Observations	43207	43207	43207	43207
Customer FE	✓	✓	✓	✓
ADM2 × year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. Past business usage is the same measure for the last season. The vegetation shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. The proportion of business days and the respective outcome are measured during the first harvest season that follows the growing season. Past business usage refers to the last harvest season in the sample. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * p< 0.1, ** p< 0.05, *** p< 0.01

Table 2.A.16: Non-farmer business usage and repayment

Outcome in harvest season	Proportion of business use days	*asinh* System -off days	*asinh* Average no. charged days per payment
Vegetation shock in growing season	0.000	-0.032	-0.019
	(0.003)	(0.038)	(0.030)
Observations	10555	10555	10139
Customer FE	✓	✓	✓
ADM2 × year FE	✓	✓	✓
Controls	✓	✓	✓

Notes: The sample consists of non-farmers only. Business usage measures the proportion of days during the harvest season with an electricity usage that results in a predicted business probability above 75%. The drought shock is a standardized measure of the deviation of the average seasonal NDVI level from its long-term mean. ADM2-year fixed effects are set at the district level. Further controls include the *asinh* harvest season length, the *asinh* number of weeks the customer had the solar panel system during the harvest season and the average ward-level cloud cover during the harvest season. Standard errors, in parentheses, are clustered on the level of districts (ADM2). * p< 0.1, ** p< 0.05, *** p< 0.01

Chapter 3

Your Pain, My Gain? Estimating the Trade Relocation Effects from Civil Conflict

joint work with Tobias Korn

Abstract We derive a novel estimation approach to quantify three-party relocation effects in a dyadic framework. Applied to the effects of civil conflicts on trade, we find robust evidence that importers substitute away from exporters that are engaged in conflict. This trade relocation persists after the resolution of a conflict. As a potential explanation for the longevity of this effect, we provide evidence that trade relocation increases the likelihood of the two countries signing a Preferential Trade Agreement, which persistently decreases their bilateral trade costs. A heterogeneity analysis suggests that trade relocation does not occur in the fuels sector, and that highly integrated supply chains are less likely to relocate. We derive our estimation approach from the structural gravity model of international trade, translating the triadic relationship between a conflict country and an exporter-importer pair into an estimable dyadic relationship. Our estimation approach can be adapted to either cover alternative unilateral shocks, e.g. natural disasters, or to analyze other bilateral dependent variables, e.g. migration or FDI flows.

Keywords: Conflict and trade, trade diversion, gravity estimation, general equilibrium

JEL Classification: F14, D74, N41

3.1 Introduction

In a globalized world, bilateral decisions do not happen in isolation. For instance, realized trade flows are outcomes of competition, financial flows depend on expected returns of different investment opportunities, and migration flows are shaped by distances and the attractiveness of alternative options. International trade theory subsumes such effects under the term of multilateral resistances (Anderson and van Wincoop, 2003); each importer's and each exporter's average access to all other trade partners determine the value of bilateral trade. Hence, events that change the competitiveness of any relevant third party can have significant effects on bilateral trade flows between two countries. Estimating such indirect effects requires empirically isolating meaningful events in third countries that are likely to influence the dyadic relationship of two other countries. In this paper, we develop a novel strategy to estimate the effect of third party events on bilateral outcomes and use it to estimate the trade relocation effects of civil conflicts. This estimation procedure can easily be adapted to other unilateral shocks as well as alternative bilateral outcome variables.

Violent conflicts are known to displace people and heavily interrupt national production chains (see, e.g., Blattman and Miguel, 2010; Verwimp et al., 2019). Similarly, evidence abounds that civil conflicts significantly hurt countries' exports (see, e.g., Martin et al., 2008a; Novta and Pugacheva, 2021). In this paper, we estimate the trade relocation effects of such unilateral economic disruptions by investigating how civil conflicts shift global trade networks. In essence, we analyze whether and under which circumstances importers divert their demand from a conflict country to another, peaceful country. Our main results are based on Partial Equilibrium (PE) structural gravity estimations using bilateral trade data for over 150 countries during the period from 1995 to 2014. To augment the typical dyadic gravity specification by variation from a third country, we define a "relocation propensity" variable that indicates whether a dyad is likely to be subject to trade relocation from conflict in another country. This indicator variable combines yearly information on the relationship between any conflict country and the two countries in a given dyad. A given dyad is considered as likely to be affected by trade relocation if (1) a conflict country used to be a relevant exporter for the dyad's importer, and (2) that dyad's exporter offers a variety of goods similar to that of the conflict country.

On average, we find bilateral trade to increase by up to 6% in response to civil conflict in another country. Our analysis further reveals a significant heterogeneity with respect to the traded sector. We find that trade in agricultural, mineral, and manufacturing goods exhibit a trade relocation effect of up to 13%, whereas fuel exports do not respond at all. The fact that the international fuel trade does not react to civil conflict reflects the dependence of importers on specific suppliers of fuel exports. This was very well demonstrated recently by the European Union's hesitation to stop oil and gas imports from Russia in the light of Russia's invasion of Ukraine.[1] Similarly, oil and gas exports are of special financial importance for belligerents on either side of a conflict, who have an interest in maintaining fuel shipments.[2] Second, in the agricultural and manufacturing sectors, trade relocation only occurs if the prior value chain integration via Foreign Direct Investment (FDI) was negligible. However, the effect is the opposite in the minerals sector, where large amounts of FDI are associated with higher trade relocation. This difference may be driven by the mining sector's vulnerability to civil conflict, as especially foreign-owned mines attract violence (Berman et al., 2017). A final heterogeneity is the timing of the relocation decision. We find that in the minerals and manufacturing sectors, relocation is stronger after long conflict spells. This finding is in line with recent research in the business literature. Especially Multinational Enterprises (MNEs), who incorporate the threat of political tensions in their location decision of FDI, must weigh the costs from staying versus the costs of relocating. Depending on their vulnerability to conflict and local advantages for production, resuming production in a conflict zone can be the better option (Dai et al., 2017). For some, the possibility to stay is even worth investments to promote peace (Oetzel and Miklian, 2017).

Once a firm relocates its production sites or finds new providers of (intermediary) goods in another country, it has economic incentives to lobby for better and cheaper market access. Hence, trade relocation may persist after the end of a civil conflict if trading costs remain decreased. In a recent study, Freund et al. (2021) provide case study evidence for this argument for the automotive sector in response to Japan's 2011 earthquake. In our generalized setting, we find that trade flows remain relocated for up to nine years after the end of a civil conflict. This effect is mostly driven by the manufacturing sector. As a possible channel to explain this result, we find that a civil conflict in one country makes its main importers more likely to form Preferential Trade Agreements (PTAs) with alternative exporters. This supports the intuition that the persistent relocation is fostered by deeper market integration, and follows the idea of endogenous RTA formation (see, e.g Egger et al., 2008). In the end, international markets find themselves in a new equilibrium (Allen and Donaldson, 2020). Our findings suggest that civil conflicts can harm economic development in the long-run as trade flows remain diverted away from the conflict country, underlining the view of civil

[1] https://www.economist.com/the-economist-explains/2022/02/26/if-the-supply-of-russian-gas-to-europe-were-cut-off-could-lng-plug-the-gap (last accessed March 23, 2022).

[2] See https://www.economist.com/middle-east-and-africa/2014/11/01/a-sticky-problem as an example (last accessed February 15, 2022).

conflict as "development in reverse" (Collier et al., 2003).

Finally, we conduct a General Equilibrium (GE) analysis based on the recent civil war episodes in Colombia, Ukraine, and Turkey as case studies.[3] These case studies confirm our PE findings and indicate that importers who used to rely heavily on shipments from the conflict countries switch to shipments from alternative exporters. What is more, we estimate changes in overall national welfare measured by total consumption expenditures in response to these conflicts. Here, we find that national welfare decreases for almost all countries involved, even for those exporters on the receiving end of the relocated trade flows. This suggests that trade relocation cannot fully offset the global loss in economic activity.

Our paper contributes to several strands of the literature. First, we contribute to the literature investigating trade relocation effects. Since Anderson and van Wincoop (2003) pointed out the importance of multilateral resistance terms in the structural gravity framework, it is widely accepted that international competition is a decisive determinant of bilateral trade. The empirical trade literature provides various insights into the trade relocation effects of PTAs. Several papers provide evidence that PTAs increase trade flows between signees ("trade creation") while decreasing trade between any signee and non-signees ("trade diversion"). Among others, Dai et al. (2014) and Cheong et al. (2015) analyze how PTAs shift international trade flows by focusing on dyads in which one country joined a PTA and the other did not. We go one step further and measure trade relocation in a *triadic* relationship. That is, we estimate the effect of country A's economic shock on bilateral trade between countries B and C. The empirical specification we develop allows to include unilateral shocks that occur outside an observed dyad. While we apply this strategy to civil conflict as a shock and bilateral trade as an outcome variable, the same specification can be applied to alternative bilateral dependent variables like migration or financial flows, as well as to different unilateral shocks such as climate shocks (Jones and Olken, 2010), resource windfalls (Bahar and Santos, 2018), taxes and regulations (Grubert and Mutti, 1991; Emran, 2005), or currency devaluations (Krugman and Taylor, 1978; Rose, 2018).

Second, we add to the evidence of how civil wars affect the international economy. Recent findings emphasize that civil wars depress the quantity and prices of exported goods (Ksoll et al., 2018; Ahsan and Iqbal, 2020). These effects are not bound to the conflict country but often spill over to neighboring countries (Qureshi, 2013; De Sousa et al., 2018). Especially in the case of transnational terrorism, protective countermeasures persistently complicate the exchange of goods, multiplying the direct effects of

[3]We selected these case studies as they constitute the most significant spikes in violence according to UCDP data which have clear start and/or end points during our period of observation.

violence (Mirza and Verdier, 2014). Similarly, international wars as well as non-violent disputes between countries reduce bilateral trade (Fuchs and Klann, 2013; Garfinkel et al., 2020a). These effects further persist when conflict erodes trust between parties (Rohner et al., 2013). However, improved trade relationships can decrease the likelihood that international wars break out as gains from trade increase the opportunity-costs of starting a war (Martin et al., 2008a,b, 2012; Garfinkel et al., 2020b). Trade restrictions and competition can even foster political violence (Amodio et al., 2020). We extend this line of the literature by considering the general equilibrium effects of civil conflict. As international markets are tightly linked, civil conflicts are hardly a unilateral or bilateral phenomenon. By providing evidence that civil wars can affect trade flows between other, peaceful countries and provoke shifts in the international equilibrium, we consider new economic consequences from political violence.

Finally, our findings add to the discussion about the persistence of the economic consequences of civil violence. According to economic theory, an economic shock should affect nations only in the short-run, while their economy rapidly recovers after the conflict is resolved (Barro and Sala-i-Martin, 1992; Mankiw et al., 1992; Blattman, 2012). These theoretical considerations receive support from several empirical findings (see, e.g., Davis and Weinstein, 2002; Brakman et al., 2004; Miguel and Roland, 2011). However, recent micro-level evidence points toward a persistent effect of civil conflict on affected individuals (Akresh et al., 2012; Justino and Verwimp, 2013; Brück et al., 2019; Tur-Prats and Valencia Caicedo, 2020). We contribute to this literature by pointing out that general equilibrium effects can cause the effects of civil conflict to persist. Our findings suggest that temporary trade relocation fosters market integration via PTAs, which in turn leads to persistent trade diversion away from the (former) conflict country.

The rest of the paper is structured as follows. In the next section, we derive our empirical specifications from the structural gravity model of international trade and introduce our dataset. Afterwards, section 3 discusses our main results. Section 4 presents several extensions to our main estimations. Finally, we will discuss a number of robustness checks in section 5, before section 6 concludes.

3.2 Estimation and Data

Our analysis follows the structural gravity model of international trade derived in Anderson and van Wincoop (2003) and Anderson (1979), based on Armington (1969). We follow Anderson et al. (2018b) and describe the exports of a variety of goods in sector s from country i to country j in year t with the equation:

$$X_{ijs,t} = \frac{Y_{is,t}E_{js,t}}{Y_{Ws,t}} \cdot \left[\frac{t_{ij,t}}{\Pi_{is,t}P_{js,t}}\right]^{1-\sigma} \tag{3.2.1}$$

Exports $X_{ijs,t}$ are positively related to the product of the exporter's level of production $Y_{is,t}$ and the importer's consumption expenditures $E_{js,t}$, relative to total world output $Y_{Ws,t}$ in that sector. Trade flows further depend on the bilateral "iceberg costs" of trade, denoted by $t_{ij,t}$. This term covers, among other things, the distance between two countries or the amount of tariffs paid on shipments. With the elasticity of substitution across varieties $\sigma > 1$, bilateral exports $X_{ijs,t}$ are negatively linked to the trade costs $t_{ij,t}$. Finally, bilateral trade depends on the multilateral resistances faced by the exporter and importer, respectively. The outward multilateral resistance Π_{is} describes the exporter's average (inverse) market access to all potential importers. The inward multilateral resistance P_{js} similarly describes the importer's average (inverse) market access to all potential exporters. Both these variables can be thought of as the competition on international markets in sector s that either i or j face with any other country to trade with country j or i, respectively.

We follow Head and Mayer (2014) in transforming equation 3.2.1 into an estimating equation and include exporter-sector-year fixed effects $\pi_{is,t}$, importer-sector-year fixed effects $\lambda_{js,t}$, and sector-dyad fixed effects μ_{ijs}. We further decompose the iceberg trade costs into a time-varying and a time-invariant component: $t_{ij,t} = \bar{t}_{ij} + \tau_{ij,t}$. Conditional on the fixed effects, the time-varying component $\tau_{ij,t}$ is the only remaining variation in equation 3.2.1 that affects bilateral trade between countries i and j. We control for bilateral trade agreements and sanctions, which are two important components of $\tau_{ij,t}$ as recently advocated in the literature (see, e.g., Dai et al., 2014; Felbermayr et al., 2019b). Finally, we add an indicator $TR_{ijs,t}$ to identify dyad-sector-year observations that are likely to be affected by trade relocation effects. We arrive at the estimating equation:

$$X_{ijs,t} = exp\left[\pi_{is,t} + \lambda_{js,t} + \mu_{ijs} + \beta \cdot TR_{ijs,t} + \gamma \cdot Z_{ij,t}\right] + \eta_{ijs,t} \tag{3.2.2}$$

where $Z_{ij,t}$ is a vector of dyadic control variables, including bilateral trade agreements and sanctions, and $\eta_{ijs,t}$ accounts for the remaining variation in $X_{ijs,t}$ that is not explained by the fixed effects and independent variables. The challenge is to incorporate civil conflicts that take place in a third country k in the variable $TR_{ijs,t}$. In theory, we expect a civil war in country k to enter bilateral trade between countries i and j via the multilateral resistance terms P_{js} or Π_{is} in Equation 3.2.1. Since multilateral resistance terms are not observable, we develop a quantifiable proxy measure for the relocation propensity, which can be derived via $P_{js,t}$ or $\Pi_{is,t}$ based on the structural gravity model of international trade (see Appendix 3.A.4 for the formal derivation). We arrive at the following estimation equation for the trade relocation effect from conflict in country k on the dyad ijs:

$$X_{ijs,t} = exp\left[\pi_{is,t} + \lambda_{js,t} + \mu_{ijs} + \beta \cdot \sum_k \left(R_{jks,t-2} \times S_{ik,t-2} \times C_{k,t-1}\right) + \gamma \cdot Z_{ij,t}\right] + \eta_{ijs,t}. \tag{3.2.3}$$

Our coefficient of interest β indicates how bilateral trade between countries i and j in sector s reacts to conflict in another country k. We approximate relocation propensity as an interaction of three variables: (i) the conflict status of any country $k \neq i, j$, denoted by $C_{k,t-1}$, (ii) the relevance of country k as an exporter for country j in sector s, $R_{jks,t-2}$, and (iii) the similarity between exporters i and k, $S_{ik,t-2}$. Note that we lag conflict by one year and the relevance and similarity conditions by two years to (i) leave time for trade relocation effects to materialize and (ii) use country characteristics before the conflict in country k.[4]

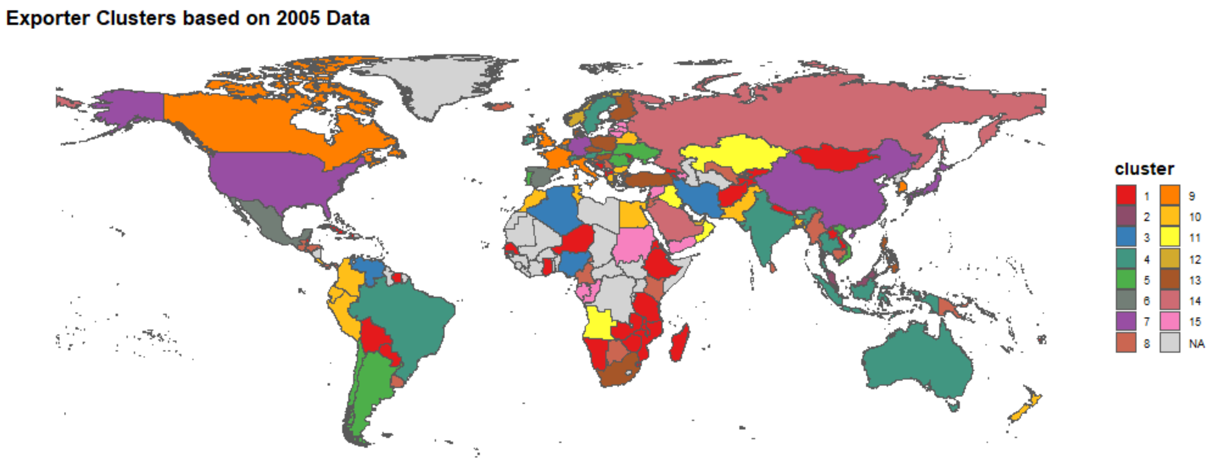
The relevance characteristic $R_{jks,t-2}$ indicates whether country j used to import relatively large amounts from country k in sector s prior to the conflict. We start by defining $R_{jks,t-2}$ broadly, indicating whether country k was among the top 7 exporters to country j in sector s.[5] Other measures, for instance top 5 or 10 exporters, are used as robustness tests. For an indicator of similarity $S_{ik,t-2}$, we leverage different variables to identify whether two countries i and k were exporters of similar goods before the conflict broke out in country k. All variables are based on disaggregated export data for 61 sectors (SITC classification). First, we construct clusters of countries with similar export structures. We apply a K-Means clustering algorithm as developed by Hartigan and Wong (1979), which allocates countries according to their similarity in production to a pre-defined number of clusters. For our preferred specifications, we divide all exporting countries in a given year into 15 or 20 different clusters. Our method is similar to the process applied by Kim et al. (2020), who assign trade-dyads to clusters according to the similarity in the sectoral composition of their trade flows. As this method of assigning export clusters inherits some degree of randomness, and there is no clear candidate for a "perfect" number of clusters, we test for robustness across different cluster sizes. Figure 3.1 below illustrates the allocation of clusters for the year 2005, where the indicator of similarity equals 1 for countries in the same cluster. To test whether our similarity measure is robust to other specifications, we additionally construct an export similarity index following Benedictis and Tajoli (2007a,b), which

[4] As a robustness test, we also examine trade flows in the same year as the conflict takes place.

[5] In our sample, the top 7 exporters are on average responsible for the first quartile (25%) of a country's imports.

mirrors the correlation between sectoral exports across countries.

Figure 3.1: Export similarity clusters



Notes: This graph shows the distribution of export clusters based on a K-Means clustering algorithm with 15 random centers in the year 2005. The algorithm randomly picks 15 countries as centers and then assigns all remaining countries to the center with the most similar export structure. Similarity is computed based on export-data for 61 sectors.

We then determine relocation propensity via the triple interaction of the indicators for relevance $R_{jks,t-2}$, similarity $S_{ik,t-2}$, and conflict $C_{k,t-1}$. In other words, whenever *all* three indicator variables take the value of one for *any* country k, our relocation propensity variable takes the value of one for the ijs dyad.[6] We hence identify a positive relocation propensity as the specific case that conflict country k was a significant trading partner of importing country j in sector s in the past, *and* this same country k used to offer a similar variety of goods to exporter i.

A causal interpretation of our results requires the unexplained variation captured by the error term $\eta_{ijs,t}$ to be uncorrelated with our relocation propensity variable, conditional on our control variables and fixed effects. We hence must rule out that unobserved, non-random characteristics captured by $\eta_{ijs,t}$ are associated with a higher likelihood that our relocation propensity variable takes the value of one. For this, it is important to note that none of the three ingredients to our relocation propensity variable is dyad-year-specific. First, the incidence of civil conflict in country k, $C_{k,t-1}$, is an event observed by all dyads in a given year t and hence controlled for by year fixed effects. Second, the relevance characteristic $R_{jks,t-2}$ is specific to a dyad's importer only and hence does not vary across an importer's export partners in a given year. Hence, importer-sector-year fixed effects account for all characteristics that make

[6]Note that this constitutes the extensive margin, coding a dyad as subject to trade relocation if they are affected by at least one conflict. In the Appendix, we provide results for the intensive margin, using the number of identified relocation possibilities as the explanatory variable. While the results are very similar, we prefer the extensive margin due to the easier interpretation of the results.

an importer more likely to experience trade relocation from a given conflict country k. The same argumentation holds for the similarity condition $S_{ik,t-2}$ at the exporter side, which is controlled for by exporter-sector-year fixed effects. Finally, characteristics that are specific to a given dyad and might increase its average propensity that both $R_{jks,t-2}$ *and* $S_{ik,t-2}$ are one is accounted for by dyad fixed effects. A potential bias in our estimates hence requires the presence of unobserved characteristics that vary at the dyad-sector-year level and correlate with the interaction of our relevance and similarity conditions. One potential caveat could be, for example, that our results are mainly driven by one of the two variables, while the other only generates minimal identifying variation. In Appendix 3.A.7, we therefore provide an in-depth discussion of the determinants and variation of both the relevance and similarity conditions and demonstrate that both variables exhibit sufficient variation. Furthermore, we provide several robustness checks below which demonstrate that both conditions are required together to estimate a significant relocation effect. Another caveat could be that during years when a conflict is active in a country that is relevant to a dyad's importer and similar to its exporter, the dyad's preferences for trading with each other systematically increase for reasons other than the civil conflict in the third country. One such possibility could be that importers apply bilateral sanctions to countries that are linked to the conflict country k, but strategically spare countries they identified as potential export substitutions for k. Here, it is reassuring that controlling for various types of sanctions leaves our results qualitatively unchanged. Our results are further not sensitive to controlling for pre-existing observable trade preferences in the form of PTAs. The non-sensitivity of our results to the inclusion of these bilateral, time-varying control variables makes us confident that the likelihood that unobserved characteristics are correlated with our trade relocation variable is low. Finally, we view reverse causation as an unlikely threat to our identification. Reverse causation would require that bilateral trade flows between two countries are significantly linked to the likelihood that a civil conflict emerged in another country that is relevant to the dyad's importer and similar to its exporter two years prior. While there is evidence that the US staged coups to increase trade with conflict countries (Berger et al., 2013), we are not aware of any evidence or anecdotes that governments stage civil wars in third countries to increase exports to or imports from a specific other, non-conflict country.

Our empirical analysis draws from various data sources related to civil conflict and international trade. For our main analysis, we include trade data for the manufacturing and primary sectors. Addressing the primary sector separately is important as civil conflicts predominantly erupt in resource-abundant countries (Ross, 2015).[7]

[7]The relationship between natural-resource abundance and the likelihood of conflict depends on several factors, such as political stability, inequality, or type of resources (Basedau and Lay, 2009;

Manufacturing data come from the Comtrade dataset, which includes bilateral trade flows between 1980 and 2018 of approximately 180 countries.[8] As a measure for trade in primary goods, we use commodity trade data from CEPII's BACI dataset, which consists of yearly bilateral trade-flows at the 6-digit HS level. According to recent advancements in the international trade literature, bilateral trade flows alone are not sufficient for a reliable empirical analysis. As Yotov (2021) shows, international trade flows need to be complemented with intra-national trade data to obtain unbiased and consistent estimates within the gravity framework. Unfortunately, the availability of consistent internal production data is still limited. Therefore, we combine several data sources to maximize the coverage across countries, sectors, and years. We follow the literature in computing internal trade flows (Baier et al., 2019). For the manufacturing sector, we compute internal trade as the difference between total manufacturing production and total manufacturing exports. To quantify total manufacturing production, we draw on data from the INDSTAT database. We proceed similarly to compute internal trade flows in the primary sector. Here, we use commodity production data from Fally and Sayre (2018). The authors combine production data of minerals, agricultural commodities and fuels from the British Geological Survey, the FAO and the Global Trade Analysis Project (GTAP). Based on these data, we compute internal trade flows for about 200 countries and across 169 commodities between 1995 and 2014. We complement our dataset with information on PTAs from CEPII's Gravity database.[9]

To identify civil conflict, we use the UCDP/PRIO Armed Conflict Dataset version 19.1 (Sundberg and Melander, 2013). We follow the established definition and code a country to experience a civil war in a given year if it has experienced violent events between government troops and a non-governmental entity, and if the number of battle deaths exceeded the threshold of 25 casualties. Our main dataset comprises 179 countries over the years 1995–2014. Table 3.A.1 reports descriptive statistics of our main variables.

3.3 Main Results

Table 3.1 presents our main results, which are obtained by estimating equation 3.2.3. Panel A provides results based on 15 clusters for the similarity definition and Panel B on 20 clusters. The results in column 1 are based on estimations across all sectors within a respective dyad, which include exporter-sector-year, importer-sector-year and

Brunnschweiler and Bulte, 2009; Bazzi and Blattman, 2014; Farzanegan et al., 2018).

[8]Trade values are primarily measured through imports, as these are usually more precisely computed. We complement missing import data with exports between the same dyad and year to maximize coverage.

[9]The PTA variable is based on the RTA-IS dataset of the World Trade Organization (WTO) and is constructed out of information on Partial Scope Agreements (PSA), Free Trade Agreements (FTA), Customs Unions (CU) and Economic Integration Agreements (EIA).

exporter-importer-sector fixed effects and control for bilateral trade agreements and sanctions on the exporter side. We find a statistically significant and positive trade relocation effect. On average, civil conflict increases trade between two other countries by 6%.[10] In columns 2-5, we investigate relocation effects by sector. Notably, we do not find any evidence of trade relocation in the fuels sector (Column 4). This finding is intuitive as fuel exports commonly do not decrease during civil conflict in the first place. As we discuss in more detail in Appendix 3.A.6, we find that civil conflicts depress exports in all sectors but in fuels, which confirms prior empirical and anecdotal evidence that warring parties have a joint interest of keeping up oil exports to finance their war efforts (Bazzi and Blattman, 2014). In the three other sectors, civil conflict in country k has a robust and significant effect on exports from country i to j, increasing bilateral shipments between 7% (manufacturing) and 13% (mining and agriculture).

Table 3.1: Trade relocation main results - conflict in top 7 trading partner countries

Dependent:	Exports from country i to country j				
	Pooled	Agricult.	Minerals	Fuels	Manufact.
	(1)	(2)	(3)	(4)	(5)
Panel A:	15 clusters				
Conflict in country k	0.06***	0.12***	0.13***	-0.02	0.07**
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Panel B:	20 clusters				
Conflict in country k	0.05**	0.13***	0.10***	-0.02	0.07**
	(0.02)	(0.04)	(0.03)	(0.03)	(0.03)
Observations	1269742	366662	322631	89446	491003
Exporter × sector × year	✓	✓	✓	✓	✓
Importer × sector × year	✓	✓	✓	✓	✓
Exp. × Imp. × sector	✓	✓	✓	✓	✓
PTA	✓	✓	✓	✓	✓
Sanctions	✓	✓	✓	✓	✓

Notes: This table reports estimates of the effects of conflict in country k on exports from country i to country j, pooled over all sectors in column 1 and disaggregated by sectors in columns 2-5. The explanatory variables take a value of 1 if (i) country k had a conflict in the previous year, (ii) country k was a top-7 exporter for country j in the pre-conflict-year and (iii) country k and country i were similar exporters in the pre-conflict-year. Similarity is measured by being in the same exporter-cluster, with a total number of 15 clusters in Panel A and 20 clusters in Panel B. We estimate all specifications with the PPML estimator and include exporter-sector-year, importer-sector-year and exporter-importer-sector fixed effects. Standard errors in parentheses are clustered at the dyad-sector level, * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

[10] The results are basically identical if we do not control for PTAs or sanctions (not shown for brevity).

We conclude from these findings that outside the fuels sector, civil conflicts provoke sizeable trade relocation effects. Next, we investigate how long these effects persist. Do trade flows return back to normal when the conflict is resolved?

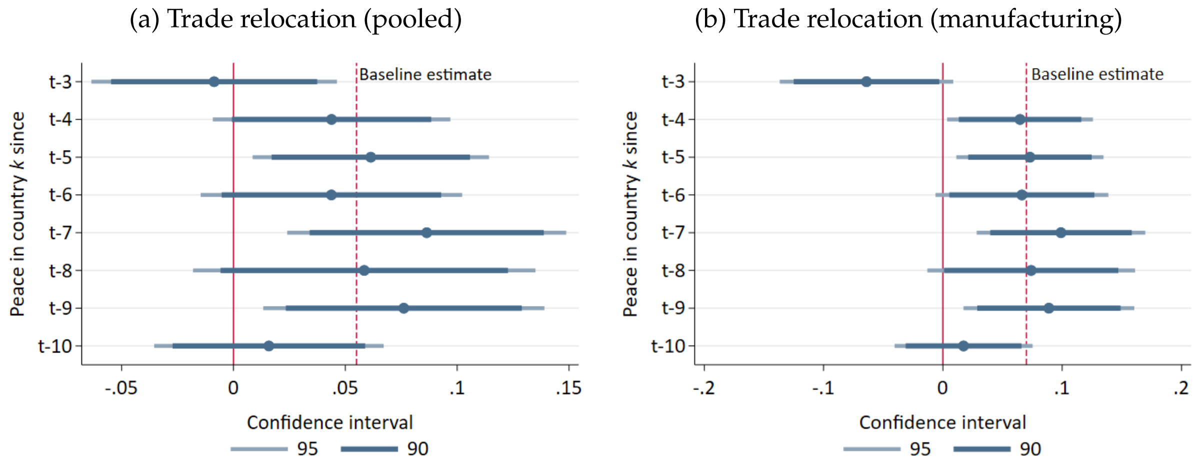
Considering the various micro-economic mechanisms that play out when trade flows relocate from one country to another, there is reason to expect that a temporary relocation can become persistent. As soon as retailers or producers of country j start importing their goods from country i instead of country k, they establish new connections and trade networks with exporting firms in country i. Companies in countries i and j integrate their supply chains and establish international branches via FDI. These newly established connections may, in turn, induce national governments to sign new trade agreements with each other. This re-drawing of international cooperation may persistently decrease bilateral trade costs. According to the dynamic equilibrium theory, a one-time shock can hence alter allocations and bilateral preferences such that economies end up converging to a new long-run equilibrium (Allen and Donaldson, 2020). In our case, this means that new supply chains and trade agreements tend to stay in place when a conflict ends, and trade relationships are unlikely to return to pre-conflict levels once country k resolves its conflict. Such a restructuring of international trade flows can hence exacerbate the conflict trap by pushing countries into the fringe of international trade, which is one explanation why conflict-ridden countries lack economic development in the long-run (Collier et al., 2003).

We analyze relocation persistence by estimating specifications similar to those presented in equation 3.2.3. Instead of an indicator variable for country k being at war, we code how many years back exporter k's civil war ended. Moreover, to consistently define the similarity and relevance conditions over time, we use the values from the year prior to conflict onset in country k. Only for cases with very long conflict spells, we use the values from five years before the conflict ended, as going too far back would mask changes in countries' production structures that are unrelated to conflict.[11] We depict our results in Figure 3.2.

Panel (a) displays estimates for shipments from the beneficiary exporter i to importer j considering all sectors. The specifications include the same fixed effects and control variables as our main results in Table 3.1. We find that still up to nine years after the end of a conflict, bilateral shipments from i to j are significantly bigger than before the conflict. This effect is mainly driven by the manufacturing sector. Whereas the confidence intervals for the pooled sample are rather wide and do not always exclude zero, the persistence estimates for the manufacturing sector only, displayed in Panel (b), hint at a statistically significant increased bilateral trade value of almost one percent

[11] Take as an example Bangladesh, which was at conflict in most years during the 1990s and early 2000s, but at the same time underwent a period of large industrialization and globalization. Using its pre-conflict production portfolio to identify similar exporters in 2010 or later would likely not give a realistic picture.

Figure 3.2: Relocation persistence



Notes: This figure shows the coefficients and confidence intervals from regressing exports from country i to country j to various lags of our relocation propensity variable. Results in Panel (a) stem from regressing trade in all sectors on conflict in country k that ended in year t-τ, as in column 1 of Table 3.1. Panel (b) provides similar results specifically for the manufacturing sector as in column 5 of Table 3.1. The dashed vertical lines represent the baseline estimates from columns 1 and 5 of Table 3.1, respectively. The light and dark blue lines depict 95% and 90% confidence intervals, respectively.

for most lags.[12] For $t-3$, our estimate is however insignificant and even tilts towards the negative. One possible explanation for the insignificant result in $t-3$ might be our coding of peace. For a country to be coded peaceful for $t-\tau$ years, we require the absence of violent activity in the country for τ years. Many salient conflicts follow an on-off nature, with violent attacks in some years, and no attacks in others (Walter, 2004). Such on-off conflicts will be part of our coding for rather short periods of peace of up to three years, but not for longer periods. What is more, we discuss in the next section that manufacturing trade relocates rather slowly after conflict onset. Therefore, on-off conflicts are less likely to cause trade relocation in the first place, which adds noise to our short-run persistence estimates.

Persistent trade relocation can come about as countries i and j persistently decrease their bilateral trade costs. Theoretically, during k's civil war, the two countries have an incentive to tighten their trade relationships, something we can observe via the formation of new trade agreements. We construct sector-specific relocation propensity measures as in our main analysis and run bilateral OLS regressions with an indicator variable for newly established PTAs as the dependent variable. We report the results in Table 3.2, where we multiply the coefficients by 100 to ease display. We find a significant increase in the likelihood of entering a PTA if countries experience trade relocation

[12] We do not find a statistically significant effect of relocation persistence in the other sectors (not shown). As the time frame of our data is relatively short, the occurrence of long time-spells after conflicts is however limited.

Table 3.2: Linear probability model: Forming new Prefrential Trade Agreements

Dependent:	Likelihood of PTA between country i and country j (0-100)							
	Agricult.		Minerals		Fuels		Manufact.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conflict in country k	0.23	0.23	0.36**	0.35**	-0.06	-0.06	1.15***	1.15***
	(0.17)	(0.17)	(0.18)	(0.18)	(0.21)	(0.21)	(0.25)	(0.25)
Observations	546129	546129	490379	490379	356116	356116	519129	519129
Exporter × year	✓	✓	✓	✓	✓	✓	✓	✓
Importer × year	✓	✓	✓	✓	✓	✓	✓	✓
Exp. × imp.	✓	✓	✓	✓	✓	✓	✓	✓
Controls		✓		✓		✓		✓

Notes: This table reports results from Linear Probability Models with the likelihood that country i and country j enter a trade agreement as the dependent variable. The explanatory variables is constructed as in Table 3.1, with the top-7 exporters defining relevance, and using 20 clusters to define similarity. All estimations include exporter-year, importer-year and exporter-importer fixed effects. We control for prior trade agreements and bilateral sanctions. Standard errors, in parentheses, are clustered at the exporter-importer level, * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

in the minerals or manufacturing sector. If relevant exporters of manufacturing goods suffer from civil war, the chances to sign a PTA with another, similar exporter increases by up to 1.15%, with a similar increase of around 0.36% for mineral exporters. These effects are sizeable compared to an average likelihood of given dyad trading under a PTA in a given year of around 12.5%. For the agricultural and fuels sectors, we do not find significant evidence that trade relocation fosters market integration.

3.4 Extensions

To better grasp the mechanisms that lead to (persistent) trade relocation, we consider various extensions to our baseline estimates.

First, we investigate conflict duration. In Table 3.3, we extend our main specification by an indicator variable for trade relocation propensity from a conflict that already lasted more than ten years. The results reveal a noticeable difference across sectors. For minerals and manufacturing, we see that trade relocation occurs especially after long conflict periods, while the opposite is true for the agricultural sector. We interpret this as evidence that in the former two sectors, firms try to keep their supply chains intact during short periods of violence. Only when violence persists, firms move their production facilities to other countries. This finding is further in line with our result that trade relocation only fosters market integration via PTAs in the manufacturing and minerals sectors. It is fair to assume that firms optimizing their supply chains would lobby for cheaper access to alternative trading partners before relocating their supply chains. In agriculture, shifting supply chains may well be cheaper and easier than in the other sectors.

Table 3.3: Relocation heterogeneity: Conflict duration

Dependent:	Sectoral exports from country i to country j			
	Agricult.	Minerals	Fuels	Manufact.
Conflict in country k	0.11** (0.05)	0.02 (0.03)	0.03 (0.03)	0.04 (0.03)
Conflict is longer than 10 years	0.04 (0.04)	0.17*** (0.04)	-0.26*** (0.10)	0.11** (0.05)
Observations	366662	322631	89446	491003
Exporter × year	✓	✓	✓	✓
Importer × year	✓	✓	✓	✓
Exp. × Imp.	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: This table reports PPML results from regressing bilateral exports on relocation propensity. The explanatory variable "Conflict in country k" is constructed as in Table 3.1, with the top-7 exporters defining relevance, and using 20 clusters to define similarity. We add another relocation propensity indicator for relocation from conflicts that lasted more than 10 years. If there are multiple conflict countries k, we use the shortest duration. All estimations include exporter-year, importer-year and exporter-importer fixed effects. We control for trade agreements and bilateral sanctions. Standard errors, in parentheses, are clustered at the exporter-importer level, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In Table 3.4, we analyze whether trade relocation varies conditional on the importance of country k as an FDI destination for firms from importer j. We would expect that substantial amounts of capital invested in conflict country k would reduce the incentive to switch trade partners. We define country k as an important FDI destination if it received more than 10% of importer j's total FDI prior to the civil conflict. In odd columns, we report trade relocation estimates for important FDI-destinations, while even columns focus on trade relocation away from countries without a significant share of FDI. In the agricultural and manufacturing sectors, we find the expected effect that only less relevant FDI destinations cause trade relocation. In the minerals sector, we find the opposite result; if conflict country k received significant FDI inflows from firms in country j, imports are *more likely* to relocate to another exporter i. This result might hint at the vulnerability of mining-sector FDI to civil conflict. Recent evidence suggests that natural resource mines are preferred targets of violent groups (Berman et al., 2017). The destruction of foreign-held capital together with a more insecure environment for (new) investments might hence encourage firms to divert both FDI and imports to other countries. Furthermore, investments in the mining sector are more mobile. Whereas agricultural and manufacturing FDI usually involves acquiring land and building plants, mining-FDI often focuses on mining equipment which can easily be moved across borders.

Table 3.4: Relocation heterogeneity: FDI destination

Dependent:	Sectoral exports from country i to country j							
	Agricult.		Minerals		Fuels		Manufact.	
Sign. FDI (j to k):	No	Yes	No	Yes	No	Yes	No	Yes
Conflict in country k	0.08*** (0.02)	0.05 (0.05)	0.02 (0.02)	0.25*** (0.09)	-0.01 (0.03)	-0.13 (0.12)	0.06*** (0.02)	-0.03 (0.03)
Observations	212285	189594	188137	167925	49705	43403	277556	226926
Exporter × year	✓	✓	✓	✓	✓	✓	✓	✓
Importer × year	✓	✓	✓	✓	✓	✓	✓	✓
Exporter × Importer	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports OLS estimates of the effects of conflict in country k on exports from country i to country j. The explanatory variables is constructed as in Table 3.1, with the top-7 exporters defining relevance, and 20 clusters defining similarity. To analyze the heterogeneity w.r.t FDI, we interact the explanatory variable with a dummy indicating that importer j has at least 10% of its FDI value in country k. This interaction variable takes the value of 1 in odd columns and 0 in even columns. All estimations include exporter-time, importer-time and exporter-importer fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors, in parentheses, are clustered at the exporter-importer level, * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

In Table 3.A.2 in the Appendix, we test for heterogeneity with respect to the characteristics of traded goods. In Panel A, we distinguish between exports of commodities that are common (exported by several countries) or rare (exported by only a handful of countries) in the agricultural and minerals sectors.[13] In the minerals sector, trade relocation is only observable for very common commodities. In the agricultural sector, no coefficient is statistically significant, possibly due to a lack of variation in our independent variable as (very) common agricultural commodities are exported by almost all countries.

In Panel B, we differentiate between intermediate and final goods. In the agriculture and mining sectors, we only find significant evidence for trade relocation among intermediate goods. For final goods, our estimates yield relatively precise zeroes. Likely, global value chains that rely on agricultural and mining commodities have the capacity and/or economic interest to pursue a quick substitution of export partners. For manufacturing goods on the other hand, we estimate a large and highly significant trade relocation effect for final goods, but an insignificant effect for intermediate goods. A likely explanation for this finding is that the final process in the manufacturing supply chain, which mainly consists of assembling ready-made parts, can more easily be offshored if civil conflict mandates relocation.

Finally, we conduct general equilibrium welfare computations for three case studies. We look at the recent peaks of civil violence in Colombia, Ukraine and Turkey, and esti-

[13] Restricted data availability in the manufacturing sector does not allow to make the same comparison for this sector.

mate (i) changes in worldwide bilateral trade flows and (ii) changes in countries' overall welfare. We discuss the general equilibrium analysis in more detail in Appendix 3.A.8. Based on the methodology discussed in Baier et al. (2019), we first estimate how civil conflict affected the conflict country's overall exports, and then use this estimate to compute hypothetical trade flows in case the respective conflict never happened. Deriving overall consumption from (hypothetical) internal and international trade flows, we further receive a proxy for countries' overall welfare levels. A comparison of actual to hypothetical trade flows and welfare levels then sketches the general equilibrium effects of the respective conflict.

The estimated welfare changes help us interpret the global effects of civil conflict. As depicted in Figure 3.A.1, for basically every country in our sample, welfare levels are smaller relative to the hypothetical scenario where a given conflict had not occurred. While it is of little surprise that the conflict countries themselves as well as their main importers experience the largest welfare reductions, even those countries that experience bilateral export increases thanks to trade relocation are overall worse off. Indeed, we only estimate a slight welfare increase for Macao in response to the civil conflict in Colombia. Apparently, trade relocation can only partially offset the welfare losses countries encounter due to increased trading costs with the conflict country. Hence, even though trade relocation helps mitigate some of the global loss in trade and welfare due to civil wars, all members of the world economy are individually worse off compared to a world at peace.

3.5 Robustness

We estimate various alternative specifications to test our results for robustness. A first concern of our estimation approach is our selection of cut-offs to code our relevance and similarity conditions. Figure 3.3 provides results for our main specification using alternative thresholds to classify relevant and similar exporters, respectively. Panel (a) to the left varies the number of export partners we classify as relevant for a dyad's importer. This exercise suggests that our results are sensitive to the cut-off we choose to classify exporters as relevant. Only when we consider anything between the top seven and ten trade partners as relevant, we find significant trade relocation effects.[14] This finding fits the intuition behind our estimation approach. If we consider a too small number of trade-partners, we miss out on relevant relocation cases. This, in turn, results in only a very small number of dyads we code as subject to trade reloca-

[14] A disaggregation into sectors in Figure 3.A.4 in the Appendix shows that a smaller number of trade-partners in the agricultural sector yields stronger results than in the minerals and manufacturing sectors. The coefficients for the fuels sector remain statistically insignificant, regardless of the number of trade-partners.

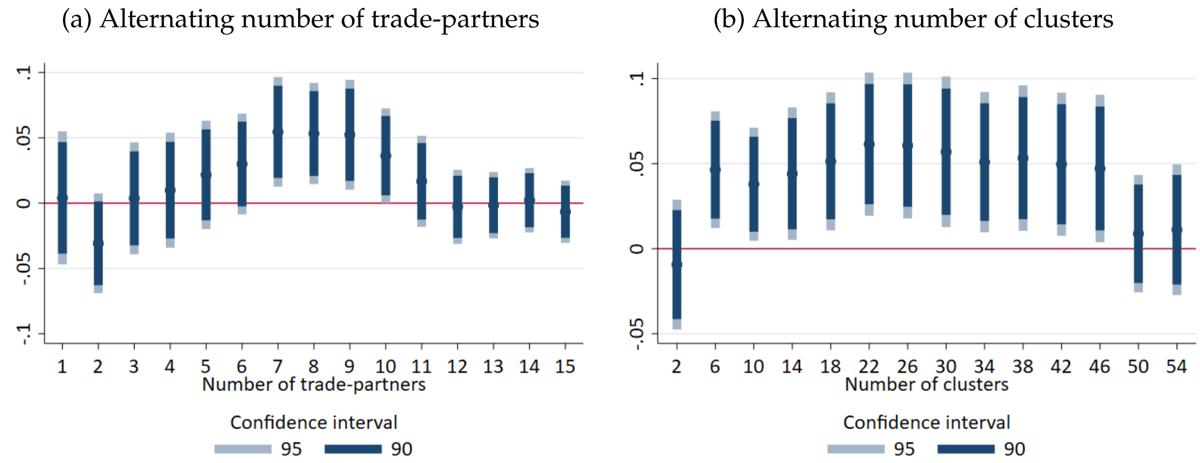
tion, while many potential relocation cases end up in our control group, biasing the results towards zero. Similarly, a too broad classification of relevant trade-partners adds numerous cases which we would code as subject to trade relocation even though the actual propensity for trade relocation is very low. According to our raw data, trade partners that are ranked tenth or higher are responsible for less than one percent of a country's overall imports, on average. This again biases our results towards zero as cases where no trade relocation is to be expected end up in our treatment group. In Panel (b) to the right, we conduct a similar robustness test and vary the number of clusters we use to code exporters as similar. Similarly as above, for a very broad categorization into e.g. only two clusters or very narrow classification into fifty clusters or more, our results turn insignificant.[15] Hence, our results are much less sensitive to the number of clusters we select to code our similarity condition. This resembles the fact that the variation in the similarity classification is rather low across intermediate numbers of clusters. For example, countries that rely mostly on agricultural production will almost always end up together in the same cluster, no matter whether the world is divided into five or forty production clusters. Still, another concern inherent to our estimation approach is that a single cluster might drive our results. To check for this possibility, we conduct leave-one-out regressions, where we repeat our main estimations but drop one cluster at a time. As we show in Figure 3.A.3 in the Appendix, our results are basically identical regardless of which countries we drop from our sample.

Two further robustness checks specifically concern our coding of exporter similarity. First, instead of clusters we construct a similarity index following Benedictis and Tajoli (2007a). This index measures the correlation of sectoral export values between two countries relative to other countries. We define countries i and k as similar if their export similarity is higher than 0.5, where 1 refers to identical and 0 to non-overlapping export patterns. Second, we change the input to our cluster calculations to allow for importer-specific considerations of which exporters they would treat as similar. Here, we classify all available exporters for each importer separately and include additional variables as inputs to the cluster algorithm. In addition to sectoral production shares, we also include various dyadic determinants of trade costs. Among other things, these are bilateral distance, common official language, and colonial heritage. Arguably, if importers search for substitution possibilities in response to a civil war in one of their main export origins, these cost factors may be as relevant as a country's production capabilities to make a trade relocation decision. We present the results of both alternative specifications in Table 3.A.5. Our main results remain qualitatively unchanged.

Next, we want to rule out the possibility that instead of the interaction of the similarity

[15]The agricultural sector is the least sensitive to the number of clusters, while for the minerals and manufacturing sectors the estimated coefficient is more often insignificant (Figure 3.A.5 in the Appendix). Again, the coefficients for the fuels sector remain statistically insignificant throughout.

Figure 3.3: Alternative relevance and similarity cut-Offs



Notes: This figure displays the coefficients of our diversion propensity as defined in Table 3.1 with alternating numbers of trade-partners (20 clusters) in the left panel and alternating number of clusters (7 trade-partners) in the right panel. All estimations are run with the PPML estimator and include exporter-sector-time, importer-sector-time and exporter-importer-sector fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors are clustered on the exporter-importer-sector level. The light and dark blue lines depict 95% and 90% confidence intervals, respectively.

and relevance conditions, one of these conditions alone produces our results. Theoretically, our identification approach might mechanically single-out much-trading dyads or countries exporting specific goods via the relevance or similarity classification, respectively. While each indicator variable alone is controlled for by our fixed effects, there remains the possibility that due to missing variation in either one of the two conditions, the other may alone drive the effects. As our results are less sensitive to the number of clusters as shown in Figure 3.3 above, one concern could be that the condition of similarity is redundant. To check whether indeed the interaction of both variables is generating our results, we invert either the similarity or the relevance classification and repeat our main estimations. Table 3.A.6 reports the results. In Panel A, we use the seven least important trade-partners to importer j, while still using the original similarity classification between conflict country k and exporter i based on twenty clusters. We find no evidence that conflict in less-important trade partners leads to trade relocation. In Panel B, we retain the original relevance classification of the top seven trading partners, but turn around our similarity classification to include all exporters with an exporter similarity index below 0.1. Again, we do not find significant trade relocation effects. Overall, we conclude from this falsification test that our identification approach indeed captures trade relocation propensity, as both relationships to the conflict country, i.e. the exporter's similarity as well as the importer's relevance, are needed together to produce our main results.

Finally, Table 3.A.7 presents additional results in which we slightly change our main specification. In Panel A, in addition to our standard relocation propensity variable, we include a similar indicator for relocation propensity based on large conflicts with more than 1000 battle deaths. Our results mostly stem from small conflicts. The coefficients for our main indicator variable considering all conflicts together remains robustly positive, whereas the indicator variable based on big conflicts yields insignificant or even negative coefficients. These results should be treated with caution though, as the number of large conflicts in our sample is relatively small. In Panel B, we use the number of conflict countries that fulfill the relevance and similarity conditions instead of an indicator that the conditions are fulfilled for *any* country to estimate the intensive margin of trade relocation. The coefficients are almost identical to our main results, only in the minerals sector the effect is more precisely estimated. This may be a hint that in this sector, import demand is more likely to spill over from several conflict countries to some specific (peaceful) exporters. In Panel C, we estimate trade-flows in the same year as the conflict in country k. The weaker results for the minerals and manufacturing sectors indicate that trade-flows need some time to adjust. Looking at international instead of domestic wars, Panel D reports no statistically significant effects. This is most likely driven by the very small number of international wars during our sample period.

3.6 Conclusion

This paper introduces a novel estimation approach for trade relocation effects that result from economic shocks in third countries. According to the structural gravity model of international trade, unilateral economic shocks affect bilateral trade between other countries via changes to the overall competition on international markets (Anderson and van Wincoop, 2003). In the short-run, a reduced competitiveness of one country can thus increase trade between other countries. If such short-run trade increases market integration, e.g. via signing PTAs, bilateral trade costs remain lower than before the shock, which in turn provokes a persistent relocation of international trade.

We apply the estimation approach to civil conflicts, which have been shown to significantly depress countries' export capacity (Novta and Pugacheva, 2021). On average, we find that dyads increase bilateral trade flows by 6% in response to civil conflict in a third country. The agricultural, manufacturing and minerals sectors exhibit a trade diversion effect of up to 13%, whereas we find no trade diversion in the fuels sector. What is more, we find that in the manufacturing sector trade relocation persists still nine years after the end of a civil conflict due to reduced bilateral trade costs via PTAs. Hence, civil conflicts can induce long-term economic losses for affected countries as international markets end up in a new equilibrium.

This paper is the first to study the short- and medium-run trade relocation effects of unilateral shocks like civil conflicts. Our results add to prior findings that civil conflicts depress the international trade flows of conflict countries (Martin et al., 2008a) and their neighbors (Qureshi, 2013). Our findings are furthermore relevant for the design of post-conflict recovery policies. After a country resolves its internal disputes, it faces a different network of international trade with increased competition due to persistent shifts in the trade relationships of former trading partners. To reintegrate the now peaceful country back into international markets and support post-conflict recovery, improving the terms of trade, e.g. via the quick resolution of (temporary) preferential tariff margins, may constitute valuable policy measures. Similarly, conflict-countries themselves may prioritize foreign policy to improve bilateral trade and hence spur the recovery of local production capacities.

Our estimation approach can easily be adapted to other settings. To analyze relocation effects, we construct a relocation propensity indicator variable, which translates the triadic relationship between a conflict country and any trading dyad into a dyadic observation. Besides civil conflicts, the approach can be applied to any other unilateral shock that can significantly alter a country's international competitiveness. Moreover, our estimation approach can be adapted to other bilateral outcome variables like migration or FDI by formulating similarity and relevance conditions that apply to the outcome variable of interest.

3.A Appendix

3.A.1 Acknowledgements

We thank Axel Dreher, Martin Gassebner, Arevik Gnutzmann-Mkrtchyan, Daniel Mirza, Holger Strulik, and Yoto Yotov as well as participants at the GSIPE workshop 2021, the RIEF Doctoral Workshop 2021 and the HiCN Workshop 2021 for valuable feedback. Tobias Korn's work on this project was in part funded by the German Research Foundation (DFG, grant BL-1502/1-1). This study was furthermore funded by the DFG - project RTG 1723 in the framework of the research training group on "Globalization and Development".

3.A.2 Additional Tables

Table 3.A.1: Descriptive statistics

Variable	Mean (1)	SD (2)	Min. (3)	Max. (4)	Obs. (5)
Panel A: General variables					
Export value (Mio. USD)	94.80	1871.00	0.00	4.9e+05	2175061
Production value (Mio. USD)	5.9e+07	4.2e+08	0.00	1.6e+10	12506
Trade agreement	0.12	0.33	0.00	1.00	2187567
Exporter sanctioned	0.08	0.27	0.00	1.00	2187567
Internal conflict (>25 deaths)	0.13	0.34	0.00	1.00	2187567
Large internal conflict (>1000 deaths)	0.03	0.17	0.00	1.00	2187567
Panel B: Cluster variables					
Conflict, 15 cluster, top 7 TP	0.04	0.19	0.00	1.00	2024615
Conflict, 15 cluster, top 10 TP	0.06	0.23	0.00	1.00	2024615
Conflict, 20 cluster, top 7 TP	0.03	0.18	0.00	1.00	2024615
Conflict, 20 cluster, top 10 TP	0.05	0.21	0.00	1.00	2024615
Panel C: Similarity variables					
Conflict, 0.50 similarity, top 7 TP	0.03	0.17	0.00	1.00	2024615
Conflict, dyadic cluster, top 7 TP	0.24	0.43	0.00	1.00	2024615

Notes: This table provides descriptive statistics of the main variables. The variables in Panels B and C depict the indicator variable as described in equation 3.2.3. Variables take a value of 1 if exporter k has a conflict, countries i and k were in the same cluster (Panel B) or similar (Panel C), and country k was a top trade-partner (TP) of country j. The sample consists of 180 countries and the sectors agriculture, minerals, fuels and manufacturing between 1995-2014.

Table 3.A.2: Trade relocation heterogeneity: Commodity characteristics

Dependent:	Sectoral exports from country i to country j					
Panel A:	Agriculture			Minerals		
	Very common	Common	Rare	Very common	Common	Rare
Conflict in country k	0.02 (0.02)	0.06 (0.04)	-0.08 (0.09)	0.06^{***} (0.02)	-0.03 (0.05)	0.27 (0.31)
Observations	324148	303658	80708	289851	216409	17951
Panel B:	Agriculture		Mining		Manufacturing	
	Final goods	Inter-mediates	Final goods	Inter-mediates	Final goods	Inter-mediates
Conflict in country k	0.01 (0.03)	0.09^{**} (0.04)	-0.03 (0.19)	0.08^{***} (0.03)	0.17^{***} (0.06)	0.02 (0.09)
Observations	312996	313568	45709	322052	111595	111664
Exporter × year	✓	✓	✓	✓	✓	✓
Importer × year	✓	✓	✓	✓	✓	✓
Exp. × Imp.	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓

Notes: This table reports estimates of the effects of conflict in country k on exports from country i to country j. In Panel A, the agriculture and minerals sector are disaggregated into 'very common' (top 10% traded commodities), 'common' (middle 80% traded commodities) and 'rare' (least 10% traded commodities). In Panel B exports from the agriculture, minerals and manufacturing sectors are disaggregated into into intermediate and final goods based on the BEC classification. The explanatory variables is constructed as in Table 3.1, with the top-7 exporters defining relevance, and 20 clusters defining similarity. All estimations include exporter-time, importer-time and exporter-importer fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors, in parentheses, are clustered at the exporter-importer level, * p< 0.1, ** p< 0.05, *** p< 0.01

Table 3.A.3: Trade relocation heterogeneity: Market share of conflict country

Dependent:	Sectoral exports from country i to country j							
	Agricult.		Minerals		Fuels		Manufact.	
Country k has above 5% market share:	No	Yes	No	Yes	No	Yes	No	Yes
Conflict in country k	0.14***	-0.00	0.11***	-0.06	0.01	-0.20**	0.08**	-0.02
	(0.04)	(0.05)	(0.03)	(0.08)	(0.03)	(0.09)	(0.03)	(0.03)
Observations	366662	366662	322631	322631	89446	89446	491003	491003
Exporter × year	✓	✓	✓	✓	✓	✓	✓	✓
Importer × year	✓	✓	✓	✓	✓	✓	✓	✓
Exp. × Imp.	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports estimates of the effects of conflict in country k on exports from country i to country j. The explanatory variables is constructed as in Table 3.1, with the top-7 exporters defining relevance, and 20 clusters defining similarity. To analyze the heterogeneity w.r.t the market share, we interact the explanatory variable with a dummy indicating that country k has a market share of at least 5% in the respective sector. This interaction variable takes the value of 0 in odd columns and 1 in even columns. All estimations are run with the PPML estimator and include exporter-time, importer-time and exporter-importer fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors, in parentheses, are clustered at the importer-exporter level, * p< 0.1, ** p< 0.05, *** p< 0.01

Table 3.A.4: PE results for GE computation

	Exports from country i to j		
	(1)	(2)	(3)
Peace × International	0.687***	0.410***	0.888***
	(0.149)	(0.139)	(0.127)
N	150719	150719	150719
Country	Colombia	Ukraine	Turkey
FTA-Control	✓	✓	✓

Standard errors in parentheses, * $p < .1$, ** $p < .05$, *** $p < .01$
Results from PPML Regressions.
All Regressions control for PTAs and include the typical fixed effects.

Table 3.A.5: Trade relocation robustness: Alternative similarity and relevance definitions

Dependent:	Exports from country i to country j			
	Agricult.	Minerals	Fuels	Manufact.
Panel A:	Export similarity >0.5			
Conflict in country k	0.08** (0.03)	0.06** (0.03)	-0.06** (0.03)	0.06** (0.02)
Panel B:	Dyadic clusters			
Conflict in country k	0.04** (0.02)	0.06*** (0.01)	0.01 (0.03)	0.09*** (0.02)
Observations	366662	322631	89446	491003
Exporter × year	✓	✓	✓	✓
Importer × year	✓	✓	✓	✓
Exporter × Importer	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: This table reports estimates of the effects of conflict in country k on exports from country i to country j, disaggregated by sectors. The explanatory variables is constructed as in Table 3.1, with similarity being defined as the two countries having an abvove 0.5 similarity index, as defined by Benedictis and Tajoli (2007a,b) in Panel A, and the two countries being in the same dyadic cluster in Panel B, and relevance as the top-7 exporter countries. All estimations are run with the PPML estimator and include the trade exporter-time, importer-time and exporter-importer fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors, in parentheses, are clustered on the exporter-importer level, * p< 0.1, ** p< 0.05, *** p< 0.01

Table 3.A.6: Trade relocation robustness: Wrong similarity and relevance conditions

Dependent:	Sectoral exports from country i to country j							
	Agricult.		Minerals		Fuels		Manufact.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A:	20 clusters, bottom 7 or non-trading partner countries							
Conflict in country k	0.02 (0.02)	0.02 (0.02)	0.00 (0.02)	0.00 (0.02)	0.04 (0.03)	0.04 (0.03)	-0.07* (0.04)	-0.07* (0.04)
Panel B:	Dissimilar countries (<10% similarity)							
Conflict in country k	0.02 (0.02)	0.02 (0.02)	-0.04** (0.02)	-0.04** (0.02)	-0.01 (0.04)	-0.01 (0.04)	-0.12*** (0.03)	-0.12*** (0.03)
Observations	366662	366662	322631	322631	89446	89446	491003	491003
Exporter × year	✓	✓	✓	✓	✓	✓	✓	✓
Importer × year	✓	✓	✓	✓	✓	✓	✓	✓
Exp. × Imp.	✓	✓	✓	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓	✓	✓	✓

Notes: This table reports a placebo study to the previous estimations. It shows effects of conflict in country k on exports from country i to country j. The explanatory variables is constructed as in Table 3.1, but, in Panel A, relevance is measured with the 7 countries with smalles (or zero) exports, and, in Panel B, the similarity is measured with a below 0.1 similarity index, as defined by Benedictis and Tajoli (2007a,b). All estimations are run with the PPML estimator and include the trade exporter-time, importer-time and exporter-importer fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors, in parentheses, are clustered on the exporter-importer level, * p< 0.1, ** p< 0.05, *** p< 0.01

Table 3.A.7: Trade relocation: Various robustness checks

Dependent:	Sectoral exports from country i to country j			
	Agricult.	Minerals	Fuels	Manufact.
Panel A:	Large conflicts			
Any conflict in country k	0.14*** (0.04)	0.12*** (0.03)	-0.02 (0.03)	0.07** (0.03)
Big conflict in country k	-0.07* (0.04)	-0.20** (0.08)	0.02 (0.07)	-0.13 (0.12)
Panel B:	Intensive margin			
Number of country k conflicts	0.11*** (0.04)	0.12*** (0.03)	-0.00 (0.02)	0.08** (0.03)
Panel C:	Conflict in same year			
Conflict in country k	0.13*** (0.03)	0.01 (0.02)	-0.02 (0.03)	0.06* (0.03)
Panel D:	International wars			
International conflict in country k	0.02 (0.03)	-0.01 (0.01)	-0.03 (0.02)	0.00 (0.01)
Observations	366662	322631	89446	491003
Exporter × year	✓	✓	✓	✓
Importer × year	✓	✓	✓	✓
Exporter × Importer	✓	✓	✓	✓
Controls	✓	✓	✓	✓

Notes: This table reports estimates of the effects of conflict in country k on exports from country i to country j, disaggregated by sectors. In Panel A, the explanatory variables are indicator variables which counts the occurrences of (i) country k having had any conflict or large conflicts in the previous year, (ii) country k being a top-7 exporter for country j in the pre-conflict-year and (iii) country k and country i being similar exporters in the pre-conflict-year. In Panel B, the explanatory variables is a continuous variable which counts the occurrences of our diversion propensity indicator for each exporter-importer pair. In Panel C, the explanatory variable is an indicator variable but with conflict measured in the same year as exports. Panel D uses international instead of internal wars. Throughout, similarity is measured by being in the same exporter-cluster, with a total number of 20 clusters. All estimations are run with the PPML estimator and include exporter-time, importer-time and exporter-importer fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors, in parentheses, are clustered on the exporter-importer level, * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

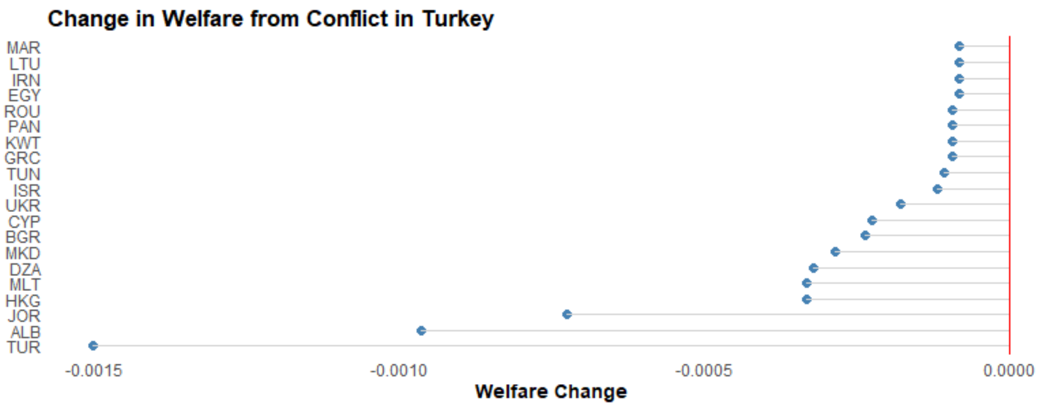
3.A.3 Additional Figures

Figure 3.A.1: GE Results: Welfare changes

(a) Welfare changes, conflict in Colombia

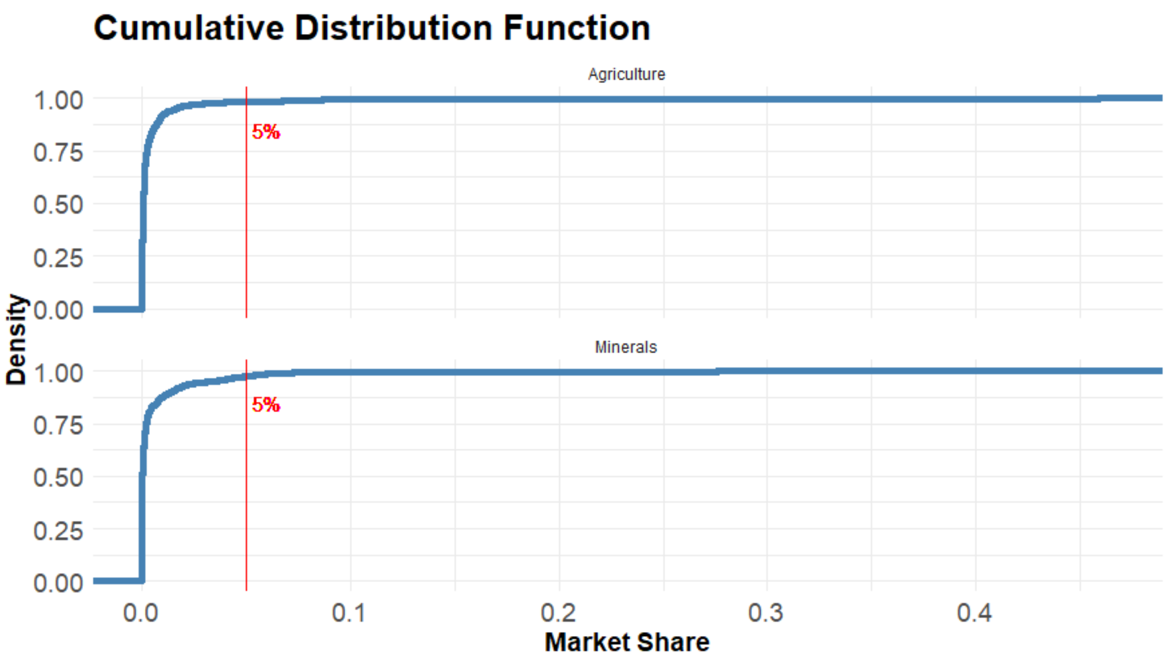
(b) Welfare changes, conflict in Ukraine

(c) Welfare changes, conflict in Turkey



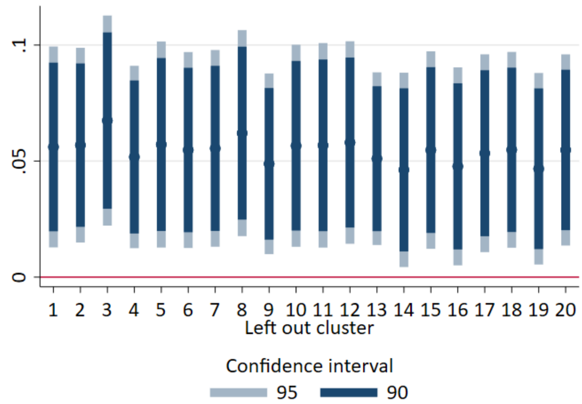
Notes: The graphs report the estimated welfare changes in the general equilibrium due to the civil wars in Colombia (Panel a), Ukraine (Panel b), and Turkey (Panel c). Each panel reports the 15 countries for whom our estimations reported the largest welfare changes. All estimates are derived based on a PE Regression of exports on peace, comparing the estimated trade flows during peace time to the actual trade flows during the civil war. See Table 3.A.4 for the respective PE results.

Figure 3.A.2: Distribution of market power



Notes: This figure shows the cumulative distribution functions for the market share measures computed based on commodity-level bilateral trade data. We compute for each exporter-commodity-year observation the market share a given observation occupies in the year's total market for a given commodity. The red line indicates the threshold used in Table 3.A.3 to identify market leading countries in exports of a specific commodity. Hence, all observations on the right hand side of the red line constitute market leaders in our heterogeneity regressions. The light and dark blue lines depict 95% and 90% confidence intervals, respectively.

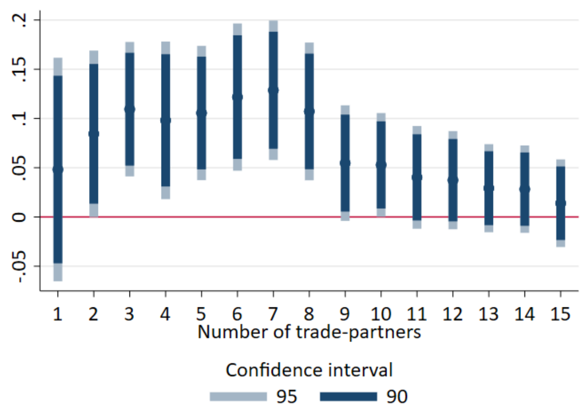
Figure 3.A.3: Leave-one-out



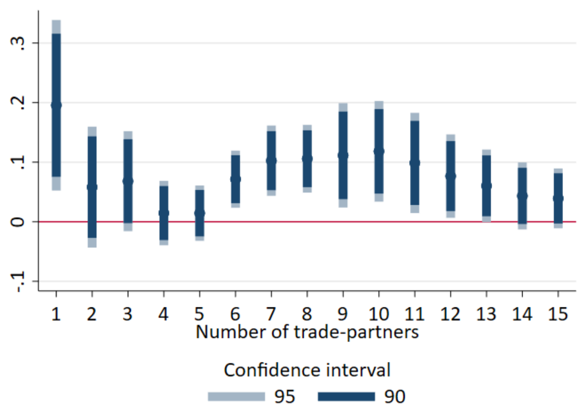
Notes: This figure displays the coefficients of relocation propensity as defined in Table 3.1 with similarity based on 20 clusters and relevance on the top 7 trade-partners. Each coefficient represents a regression leaving out one cluster. All estimations are run with the PPML estimator and include exporter-sector-time, importer-sector-time and exporter-importer-sector fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors are clustered on the exporter-importer-sector level.

Figure 3.A.4: Number of trade-partners - Sector disaggregation

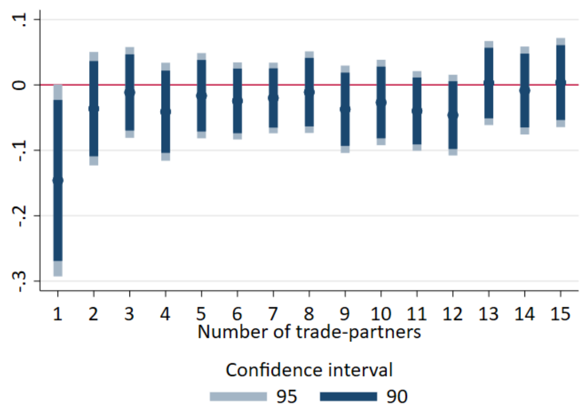
(a) Agriculture



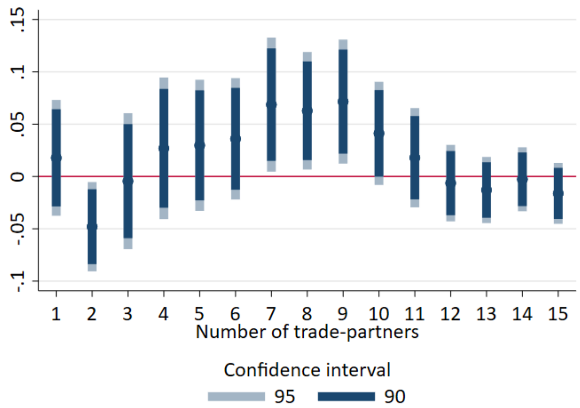
(b) Minerals



(c) Fuels

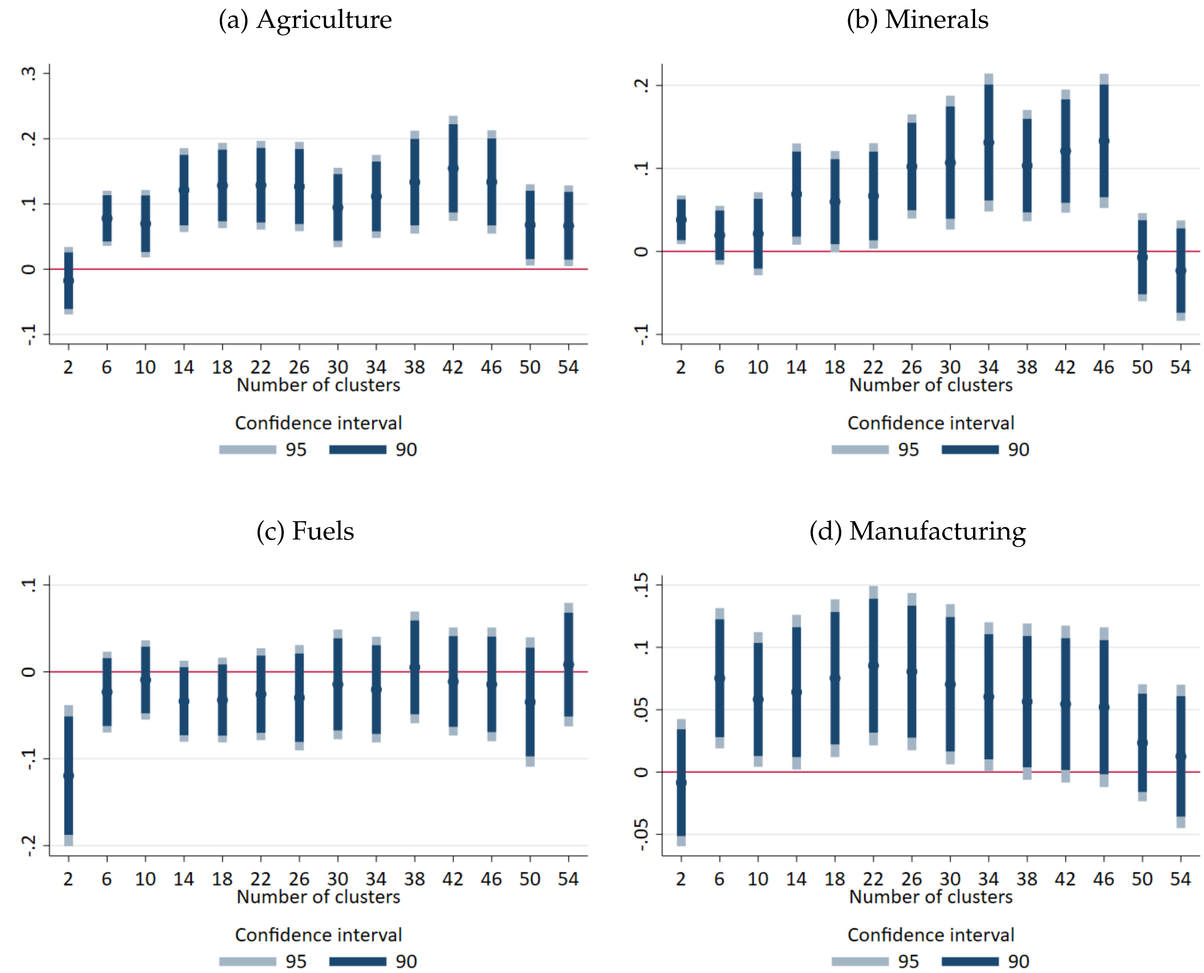


(d) Manufacturing



Notes: This figure displays the coefficients of relocation propensity as defined in Table 3.1 with similarity based on 20 clusters and relevance on a varying number trade-partners. All estimations are run with the PPML estimator and include exporter-sector-time, importer-sector-time and exporter-importer-sector fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors are clustered on the exporter-importer-sector level. The light and dark blue lines depict 95% and 90% confidence intervals, respectively.

Figure 3.A.5: Number of clusters - Sector disaggregation



Notes: This figure displays the coefficients of relocation propensity as defined in Table 3.1 with similarity based on a varying number of clusters and relevance on the top 7 trade-partners. All estimations are run with the PPML estimator and include exporter-sector-time, importer-sector-time and exporter-importer-sector fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors are clustered on the exporter-importer-sector level. The light and dark blue lines depict 95% and 90% confidence intervals, respectively.

3.A.4 Derivation of Relocation Estimation

We leverage a measure of relocation propensity to derive an estimating equation for trade relocation considering the inward multilateral resistance $P_{js,t}$.[16] Following Anderson et al. (2018b), we can define $P_{js,t}$ as:

$$P_{js,t} = \sum_k \frac{t_{kj,t}}{\Pi_{ks,t}} \cdot \left[\frac{Y_{ks,t}}{Y_{Ws,t}}\right]^{\frac{1}{1-\sigma}} \tag{3.A.1}$$

According to equation 3.A.1, importer j's inward multilateral resistance corresponds to its average access to exports from all other countries k. Theoretically, civil conflict in any other country k can enter equation 3.A.1 either via $Y_{ks,t}$, if war and destruction decrease country k's overall production in sector s, or via $t_{kj,t}$, if civil violence leads to a tightened security situation and therefore increases bilateral shipping costs. W.l.o.g., we assume that civil conflict works via a decrease in overall production $Y_{ks,t}$, while the same argumentation holds for $t_{kj,t}$. Then, we can rewrite the production of each country k to incorporate a potential conflict-shock as $Y_{ks,t} = \bar{Y}_{ks,t} \cdot (1 - \Delta^Y_{ks,t})$, where $\Delta^Y_{ks,t}$ denotes the share of production lost due to civil conflict and $\bar{Y}_{ks,t}$ represents the level of production absent conflict. Next, note that this general way of specifying $P_{js,t}$ allocates the same weight to any exporter k affecting the dyad ij – i.e. a given shock $\Delta^Y_{ks,t}$ affects the inward multilateral resistance by the same amount across all countries k and for all export partners i. We however argue that two bilateral relationships, first between importer j and conflict country k, and second between the two exporters i and k, must be taken into account. While the standard gravity equation suggests that a change in P_{js} might lead to trade relocation from *any* conflict country k to *any other* non-conflict country i in sector s, we argue that the realized trade relocation actually depends on the relocation propensity inherent to the (sector-specific) triad ijk, which we pin down to two important bilateral characteristics underlying (i) the kj-dyad and (ii) the ik-dyad. To see this, we augment equation 3.A.1 to represent the exporter-specific inward resistance by adding two weight matrices that indicate the relationships between k and j, and k and i, respectively, while also including the conflict shock to country k's production in sector s. Both weight matrices are lagged by one period to focus on the country characteristics before conflict emerged in country k (and potentially altered its characteristics). We arrive at the equation:

$$P_{ijs,t} = \sum_k \frac{t_{kj,t}}{\Pi_{ks,t}} \cdot \left[\boldsymbol{W}^R_{jks,t-1} \cdot \boldsymbol{W}^S_{ik,t-1} \cdot \frac{\bar{Y}_{k,t} \cdot (1 - \Delta^Y_{k,t})}{Y_{W,t}}\right]^{\frac{1}{1-\sigma}} \tag{3.A.2}$$

[16] Note that the same argument holds from the importer side via the outward multilateral resistance $\Pi_{is}^{1-\sigma} = \sum_l \frac{E_{ls}}{Y_{Ws}} \cdot \left[\frac{t_{ils}}{P_{ls}}\right]^{1-\sigma}$.

The first weight matrix $\boldsymbol{W}^R_{jks,t-1}$ refers to the relevance of each country k as an exporter for importer j.[17]. We expect that the realized trade relocation effect is larger if there is a bigger trade value to be relocated, i.e. if dyad kj used to trade a lot in sector s before the conflict emerged in country k. Note that for each (sector-specific) importer j, only the j'th row of the matrix $\boldsymbol{W}^R_{jks,t-1}$ will affect $P_{ijs,t}$, which essentially reduces the weight matrix to the js-specific weight vector $\boldsymbol{w}^R_{\bar{j}ks,t-1}$. The second weight matrix $\boldsymbol{W}^S_{iks,t-1}$ denotes each country k's similarity to a dyad's exporter i, which is not sector-specific. We argue that not all countries are equally suited to "fill in" the gap left by the diminished exports from country k to country j. In theory, we usually assume that countries trade with each other because of the specific varieties of goods that each exporter i has to offer (Armington, 1969). Hence, as country k can provide less of its varieties, country j will turn to country i only if it offers a variety of goods similar to those of conflict country k. Therefore, the relocation propensity arguably depends on exporter i exhibiting a similar export composition as conflict country k. Note here that the variation brought in from the weight matrix $\boldsymbol{W}^S_{ik,t-1}$ is the same for all importers j and hence does not affect $P_{ijs,t}$ differentially across importers. For simplicity, assume that the elements of both matrices only take the values 0 and 1, where 1 indicates that country k is relevant for importer j or similar to exporter i, respectively. Then, rewriting 3.A.2 to represent the remaining variation $\hat{P}$ in P when the importer-sector-year and exporter-sector-year fixed effects of the gravity equation are accounted for yields:

$$\hat{P}_{ijs,t} = \sum_k \left(\boldsymbol{w}^R_{\bar{j}ks,t-1} \cdot \boldsymbol{W}^S_{ik,t-1} \cdot \bar{Y}_{k,t} \cdot (1 - \Delta^Y_{k,t}) \right)^{\frac{1}{1-\sigma}} \tag{3.A.3}$$

equation 3.A.3 demonstrates two things. First, by specifying the gravity equation with the correct fixed effects as outlined in Head and Mayer (2014), the triple-interaction of the (i) conflict-shock to a country k, (ii) the similarity condition for countries i and k, and (iii) the relevance condition for countries j and k in sector s is the only variation left in the multilateral resistance term. Second, it follows from the negative income shock to country k, $-\Delta^Y_{k,t}$, and the elasticity of substitution $\sigma > 1$, that a conflict-shock to any country k increases importer j's inward multilateral resistance P (i.e. $\frac{d\hat{P}_{ijs,t}}{d\Delta^Y_{k,t}} > 0$). Finally, we can separate the general part of the inward multilateral resistance $\bar{P}_{js,t}$, which can be accounted for by fixed effects, from the remaining variation outlined in equation 3.A.3 and insert both into equation 3.2.1 to arrive at:

$$X_{ijs,t} = \frac{Y_{is,t} E_{js,t}}{Y_{Ws,t}} \cdot \left[\frac{t_{ij,t}}{\Pi_{is,t} \cdot \bar{P}_{js,t} \cdot \hat{P}_{ijs,t}} \right]^{1-\sigma} \tag{3.A.4}$$

[17] A typical element of matrix $\boldsymbol{W}^R_{jks,t-1}$ would be $\frac{X_{kjs}}{E_{js}}$, which denotes the share of imports from country k in country j's overall consumption expenditures.

Proceeding as above by adding fixed effects and taking logs, we arrive at our main estimating specification:

$$X_{ijs,t} = exp\left[\pi_{is,t} + \lambda_{js,t} + \mu_{ijs} + \beta_2 \cdot \sum_k \left(R_{jks,t-2} \times S_{ik,t-2} \times C_{k,t-1}\right) + \gamma \cdot Z_{ij,t}\right] + \eta_{ijs,t} \tag{3.A.5}$$

3.A.5 Construction of GE Dataset

Our GE estimates require a symmetric dataset which also includes internal trade flows of all sample countries. We calculate internal trade flows by subtracting a country's exports from its total production. In the next step, we construct a symmetric dataset. This is, we require bi-directional trade flows between all available exporters and importers in the sample as well as non-negative internal trade flows for each country and in every year. Due to differing data availability across years, we restrict our sample to the manufacturing sector and the years 1992-2016. Additionally, we reduce the number of countries to 68 importers and exporters. As a decision rule for our sample construction, we decided to only keep years or countries whose numbers of observations amount to at least 80% of the year and 80% of the importer/exporter with the most observations, respectively. Our results remain unchanged for stricter and looser restrictions.

3.A.6 Direct Effects

A prerequisite for finding significant trade relocation effects of civil conflicts is that conflict countries decrease their amount of exports. Prior findings emphasize that civil wars depress international trade (see, e.g., Bayer and Rupert, 2004; Long, 2008; Qureshi, 2013). To replicate these findings with our data and adapt the empirical strategy to the gravity framework of international trade, we follow Head and Mayer (2014) and extend equation 3.2.1 accordingly. When we include the usual fixed effects, the effect of civil conflict in a country i on that same country's exports cannot directly be estimated as the variable is collinear with the exporter-year fixed effects $\pi_{is,t}$. We therefore follow Yotov et al. (2016a) and include intranational trade flows along with bilateral trade flows in our dataset.[18] This allows estimating the effect of a unilateral shock like civil conflict on bilateral trade by interacting the variable of interest with an indicator variable for international trade flows (Beverelli et al., 2018). We arrive at an estimating equation of the form:

[18]Yotov (2021) provides an extensive overview of the benefits of adding intranational trade in bilateral trade estimations.

$$X_{ijs,t} = exp\left[\pi_{is,t} + \lambda_{js,t} + \mu_{ijs} + \beta_1 \cdot (C_{i,t} \times I_{ijs}) + gamma \cdot Z_{ij,t}\right] + \eta_{ijs,t} \quad (3.A.6)$$

where $\eta_{ijs,t}$ accounts for the remaining variation in $X_{ijs,t}$ not explained by the fixed effects and control variables. The variable $C_{i,t}$ indicates the presence of civil conflict in country i at year t, and I_{ijs} indicates international trade flows (i.e. that $i \neq j$). This form of the gravity specification affects the interpretation of the coefficient β_1. Here, β_1 constitutes the elasticity of exports from origin i to destination j in sector s relative to internal consumption of country i to civil conflict emerging in country i.

Table 3.A.8: Direct effects: Internal conflicts hurt exports

Dependent:	Total exports from country i					
	All sectors		Agri culture	Minerals	Fuels	Manuf acturing
	(1)	(2)	(3)	(4)	(5)	(6)
Conflict (t-1)	-0.05	-0.06**	-0.07*	-0.10**	0.07	-0.06*
× international trade	(0.03)	(0.03)	(0.04)	(0.04)	(0.12)	(0.03)
Conflict (t-2)	-0.07**	-0.08**	0.02	0.06	-0.09	-0.09***
× international trade	(0.03)	(0.03)	(0.04)	(0.04)	(0.12)	(0.04)
Observations	1290234	1290234	354750	314425	88550	532509
Exporter × sector × year	✓	✓	✓	✓	✓	✓
Importer × sector × year	✓	✓	✓	✓	✓	✓
Exp. × imp. × sector	✓	✓	✓	✓	✓	✓
Controls		✓	✓	✓	✓	✓

Notes: This table reports estimates of the effects of civil conflict on a country's exports. We interact a dummy variable for lagged civil conflict with an indicator variable for international trade flows. Coefficients must hence be interpreted as change in exports relative to a country's internal trade. All estimations are run with the PPML estimator, exporter-sector-time, importer-sector-time and exporter-importer-sector fixed effects. Controls are indicators for trade agreements and bilateral sanctions on the exporter side. Standard errors are clustered at the exporter-importer-sector level, * $p< 0.1$, ** $p< 0.05$, *** $p< 0.01$

We use the Pseudo-Poisson Maximum Likelihood (PPML) estimator as suggested by Santos Silva and Tenreyro (2006). To account for differences in the duration and velocity of the effect, we lag civil conflict by one as well as two years. The results are presented in Table 3.A.8 and confirm our priors based on the literature. Columns (1) & (2) consider trade data across all four sectors and include sector fixed effects to account for sector-specific shocks. Column (1) does not include any bilateral control variables, while we control for bilateral trade agreements and sanctions starting from column (2). Without control variables, we only find a significantly negative effect of conflict on bilateral exports if conflict is lagged by two years, whereas both lags are statistically significant when the bilateral control variables are included. On average, a conflict country's exports decrease by around 6% and 8% relative to the country's internal consumption one and two years after civil conflict, respectively.[19] In columns (3)–(6), we test for heterogeneity across sectors by restricting the sample to trade flows from the respective sector.[20] Overall, the effect of civil conflict on international trade is quite heterogeneous. Agricultural exports only suffer slightly one year after conflict with an effect that is barely statistically significant. Exports of mineral goods, however, are significantly reduced by around 10% one year after conflict, whereas the second lag is not statistically different from zero. For manufacturing exports, we find significant reductions for both lags of the conflict variable. Interestingly, fuel exports do not appear to decline at all during civil conflict. This could, on the one hand, indicate that importers are so dependent on fuel imports that trade-flows continue even in the presence of civil unrest. On the other hand, fuel exports are an important financing tool for civil wars (Bazzi and Blattman, 2014; Andersen et al., 2017). Therefore, the government as well as the rebels are eager to maintain fuel exports during conflict. Hence, our results suggest that, on average, ongoing civil conflicts depress national exports *relative to internal consumption*.[21] This effect is most immediate in the minerals sector and longer lasting in manufacturing trade, while it does not seem to occur in the fuels sector. Note however that all these estimates likely constitute lower-bound estimates of the actual effect, since we estimate reductions in international trade *relative to* internal trade. Hence, as internal trade is likely to also be negatively affected by civil conflict, our results mirror the additional deterioration of international trade flows with respect to the themselves as well internal trade flows.

[19] According to the formula $(e^{\beta} - 1) \times 100\%$.

[20] Note that the gravity equation is separable by sectors as outlined in Yotov et al. (2016a) and hence equation 3.A.6 can be applied separately by sector.

[21] Note again that the interaction term in equation 3.A.6 mandates this interpretation.

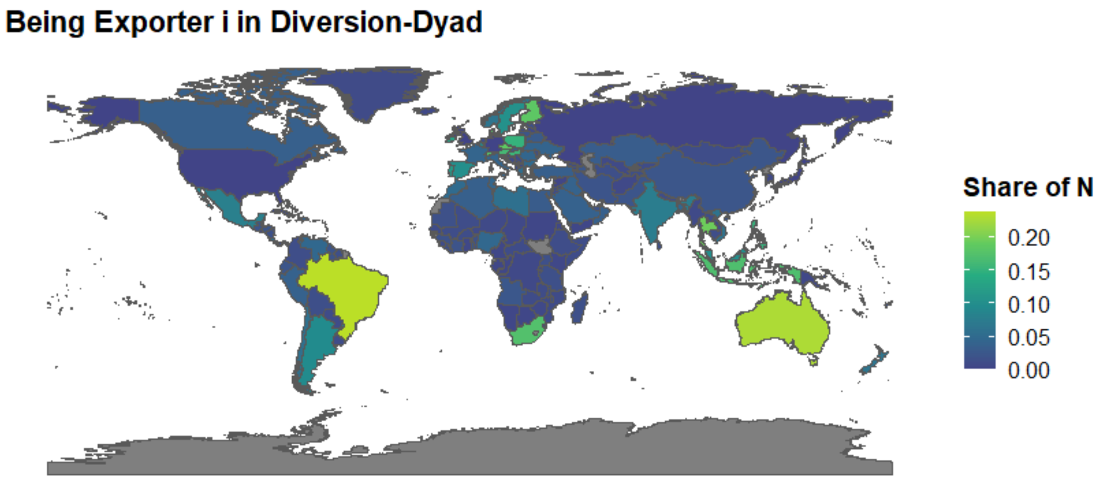
3.A.7 Relocation Propensity

Figure 3.A.6 gives some intuition to the distribution of our relocation propensity variable. The two maps report the geographic distribution of the likelihood to appear as exporter i or importer j in a relocation dyad. The odds of being affected as an importer, i.e. having a relevant trade partner starting a civil war, are distributed quite homogeneously across the globe. While East Africa and the Middle East stick out with a slightly higher propensity and Europe appears only rarely affected, the overall propensity is fairly equally distributed across all regions. The likelihood that in at least one of a country's trading sectors a relevant exporter starts a civil war for most countries lies close to 5 percent. The picture is different when looking at the likelihood of being an affected exporter, i.e. the odds that a country with a similar export structure starts a civil war. Here, Brazil and Australia stick out with a very high likelihood of around 20 percent, followed by South Africa, Argentina, Indonesia, Eastern Europe and Scandinavia. On the other hand, the USA and several other countries, especially in Africa, Asia and Central Europe, are almost never coded as exporters benefiting from trade relocation.

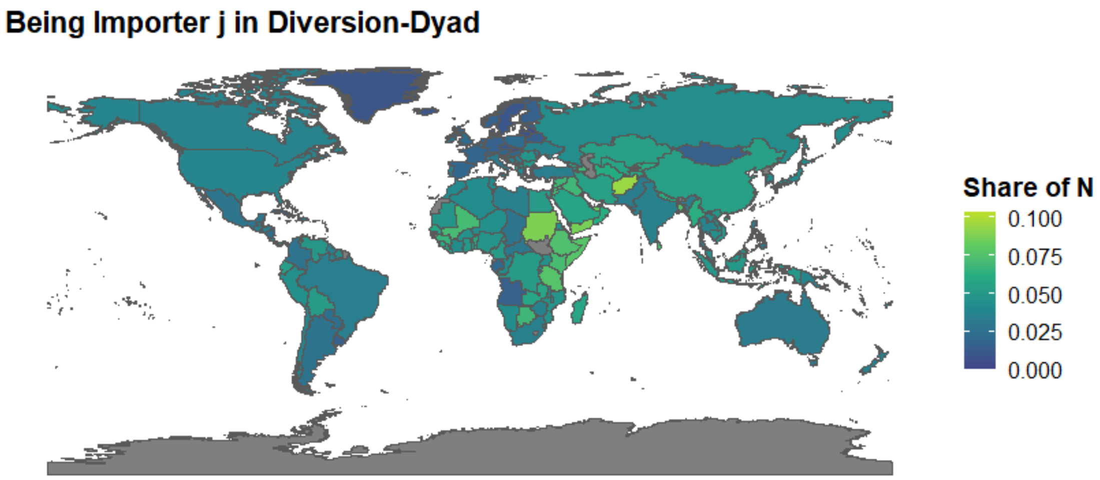
In Figure 3.A.7, we further investigate the determinants of the similarity and relevance characteristics. Here, we regress the likelihood that a country is a similar exporter (Panel (a)) or a relevant importer (Panel (b)) to a conflict country k on the common gravity variables. As is to be expected, these variables only play little role for the similarity characteristic. Among the bilateral variables, only inverse distance and an indicator for common legal origins are significantly positive, which likely mirrors local clusters of resources or similar production techniques based on the legal environment. Furthermore, conflict countries have on average a lower GDP, while beneficiary exporters are more likely to be WTO members. For the relevance characteristic however, most gravity variables turn out highly significant and with the expected sign. Important trade partners of a conflict country are on average closer and have the same official language or colonial history. Similarly, higher economic masses of both countries j and k as well as an existing Regional Trade Agreement between the two are significant determinants of the relevance characteristic. This emphasizes that, as is to be expected by construction, the relevance characteristic we identify is strongly related to the classical determinants of bilateral trade.

Figure 3.A.6: Geographic distribution of diversion propensity

(a) Exporter *i*



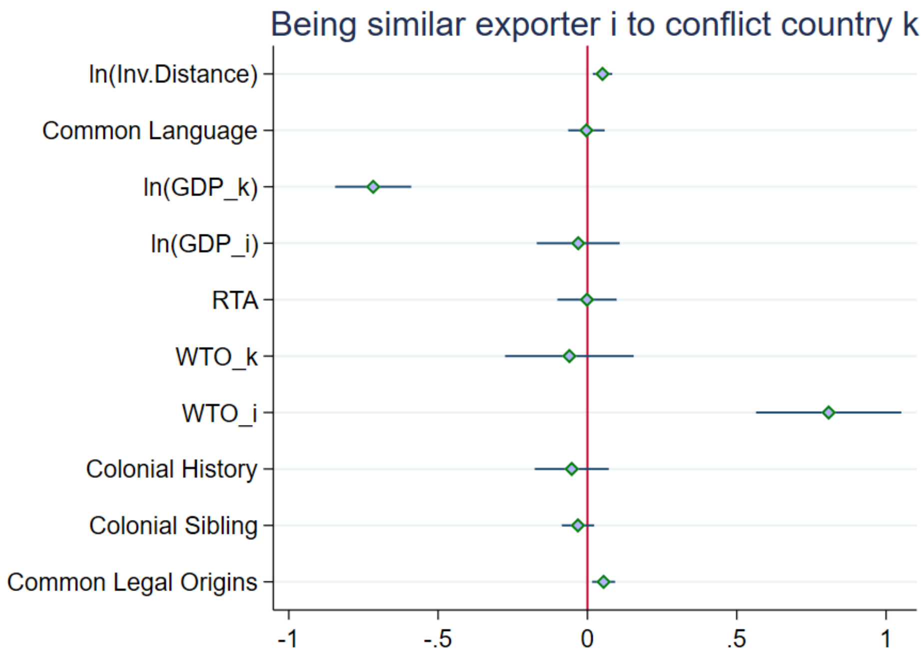
(b) Importer *j*



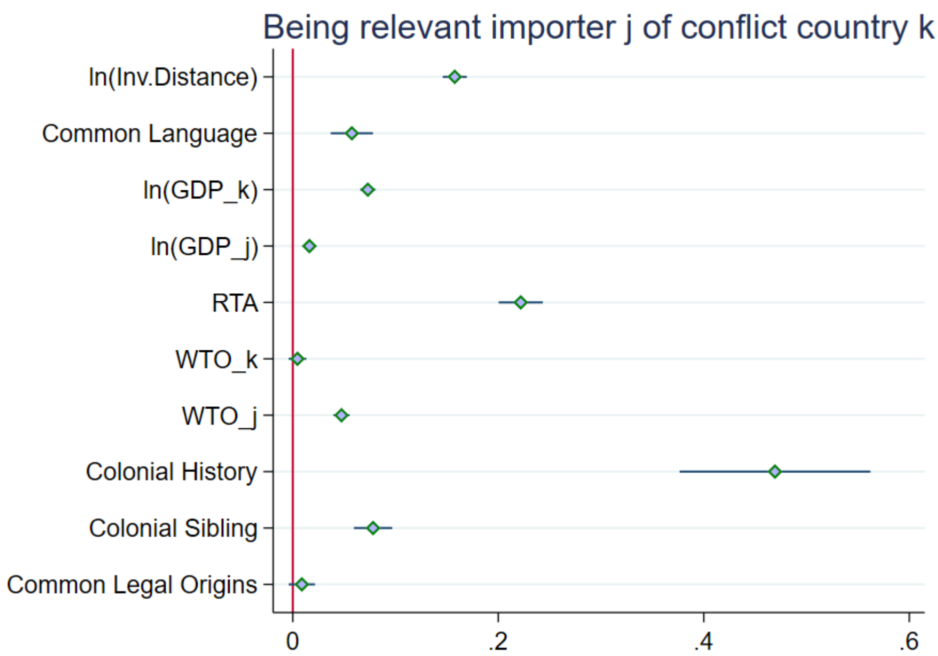
Notes: This figure shows a country's likelihood to appear as an exporter *i* or importer *j* in a dyad with positive trade relocation propensity. Panel (a) shows the geographic distribution of the likelihood to be an exporter affected by trade relocation, while panel (b) plots the same distribution for importers. The different shades display the share of a country's observations that it is coded as having a positive relocation propensity. For example, in panel (a) a share of 0.1 means that 10 percent of a country's export observation across all sample years and all importers are coded as being an exporter profiting from trade relocation due to civil conflict in some country *k*.

Figure 3.A.7: Explaining propensity of being i or j

(a) Being similar exporter *i*



(b) Being relevant importer *j*



Notes: This figure reports the results from regression the status of being a similar exporter *i* (Panel a) or a relevant importer *j* (Panel b) for conflict country *k* on the most common gravity variables All regressions include importer, exporter, and year fixed effects. Standard errors are clustered at the dyad level. Lines depict 95% Confidence Intervals.

3.A.8 General Equilibrium

We analyse three case studies in a General Equilibrium (GE) framework. These case studies allow us to focus on specific conflicts and accurately trace the relocation effects. As recent examples of significant violent episodes, we focus on (i) the peak of clashes between the FARC rebels and AUC paramilitary forces in Colombia in the 1990s and until 2005, (ii) the Ukrainian civil war from 2014 to present, and (iii) the violent 1990s in Turkey where the PKK fought for local independence. The case studies were selected based on the significance of the respective conflict shocks (at least two years of violence with more than 1000 battle deaths) among a handful of countries where international and internal trade data were available during and before or after the conflict period. For these three cases, we proceed in two steps. First, we construct an indicator variable for each case that takes the value of one for all dyads that include the respective conflict country as an exporter during years of peace. We then regress trade on the interaction of this variable with an indicator variable for international trade flows including country-year and dyad fixed effects similar to equation 3.A.6. From this, we receive an estimate for the effect of peace on the respective country's exports relative to its internal consumption of self-produced goods.[22]

Second, we use the respective estimates and compute hypothetical trade changes in the general equilibrium during a conflict-year. Following Baier et al. (2019), we apply a one sector Armington-CES model, assuming a constant trade elasticity of $\theta = 4$.[23] This computation generates counterfactual trade flows for all sample countries in case the civil war in either Colombia, Ukraine or Turkey had not happened. Finally, the comparison of hypothetical to actual trade flows provides an estimate for the effect of one country's civil war on its and *all other countries'* trade. These computations require a symmetric dataset; i.e. trade flows must be provided for all potential dyads in the sample in every year and always in both directions. Further, for all countries, information on positive internal trade flows must be included in every year. We follow Baier et al. (2019) to adjust our main dataset accordingly.[24] In the end, we receive a dataset that contains 81 countries and covers the years 1993–2015.

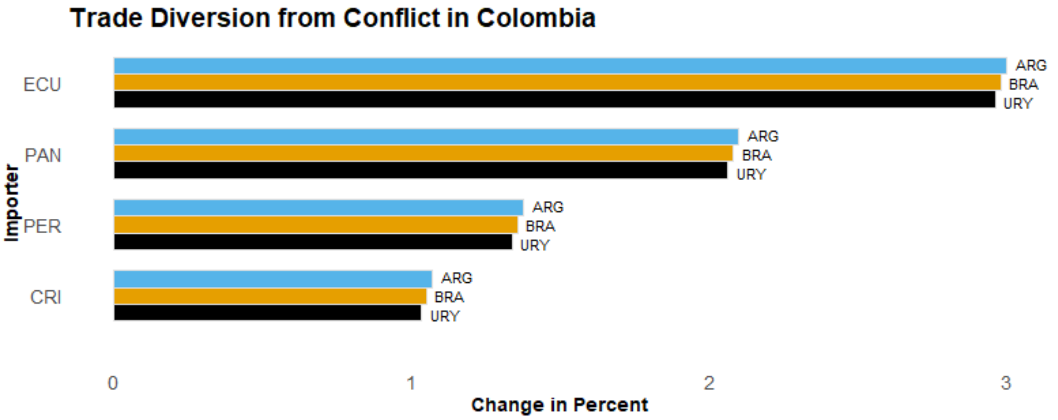
[22] The results are equivalent when estimating the effect of conflict and then computing hypothetical trade flows during a peace year. However, when comparing the hypothetical conflict scenario to the actual peace outcomes, the resulting diversion estimates would have to be inverted to show the trade diversion effects from conflict as opposed to the trade diversion effects from peace. To present the unchanged results, we hence estimate the effect of peace instead of conflict for our GE computations.

[23] We run these computations via the "ge_gravity" Stata Command provided by Thomas Zylkin and discussed in Baier, Yotov and Zylkin (2019).

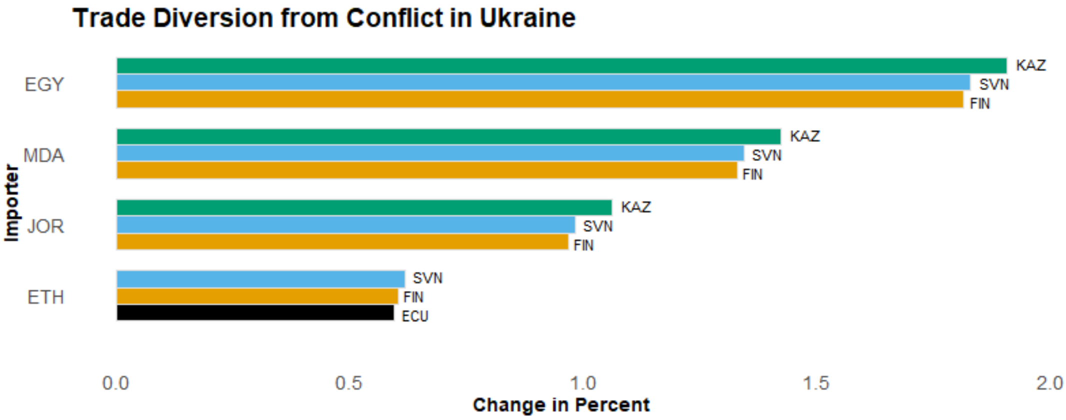
[24] Appendix 3.A.5 provides more details on the dataset construction.

Figure 3.A.8: GE results: Trade diversion

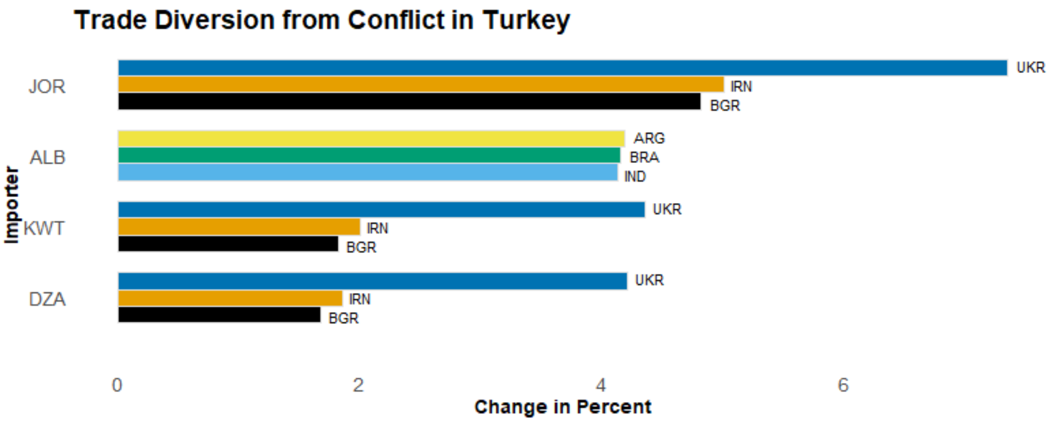
(a) Trade changes, conflict in Colombia



(b) Trade changes, conflict in Ukraine



(c) Trade changes, conflict in Turkey



Notes: The graphs report the estimated trade changes in the general equilibrium due to the civil wars in Colombia (Panel a), Ukraine (Panel b), and Turkey (Panel c). See Table 3.A.4 for the respective PE results.

Figure 3.A.8 presents the results of the GE analyses for our three case studies. Each panel of Figure 3.A.8 reports, for each of the four importers most affected by trade diversion, the export changes for the three origins with the largest export changes. In

the case of Colombia, for example, primarily its neighbors Ecuador, Panama and Peru increased imports from various countries by up to three percent. The picture of the beneficiary exporters for Colombia is quite homogeneous. To all four destinations, Argentina increased its shipments the most, closely followed by Brazil and Uruguay. The effects of the civil wars in Ukraine and Turkey had a larger geographic reach, as both countries are important exporters for Northern African and Middle Eastern countries. During the civil war in Ukraine, mainly other regional exporters increased their shipments to the former destinations of Ukrainian exports. Kazakhstan, Slovenia, and Finland register the largest export increases to Egypt, Moldova and Jordan. In response to the civil war in Turkey, Jordan, Albania, Kuwait and Algeria registered the largest trade diversion effects. Here however, the group of affected origins is more heterogeneous. Jordan, Kuwait and Algeria mainly turned towards Ukraine, Iran and Bulgaria to substitute for Turkish shipments. Albania, on the other hand, instead increased its shipments rather from large but non-regional suppliers, i.e. Argentina, Brazil, and India.

Trade diversion does however not totally mitigate the welfare loss from civil conflict. In Figure 3.A.1, we report the estimated international welfare changes in response to the civil conflicts in Colombia, Ukraine and Turkey. We mostly find negative welfare changes, with the biggest losses borne by the conflict-countries as well as the importers mainly affected. Indeed, even the benefiting exporters bear welfare losses, meaning that their increase in exports could not offset the loss of imports from and exports to the conflict countries. Overall, this emphasizes that, even though trade diversion can mitigate the effects of conflict, global welfare still decreases.

Bibliography

Abid, M., Ali, A., Raza, M., Mehdi, M., et al., 2020. Ex-ante and Ex-Post Coping Strategies for Climatic Shocks and Adaptation Determinants in Rural Malawi. Climate Risk Management 27, 100200.

Acemoglu, D., Akcigit, U., Kerr, W., 2016. Networks and the Macroeconomy: An Empirical Exploration. NBER Macroeconomics Annual 30, 273–335.

Acemoglu, D., Lelarge, C., Restrepo, P., 2020. Competing with Robots: Firm-Level Evidence from France. AEA Papers and Proceedings .

Acemoglu, D., Restrepo, P., 2018. Low-Skill and High-Skill Automation. Journal of Human Capital 12, 204–232.

Acemoglu, D., Restrepo, P., 2019a. Artificial Intelligence, Automation, and Work. University of Chicago Press.

Acemoglu, D., Restrepo, P., 2019b. Automation and New Tasks: How Technology Displaces and Reinstates Labor. Journal of Economic Perspective Spring 2019, 3–30.

Acemoglu, D., Restrepo, P., 2020. Robots and Jobs: Evidence from US Labor Markets. Journal of Political Economy 128, 2188–2244.

Acevedo, S., Mrkaic, M., Novta, N., Pugacheva, E., Topalova, P., 2020. The Effects of Weather Shocks on Economic Activity: What are the Channels of Impact? Journal of Macroeconomics 65, 103207.

Adão, R., Kolesár, M., Morales, E., 2019. Shift-Share Designs: Theory and Inference. The Quarterly Journal of Economics 134, 1949–2010.

Aghion, P., Antonin, C., Bunel, S., Jaravel, X., 2020. What Are the Labor and Product Market Effects of Automation? New Evidence from France .

Ahsan, R.N., Iqbal, K., 2020. How Does Violence Affect Exporters? Evidence from Political Strikes in Bangladesh. Review of International Economics 28, 599–625.

Aker, J.C., Mbiti, I.M., 2010. Mobile Phones and Economic Development in Africa. Journal of Economic Perspectives 24, 207–32.

Akresh, R., Bhalotra, S., Leone, M., Osili, U.O., 2012. War and Stature: Growing Up During the Nigerian Civil War. American Economic Review 102, 273–277.

Albanesi, S., Vamossy, D.F., 2019. Predicting Consumer Default: A Deep Learning Approach. Working Paper 26165. National Bureau of Economic Research.

Alesina, A., Miano, A., Stantcheva, S., 2018. Immigration and Redistribution .

Allen, T., Atkin, D., 2016. Volatility and the Gains from Trade. Technical Report. National Bureau of Economic Research.

Allen, T., Donaldson, D., 2020. Persistence and Path Dependence in the Spatial Economy. National Bureau of Economic Research Working Paper Series .

Almeida, R.K., Corseuil, C.H.L., Poole, J.P., 2017. The Impact of Digital Technologies on Routine Tasks: Do Labor Policies Matter? The World Bank.

Alobo Loison, S., 2015. Rural Livelihood Diversification in Sub-Saharan Africa: a Literature Review. The Journal of Development Studies 51, 1125–1138.

Alvarez, J., Benguria, F., Engbom, N., Moser, C., 2018. Firms and the decline in earnings inequality in Brazil. American Economic Journal: Macroeconomics 10, 149–189.

Amiti, M., Konings, J., 2007. Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia. American Economic Review 97, 1611–1638.

Amodio, F., Baccini, L., Di Maio, M., 2020. Security, Trade, and Political Violence. Journal of the European Economic Association 19, 1–37.

Andersen, J.J., Nordvik, F.M., Tesei, A., 2017. Oil and Civil Conflict: On and Off (Shore). CESifo Group Munich .

Anderson, J.E., 1979. A Theoretical Foundation for the Gravity Equation. American Economic Review 69, 106–116.

Anderson, J.E., Larch, M., Yotov, Y.V., 2018a. GEPPML: General Equilibrium Analysis with PPML. The World Economy 41, 2750–2782.

Anderson, J.E., Larch, M., Yotov, Y.V., 2018b. GEPPML: General Equilibrium Analysis with PPML. The World Economy 41, 2750–2782.

Anderson, J.E., van Wincoop, E., 2003. Gravity with Gravitas: A Solution to the Border Puzzle. American Economic Review 93, 170–192.

Antwi-Agyei, P., Dougill, A.J., Stringer, L.C., Codjoe, S.N.A., 2018. Adaptation Opportunities and Maladaptive Outcomes in Climate Vulnerability Hotspots of Northern Ghana. Climate Risk Management 19, 83–93.

Anyamba, A., Tucker, C.J., Mahoney, R., 2002. From El Niño to La Niña: Vegetation Response Patterns over East and Southern Africa during the 1997–2000 Period. Journal of climate 15, 3096–3103.

Apella, I., Zunino, G., 2017. Technological Change and the Labor Market in Argentina and Uruguay: a Task Content Analysis. World Bank Policy Research Working Paper .

Armington, P.S., 1969. A Theory of Demand for Products Distinguished by Place of Production. Staff Papers (International Monetary Fund) 16, 159–178.

Artuc, E., Bastos, P., Rijkers, B., 2019a. Robots, Tasks and Trade. World Bank Policy Research Working Paper .

Artuc, E., Christiaensen, L., Winkler, H.J., 2019b. Does Automation in Rich Countries Hurt Developing Ones?: Evidence from the US and Mexico .

Asfaw, S., Scognamillo, A., Di Caprera, G., Sitko, N., Ignaciuk, A., 2019. Heterogeneous Impact of Livelihood Diversification on Household Welfare: Cross-Country Evidence from Sub-Saharan Africa. World Development 117, 278–295.

Autor, D.H., Dorn, D., Hanson, G.H., 2013. The China Syndrome: Local labor Market Effects of Import Competition in the United States. American Economic Review 103, 2121–2168.

Bahar, D., Santos, M.A., 2018. One More Resource Curse: Dutch Disease and Export Concentration. Journal of Development Economics 132, 102–114.

Baier, S.L., Yotov, Y.V., Zylkin, T., 2019. On the Widely Differing Effects of Free Trade Agreements: Lessons from Twenty Years of Trade Integration. Journal of International Economics 116, 206–226.

Banerjee, A.V., Duflo, E., 2019. Good Economics for Hard Times. PublicAffairs.

Barrett, C.B., Reardon, T., Webb, P., 2001. Nonfarm Income Diversification and Household Livelihood Strategies in rural Africa: Concepts, Dynamics, and Policy Implications. Food Policy 26, 315–331.

Barrios, S., Ouattara, B., Strobl, E., 2008. The Impact of Climatic Change on Agricultural Production: Is it Different for Africa? Food Policy 33, 287–298.

Barro, R.J., Sala-i-Martin, X., 1992. Convergence. Journal of Political Economy 100, 223–251.

Barrot, J.N., Sauvagnat, J., 2016. Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks. The Quarterly Journal of Economics 131, 1543–1592.

Basedau, M., Lay, J., 2009. Resource Curse or Rentier Peace? The Ambiguous Effects of Oil Wealth and Oil Dependence on Violent Conflict. Journal of Peace Research 46, 757–776.

Battaglia, M., Gulesci, S., Madestam, A., 2018. Repayment Flexibility and Risk Taking: Experimental Evidence from Credit Contracts. Centre for Economic Policy Research.

Bayer, P., Kennedy, R., Yang, J., Urpelainen, J., 2020. The Need for Impact Evaluation in Electricity Access Research. Energy Policy 137, 111099.

Bayer, R., Rupert, M., 2004. Effects of Civil Wars on International Trade, 1950-92. Journal of Peace Research 41, 699–713.

Bazzi, S., Blattman, C., 2014. Economic Shocks and Conflict: Evidence from Commodity Prices. American Economic Journal: Macroeconomics 6, 1–38.

Beegle, K., Dehejia, R.H., Gatti, R., 2006. Child Labor and Agricultural Shocks. Journal of Development Economics 81, 80–96.

Beg, S., 2021. Digitization and Development: Property Rights Security, and Land and Labor Markets. Journal of the European Economic Association .

Benedictis, L.D., Tajoli, L., 2007a. Economic Integration and Similarity in Trade Structures. Empirica 34, 117–137.

Benedictis, L.D., Tajoli, L., 2007b. Openness, Similarity in Export Composition, and Income Dynamics. The Journal of International Trade & Economic Development 16, 93–116.

Benguria, F., Ederington, J., 2017. Decomposing the Effect of Trade on the Gender Wage Gap .

Berger, D., Easterly, W., Nunn, N., Satyanath, S., 2013. Commercial Imperialism? Political Influence and Trade during the Cold War. American Economic Review 103, 863–96.

Berman, N., Couttenier, M., Rohner, D., Thoenig, M., 2017. This Mine Is Mine! How Minerals Fuel Conflicts in Africa. American Economic Review 107, 1564–1610.

Bessen, J.E., Goos, M., Salomons, A., van den Berge, W., 2019. Automatic Reaction-What Happens to Workers at Firms that Automate? Boston Univ. School of Law, Law and Economics Research Paper .

Beverelli, C., Keck, A., Larch, M., Yotov, Y., 2018. Institutions, Trade and Development: A Quantitative Analysis. Drexel University Working Paper Series, WP 2018-03 .

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., Jones, Z.M., 2016. mlr: Machine Learning in R. The Journal of Machine Learning Research 17, 5938–5942.

Blakeslee, D., Fishman, R., Srinivasan, V., 2020. Way Down in the Hole: Adaptation to Long-Term Water Loss in Rural India. American Economic Review 110, 200–224.

Blattman, C., 2012. Post-conflict Recovery in Africa: The Micro Level. Oxford Companion to the Economics of Africa , 124–130.

Blattman, C., Miguel, E., 2010. Civil War. Journal of Economic Literature 48, 3–57.

Boehm, C.E., Flaaen, A., Pandalai-Nayar, N., 2019. Input Linkages and the Transmission of Shocks: Firm-level Evidence from the 2011 Tōhoku Earthquake. Review of Economics and Statistics 101, 60–75.

Bonfiglioli, A., Crino, R., Gancia, G., Papadakis, I., 2021. Robots, Offshoring and Welfare. Working Paper .

Borusyak, K., Hull, P., Jaravel, X., 2018. Quasi-Experimental Shift-Share Research Designs. National Bureau of Economic Research .

Brakman, S., Garretsen, H., Schramm, M., 2004. The Strategic Bombing of German Cities During World War II and its Impact on City Growth. Journal of Economic Geography 4, 201–218.

Brambilla, I., César, A., Falcone, G., Gasparini, L., Lombardo, C., 2021. The Risk of Automation in Latin America. Documentos de Trabajo del CEDLAS .

Brambilla, I., Tortarolo, D., 2018. Investment in ICT, Productivity, and Labor Demand: the Case of Argentina. World Bank Policy Research Working Paper .

Branco, D., Féres, J., 2021. Weather Shocks and Labor Allocation: Evidence from Rural Brazil. American Journal of Agricultural Economics 103, 1359–1377.

Brunnschweiler, C.N., Bulte, E.H., 2009. Natural Resources and Violent Conflict: Resource Abundance, Dependence, and the Onset of Civil Wars. Oxford Economic Papers 61, 651–674.

Bruno, V., Shin, H.S., 2015. Cross-border Banking and Global Liquidity. The Review of Economic Studies 82, 535–564.

Bryan, E., Deressa, T.T., Gbetibouo, G.A., Ringler, C., 2009. Adaptation to Climate Change in Ethiopia and South Africa: Options and Constraints. Environmental Science & Policy 12, 413–426.

Brück, T., Di Maio, M., Miaari, S.H., 2019. Learning The Hard Way: The Effect of Violent Conflict on Student Academic Achievement. Journal of the European Economic Association 17, 1502–1537.

Burgess, R., Donaldson, D., 2010. Can Openness Mitigate the Effects of Weather Shocks? Evidence from India's Famine Era. American Economic Review 100, 449–53.

Burke, M., Bergquist, L.F., Miguel, E., 2019. Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets. The Quarterly Journal of Economics 134, 785–842.

Bustos, P., Garber, G., Ponticelli, J., 2020. Capital Accumulation and Structural Transformation. The Quarterly Journal of Economics 135, 1037–1094.

Caliendo, L., Dvorkin, M., Parro, F., 2015. The Impact of Trade on Labor Market Dynamics. National Bureau of Economic Research .

Caliendo, L., Parro, F., 2015. Estimates of the Trade and Welfare Effects of NAFTA. The Review of Economic Studies 82, 1–44.

Call, M., Gray, C., Jagger, P., 2019. Smallholder Responses to Climate Anomalies in Rural Uganda. World Development 115, 132–144.

Carbonero, F., Ernst, E., Weber, E., 2018. Robots Worldwide: The Impact of Automation on Employment and Trade. ILO Research Working Paper .

Caselli, M., Fracasso, A., Traverso, S., 2019. Globalization, Robotization and Electoral Outcomes: Evidence from Spatial Regressions for Italy. Journal of Regional Science .

Castro Iragorri, C., Garcia, K., 2014. Default Risk in Agricultural Lending, the Effects of Commodity Price Volatility and Climate. Agricultural Finance Review 74, 501–521.

Cattaneo, C., Peri, G., 2016. The Migration Response to Increasing Temperatures. Journal of Development Economics 122, 127–146.

Chantarat, S., Mude, A.G., Barrett, C.B., Carter, M.R., 2013. Designing Index-based Livestock Insurance for Managing Asset Risk in Northern Kenya. Journal of Risk and Insurance 80, 205–237.

Chen, T., Guestrin, C., 2016. Xgboost: A Scalable Tree Boosting System. Proceedings of the 22nd acm sigkdd International Conference on Knowledge Discovery and Data Mining .

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., contributors, X., 2020. R Package: xgboost Https://cran.r-project.org/web/packages/xgboost/index.html.

Cheong, J., Kwak, D.W., Tang, K.K., 2015. It Is Much Bigger Than What We Thought: New Estimate of Trade Diversion. The World Economy 38, 1795–1808.

Chuang, Y., 2019. Climate Variability, Rainfall Shocks, and Farmers' Income Diversification in India. Economics Letters 174, 55–61.

Cilekoglu, A., Moreno, R., Ramos, R., 2021. The Impact of Robot Adoption on Global Sourcing. Working Paper .

Collier, B., Katchova, A.L., Skees, J.R., 2011. Loan Portfolio Performance and El Niño, an Intervention Analysis. Agricultural Finance Review 71, 98–119.

Collier, P., Hegre, H., Hoeffler, A., Reynal-Querol, M., Sambanis, N., 2003. Breaking the Conflict Trap: Civil War and Development Policy. World Bank Publications.

Colmer, J., 2021. Rainfall Variability, Child Labor, and Human Capital Accumulation in Rural Ethiopia. American Journal of Agricultural Economics 103, 858–877.

Cortes, G.M., Morris, D.M., 2020. Are Routine Jobs Moving South? Evidence from Changes in the Occupational Structure of Employment in the USA and Mexico. volume 2020/11. WIDER Working Paper.

Costa, F., Garred, J., Pessoa, J.P., 2016. Winners and Losers from a Commodities-for-Manufactures Trade Boom. Journal of International Economics 102, 50–69.

Czaika, M., De Haas, H., 2014. The Globalization of Migration: Has the World Become More Migratory? International Migration Review 48, 283–323.

Czura, K., 2015. Do Flexible Repayment Schedules Improve the Impact of Microcredit? Evidence from a Randomized Evaluation in Rural India. Munich Discussion Paper .

Dai, L., Eden, L., Beamish, P.W., 2017. Caught in the Crossfire: Dimensions of Vulnerability and Foreign Multinationals' Exit from War-Afflicted Countries. Strategic Management Journal 38, 1478–1498.

Dai, M., Yotov, Y., Zylkin, T., 2014. On the Trade-diversion Effects of Free Trade Agreements. Economics Letters 122, 321–325.

Dauth, W., Findeisen, S., Südekum, J., Woessner, N., 2017. German Robots-The Impact of Industrial Robots on Workers .

Davis, D.R., Weinstein, D.E., 2002. Bones, Bombs, and Break Points: The Geography of Economic Activity. American Economic Review 92, 1269–1289.

Davis, R.J., Holladay, J.S., Sims, C., 2021. Coal-Fired Power Plant Retirements in the U.S. Working Paper 28949. National Bureau of Economic Research.

De Sousa, J., Mirza, D., Verdier, T., 2018. Terrorism Networks and Trade: Does the Neighbor Hurt? European Economic Review 107, 27–56.

Dercon, S., Krishnan, P., 1996. Income Portfolios in Rural Ethiopia and Tanzania: Choices and Constraints. The Journal of Development Studies 32, 850–875.

Dercon, S., Krishnan, P., 2007. Vulnerability, Seasonality and Poverty in Ethiopia. The Journal of Development Studies 36, 25–53.

Dinkelman, T., 2011. The Effects of Rural Electrification on Employment: New Evidence from South Africa. American Economic Review 101, 3078–3108.

Dix-Carneiro, R., Kovak, B.K., 2015. Trade Liberalization and the Skill Premium: A Local Labor Markets Approach. American Economic Review 105, 551–557.

Dix-Carneiro, R., Kovak, B.K., 2017. Trade Liberalization and Regional Dynamics. American Economic Review 107, 2908–2946.

Dreher, A., 2006. Does Globalization Affect Growth? Evidence from a New Index of Globalization. Applied Economics 38, 1091–1110.

Dreher, A., Langlotz, S., 2020. Aid and Growth: New Evidence Using an Excludable Instrument. Canadian Journal of Economics/Revue Canadienne d'Économique 53, 1162–1198.

Dunning, C.M., Black, E.C.L., Allan, R.P., 2016. The Onset and Cessation of Seasonal Rainfall over Africa. Journal of Geophysical Research: Atmospheres 121, 11,405–11,424.

Dustmann, C., Vasiljeva, K., Piil Damm, A., 2019. Refugee Migration and Electoral Outcomes. The Review of Economic Studies 86, 2035–2091.

Eaton, J., Kortum, S., 2002. Technology, Geography, and Trade. Econometrica 70, 1741–1779.

Egger, H., Egger, P., Greenaway, D., 2008. The Trade Structure Effects of Endogenous Regional Trade Agreements. Journal of International Economics 74, 278–298.

Emerick, K., 2018. Trading Frictions in Indian Village Economies. Journal of Development Economics 132, 32–56.

Emran, M.S., 2005. Revenue-increasing and Welfare-enhancing Reform of Taxes on Exports. Journal of Development Economics 77, 277–292.

Etea, B.G., Zhou, D., Abebe, K.A., Sedebo, D.A., 2019. Is Income Diversification a Means of Survival or Accumulation? Evidence from Rural and Semi-Urban Households in Ethiopia. Environment, Development and Sustainability , 1–19.

Faber, M., 2020. Robots and Reshoring: Evidence from Mexican Labor Markets. Journal of International Economics 127, 103384.

Fally, T., Sayre, J., 2018. Commodity Trade Matters. National Bureau of Economic Research .

Farzanegan, M.R., Lessmann, C., Markwardt, G., 2018. Natural Resource Rents and Internal Conflicts: Can Decentralization Lift the Curse? Economic Systems 42, 186–205.

Felbermayr, G., Syropoulos, C., Yalcin, E., Yotov, Y., 2019a. On the Effects of Sanctions on Trade and Welfare: New Evidence Based on Structural Gravity and a New Database. LeBow College of Business Working Paper Series, Drexel University .

Felbermayr, G., Syropoulos, C., Yalcin, E., Yotov, Y., 2019b. On the Effects of Sanctions on Trade and Welfare: New Evidence Based on Structural Gravity and a New Database. LeBow College of Business Working Paper Series, Drexel University .

Fink, G., Jack, B.K., Masiye, F., 2020. Seasonal Liquidity, Rural Labor Markets, and Agricultural Production. American Economic Review 110, 3351–3392.

Freund, C., Mattoo, A., Mulabdic, A., Ruta, M., 2021. Natural Disasters and the Reshaping of Global Value Chains. Policy Research Working Paper Series 9719. The World Bank.

Frey, C.B., Osborne, M.A., 2017. The Future of Employment: How Susceptible are Jobs to Computerisation? Technological Forecasting and Social Change 114, 254–280.

Fuchs, A., Klann, N.H., 2013. Paying a Visit: The Dalai Lama Effect on International Trade. Journal of International Economics 91, 164–177.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., 2015. The Climate Hazards Infrared Precipitation with Stations—A new Environmental Record for Monitoring Extremes. Scientific Data 2, 1–21.

Garfinkel, M., Syropoulos, C., Zylkin, T., 2020a. Prudence vs. Predation and the Gains from Trade. Drexel University Working Paper Series, WP 2020-06 .

Garfinkel, M.R., Syropoulos, C., Yotov, Y.V., 2020b. Arming in the Global Economy: The Importance of Trade with Enemies and Friends. Journal of International Economics 123, 103295.

Gasparini, L., Brambilla, I., Falcone, G., Lombardo, C., César, A., 2021. Routinization and Employment: Evidence for Latin America Leonardo. Documentos de Trabajo del CEDLAS .

Giuliano, P., Ruiz-Arranz, M., 2009. Remittances, Financial Development, and Growth. Journal of Development Economics 90, 144–152.

Goldberg, P.K., Pavcnik, N., 2007. Distributional Effects of Globalization in Developing Countries. Journal of Economic Literature 45, 39–82.

Goldsmith-Pinkham, P., Sorkin, I., Swift, H., 2020. Bartik Instruments: What, When, Why, and How. American Economic Review 110, 2586–2624.

Graetz, G., Michaels, G., 2018. Robots at Work. Review of Economics and Statistics 100, 753–768.

Gray, C.L., Mueller, V., 2012. Natural Disasters and Population Mobility in Bangladesh. Proceedings of the National Academy of Sciences 109, 6000–6005.

Grimm, M., Lenz, L., Peters, J., Sievert, M., 2020. Demand for Off-Grid Solar Electricity: Experimental Evidence from Rwanda. Journal of the Association of Environmental and Resource Economists 7, 417–454.

Grohmann, A., Herbold, S., Lenel, F., 2020. Repayment under Flexible Loan Contracts: Evidence from Tanzania .

Grubert, H., Mutti, J., 1991. Taxes, Tariffs and Transfer Pricing in Multinational Corporate Decision Making. The Review of Economics and Statistics , 285–293.

Hakobyan, S., McLaren, J., 2016. Looking for local labor market effects of NAFTA. Review of Economics and Statistics 98, 728–741.

Hallegatte, S., Bangalore, M., Bonzanigo, L., Fay, M., Kane, T., Narloch, U., Rozenberg, J., Treguer, D., Vogt-Schilb, A., 2015. Shock Waves: Managing the Impacts of Climate Change on Poverty. The World Bank.

Hardy, W., Keister, R., Lewandowski, P., 2018. Educational Upgrading, Structural Change and the Task Composition of Jobs in Europe. Economics of Transition 26, 201–231.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society. Series (Applied Statistics) 28, 100–108.

Head, K., Mayer, T., 2014. Gravity Equations: Workhorse, Toolkit, and Cookbook. Handbook of International Economics 4, 131–195.

Helpman, E., Itskhoki, O., Muendler, M.A., Redding, S.J., 2017. Trade and Inequality: From Theory to Estimation. The Review of Economic Studies 84, 357–405.

Hirata, G., Soares, R.R., 2016. Competition and the Racial Wage Gap: Testing Becker's Model of Employer Discrimination .

Hjort, J., Poulsen, J., 2019. The Arrival of Fast Internet and Employment in Africa. American Economic Review 109, 1032–79.

Hyland, M., Russ, J., 2019. Water as Destiny–The Long-Term Impacts of Drought in Sub-Saharan Africa. World Development 115, 30–45.

IEA, 2019. Africa Energy Outlook 2019 .

International Federation of Robotics, 2018. World Robotics: Industrial Robots .

IPCC, 2019. Climate Change and Land: an IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems. [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press.

IPCC, 2021. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press. In Press.

Jayachandran, S., 2006. Selling Labor Low: Wage Responses to Productivity Shocks in Developing Countries. Journal of Political Economy 114, 538–575.

Jenkins, R., 2015. Is Chinese Competition Causing Deindustrialization in Brazil? Latin American Perspectives 42, 42–63.

Jessoe, K., Manning, D.T., Taylor, J.E., 2018. Climate Change and Labour Allocation in Rural Mexico: Evidence from Annual Fluctuations in Weather. The Economic Journal 128, 230–261.

Jones, B.F., Olken, B.A., 2010. Climate Shocks and Exports. American Economic Review 100, 454–59.

Joshi, K., 2019. The Impact of Drought on Human Capital in Rural India. Environment and Development Economics 24, 413–436.

Justino, P., Verwimp, P., 2013. Poverty Dynamics, Violent Conflict, and Convergence in Rwanda. Review of Income and Wealth 59, 66–90.

Kaminski, J., Christiaensen, L., Gilbert, C.L., 2016. Seasonality in Local Food Markets and Consumption: Evidence from Tanzania. Oxford Economic Papers 68, 736–757.

Karnieli, A., Agam, N., Pinker, R.T., Anderson, M., Imhoff, M.L., Gutman, G.G., Panov, N., Goldberg, A., 2010. Use of ndvi and land surface temperature for drought assessment: Merits and limitations. Journal of Climate 23, 618–633.

Kim, I.S., Liao, S., Imai, K., 2020. Measuring Trade Profile with Granular Product-Level Data. American Journal of Political Science 64, 102–117.

Kis-Katos, K., Pieters, J., Sparrow, R., 2018. Globalization and Social Change: Gender-specific Effects of Trade Liberalization in Indonesia. IMF Economic Review 66, 763–793.

Kjellstrom, T., Briggs, D., Freyberg, C., Lemke, B., Otto, M., Hyatt, O., 2016. Heat, Human Performance, and Occupational Health: a Key Issue for the Assessment of Global Climate Change Impacts. Annual Review of Public Health 37, 97–112.

Koch, M., Manuylov, I., Smolka, M., 2019. Robots and Firms. Working Paper .

Kotir, J.H., 2011. Climate Change and Variability in Sub-Saharan Africa: a Review of Current and Future Trends and Impacts on Agriculture and Food Security. Environment, Development and Sustainability 13, 587–605.

Kovak, B.K., 2013. Regional Effects of Trade Reform: What is the Correct Measure of Liberalization? American Economic Review 103, 1960–1976.

Krenz, A., Prettner, K., Strulik, H., 2021. Robots, reshoring, and the lot of low-skilled workers. European Economic Review 136, 103744.

Krugman, P., Taylor, L., 1978. Contractionary Effects of Devaluation. Journal of International Economics 8, 445–456.

Ksoll, C., Macchiavello, R., Morjaria, A., 2018. Guns and Roses: Flower Exports and Electoral Violence in Kenya. Global Poverty Research Lab Working Paper No. 17–102.

Kunst, D., 2020. Premature Deindustrialization through The Lens of Occupations: Which Jobs, Why, and Where? SSRN Scholarly Paper .

Lee, K., Miguel, E., Wolfram, C., 2020a. Does Household Electrification Supercharge Economic Development? Journal of Economic Perspectives 34, 122–144.

Lee, K., Miguel, E., Wolfram, C., 2020b. Experimental Evidence on the Economics of Rural Electrification. Journal of Political Economy 128, 1523–1565.

Lemaire, X., 2018. Solar Home Systems and Solar Lanterns in Rural Areas of the Global South: What Impact? Wiley Interdisciplinary Reviews: Energy and Environment 7, e301.

Lenz, L., Munyehirwe, A., Peters, J., Sievert, M., 2017. Does Large-Scale Infrastructure Investment Alleviate Poverty? Impacts of Rwanda's Electricity Access Roll-Out Program. World Development 89, 88–110.

Liebmann, B., Bladé, I., Kiladis, G.N., Carvalho, L.M.V., B. Senay, G., Allured, D., Leroux, S., Funk, C., 2012. Seasonality of African Precipitation from 1996 to 2009. Journal of Climate 25, 4304–4322.

Lipscomb, M., Mobarak, A.M., Barham, T., 2013. Development Effects of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil. American Economic Journal: Applied Economics 5, 200–231.

Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate Trends and Global Crop Production since 1980. Science 333, 616–620.

Long, A.G., 2008. Bilateral Trade in the Shadow of Armed Conflict. International Studies Quarterly 52, 81–101.

Lybbert, T.J., McPeak, J., 2012. Risk and Intertemporal Substitution: Livestock Portfolios and Off-Take among Kenyan Pastoralists. Journal of Development Economics 97, 415–426.

Macours, K., Premand, P., Vakis, R., 2012. Transfers, Diversification and Household Risk Strategies: Experimental Evidence with Lessons for Climate Change Adaptation. World Bank Policy Research Working Paper .

Maloney, W.F., Molina, C., 2016. Are Automation and Trade Polarizing Developing Country Labor Markets, too? World Bank Policy Research Working Paper .

Mankiw, N.G., Romer, D., Weil, D.N., 1992. A Contribution to the Empirics of Economic Growth. The Quarterly Journal of Economics 107, 407–437.

Martin, P., Mayer, T., Thoenig, M., 2008a. Civil Wars and International Trade. Journal of the European Economic Association 6, 541–550.

Martin, P., Mayer, T., Thoenig, M., 2008b. Make Trade Not War? The Review of Economic Studies 75, 865–900.

Martin, P., Mayer, T., Thoenig, M., 2012. The Geography of Conflicts and Regional Trade Agreements. American Economic Journal: Macroeconomics 4, 1–35.

Mattoo, A., Mulabdic, A., Ruta, M., Mattoo, A., Mulabdic, A., Ruta, M., 2017. Trade Creation and Trade Diversion in Deep Agreements. Policy Research Working Paper Series, The World Bank .

McLuhan, M., Fiore, Q., et al., 1968. War and Peace in the Global Village .

Meroni, M., Fasbender, D., Rembold, F., Atzberger, C., Klisch, A., 2019. Near Real-Time Vegetation Anomaly Detection with MODIS NDVI: Timeliness vs. Accuracy and Effect of Anomaly Computation Options. Remote Sensing of Environment 221, 508–521.

Messina, J., Silva, J., 2021. Twenty years of Wage Inequality in Latin America. The World Bank Economic Review 35, 117–147.

Micco, A., 2019. Automation, Labor Markets, and Trade. Working Paper .

Miguel, E., Roland, G., 2011. The Long-run Impact of Bombing Vietnam. Journal of Development Economics 96, 1–15.

Mirza, D., Verdier, T., 2014. Are Lives a Substitute for Livelihoods? Terrorism, Security, and US Bilateral Imports. Journal of Conflict Resolution 58, 943–975.

Mkhabela, M.S., Mkhabela, M.S., Mashinini, N.N., 2005. Early Maize Yield Forecasting in the four Agro-Ecological Regions of Swaziland using NDVI Data derived from NOAA's-AVHRR. Agricultural and Forest Meteorology 129, 1–9.

Molnar, C., 2022. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.

Moore, G.E., et al., 1965. Cramming more Components onto Integrated Circuits .

Mueller, V., Gray, C., Kosec, K., 2014. Heat Stress Increases Long-Term Human Migration in Rural Pakistan. Nature Climate Change 4, 182–185.

Muendler, M.A., Poole, J., Ramey, G., Wajnberg, T., 2004. Job Concordances for Brazil: Mapping the Classificaçao Brasileira de Ocupaçoes (CBO) to the International Standard Classification of Occupations (ISCO-88). unpublished manuscript .

Nash, J., Halewood, N., Melhem, S., 2013. Unlocking Africa's Agricultural Potential: an Action Agenda for Transformation .

NOAA STAR, 2018. Global Vegetation Health Products. Center for Satellite Applications and Research, NOAA, USA .

Novta, N., Pugacheva, E., 2021. The Macroeconomic Costs of Conflict. Journal of Macroeconomics 68, forthcoming.

OECD & FAO, 2016. Agriculture in Sub-Saharan Africa: Prospects and Challenges for the Next Decade. OECD-FAO agricultural outlook 2016-2025 .

Oetzel, J., Miklian, J., 2017. Multinational Enterprises, Risk Management, and the Business and Economics of Peace. Multinational Business Review 25, 270–286.

Oster, E., 2018. Diabetes and Diet: Purchasing Behavior Change in Response to Health Information. American Economic Journal: Applied Economics 10, 308–48.

Pape, U., Wollburg, P., 2019. Impact of Drought on Poverty in Somalia .

Pavcnik, N., Blom, A., Goldberg, P., Schady, N., 2004. Trade Liberalization and Industry wage Structure: Evidence from Brazil. The World Bank Economic Review 18, 319–344.

Pelka, N., Musshoff, O., Weber, R., 2015. Does Weather Matter? How Rainfall Affects Credit Risk in Agricultural Microfinance. Agricultural Finance Review 75, 194–212.

Peters, A.J., Walter-Shea, E.A., Ji, L., Vina, A., Hayes, M., Svoboda, M.D., 2002. Drought Monitoring with NDVI-Based Standardized Vegetation Index. Photogrammetric Engineering and Remote Sensing 68, 71–75.

Peters, J., Sievert, M., 2016. Impacts of Rural Electrification Revisited–the African Context. Journal of Development Effectiveness 8, 327–345.

Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.M., Tucker, C.J., Stenseth, N.C., 2005. Using the Satellite-Derived NDVI to Assess Ecological responses to Environmental Change. Trends in Ecology & Evolution 20, 503–510.

Porter, C., 2012. Shocks, Consumption and Income Diversification in Rural Ethiopia. Journal of Development Studies 48, 1209–1222.

Qureshi, M.S., 2013. Trade and Thy Neighbor's War. Journal of Development Economics 105, 178–195.

Rey, H., 2015. Dilemma not Trilemma: the Global Financial Cycle and Monetary Policy Independence .

Ricardo, D., 1817. From the Principles of Political Economy and Taxation .

Rodrik, D., 2016. Premature Deindustrialization. Journal of Economic Growth 21, 1–33.

Rodrik, D., 2018. New Technologies, Global Value Chains, and Developing Economies. National Bureau of Economic Research .

Roerink, G.J., Menenti, M., Soepboer, W., Su, Z., 2003. Assessment of Climate Impact on Vegetation Dynamics by Using Remote Sensing. Physics and Chemistry of the Earth, Parts A/B/C 28, 103–109.

Rohner, D., Thoenig, M., Zilibotti, F., 2013. War Signals: A Theory of Trade, Trust, and Conflict. The Review of Economic Studies 80, 1114–1147.

Roncoli, C., Ingram, K., Kirshen, P., 2001. The Costs and Risks of Coping with Drought: Livelihood Impacts and Farmers' Responses in Burkina Faso. Climate Research 19, 119–132.

Rose, A.K., 2018. Currency Wars? Unconventional Monetary Policy Does Not Stimulate Exports. National Bureau of Economic Research .

Ross, M.L., 2015. What Have We Learned about the Resource Curse? Annual Review of Political Science 18, 239–259.

de Roux, N., 2021. Exogenous Shocks, Credit Reports and Access to Credit: Evidence from Colombian Coffee Producers .

Salazar, C., Rand, J., 2020. Pesticide Use, Production Risk and Shocks. The Case of Rice Producers in Vietnam. Journal of Environmental Management 253, 109705.

Santos Silva, J.M.C., Tenreyro, S., 2006. The Log of Gravity. The Review of Economics and Statistics 88, 641–658.

Schlenker, W., Lobell, D.B., 2010. Robust Negative Impacts of Climate Change on African Agriculture. Environmental Research Letters 5, 014010.

Serdeczny, O., Adams, S., Baarsch, F., Coumou, D., Robinson, A., Hare, W., Schaeffer, M., Perrette, M., Reinhardt, J., 2017. Climate Change Impacts in Sub-Saharan Africa: from Physical Changes to their Social Repercussions. Regional Environmental Change 17, 1585–1600.

Shah, M., Steinberg, B.M., 2017. Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital. Journal of Political Economy 125, 527–561.

Simoes, A., Hidalgo, C.A., 2011. The Economic Complexity Observatory: An Analytical Tool for Understanding the Dynamics of Economic Development. Workshops at the twenty-fifth AAAI conference on artificial intelligence .

Spilimbergo, A., Srinivasan, K. (Eds.), 2019. Brazil: Boom, Bust, and Road to Recovery.

Sundberg, R., Melander, E., 2013. Introducing the UCDP Georeferenced Event Dataset. Journal of Peace Research 50, 523–532.

Suri, T., 2017. Mobile Money. Annual Review of Economics 9, 497–520.

The World Bank, 2020. World Development Indicators .

Thomas, D.R., Harish, S.P., Kennedy, R., Urpelainen, J., 2020. The Effects of Rural Electrification in India: An Instrumental Variable Approach at the Household Level. Journal of Development Economics 146, 102520.

Thomas, D.S.G., Twyman, C., Osbahr, H., Hewitson, B., 2007. Adaptation to Climate Change and Variability: Farmer Responses to Intra-Seasonal Precipitation Trends in South Africa. Climatic Change 83, 301–322.

Timmer, M.P., Dietzenbacher, E., Los, B., Stehrer, R., de Vries, G.J., 2015. An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production. Review of International Economics 23, 575–605.

Tiwari, S., Jacoby, H.G., Skoufias, E., 2017. Monsoon Babies: Rainfall Shocks and Child Nutrition in Nepal. Economic Development and Cultural Change 65, 167–188.

Topalova, P., Khandelwal, A., 2011. Trade Liberalization and Firm Productivity: The Case of India. Review of Economics and Statistics 93, 995–1009.

Tucker, C.J., 1979. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. Remote sensing of Environment 8, 127–150.

Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F., El Saleous, N., 2005. An Extended AVHRR 8-km NDVI Dataset Compatible with MODIS and SPOT Vegetation NDVI Data. International Journal of Remote Sensing 26, 4485–4498.

Tur-Prats, A., Valencia Caicedo, F., 2020. The Long Shadow of the Spanish Civil War. CEPR Discussion Paper DP15091.

van de Walle, D., Ravallion, M., Mendiratta, V., Koolwal, G., 2017. Long-Term Gains from Electrification in Rural India. The World Bank Economic Review 31, 385–411.

Vermote, E., 2015. MOD09A1 MODIS/Terra Surface Reflectance 8-Day L3 Global 500m SIN Grid V006 [Data set]. The NASA EOSDIS Land Processes DAAC .

Verwimp, P., Justino, P., Brück, T., 2019. The Microeconomics of Violent Conflict. Journal of Development Economics 141, 1–6.

Walter, B.F., 2004. Does Conflict Beget Conflict? Explaining Recurring Civil War. Journal of Peace Research 41, 371–388.

Wassie, Y.T., Adaramola, M.S., 2021. Socio-economic and Environmental Impacts of Rural Electrification with Solar Photovoltaic Systems: Evidence from Southern Ethiopia. Energy for Sustainable Development 60, 52–66.

Weisser, C., Lenel, F., Lu, Y., Kis-Katos, K., Kneib, T., 2021. Using Solar Panels for Business Purposes: Evidence Based on High-Frequency Power Usage Data. Development Engineering .

World Bank, 2015. Sustainable Energy for All 2015: Progress Toward Sustainable Energy. The World Bank.

Yotov, Y., 2021. The Variation of Gravity within Countries (or 15 Reasons Why Gravity Should Be Estimated with Domestic Trade Flows). CESifo Working Paper No. 9057 .

Yotov, Y.V., Piermartini, R., José-Antonio, Larch, M., 2016a. An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model. World Trade Organization .

Yotov, Y.V., Piermartini, R., Monteiro, J.A., Larch, M., 2016b. An advanced guide to trade policy analysis: The structural gravity model. World Trade Organization Geneva, Switzerland.

Zhuravskaya, E., Petrova, M., Enikolopov, R., 2020. Political Effects of the Internet and Social Media. Annual Review of Economics 12, 415–438.

Declaration for admission to the doctoral examination

I, Henry Stemmler, confirm

1. that the dissertation ”Essays on Technological Change and Trade in Development Economics” that I submitted was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,
2. that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,
3. that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,
   - No remuneration or equivalent compensation were provided
   - No services were engaged that may contradict the purpose of producing a doctoral thesis
4. that I have not submitted this dissertation or parts of this dissertation elsewhere.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

Date, Signature: ____________________

Author contributions

This thesis consists of three independent chapters. The contributions of each co-author to the respective chapter are summarized in the following.

Chapter 1: Automated Deindustrialization: How Global Robotization affects Emerging Economies - Evidence from Brazil

The first chapter of the thesis is single-authored. The conceptualization, the research design, the preparation and analysis of the data and the writing of the manuscript were carried out by myself.

Chapter 2: Dealing with agricultural shocks: Income source diversification through solar panel home systems

This chapter is co-authored with Krisztina Kis-Katos, Friederike Lenel and Christoph Weisser. Krisztina Kis-Katos, Friederike Lenel and I contributed to the conceptualization, the research design and the writing of the manuscript. Christoph Weisser was responsible for the machine learning algorithms. The rest of the data preparations and analyses were mainly carried out by myself.

Chapter 3: Your Pain, My Gain? Estimating the Trade Relocation Effects from Civil Conflict

This chapter is co-authored with Tobias Korn. We equally contributed to the conceptualization, the research design, the data preparation and analysis, and the writing of the manuscript.

Date, Signature: ______________________________