Title
Networks, Migration and Spillovers Across Space

Author
Egger, Dennis

Publication Date
2022

Peer reviewed|Thesis/dissertation

Networks, Migration and Spillovers Across Space

by

Dennis Egger

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Edward Miguel, Co-chair
Professor Benjamin Faber, Co-chair
Professor Christopher Walters

Spring 2022

Networks, Migration and Spillovers Across Space

Abstract

Networks, Migration and Spillovers Across Space

by

Dennis Egger

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Edward Miguel, Co-chair

Professor Benjamin Faber, Co-chair

Externalities of agents' behaviors on other individuals are a key concern of economic analysis. Moreover, from a policy perspective, the spillover effects of an intervention on those not targeted are paramount to understand its effects and evaluating its desirability. Spillovers propagate through social and economic interactions between individuals – including within the household – and through participation in common markets or institutions. The geographic clustering of social networks, markets and institutions as well as individuals' location choices through migration thus govern the spatial dispersion of externalities. In this dissertation, I study three examples of how social and economic networks shape the geography of economic interactions.

In the first chapter, joint with Daniel Auer and Johannes Kunz, we study the effects of migrant networks on the labor market integration of refugees, the performance of local firms, and the wages of their employees in Switzerland. To track outcomes of individuals and firms, we link six employer-employee matched administrative datasets covering the universe of residents (citizens, migrants, and refugees) and registered firms from 2008 to 2017. Leveraging the quasi-random placement of refugees across locations and a novel IV strategy, we show that larger local networks persistently increase employment and income of refugees. Network effects are large, accounting for 23% of the variation in incomes within nationality cohorts across cantons. In line with homophily, demographically similar networks and economically successful peers have larger positive impacts. Network effects are shaped by direct personal contacts: refugees who quasi-randomly lived in the same residential center are three times more likely to become co-workers at the same firm. Using a shift-share IV design, we then show that firms experiencing a positive shock to their employee's network hire both more migrants and natives. Their wage bill and the average wages of existing employees grow, and high-skilled natives rise within the firm hierarchy. This is consistent with referrals improving firm-worker match quality and productivity. Concerns about adverse economic impacts of

spatially concentrated immigration are not borne out in the data, suggesting that existing migration policies in Switzerland and other high-income countries may need to be reconsidered.

In the second chapter, joint with Johannes Haushofer, Edward Miguel, Paul Niehaus and Michael Walker, we study impacts of unconditional cash transfers on local economies in Kenya. Tracing out the effect of large economic stimuli on the pattern of transactions in an integrated economy, and their aggregate implications, has long been a central goal of economic analysis, but until now has not been studied experimentally. This study was designed to study the aggregate consequences of cash transfer programs while accounting for multipliers and externalities. We carried out a large-scale experiment in rural Kenya that provided one-time cash transfers worth roughly USD 1000 across 653 villages with around 280,000 people, with a large implied fiscal shock of roughly 15% of local GDP, and deliberately randomized the intensity of cash transfers across geographic sublocations. We first document large direct impacts on households that received transfers, including increases in consumption expenditures and durable assets 18 months after transfers. Enterprises in areas that receive more cash transfers also experience meaningful gains in total revenues, in line with the increased household expenditures. Untreated households, too, show large consumption expenditure gains, by an amount comparable to recipients' gains. Through monthly measurement of scores of commodities and consumer and durable goods, we document positive but minimal local price inflation (0.1% on average) in areas that received additional cash. To assess aggregate implications, we compute a local fiscal multiplier, taking advantage of data on representative samples of treated and untreated households and firms. Both income data and consumption data yield large positive estimated local fiscal multipliers of approximately 2.3 to 2.5. A speculative possibility for how local output increases, despite no meaningful local price inflation or firm investment response, is that many local enterprises are characterized by substantial 'slack' in their utilization of factors of production. Finally, we interpret the welfare implications of these results through the lens of a simple household optimization framework. In this framework, the fact observed consumption gains for untreated households are not driven by corresponding increases in labor supply, combined with a lack of local price inflation or of adverse spillovers along other non-market dimensions, suggest that non-recipients as well as recipients were made better off in this setting. This in turn suggests that some existing evaluations of cash transfer programs that ignore aggregate effects may be under-estimating overall program gains.

In the third chapter, together with Pierre Biscaye and Utz Pape, we study externalities arising not through social connections among residents of a local economy, or through their participation in the same market, but rather within the household as a result of the sharing of household economic and childcare activities by household members. We identify impact of childcare on adult labor supply in the context of COVID-19-related school closures in Kenya. We compare changes in employment after schools partially reopened in October 2020 for adults with children in a grade eligible to return against adults with children in adjacent grades. Using nationally-representative panel data, we find that a child returning to school increases adults' weekly labor hours by 22%. Contrary to evidence from high-income settings, effects are not significantly different by sex of the adult. This is explained by two offsetting mechanisms, driven by children's role as both childcare recipients and contributors to household childcare and agriculture. Women benefit relatively more from reductions in childcare burdens when children return to school, while men pick up a larger share of reduced child agricultural labor. Our results suggest policies increasing childcare accessibility could substantially increase adult labor supply in low- and middle-income countries.

While all three chapters are intended to answer a stand-alone set of research questions across different settings, they each shed light on ways in which the actions of economic agents, or the targeting of policies towards a subset of residents, affect others in their geographic vicinity.

Contents

List of Figures

List of Tables

Acknowledgments

There are so many people who have supported me throughout this project and I will not be able to mention them all. I have been blessed to stand on the shoulder of the giants in research that have come before me, inspired me, and those that have studied with me along the way. I have continuously felt supported, challenged, and guided during my years as a graduate student at Berkeley.

First, I am thankful to my committee, Ted Miguel, Benjamin Faber and Chris Walters. They have guided me from day one, pushed me to think deeper, picked me up when I needed it, and were patient and supportive throughout. Their commitment to rigorous and impactful research is an inspiration that will stay with me throughout my career.

I would also like to thank my co-authors Daniel Auer, Pierre Biscaye, Johannes Haushofer, Johannes Kunz, Ted Miguel, Paul Niehaus, Utz Pape, Michael Walker. I have learned so much from them, and they have influenced my work in so many ways. Working with them has been a great source of joy and energy over the last six years; and without the weekly calls, the recent COVID lockdowns in particular would have been so much worse.

Beyond my committee and co-authors, I have benefited from advice and support from Mathilde Bombardini, David Card, Supreet Kaur, Pat Kline, Jesse Rothstein, Nick Tsivanidis, Daniel Sturm, and Steve Redding. Working at Berkeley, and across many different research teams, I have also had the privilege to interact with and learn from many friends and colleagues in the economics community; in particular Felipe Arteaga, Livia Alfonsi, Susanna Berkouwer, Luisa Cefala, Kaveh Danesh, Madeline Duhon, Tilman Graff, Arlen Guarin, Johannes Hermle, Aleks Jakubowski, Oliver Kim, Felipe Lobel, Layna Lowe, Magdalena Larrebourre, Anya Marchenko, Gwyneth Miner, Max Mueller, Shahan Shahid Nawaz, Priscila de Oliveira, Pedro Pires, Bhumi Purohit, Tatiana Reyes, and Nick Swanson. They have made my time in graduate school stimulating and – most importantly – a happy one.

These projects would not have been possible without the work and collaboration of implementation and field partners. For our work in Switzerland, we thank the Swiss Federal Statistics Office, the State Secretariat for Economic Affairs and the State Secretariat for Migration for providing the data analyzed in this project, and Rainer Winkelmann and the University of Zurich for providing IT resources. In Kenya, the collaboration with implementing partner GiveDirectly, and field partners Vyxer Remit Kenya as well as IPA could not have been better. In particular, I would like to thank Carol Nekesa, Andrew Wabwire, Justin Owino, Sam Balongo, Erick Mugoya, Evans Omondi, and Ben Wekesa for their insight, inspiration and commitment. At the World Bank, I would particularly like to thank Javier Baraibar Molina, Antonia Delius, and Emanuele Clemente.

Funding for this project was generously provided by the World Bank, the National Science Foundation, International Growth Centre, CEPR/Private Enterprise Development in Low-Income Countries (PEDL), the Weiss Family Foundation, the Strandberg Fund, CEGA, and an anonymous donor.

Lastly, I owe endless gratitude to my partner, friends and family who have supported me through the ups and downs. You know who you are. Thank you so much!

Chapter 1

Effects of Migrant Networks on Labor Market Integration, Local Firms and Employees

1.1 Introduction

Migrants tend to sort into spatially concentrated immigrant communities (see e.g. Bartel 1989; Musterd 2005). This suggests that local networks play a role in promoting migrant well-being and labor market integration. While positive effects of network size on the labor market integration of migrants have been documented (e.g. Edin, Fredriksson, and Åslund 2003; Munshi 2003; Damm 2009), little is known about the channels through which networks operate. They may be a source of information about employment opportunities (Beaman 2011; Bayer, Ross, and Topa 2008), referrals (Dustmann et al. 2015), knowledge of institutions at the destination (Biavaschi, Giulietti, and Zenou 2021), social support (Blumenstock, Chi, and Tan 2021), financial aid (Giulietti, Wahba, and Zenou 2018), or cultural identity. Moreover, larger enclaves may benefit their members because host communities are more receptive of new immigrants in places already familiar with an immigrant group as suggested by the contact hypothesis (e.g. Allport 1954; Mousa 2020; Lowe 2021).

At the same time, there is considerable academic and political debate about the potential adverse effects of immigration. Employees in local labor markets may experience displacement and negative wage effects (for reviews of the academic evidence, see e.g. Borjas 2003; Card 2009; Dustmann, Schönberg, and Stuhler 2016). Migrant enclaves may slow civic and social integration (Lazear 1999; Danzer and Yaman 2013), and large immigrant inflows may lead to political backlash among host communities (Dustmann, Vasiljeva, and Piil Damm 2018). But are these concerns warranted? And do they outweigh the benefits of larger networks?

In this paper, we study how networks affect the labor market integration of refugees, the economic performance of local firms, and the earnings of the existing workforce. We leverage the two-stage quasi-random allocation of asylum seekers in Switzerland and comprehensive employer-employee matched administrative and survey data covering the universe of migrants, residents and firms between 2008 and 2017. First, we investigate how the size and composition of networks dynamically impact the labor market trajectories of new migrants. Employing our rich data and a novel instrumental variables identification strategy, we test and extend existing theories and evidence of refugee labor market integration in a unified framework. Next, we study how changes in migrant networks impact local firms and their employees, incorporating information sharing and job referrals as an additional lens through which to interpret wage effects of immigration on native workers. And last, we dive into the mechanisms driving these effects, and empirically assess recent theories of information sharing and referrals within networks using quasi-random variation in co-residence among refugees in the first months after arrival.

Causal identification of the effects of migrant networks is complicated by the fact that migration choices are rarely exogenous. Social networks too, form endogenously. We leverage a feature of Swiss migration policy – the two-stage quasi-random allocation of refugees – to address these challenges. After an initial hearing at a federal processing center, asylum seekers are allocated to one of Switzerland's 26 cantons, proportionally to each canton's population. By law, this allocation is *random and electronic*, unless the applicant meets one of a few tightly circumscribed legal criteria (FAA-142.31 1998; State Secretariat for Migration 2015). Due to a Supreme Court ruling, the federal government is required to document and justify all exemptions from random allocation towards receiving cantons. Beginning in 2008, our dataset contains all these justification records, allowing us to reliably identify randomly allocated individuals. Within cantons, refugees then spend the first 6-12 months in a cantonal residential center, and allocation is again quasi-random conditional on a few practical considerations. Each center houses approximately 100–150 individuals at a time and meals are generally shared, leading to substantial exposure and social connections between residents.[1]

Quasi-random assignment of refugees to cantons implies an assignment of the bundle of characteristics at the assigned location – including the existing local network of co-nationals. But not all network members are themselves randomly assigned. So, even if refugees themselves are exogenously assigned, their network is not. If some migrants select into locations based on comparative advantage or differential valuation of amenities, existing local networks may be correlated with location fundamentals. To isolate the effect of networks, we therefore construct an instrument of the existing migrant stock based on previously exogenously assigned migrants. Our in-depth knowledge of the allocation mechanism and uniquely rich data allow us to construct the theoretically expected distribution of refugee assignments to each location, taking arrival cohorts at each reception center, cantonal assignment probabilities, family structures, and exemptions from random allocation as given. This approach, inspired by recent advances in the econometrics of settings with partially exogenous treatments (Abdulkadiroğlu et al. 2017; Borusyak and Hull 2020), makes explicit the source of randomness and potential counterfactual assignments, thus allowing for credible identification and randomization inference. We verify balance of assignments with respect to refugee

[1]One of the authors, Dennis Egger, worked in a cantonal asylum center prior to his doctoral studies, and has conducted multiple interviews with past colleagues to understand the allocation process. Anecdotes based on personal experience suggest the relationships formed in cantonal centers are often within nationality groups, strong, and many persist after refugees leave the center.

and location characteristics, and show that the observed distributions in our data are consistent with those expected under our characterization of the allocation mechanism.

The allocation policy generates quasi-exogenous variation in the nationality-mix of migrant inflows. Some firms – those initially hiring certain types of migrants – therefore experience a larger shock to their employees' networks compared to others. A strength of our setting is that it yields many uncorrelated and plausibly exogenous shocks across years and nationalities ideally suited for a shift-share instrumental variables (SSIV) design that improves on existing designs relying primarily on voluntary migration (Adão, Kolesár, and Morales 2019; Borusyak, Hull, and Jaravel 2021). In the second part of the paper, we use baseline employment shares of each firm, and an SSIV approach to estimate how inflows that are better matched to a firm impact its performance, hiring, and employee wages.

We trace labor market outcomes of individuals and firms using a novel and comprehensive dataset, comprised of four administrative employer-employee matched panel registries and two large-scale population representative surveys, all matched through a unique social security identifier and enterprise ID. The data covers the universe of refugees and migrants arriving between 2000 and 2017, as well as all individuals resident in Switzerland between 2010 - 2017, and the universe of employers between 2011 and 2017. Annual census registry data provides basic demographics and locations for all individuals. The migrant registry database contains detailed information on migrants' origin and each refugee's asylum process – including a detailed residence history, allocation information, and, contrary to earlier papers, any deviations from random assignment. Earnings and labor market participation for all individuals employed, self-employed, or receiving social security benefits in Switzerland come from monthly spell-level social security data, and are matched to the Swiss business registry covering all registered enterprises in Switzerland.[2] The biennial Earnings Structure Survey captures detailed employment, compensation, education, and job title information for a third of the labor force, while the annual structural survey has information on education, language use, family structure, residence, and commuting for up to 600,000 individuals each year.

We make four main contributions. First, we show that networks substantially and persistently improve refugees' labor market outcomes. Doubling the number of co-nationals resident in their assigned canton increases their employment probability by 15pp (28%), and annual income by 5010 CHF five years after arrival (representing 36% of mean annual earnings).[3] Variation in network size alone accounts for 23% of the overall variation in long-run labor earnings within nationality-by-year arrival cohorts across cantons – a large effect.

These results are qualitatively similar, but quantitatively larger than earlier findings (e.g. Munshi 2003; Edin, Fredriksson, and Åslund 2003; Damm 2009; Beaman 2011).

[2]Existing studies using Swiss data rely only on the migrant registry (Martén, Hainmueller, and Hangartner 2019). This does not contain wages, and the employment indicator becomes less accurate over time compared to the social security registry data as it is not regularly updated, and generally not updated at all after migrants leave the asylum system.

[3]The conversion rate between Swiss Francs (CHF) and US$ is approximately 1:1 in our study period. The population average annual income in Switzerland was 57,900 US$ PPP. in 2010 and 67,870 US$ PPP. in 2017. The equivalent earnings for the refugee population have been 41,790 US$ PPP (2010) and 36,410 US$ PPP (2017) conditional on being employed. See Figure 1.4 for a descriptive overview.

However, a key challenge in the existing literature has been to identify credibly exogenous variation in migrant destinations. Although refugee dispersal policies are promising, they typically leave substantial discretion to allocation officers for practical and ethical reasons. Both in Denmark (Damm 2009; Sale 2021), and in the US (Beaman 2011), observable refugee demographics are statistically significantly correlated with networks at the assigned destinations. The approach in the existing literature has been to control for these observables, yet concerns remain about potential imbalance on unobservables. Switzerland is unique in that allocation is explicitly random by law and independent across applicants. In addition, our uniquely rich data on communications between reception centers and allocation officers allows us to reliably identify exogenously allocated individuals, as demonstrated by balance tests for scores of refugee characteristics.

Moreover, existing studies have mostly not addressed the potential endogeneity of existing local networks at the destination, except for including a destination fixed effect (e.g., Edin, Fredriksson, and Åslund 2003 in Sweden and Martén, Hainmueller, and Hangartner 2019 in Switzerland).[4] Damm (2009) is the first to use past allocations as an instrument for current migrant stocks, but in constructing the instrument does not account for clustered allocations, placement officer discretion and sorting on observables.[5] In contrast, our instrument is based closely on our detailed knowledge of the assignment mechanism and transparently isolates the exogenous component of local networks.

Second, we investigate what channels drive network effects, looking at dynamics, heterogeneity, as well as network composition. Effects increase with refugee's time since arrival, are more pronounced for male, younger individuals, and for origin countries that are ethnically more homogeneous. Turning to network composition, we find that network members who arrived through the asylum system, those that are more similar in terms of demographics (age, sex and education), and more economically successful have larger impacts. This is consistent with network formation based on homophily (e.g. Currarini, Jackson, and Pin 2009), and strong ties providing more relevant employment-related information and support (Giulietti, Wahba, and Zenou 2018), more so than alternative views, according to which a networks' value is primarily based on its 'quality' alone (Calvó-Armengol and Jackson 2004), or where weak ties and diversification of information increase a network's value (Granovetter 1973).

[4]Edin, Fredriksson, and Åslund (2003), Damm (2009), Sale (2021), and Martén, Hainmueller, and Hangartner (2019) also include separate fixed effects for origin country, and year. A potential concern with this strategy is that within an origin country, later arrivals may be systematically different from early arrivals, and by definition encounter larger networks on average. Beaman (2011) therefore includes origin-by-arrival-year fixed effects, using variation only *within* arrival-year cohorts of each nationality. Our paper goes one step further, concentrating on variation within nationality-by-year-by-reception center, thus allowing for differential selection of migration routes.

[5]Sale (2021) uses the same IV strategy as Damm (2009), additionally accounting for differential effects between refugees placed at the same time and those placed further apart, building on Beaman (2011). A main contribution of Sale (2021) is to use these insights to characterize the dynamically optimal path of refugee allocations, a point we return to in the conclusion. The instrument in Damm (2009) and Sale (2021) is the total number of previously assigned refugees. This does not take into account the documented sorting of refugees allocated previously due to allocation officer discretion, or the potential for serially correlated assignments based on the Danish policy of assigning groups of co-nationals jointly while intending to balance overall numbers across multiple years.

Third, we isolate the contribution of information sharing and referrals. We show that within cantons, migrants that quasi-exogenously overlap in the same cantonal residential center for the first 6-12 months after arrival are three times more likely to end up working for the same employer *after* leaving the center. Corroborating this interpretation, firms exposed to a larger shock to their employees' network – those that previously had a higher share of employees from an origin – are significantly more likely to hire additional workers from that origin relative to other firms.

While the existing evidence of migrant network effects has been interpreted through an information / referral channel, direct causal evidence has not been established. Earlier studies document ethnic clustering within employers (e.g., Damm 2009; Martén, Hainmueller, and Hangartner 2019), and survey data suggests that referrals are a particularly prevalent among immigrant networks (Dustmann et al. 2015). Beaman (2011), moreover, finds that a large network arriving *at the same time* initially dampens labor market chances, consistent with a model where information sharing in networks leads to competition for a limited number of job openings. Yet, these facts may also be driven by differences in the skill-mix across migrant communities, firms specializing in certain types of labor, and general equilibrium effects in the local labor market. Our design overcomes this by leveraging quasi-random *within*-nationality variation in social connections, and a design that exploits *firm-level* shocks to employee networks.

And fourth, we provide evidence on how networks impact local firms and workers. Concerns about displacement of locals within firms are not borne out in our data. On the contrary, firms with a better matched migrant inflow hire *both* more migrants and more non-migrant workers. Total employment and the wage bill increases. Native workers (and in particular high-skilled ones) benefit, experiencing wage increases and promotions upwards in the firm hierarchy. Beyond corroborating the referral/information sharing channel of migrant networks, this is direct evidence that migrant networks not only benefit migrants themselves, but also improve the match-quality between firms and migrants and increase firm productivity.[6]

Our approach directly tests the firm-level implications of recent models of job referrals in a setting with quasi-exogenous shocks and full-population data. Kramarz and Skans (2014), Pallais (2014), and Barwick et al. (2019) show that referrals are particularly important for early labor market entrants. Because new entrants have few observable signals of quality, uncertainty in the match process is higher, and this leads to inefficient hiring of inexperienced workers. This channel may be particularly important for immigrants, whose productivity is more difficult to observe for local employers. But while Pallais and Sands (2016) show that referrals lead to more efficient firm-worker matches and increased productivity in an online labor market, over-reliance on networks in hiring may also lead to nepotism, and even reduce firm size and productivity (Chandrasekhar, Morten, and Peter 2020).[7] In our context, the former view seems to be quantitatively more important.

[6] This validates aggregate level evidence of the positive productivity impacts of immigration (Peri 2012).
[7] Witte (2021) finds empirical evidence of this reduction in productivity in Ethiopia.

Consequently, our results highlight the role of information sharing in networks as an important mechanism through which migrant inflows affect the wages of local employees, bridging the gap between the aforementioned literature on referrals and the literature on the wage impacts of immigration. Many studies using shift-share designs have found conflicting results at the labor market level (see Card 2009 and Dustmann, Schönberg, and Stuhler 2016 for reviews).[8] Our approach complements earlier work by focusing not on aggregate-level immigration shocks, but instead on firm-specific shocks to employee networks, and variation in within-migrant origin composition. Holding overall immigration fixed allows us to abstract from general equilibrium considerations, and cleanly identify the impacts of networks through referrals. In doing so, we contribute to a small literature on the firm-level impacts of immigration (Dustmann and Glitz 2015; Mitaritonna, Orefice, and Peri 2016), and shed light on the *within*-firm substitutability of migrants and natives as well as firm-level wage setting (Manning and Amior 2021).

Our findings have implications for immigration policy. Concerns about immigration are particularly salient in the case of refugees who tend to integrate slower than self-selected migrants (Brell, Dustmann, and Preston 2020). In the last decade, the global number of refugees has doubled and large inflows have created considerable political backlash (Dustmann, Vasiljeva, and Piil Damm 2018). The term 'refugee crisis' re-emerged, as in the case of Syrian refugees arriving in Europe in 2015. Host country immigration policies have often been motivated by concerns about migration rather than its upsides. Particularly in Europe and the United States, countries have adopted dispersal policies that aim to reduce spatial concentration of refugees, and policies limiting employment opportunities for refugees upon arrival. In light of our results, these policies may need to be reconsidered.

Three caveats are worth mentioning: First, our empirical design uses variation generated *within* an existing dispersal policy. Results may therefore not generalize to contexts where ethnic concentration is far beyond what is observed in our setting. In such cases, incentives for integration may be non-linear in group size (e.g. as in Bazzi et al. 2019). Second, a strength of our design is that it holds the overall migrant inflow across locations roughly constant. While general equilibrium effects are therefore unlikely to confound our results, this also implies that our setting is less well-suited for quantifying these effects, which may play an important role in context with larger immigration shocks (e.g. as in Card 1990). And third, there may be a trade off between economic and civic integration. In ongoing work, we use our data on language spoken at home, residential segregation, intermarriage, female labor force participation, and a novel dataset on all first names of new-born babies – matched to the migration registry – to study how migrant networks affect integration along those dimensions.

[8] Foged and Peri (2016) use Danish dispersal policy to show that immigration into a labor market leads to skill-upgrading of locals moving across firms, and increases in local workers' wages, consistent with our evidence.

1.2 Institutional Context and Background

Switzerland receives approx. 30,000 asylum requests each year, and about 60% of those remain in Switzerland for at least a year (cf. Figure A.1.1). In 2015 asylum seekers represented 0.8% of the Swiss population. This is one of the highest shares in Europe and among high-income countries in general (e.g., in 2015 refugees were 1.4% of the population in Sweden, 0.3% in Germany, 0.2% in the UK, and 0.08% in the U.S.). Figure 1.1 plots the number of refugees newly registered and present in Switzerland at the end of each year.[9] Their primary origins were countries of the former Yugoslavian Republic in the early 2000s, then shifting towards the Middle East, particularly Syria, Afghanistan, and Iraq, in the late 2010s. Throughout the period, there is a substantial share arriving from Sub-Saharan Africa, and Eritrea, Somalia, and Nigeria in particular.

Swiss asylum law sets the rules for allocation of newly arriving refugees, their asylum process, and regulations regarding residence and employment. Initially, refugees requesting asylum are transferred directly to one of several federal processing centers operated by the State Secretariat for Migration, usually located at the border or at airports (see Figure 1.3). Typically, this will be the closest center with availability of accommodation from their point of immigration. It is therefore likely that refugees choosing different migration routes systematically sort into different centers. In these processing centers, the identity and main characteristics of the individuals is recorded and they have a medical check and their first asylum hearing. The main purpose of this meeting is to evaluate a refugee's reasons for demanding asylum, as well as any legal requests for exceptions from random allocation to cantons (see below).

After a period of maximum 140 days refugees whose asylum request has not been rejected immediately (e.g., on formal grounds) are allocated to one of the country's 26 cantons (equivalent to federal states). In practice, allocation is much faster – the median in our data is 15 days, and 90% of refugees are allocated to cantons less than 42 days after the recorded immigration date.

By law, this allocation is electronic, random and proportional to the residence population in the cantons (FAA-142.31 1998; State Secretariat for Migration 2015). A group of allocation officers located the the State Secretariat for Migration's headquarters assigns refugees based on rudimentary information that suffices for identifying each individual (see Figure 1.2 for a schematic illustration of the cantonal allocation process). These allocation officers never meet individual asylum seekers in person, and allocate 100s of asylum seekers each day on average. All they receive is a file, containing a partial extract of each asylum dossier: throughout our study period, hard copy asylum dossiers were not fully digitized, and allocation officers never saw the full dossier. We have obtained the corresponding data set including all information available to the allocation officers.

Importantly, the law stipulates that random allocation can be suspended for a tightly circumscribed set of reasons. First, refugees that can demonstrate that an immediate family

[9] This is less than the total number of asylum requests, since some refugees leave Switzerland before the end of the year.

member (spouse, child, or parent) are already residing in Switzerland have the legal right to family reunification. Second, if a refugee's medical conditions can only be treated in certain cantons, this can be taken into account in the allocation. And third, an allocation decision may be non-random if there is any *serious* threat to the life or safety of a refugee, or th general population. This may occur, for example, if a refugee has been subject to human trafficking, and traffickers are active in a certain region of Switzerland, or if a refugee has committed a crime and is being detained in a specific canton. In addition, a few practical reasons *may in rare cases* result in a suspension of random allocation. The most important case are asylum seekers who have already requested asylum in another European country. Under the Dublin treaty, individuals may only request asylum in one treaty county, and are sent back to the first request country for any subsequent requests. In such cases, refugees are sometimes allocated to the same or a nearby canton in which the reception center is located for the first 140 days, and transferred to their final canton later in case deportation has not yet occurred.

These reasons are evaluated and recorded as part of the first asylum hearing by immigration officers, and entered into one of two fields within the data sent to the allocation officer: first, a direct request field containing the requested canton; second, a free text comment field containing additional details necessary for the allocation process. Importantly, several rulings by the Federal Supreme Court of Switzerland state require the Department for Migration to be able to justify in court any deviation from random assignment towards receiving cantons (State Secretariat for Migration 2015). Beginning in 2008, any requests are therefore reliably recorded in the allocation dataset we obtained for this purpose. Requests that do not have any legal merit according to the immigration official are not recorded.

We apply a machine learning algorithm to aggregate every free text entry into any mentions of cantons, any legally required exceptions to random allocation, or any asylum process details relevant to the allocation in any of the three official Swiss languages (e.g. any mention of immediate family, medical reasons, prison).[10] In our analyses, we consider all refugees where neither the request field, nor the free text comment contains any mention of a canton in any language as randomly allocated. When multiple family members apply as part of the same dossier, we consider comments for *all* family members, i.e. if one family member is considered non-random, so are all the others. Overall, roughly two thirds of all individuals are randomly allocated.

Table A.1.1 shows that women (who often arrive later to join their partners), more highly educated refugees, those arriving in larger families with children, as well as those more likely to be accepted, are more likely to be non-randomly allocated. In line with this, family reunion is the quantitatively dominant predictor of the suspension of random allocation.

[10]Specifically: We run the standard Latent Dirichlet Allocation algorithm (Blei, Ng, and Jordan 2003), yielding $\approx$ 350 common features of requests. These include such as 'brother', 'medical', etc. We then categorise these features into topics, i.e. 'brother, mother' etc. are classified into 'core family', the main valid reason for suspension of the random allocation. We extract these and other non-valid reasons in the form of dummy indicators to test whether any of the non-valid reasons is related to the allocation decision and find no evidence thereof (see Table A.1.1). In a final step we assess accuracy using random forest models (Breiman 2001) to assure that no important feature was excluded from the extraction; more details can be found in Auer and Kunz (2021).

We verify that scores of observable characteristics of individuals – both those potentially directly observed by allocation officers and others – are balanced across cantons (Table A.1.1). Interestingly, both random and non-randomly allocated refugees look reasonably balanced, though the imbalance is larger for non-randomly allocated individuals. This is in line with exceptions from random allocation being tightly circumscribed, and the observation that for most family reunions, the family already present in Switzerland are likely to have been randomly allocated to their canton at some point in the past. Appendix A.2 contains more details allocation officers, and validity checks on the randomness of this allocation.

Refugees allocated to a specific canton must reside there until their asylum request has been granted and a residence permit has been issued. The median time for this is 361 days (see Figure A.1.1). During this period, the allocated canton is responsible for housing and administration of assigned asylum seekers. Typically, asylum seekers spend the first 6-12 months in a cantonal residential asylum center while awaiting their asylum decision. These are dorm-style residential buildings, where asylum seekers eat communally and receive language and civic education classes. This second-stage within-canton allocation to residential centers is under the jurisdiction of cantons. While processes differ slightly across cantons, the allocation to residential centers is typically haphazard, based on limited information contained in asylum dossiers as well as well as a few practical considerations such as availability of rooms, child and disability friendliness of accommodations, safety concerns, etc. We consider this second-stage allocation exogenous conditional on those observables, and exploit conditionally random co-residence over the first few months to isolate the impact of social interactions between refugees (see Sections 1.3 and 1.6). After leaving a cantonal residential center, refugees may reside anywhere within the canton, though in practice cantonal authorities in collaboration with municipalities assist in finding and providing subsidized housing.

How binding is the initial allocation? In case the asylum request is rejected but the person cannot be sent back to the origin country, residence remains restricted to the allocated canton. The same restriction also applies to refugees with a positive asylum decisions who are dependent on social assistance (which is provided by the allocated canton). Accepted refugees may migrate to anywhere in Switzerland, although this is rare in our data. Even 9 years after arrival, approx. 90% of refugees remain resident in their initially assigned canton (see Figure A.1.1).

Overall, refugees are allocated according to a quasi-random two-step process, and their residence and employment is effectively tied to a small geographic location for a substantial amount of time after arrival.

1.3 Data and Empirical Strategy

In this section, we describe our dataset, and outline our empirical strategy to estimate the effects of networks on refugee's employment trajectories, local firms and their employees as well as our methodology to isolate the referral / information channel of network effects.

Data

We trace labor market outcomes for individuals and firms using a novel and comprehensive dataset, linking six administrative employer-employee matched panel registries and population representative surveys covering the universe of refugees and migrants arriving between 2000 and 2017, the universe of individuals resident in Switzerland between 2010 and 2017, and the universe of employers between 2011 and 2017.

Annual census registry data provides basic demographics, immigration status and location at the zip-code level for all individuals resident in Switzerland at the end of each year. The migrant registry database contains rich demographics, legal status, details on migrants' origin as well as basic employment information for all non-citizens resident in Switzerland. For refugees, it has detailed information on their asylum process, including day and outcome of any asylum decisions and a detailed residence history. Importantly we also have access to the full list of variables created as part of the cantonal allocation process. This data is used for communication between immigration officers at federal reception centers who conduct the initial hearing and the allocation officers located at the headquarters in Bern, who never meet any refugees. It includes information on the reception center, any cantonal allocation wishes, desired departure date as well as a free text comment field that immigration officers use to communicate any allocation-relevant information (see Section 1.2 for details).[11] Beginning in 2008, >99.9% of migrants can be uniquely and completely matched to the census registry using their social security number that they receive upon immigration.

We capture labor market outcomes using three related datasets. First, any employment and income are captured at the spell level in monthly frequency by the central social security registry. The registry contains all income of individuals above age 17 living or working in Switzerland that is subject to social security contributions. This includes any income above 2300 CHF annually derived from employment, self-employment, military service, disability and unemployment insurance. Employment income is measured comprehensively, including wages, overtime compensation, tips, bonuses, non-cash benefits, etc. and is not top-coded. Average annual income conditional on employment for all refugees in our data is ≈30,000 CHF, and while there is no nationwide minimum wage in Switzerland, it is ≈ 20 CHF per hour in cantons and sectors that impose one. With this in mind, it is unlikely that this lower bound leads to significant left-censoring. This is the highest quality and most comprehensive registry data on labor income in Switzerland. However, it does not capture any capital income. These data are uniquely matched to the population and migrant registries using an anonymized social security number.

Data on firms comes from the enterprise registry that captures all registered enterprises and contains information on economic sector, legal form, total employment and the location of each firm. It is linked annually with the universe of all employees identified by their social security number. Since this link is primarily based on social security reports done at the firm-level, our data does not separate the employment in multi-location firms into its

[11] Earlier studies using data on refugee allocations from the Swiss State Secretariat for Migration have either not observed the free text field, or not used its content in their empirical design. Moreover, they have primarily relied on the migrant registry, where employment outcomes become less reliable over time, since they are no longer systematically updated after a refugee is accepted and leaves the asylum system.

different locations. In our analyses, we therefore focus on single-location firms (97% of all firms). Although social security records do not contain an identifier for the employer, we are able to match these two datasets based on the social security reporting number. This number contains information about the type of income (e.g., employment earnings, self-employment earnings, social security income, disability insurance, etc.) and the reporting compensation office. Compensation offices are unique for large employers, while multiple small employers in the same location/sector sometimes use a single office for reporting. Individuals with only one employer within a year match uniquely; for others, we use an iterative match procedure based on the reporting number and achieve an unambiguous match rate of over 95%.

These registry data are further augmented by two large-scale representative surveys that are linked to registry data using social security numbers. The biennial Earnings Structure Survey has detailed employment, compensation, education, and job title information for a third of the labor force in each year, and roughly 50% of all individuals at least once over our observation period. The structural survey covers census-type variables such as education, language usage, family structure, residence, and commuting for a repeated cross-section of 200,000 households and 600,000 individuals annually between 2010 - 2017.

Estimating Network Effects on Refugee Labor Market Outcomes

First, we are interested in estimating the impact of characteristics of existing migrant networks, such as their size and composition, on refugee outcomes. We define a refugee's ethnic network as all individuals from the same origin nationality living in the same canton in the year before each refugee's arrival. Social networks based on nationality are highly predictive of social interactions in our context: Cantons are generally small (out of 26 cantons, 8 have less than 100,000 residents, 17 less than 300,000) and typically have only one or two urban centers. Among migrants living in the same canton, co-nationals are 40 times more likely to live in the same household, 30% more likely to live in the same zip code, and more than twice as likely to work for the same employer compared to two randomly chosen migrants.

The key challenge in estimating impacts of existing networks is that they might be endogenous. Suppose the following model:

$$y_{iod,t} = \beta \text{ network}_{od,\underline{t}-1} + \alpha_{o\underline{t}} + \gamma_{o,t-\underline{t}} + \delta_d + \varepsilon_{iod,t} \tag{1.1}$$

where $y_{iod,t}$ is an outcome for refugee i from origin country o assigned to canton d in year t. network$_{iod,\underline{t}-1}$ is a measure of the network of migrants from origin country o in canton d at time $\underline{t}-1$ and $\underline{t}$ is the immigration year of refugee i. $\alpha_{o\underline{t}}$ is an origin-by-arrival year cohort fixed effect. Including these is important, since refugee inflows within origin may be selected over time, if e.g., individuals with certain characteristics leave a conflict zone later, and encounter larger networks on average.[12] $\gamma_{o,t-\underline{t}}$ is a cohort-by-years-since-arrival fixed effect, absorbing any differential integration trajectories across origin nationalities that are common across all cantons. δ_t is a canton-of-assignment fixed effect, absorbing any systematic differences in networks and migrant outcomes across cantons.

[12]Among the existing literature, only Beaman (2011) includes origin-by-arrival-year fixed effects.

The parameter of interest is β. Causal identification requires network$_{od,\underline{t}-1}$ to be exogenous, i.e., $\text{Cov}(\varepsilon_{iod,t}, \text{network}_{od,\underline{t}-1}) = 0$. This may fail for two reasons: First, if migrants choose their destination, economically successful individuals may self-select into different networks, causing omitted variable bias (selection of refugee destinations). Second, *other* individuals may self-select into a location based on its attractiveness, and if different nationalities sort into different locations based on their comparative advantage or relative valuation of location amenities, existing migrant networks may be correlated with these location features (selection of other network members into destinations).

We address the first concern by focusing on quasi-exogenously allocated refugees, which we define as all individuals where no canton is requested in any communications between immigration officials at their reception center and the allocation officer located in the headquarters of the State Secretariat for Migration. As described in section 1.2, for a refugee to be non-randomly allocated, valid legal reasons have to be evaluated by an immigration official, and communicated to the allocation officer (who never meets any refugees) for execution. By law, any such deviation from randomness needs to be justified and recorded, and our data contains all such records. For quasi-randomly allocated individuals (approx. 2/3 of all refugees), their assignment location is therefore exogenous. Table A.1.1 verifies that assigned network size is indeed uncorrelated with refugee characteristics.

Second, even if individuals are completely randomly assigned to a destination d, existing networks at that destination may still not be exogenous. Some members of migrant networks may have arrived through channels other than the asylum system, some may have been granted an exception from random allocation, and some may have relocated after initial random assignment. Suppose location d is particularly attractive to migrants from origin o, e.g., because of the skill-complementarity between o-types and the industry mix at destination d, or because o-types particularly value d's amenities. While the inclusion of a destination fixed effect α_d controls for any systematic differences in locations affecting *all* refugees (i.e., absolute advantage of a location), it does not account for such group-specific sorting (i.e., comparative advantage). We therefore construct an instrument for the existing migrant stock based on past quasi-randomly allocated refugees.[13]

An Instrument for Ethnic Networks

Inspired by recent advances in the econometrics of settings with partially random mechanisms, our instrument builds on our detailed understanding and unique data on the allocation process to carefully isolate the random component of past

[13] This idea was first proposed by Damm (2009). However, compared to Switzerland, the allocation mechanism in Denmark leaves more discretion to allocation officers, and groups of refugees are assigned at the same time, based on integration offices moving across the country, and allocations are conditional on observable characteristics. No requests for family reunions or other exceptions from random allocation are observed and balance tests show that assigned network sizes are significantly correlated with refugee characteristics. In Switzerland, allocations random by law, independent across refugees (unless they are part of the same nuclear family), allocation officers never meet refugees and any exceptions are tightly circumscribed by the law, recorded in the dossier, and observed in our data (see section 1.2).

allocations (Abdulkadiroğlu et al. 2017; Borusyak and Hull 2020).[14] Past allocations of refugees are quasi-exogenous, and predict existing migrant networks, thus satisfying the requirements for an instrumental variables approach.

Allocation officers observe a minimal amount of information on each refugee, including their nationality, family structure (i.e., everyone applying for asylum in the same dossier), the federal reception center where their initial hearing was conducted, and any allocation requests recorded either as a direct request, or in a free-text comment by the immigration officer at the federal reception center. By law, the allocation officers' task is to electronically and randomly allocate these refugees to different cantons, proportional to each canton's population. Any deviation from randomness must be recorded in the comment field. Which refugees ultimately end up allocated to each canton therefore is partly endogenous – i.e., it depends on the legally valid requests within each cohort and canton, the distribution of arrivals at each reception center as well as the cantons' population-based quota – and partly exogenous. Our strategy is to model this allocation process carefully, and define the instrument as the deviation of realized allocations from expected allocations under this partially random process. See Figure 1.2 for a schematic of this process. Our instrument is:

$$\text{network}_{od,\underline{t}-1} = \sum_{l=1}^{L} \text{assigned\_network}_{od,\underline{t}-l} - E\Big(\sum_{l=1}^{L} \text{assigned\_network}_{od,\underline{t}-l}\Big) \tag{1.2}$$

Expected allocations capture any systematic differences in assignments (e.g. due to allocation requests, differences in immigration patterns and allocations by reception centers), and deviations from this expectation should therefore be exogenous.

We model the allocation process as follows: Nationality-by-year arrival cohorts in each reception center z are taken as given (n^z_{ot}), allowing for the potential endogenous selection of refugees into different reception centers. First, allocation officers assign any refugees with legally validated placement requests. Next, they assign those without a placement request to fill the cantonal quota: The overall assignment probabilities of the remaining quasi-randomly allocated refugees from each reception center z to each canton d in each year t are taken from our data and denoted p^z_{dt}. Our approach therefore takes into account overall cantonal quotas, and any differential representation of cantons among those making a successful placement request, and allows for any differential allocation of refugees from different centers to specific cantons (e.g., for legal or practical reasons, or because the language of the dossier may be observed and correspond to a reception center). Taking into account family structure, allocation officers then jointly allocate each migrant family without a placement request to one of 26 cantons according to a Multinomial distribution with assignment probabilities p^z_{dt}. The number of refugees from origin o and reception center z assigned to each canton d follows a Binomial distribution with probability density function $B(n^z_{ot}, p^z_{dt})$, with the expected number of assigned individuals therefore given by $E(n^z_{odt}) = p^z_{dt} \cdot n^z_{ot}$.

This characterization of the assignment process is testable. First, we simulate the assignment process 999 times, and construct the deviation from the expected number of as-

[14]To our knowledge, these methods – developed initially to evaluate school allocation mechanisms – have not been employed to study network effects on labor market outcomes.

signed refugees as a share of the cantonal population for each counterfactual allocation. Panel A of Figure 1.3 shows that our observed distribution is very similar to the average counterfactual distribution. Moreover, we know that the standard deviation of a Binomial variable increases with the square root of the number of trials.[15]. In our case, larger arrival cohorts will have more variation in deviation from the expected number assigned to each canton. Panel B of Figure 1.3 shows that the standard deviation of assigned population shares in our data closely matches the expected relationship.

Second, a valid instrument should be uncorrelated with baseline nationality-by-canton characteristics. Cantons with initially larger networks, or networks with different compositions should not systematically have above- or below-expected number of refugees assigned to them. Table A.1.2 shows that initially larger networks – even conditional on destination fixed effects – have somewhat higher shares of non-refugees, women and older individuals, employed individuals and migrants that arrived more than 3 years ago. Although these correlations are small, they are statistically significant and suggest that migrants (other than those randomly assigned) differentially select into larger networks, highlighting the need for an IV approach. Our instrument based on quasi-exogenous inflows, on the other hand, is uncorrelated with initial network characteristics.[16]

In short, the assignment distributions we observe in the data are consistent with what would be expected under our stylized assignment mechanism. Moreover, in contrast to the migrant stock itself, it is balanced with respect to size and characteristics of initial networks, even conditional on destination fixed effects. As an added benefit, this implies that design-based randomization inference is credible in our setting (Fisher 1936). One remaining concern may be that different allocation officers have preferences over regions (e.g. due to language) and specialize in certain nationalities. In Appendix A.2, we show that destination and nationality shares are similar across all allocation officers, and provide additional validation tests of the instrument.

In our main analyses, we focus on individuals arriving between the age of 19 and 54 who are likely to be in the labor force over the entire observation period.[17] We include origin nationalities with at least 20 refugees in our sample period, and where at

[15]Specifically, the variance of the number of refugees assigned to each canton is $V(n^z_{odt}) = n^z_{ot} p^z_{dt}(1 - p^z_{dt})$. The expected (canton-share-weighted) variance of the deviation from the expected number $\bar{n}_{okt} = n_{okt} - E(n_{okt})$ across cantons is then given by:

$$E\Big(\frac{1}{1 - \sum_k (p^z_{dt})^2} \sum_k p^z_{dt} (\bar{n}^z_{odt} - \sum_d p^z_{dt} \bar{n}_{odt})^2\Big) \approx p^z_{dt} \cdot f(\{p^z_{dt}\})$$

The closed form approximation holds true exactly only under independent allocation across cantons, while in reality, there should be a small correlation across cantons, which is taken into account in our simulations.

[16]A third test is whether the reduced-form effect in Equation 1.1 is sensitive to the inclusion of different controls. Our instrument should not be correlated with refugee or destination characteristics, and Table A.1.4 that our estimated reduced-form coefficient is indeed stable across different sets of fixed effects.

[17]18 is the legal age of adulthood in Switzerland, and 18/19 is when most individuals finish high schools or apprenticeships.

least 10% of residents in Switzerland arrived through the asylum system since 2000. [18] Since placement requests are only recorded from 2008 onward, and because we use up to 3 years of previous allocations to instrument for the previous year's migrant stock, we focus on refugees arriving between 2011 - 2017.[19] We cluster standard errors at the nationality-by-year cohort level, corresponding to the variation in the instrument and accounting for the potential correlation of assignments within nationalities across cantons due to cantonal quotas. We also report exact p-values from randomization inference based on 999 iterations of the allocation algorithm described in Section 1.3.

This approach is valid for any measure of a refugee's network upon arrival, including measures of network size or composition. For our main specification, we use network size, defined as the inverse hyperbolic sine transformation of the number of co-nationals resident in the assigned canton in the previous year: ArcSinh(no. of co-nationals$_{od,\underline{t}-1}$).[20] Differences in the overall population across cantons are absorbed in canton-FE, and we interpret β as the effect of a proportional change in a migrant's network size within the same canton. To ease interpretation, we scale our instrument relative to the network size in the previous year, such that instrument and endogenous variable are denoted in the same proportional units. [21] This instrument is highly predictive of migrant stocks, with a first-stage coefficient of 0.79 (SE=0.13) and a first-stage F-statistic of 36 for the full sample, and a minimum F-stat of 19 across all specifications and samples presented.[22]

Estimating Network Effects on Local Firms and Employees

We use a similarly constructed instrument and a shift-share instrumental variables approach (SSIV) to estimate the impact of a migrant networks on local firms and employees. We would like to understand how a change in the network of a firm's employees affects the firm, and its employees. The migrant network of a firm's em-

[18]We chose these thresholds *a priori*. However, results vary in predictable ways when relaxing these thresholds. The higher the threshold, the stronger the first-stage, as asylum allocations become more predictive of overall networks for nations where a higher share immigrates through the asylum system. However, increasing threshold also reduces our overall sample size, and leads to reduced power in our analyses.

[19]We will soon receive additional years of outcome data, so that we can extend our dynamic results even further with an increased sample size.

[20]Results are robust to using alternative specifications, including the population share of each community.

[21]Specifically, the instrument is defined as:

$$\text{ArcSinh}\left(\sum_{l=1}^{3}\Big(\text{no. assigned co-nationals}_{od,\underline{t}-l}\Big) + \text{no. of co-nationals}_{od,\underline{t}-4}\right)$$
$$-E\left(\text{ArcSinh}\left(\sum_{l=1}^{3}\Big(\text{no. assigned co-nationals}_{od,\underline{t}-l}\Big) + \text{no. of co-nationals}_{od,\underline{t}-4}\right)\right)$$

[22]A first-stage coefficient slightly below 1 ($p = 0.11$) implies that an inflow of randomly assigned refugees leads to slightly fewer voluntary migrations to the same canton, at least over a period of 3 years. This is consistent with Beaman (2011), suggesting that a large influx of new migrants into a network may be temporarily negative in the short-run due to competition for a limited amount of jobs, while turning positive in the long-run (see Figure A.1.3).

ployees is proxied by the number of co-nationals resident in the same canton, and we aggregate this measure across all origin nations, using baseline employment shares

$$\begin{aligned}\Delta y_{jdkt} =& \beta\Big(\sum_o s_{oj}^{BL} \cdot \Delta_L \text{ArcSinh}(\text{no. of co-nationals}_{odt})\Big) \\ &+ \gamma \mathbf{1}(\text{any refugee employees})_{BL} + \alpha_k + \phi_d + \delta y_{jdk}^{BL} + \varepsilon_{jdkt}\end{aligned} \tag{1.3}$$

where y_{jdkt} is an outcome of firm j in canton d in sector k at time t, s_{oj}^{BL} is the baseline employment share (as a share of all employees from asylum-sending nations), $\mathbf{1}(anyrefugeeemployees)$ is an indicator for having any employees from asylum-sending nations at baseline, and Δ_LArcSinh(Number of co-nationals$_{odt}$) is the growth rate of the migrant stock from origin o in canton d over the last L years. We include sector fixed effects α_k to control for any sector-specific time trends, and controls of the outcome variable at baseline δy_{jdk}^{BL}. [23] ϕ_d and α_k control for any canton- and sector-specific trends respectively, as well as the overall inflow of migrants across cantons.

The parameter of interest is β, which we interpret as the differential response of firms experiencing a doubling of their employees' network size relative to other firms in the same labor market. Note that this is different from existing shift-share approaches to estimating the impacts of overall immigration at the labor market level: First, our shock is defined as a difference in the *composition* of migrants, holding overall immigration constant (both by design, and by including canton fixed effects). Second, our shock is firm-specific, thus allowing us to identify impacts across firms in the same labor market, but with different initial migrant compositions.

Importantly, the change in network size may affect firms through multiple channels. First, a larger network may imply an increase in potential job referrals or information about job opportunities. Second, a larger network implies a relative labor supply shock of migrant labor types used intensely in a firm's production function. To the extent that refugees from different origins are imperfect substitutes, firms with different initial composition of migrant employees may therefore respond differently. In section 1.5 we interpret our findings with these possibilities in mind, and in section 1.6 we use within-canton variation in cohabiting relationships to further disentangle these channels.

For causal identification, we require relative changes in migrant stocks across origins to be as good as randomly assigned. If firms with initially larger employment shares of a given nationality grow, they may attract additional immigrants from that origin in search of employment opportunities. Moreover, initial employment shares are likely endogenous and firms hiring employees from refugee nations may be systematically different from other firms. We overcome this concern by instrumenting relative population changes using cumulative allocations as above – i.e. deviations from expected allocations in of each nationality across canton-years. Because the variation in the deviation from expectation in the number of

[23]Results are robust to excluding baseline controls.

refugees from a nationality varies systematically with the cohort size (see Figure 1.3), we here use deviations in the relative number assigned to each to aggregate across nationalities. [24]

Even if initial employment shares are endogenous, β is causally identified as long as the allocation shocks are as good as randomly assigned, i.e. deviations from expected allocations should not strategically react to differential growth in employment opportunities for different nationalities (Borusyak, Hull, and Jaravel 2021; Adão, Kolesár, and Morales 2019). We verify this assumption in two ways: First, Table A.1.2 shows that the instrument is uncorrelated with the composition of baseline migrant stocks in each nationality. Second, Table A.1.3 shows that baseline firm characteristics are uncorrelated to the shift-share instrument with the exception of a small imbalance in the share of employees with tertiary education. We therefore control for average baseline values of the outcome in our main specification, though results are robust to excluding baseline controls.

In our main specification, we concentrate on long-differences of 6 years in outcomes across our entire sample period from 2011-2017.[25] Since our social-security data does not separate firm-outcomes by location for multi-location firms, we concentrate on single-location firms (97% of all firms) and exclude firms with less than three employees at baseline (60% of firms). For inference, we estimate the instrumented version of Equation 1.3 at the shock-level and cluster shocks at the nationality level to account for correlated shocks across firms with similar initial employment shares (Borusyak, Hull, and Jaravel 2021).

Isolating the referral channel

Networks may affect migrants through different channels, including information sharing, job referrals, social support and the way they impact the response of locals. Similarly, firms may be affected through the relative change in labor supply as well as any potential impact on referrals and information sharing between employers and existing employees. To isolate the referral channel, we use the within-canton quasi-random allocation of refugees to different cantonal centers. These centers house around 100–150 refugees on average, and asylum seekers typically live there for the first 6-12 months after arrival. Our strategy tests whether – within co-national ethnic networks – individuals that overlap at such a center are more likely to work for the same employer after leaving the center. Suppose

$$\mathbf{1}(\text{coworkers}_{ijt}) = \psi \mathbf{1}(\text{center-overlap}_{ij,\underline{t}}) + \alpha_i + \alpha_j + \gamma \mathbf{x}_i' \cdot \mathbf{x}_j' + \text{muni}_i \cdot \text{muni}_j + \varepsilon_{ijt} \quad (1.4)$$

where $\mathbf{1}(\text{coworkers}_{ijt})$ is an indicator variable equal to 1 if i and j work for the same employer at time t, $\text{center-overlap}_{ij,\underline{t}}$ is an indicator for whether i and j overlapped in a cantonal

[24] Specifically, our instrument is defined as:

$$\sum_o s_{oj}^{BL} \cdot \left(\text{ArcSinh}\Big(\sum_{l=1}^{L-1} \text{assigned refugees}_{od,t-l} \Big) - E\Big(\text{ArcSinh}\Big(\sum_{l=1}^{L-1} \text{assigned refugees}_{od,t-l} \Big) \Big) \right)$$

where we focus only on refugees without valid legal allocation requests as above.

[25] Given the long lag in network effects, long differences are preferable. Once we obtain additional years of data, we will explore including additional lags.

center in the months after arrival, α_i and α_j are individual fixed effects, $\mathbf{x}_i' \cdot \mathbf{x}_j'$ is a full set of interactions between i and j characteristics potentially observed by the cantonal allocation officer, and $\text{municipality}_i \cdot \text{municipality}_j$ controls bilaterally for the municipality of residence at time t. We consider all pairs assigned to the same canton and use only employment outcomes in 2017 (the last period in our data) for simplicity. Standard errors are clustered at the level of the network, i.e. at the assignment-canton-by-nationality level.

ψ is the difference in the probability of working for the same employer between individuals that lived in the same center and those that did not but were assigned to the same canton. Since assignment to residential centers within cantons happens upon arrival (when cantonal officials do not know much about refugees beyond their dossier), is typically haphazard, based on availability of accommodation, and quasi-random conditional on observables, we interpret this difference as the causal impact of living together after arrival. Since refugees live in cantonal centers for a few months, and centers only house 100–150 individuals at at a time, cohabiting leads to substantial social interaction, and thus a higher likelihood of 'strong ties'. We exclude pairs within the same family and any individuals still living in a cantonal center to avoid differences in co-working based on family relationships or location only.

Beyond social relationships, living together may affect co-working either through subsequent residence choices, i.e. refugees allocated to the same center may reside closer to that center. After controlling for a full set of municipality effects, the remaining difference should isolate the effect purely driven through refugees knowing each other, sharing information and potential referrals. Since we do not observe how individuals obtained their job, we cannot directly distinguish referrals from information sharing more broadly.[26]

1.4 Network effects on the labor market trajectories of migrants

We now turn to tracing out the dynamic effects of networks on refugees' labor market trajectories. We consider an individual employed in a given year if there is any employment or self-employment income in the social security registry.[27] Employment earnings are broadly defined, and include any wages (including overtime), bonuses, tips, as well as any in-kind remuneration. For monthly incomes, we divide the total annual earnings by the number of months this individual paid any social security contributions from employment or self-employment.

As shown in Figure 1.4, refugee integration trajectories are much slower on average than those of other migrants – reflecting both the nature of their migration decision, their origin, and that their immigration authorization is not based on employment. Increases in labor force participation begin to level out only 8-9 years after arrival, while wages continue to increase (Panel A). Consistent with refugees representing a particularly vulnerable population regardless of their origin, refugees from different regions have remarkably similar integration trajectories (Panel C), with the excep-

[26]In ongoing work, we estimate Equation 1.4 separately for co-workers that found their job at the same time vs. sequentially to further disentangle referrals from information sharing.

[27]Self-employment is rare over the first 6 years after arrival. Only 0.1% of working age refugees arriving after 2011 are self-employed in any year in our observation period.

tion of South and East Asian refugees, primarily from Sri Lanka, China (Tibet) and Mongolia. With these overall patterns in mind, we first look at whether the size of a refugee's network upon arrival affects labor market effects in the *long-run*.

The labor market effects of networks are large and persistent (see Table 1.1). Estimates from Equation 1.1 using our IV approach show that doubling a refugee's network size upon arrival increases labor market participation by 15pp (or 28% of the average employment rate in the sample) and annual earnings by 5010 CHF (or 36% of mean annual labor income).[28] While there is some evidence that monthly earnings among the employed increase (see Columns 7-9), these effects are small and not statistically significant. Networks, therefore, seem to improve labor market outcomes primarily on the extensive margin. However, if the marginal migrant finding employment due to network effects is negatively selected, our estimate may understate the impact of networks on job quality for each individual.[29] These results are conditional on refugees still being in Switzerland after 5 to 6 years. Table A.1.5 shows that network size does not significantly affect whether individuals remain in Switzerland, or whether they move within Switzerland after the initial assignment.

These effects represent an important component of the overall impact of quasi-randomly assigned locations. Using our preferred approach, the estimated network effects account for 23% of the overall variance in earnings across cantons within nationality-by-arrival-year cohorts, obtained from a model estimating earnings effects separately for each canton using canton-of-assignment dummies (see Appendix A.3 for details and alternative approaces).

IV estimates are substantially larger than those obtained using OLS. This may be because non-randomly assigned migrants are negatively selected. However, a more likely explanation (which we corroborate when looking at network composition below) is that our instrumental variable induces network changes of precisely the kind that are particularly valuable: those where a large share of demographically similar network members arrived recently through the asylum system.[30]

Figure 1.5 breaks the effect on labor income into annual dynamics. Our IV estimates increase continuously up to 6 years after arrival, and although the growth rate slows somewhat, effects have not yet reached a plateau. Reduced form estimates are presented in Figure A.1.3. There is some evidence that larger networks have a *negative* impact on income in the very first year after arrival. This is consistent with the interpretation in Beaman (2011), who argues that arrivals of large groups *initially* compete for limited employment opportunities and information, while effects turn positive for network members that arrived longer ago. An alternative view would be that networks may encourage members to

[28] The exact Fisher randomization p-values for these effects are 0.05 and 0.08 respectively, and Figure A.1.2 illustrates our estimate relative to 999 simulated random assignments. Figure A.1.4 moreover shows that the reduced form estimates are robust to changing the set of included fixed effects, validating our instrumental variables design.

[29] In ongoing work, we intend to investigate this possibility by investigating if there is any selection into employment based on observables due to networks, and whether monthly income results change conditional on such observable characteristics.

[30] In fact, interacting the overall network size with the share of network members that arrived through the asylum system, as in Table 1.2, suggests that asylum network members are approx. 4 times as valuable as non-asylum members, in line with the magnitude of the difference between OLS and IV.

wait until a better job opportunity becomes available by providing a higher outside option or reference point for new arrivals (e.g. as in Caldwell and Harmon 2019).

Network effects are larger for male refugees (see Figure 1.6). This is in line with a generally low female labor force participation among refugees, and a much larger gender gap in employment than among the Swiss population overall (Figure 1.4, Panel B). Moreover, effects are larger for refugees arriving at an age below 25. Together with the fact that refugees arriving as children or adolescents generally integrate faster, acquire Swiss education, and look similar to Swiss nationals in terms of labor market participation (Panel D of Figure 1.4), this suggests that networks may be a complement to refugee's skills, and help signal these skills to potential employers. This is corroborated by the fact that network effects are somewhat larger for native French speakers (primarily from Western Africa) who are quasi-randomly assigned to a canton of Switzerland, in which French is the dominant language, though this difference is not statistically significant.[31].

Thus, network size is an important driver of long-run labor market integration of refugees. But not all networks are created equal, and leading network theories have different implications for which types of networks are most useful – even conditional on their size. Granovetter (1973) argues that weak ties and less similar individuals are more useful, since they diversify an individual's information. But recent evidence from online social networks (Gee, Jones, and Burke 2017) and family connections (Kramarz and Skans 2014) suggests that strong ties are more valuable at the margin, perhaps because they are more likely to use their social capital to provide support. Moreover, information is more likely to flow between similar individuals in networks characterized by homophily (Currarini, Jackson, and Pin 2009). If an important component of a network's value lies in the distribution of information about job opportunities and referrals, we may additionally expect more economically successful networks to have the biggest impact (e.g., as in Calvó-Armengol and Jackson 2004).

Table 1.2 presents OLS estimates of Equation 1.1, including the size of each refugee's network as well as interactions for various attributes of each network.[32] Co-national refugees of the same sex (Column 2) and the same age group (Column 3) have substantially larger positive impacts. Contrary to the pure 'quality' hypothesis, peers with secondary education have somewhat *smaller* positive effect. Given that the majority of refugees did not complete secondary school themselves, this is again more in line with more similar network members having more positive impacts. Columns 5 and 6 moreover show that network members that are themselves employed and networks with a higher monthly income generate more positive effects.

Overall, these results provide evidence for the value of strong ties over weak ties, and highlight an important role for homophily. This is corroborated by the fact that refugees from ethnically more fragmented sending nations benefit

[31] See also Auer (2018) for the effect of language on refugees' employment in Switzerland

[32] We focus here on OLS estimates, as some network attributes are ex-post outcomes, and cannot separately be instrumented for by characteristics of past quasi-randomly assigned refugees. Moreover, our focus here is mostly on the network composed of other asylum seekers most of whom were also at some point randomly allocated, OLS is likely to be less biased. However, for demographics – where instrumenting is possible – IV results are consistent with those from OLS.

less from their network (see Figure 1.6). Moreover, within strong ties, 'quality' seems to matter, consistent with already employed refugees providing information about job opportunities and/or referrals to newly arriving peers.

1.5 Impacts of Networks on Local Firms and Employees

We have shown that larger, more homogeneous, and high quality networks improve long-run migrant outcomes, but how does this affect local firms, and through what channels?

To assess potential effects on firm performance, we first investigate the distribution of refugees across industries and firms. Figure A.1.4 shows that over 40% of refugees initially work in hospitality and food services. That share decreases over time to 20% 10 years after arrival, as refugees increasingly work in health and social work, admin and support services, manufacturing, and pubic administration. Consistent with potential network effects, we observe substantive sorting of refugees into specific firms within sectors. A firm employing at least 1 refugee is 10 times more likely to hire a second than a randomly chosen other firm in the same sector is to hire the first (Panel A of Figure A.1.5). Moreover, there is substantial sorting by nationality: the second refugee employee is three times more likely to come from the same origin than what would be expected if a firm hired a randomly chosen refugee in the same canton (Panel B). This is suggestive evidence that the employment effects of ethnic networks may be concentrated in firms that previously hired refugees from the same origin, and that migrant networks may play an important role in this process. In the next two sections, we unpack these descriptive statistics, and causally estimate the impact of migrant networks on firms and their employees.

We define a firm's network as the co-national network of its employees as it is existing employees that are best able to provide information about employment opportunities at the firm, and that are a source of potential referrals. Quasi-exogenous refugee assignment creates variation in relative inflows in the composition of migrants, where types are defined by their origin nationality o. Our shift share IV strategy thus exploits quasi-random variation a firm's network based on differences from the expected allocation in relative inflows of migrants from different origins into the canton where the firm is located. Exposure shares of each firm are defined as the o-type employment shares at the start of our observation period in 2011 (as a share of all employment from refugee nations). We concentrate on single-location firms with at least 3 employees in 2011 and active in all years between 2011 and 2017, and estimate the effect of inflows over the entire period on 6-year changes in firm outcomes. (See Section 1.3 and Equation 1.3 for more details.)

Two points are important to note upfront. First, a relatively larger inflow of o-types not only increases the network of employees from origin o, but also represents a relative labor supply shock of o-types relative to other migrants. If different migrant types are imperfect substitutes, firms with a high initial employment share of o-types may therefore be impacted both through a 'traditional' labor supply channel, as well as a change in the network of its employees. In fact, we view networks as one potential mechanism through which a better 'matched' migrant labor supply shock may impact local firms, e.g., through an increase in referrals. In this section, we cannot directly disentangle these channels, but we note that our firm-level results provide reduced-form causal evidence to discipline recent theories of wage setting at the firm level as well as to shed light on firm-level impacts of referrals. We turn to disentangling the referral / information channel from a pure labor supply channel in Section 1.6.

Second, refugee allocation policy varies the composition of migrants while holding constant the overall migrant inflow as a share of the local population. Moreover, refugees represent less than 1% of the local labor force in all cantons. While this limits our ability to study general equilibrium effects of migrant inflows on wages across labor markets, the upside is that such effects are also unlikely to affect our firm-level estimates. Empirically, our canton-of-assignment fixed effects absorb any remaining variation in overall immigration across cantons. We therefore interpret our results as *relative* effects between firms that are experiencing similar overall migrant inflows, but for some firms those inflows are better 'matched' to the composition of their existing immigrant employees compared to others.

Larger firm networks increase overall employment, and shift within-firm employment composition towards network members. A quasi-exogenously driven doubling of the nationality-weighted firm network leads to a 53% increase in the number of employees from asylum nations relative to other firms (Row 2 of Table 1.3). Strikingly, there is no evidence of displacement of other employees at the firm level. If anything, employment of *other migrants* (i.e. those from nations without asylum seekers) is also 13% higher, as is employment of Swiss nationals (13%), though these differences are not statistically significant. Since proportional employment effects are larger for refugee nation employees, increases in the size firm networks further increase their employment share of refugee nation employees relative to other firms, and causally increase overall sorting of migrants from the same origin into specific firms.

Importantly, total employment in firms experiencing a doubling of their network grows by 23%, and the overall wage bill increases by a statistically significant 44% relative to other firms in the same sector. Firms met with a better matched migrant inflow thus grow faster overall, consistent with increases in productivity and/or increases in demand for their products. This corroborates the finding of increases in productivity of workers hired through referrals in online labor markets (Pallais and Sands 2016), and goes against recent theoretical predictions that relying on hiring through networks may reduce productivity (Chandrasekhar, Morten, and Peter 2020). Note that, while these effects seem large, the magnitudes of differential growth rates are relatively small in practice: The 90-10 percentile range of relative network size shocks experienced by firms in the sample is 0.05 across all firms, and 0.4 across firms hiring any refugee nation employees at baseline. This implies a 90-10 percentile range of the employment effect of 1% and 9% respectively for firms in our sample.

How does this increase in hiring driven by changes in migrant networks impact wage setting? IV estimates of wage impacts are generally more positive for migrant employees, and less positive for Swiss workers. This is consistent with high wage-growth regions attracting larger voluntary migrant inflows. Focusing on the IV estimates, doubling a firm's network size increases average wages paid by the firm by a statistically significant 17% (Row 1, Table 1.4). Interestingly, this effect is primarily driven by wage increases for Swiss employees, who are least likely to be substitutes for refugees.

In fact, wages of other migrants do not change, and wages of refugee employees *fall* by 30%, though this difference is not statistically significant ($p = 0.11$).

Wage effects among all employees are similar to wage effects on stayers (Panel A vs. Panel B): Stayers in firms doubling their network size experience a 16% increase in wages compared to other firms. As above, wage impacts become more positive as the likely degree of substitutability with refugees decreases: existing refugee employees experience marginally significant declines while other migrants and Swiss natives experience wage gains – though the effect is only significant at conventional levels for native workers. Interestingly, low-skilled natives appear to have lower wage increases (or even slight decreases) than less than high-skilled natives, though we cannot reject that wage effects are zero.[33]

Panel C moreover shows that new hires in firms experiencing an increase in their network are hired at lower hierarchies. The average increase in the percentile rank of existing employees within the firm wage distribution is a statistically significant 7%. Both existing refugee employees and natives rise in the firm's wage ladder, though this rise is only significant for natives. Interestingly, we do not find clear evidence that existing co-national refugee employees are compensated by higher wages for potential referrals.

The evidence presented is in line with evidence in Kramarz and Skans 2014, who find that referral hires are often hired at lower initial wages. Moreover, it echoes the findings of Foged and Peri (2016) and Ottaviano and Peri (2012) who argue that immigration leads to skill-upgrading of native workers. We add to the existing interpretation by showing that this holds across firms within cantons experiencing a firm-specific shock to their network, and offer a new interpretation through referrals: Growing networks among a firm's existing migrant employees increase the supply of referrals. More network members are hired – with limited evidence for wage reductions – while existing employees (particularly Swiss natives) rise up in the firm ladder.[34]

1.6 Isolating the information sharing / referral channel

In the previous sections, we have shown that migrant networks benefit individual migrants, and increase network-based sorting into firms with existing migrant employees. Moreover, larger inflows into a firm's network increase overall firm employment and wages for existing employees, while new refugee employees are hired at somewhat lower wages. This evidence is consistent with networks sharing information on employment opportunities, and firms hiring newly arriving refugees through referrals or information shared within employee networks. However, although existing evidence has often been interpreted in this vain, network benefits for individual migrants may also result from support not directly tied to employment (e.g., as in Blumenstock, Chi, and Tan 2021)

[33]Note: We use education data from the labor force survey that only covers $\approx 30\%$ of the labor force in each round, and 50% in any round, explaining the reduction in sample size for this composition.

[34]In ongoing work, we investigate potential selection effect by following initial employees across firms, and controlling for observable employee characteristics. This allows us to test for differential hiring / firing of employees within cell based on observables.

or from natives responding more favourably to more familiar migrant communities. Moreover, if refugees from different origins are imperfect substitutes, differential within-firm wage responses may also be explained by monopsonistic wage setting of firms.

In this section, we use quasi-exogenous variation in social exposure through co-residence of migrants in refugee reception centers. These differences in social ties *within* an origin-by-canton network allow us to disentangle the mechanism of information sharing within social networks and job referrals from these alternative explanations. As described in Section 1.2, refugees reside in cantonal reception centers for the first 6-12 months upon arrival. Within cantons, allocation to these centers is plausibly exogenous, based on availability of beds, and conducted by a cantonal official that has usually had very limited exposure to a migrant before making an allocation decision. Each center houses 100-150 refugees at a time, meals are typically communal, as are many common spaces: living together therefore creates substantial interactions.

Our strategy is to test whether – within co-national migrant networks – social exposure is associated with a higher probability of working for the same employer. Cantonal officials have somewhat more room for discretion in their allocation across centers than at the federal level as suitable housing requires consideration of practical concerns such as the availability of rooms, child friendliness, disabilities, safety, etc. Because of this, we control for an increasingly detailed set of characteristics for *both* individuals when estimating Equation 1.4, as similar characteristics may *at least in principle* increase the likelihood of being allocated to the same center. We exclude pair of workers within the same family/household, and any refugees still living at the center to isolate the effect of social connections on *future* co-working relationships.

Two randomly chosen refugees assigned to the same canton have a 0.3% probability of working together (Column 1, Table 1.5). Co-nationals are 3 times more likely to be employed by the same employer, at an average of 1%. The difference is highly statistically significant ($p = 0.000$), reiterating the observation above that refugees strongly sort into employers based on nationality. Strikingly, refugees assigned to the same residential center in the same year are 0.2% more likely to work together if they do not share a nationality, but a full 1.8% more likely ($p = 0.000$) if they are from the same origin. This is a very large effect, tripling the baseline likelihood of working for the same employer among co-nationals. Hence, residential centers are highly predictive of later co-working relationships, particularly among co-nationals.[35]

We next focus on co-national networks only (Columns 2–5), adding increasingly rich interactions for each ij pair. Column 2 adds individual fixed effects, Column 3 a full set of bilateral interactions for demographics that potentially affect center allocation (including arrival year, sex, marital status, ethnicity, and arrival age). Column 4 adds interactions for the main categories extracted from the free text communication between allocation officer and the federal reception center (including dummies for any mentions of core family members, other family members, peers, medical conditions, linguistic preferences, prison sentences, births,

[35] This is consistent with anecdotal evidence and personal experience that even within residential centers, refugees tend to associate strongly along the lines of origin nationality and language.

pregnancies, disappearances) as well as categories of the asylum process such as whether they are a Dublin case[36], or whether their asylum claim was rejected or accepted temporarily (see section 1.2). Despite this rich set of controls, the coefficient on co-residence remains relatively stable and highly statistically significant ($p = 0.000$). This highlights that allocation to centers is indeed quasi-exogenous – if all these observables do not affect reception center assignments, it is unlikely that any other unobservable characteristics would change our conclusion. Thus, co-residence has a robust causal effect on co-working even *after* co-residence ends.

However, being assigned to the same cantonal reception center may not only affect social interactions, but also where refugees locate *after* leaving the center – either for convenience or because cantonal officials assist with housing closer to their current center. It is possible, therefore, that the impact of co-residence is primarily due to location, rather than social interaction. In Column 5, we therefore add a full set of 1406 x 1406 fixed effects for refugees' current municipality of residence (i.e. after leaving the center). This reduces the coefficient by roughly 50% – but even controlling for residence municipality, refugees are twice as likely to work for the same employer if their residence overlapped at a cantonal residence center. Thus, social interactions explain at least 50% of the overall effect of co-residence. This share may be even higher if refugees' residential choices after leaving the center are partly a result of these social interactions (e.g., refugees that get to know each other well in the center may later decide to locate closer to each other). Thus, the impacts of quasi-randomly induced social connections – 6-12 months co-residence in a dorm-style accommodation is likely to lead to 'strong ties' – are large, and persist even after leaving the residence.

This causal evidence of social connections on co-working rules out any sorting into employers based on similar characteristics within nationality cohorts, as well as any pure relative labor supply channel. The effect on sorting is large, suggesting that a substantial part of overall network effects on refugees' labor market integration and firms are likely driven by referrals or information sharing within social networks themselves. Since we do not directly observe referrals, we cannot distinguish between those last two channels. However, for many practical and policy purposes they may yield the same conclusions.

1.7 Conclusion

Using comprehensive administrative data and a novel identification strategy, we have shown that networks have large and persistent effects on the labor market trajectories of migrants. Larger migrant communities lead to substantially higher labor market participation and higher average earnings of migrants. They benefit most from being located near individuals who are demographically similar to them and economically successful. Using data on migrants' residence histories and quasi-random social exposure of refugees *within* networks, we establish that a substantial share of the overall network effects are driven by direct social connections, and information sharing / referrals within

[36] I.e., a person covered under the Dublin treaty, which states that refugees may only claim asylum in one European treaty county. Refugees claiming asylum in a second country are then usually deported to their initial asylum country, where their request is processed.

social networks. Moreover, we find no evidence supporting concerns about the negative labor market impacts of concentrated immigration. On the contrary, we show that migrant networks benefit firms with existing migrant employees by improving firm-worker matches through referrals. Firms grow, and pay higher wages. Native employees in particular experience wage *gains*, and move up within the firm hierarchy.

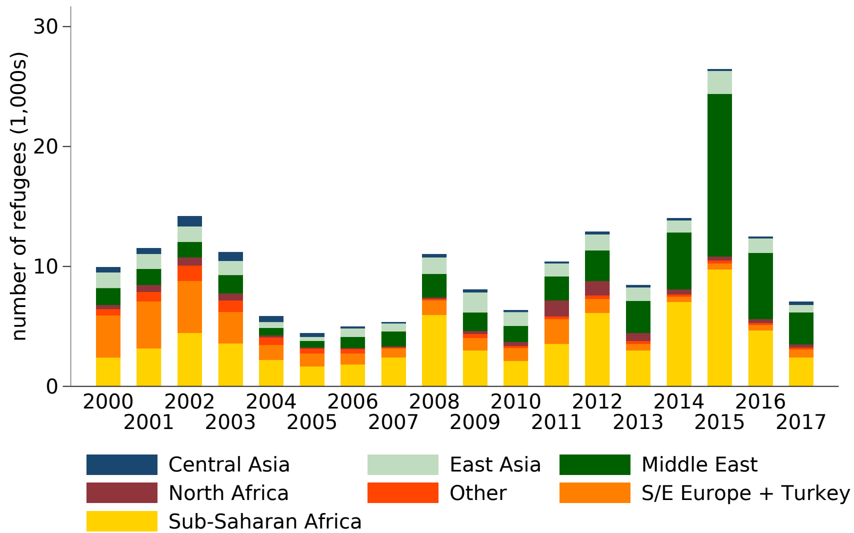
Our findings suggest that existing migration policies – particularly concerning refugees – may need to be reconsidered. Many high income countries adopted legislation that constrains refugee's labor market participation, including employment bans in the initial stages of the asylum process. Our results show that dispersal policies reduce the labor market prospects of migrants *as well as* the gains to local firms and employees through better firm-employee matches. Prolonged refugee unemployment moreover implies a higher fiscal cost of housing and financial support for this population, and may have longer term social and economic hysteresis effects (e.g. Kroft, Lange, and Notowidigdo 2013). Refugee numbers are likely to rise going forward, and aging demographics in Europe highlight a need for increased immigration. Well-designed migration policies can benefit migrants themselves, have positive effects on local firms and employees, and lower the fiscal burden from social and unemployment assistance paid to refugees.

Using machine learning algorithms, Bansak et al. (2018) and Ahani et al. (2021) show that expected refugee labor market outcomes could be improved by 30–70% when matching refugees optimally to locations based on individual characteristics. Sale (2021) further takes into account the dynamic impacts of refugees on earlier and later cohorts, finding optimal placement would increase refugee employment by 27%. Together with the findings presented in this paper, this suggests that host countries may want to consider group-based allocation rather than spreading out individual applicants separately to harness increasing returns to group size in allocation. Importantly, these existing approaches do not take into account the potential effect of refugees on local firms and employees. This paper shows that increased spatial concentration of refugees, and reductions in the barriers to employment in the initial stages of the asylum process are unlikely to negatively affect native employees. In fact, they may even have positive spillovers through migrant networks, as employed migrants help new arrivals integrate. Our results also suggest that policies aiming to harness information contained within migrant networks – e.g. through mentorship programs – may be promising.

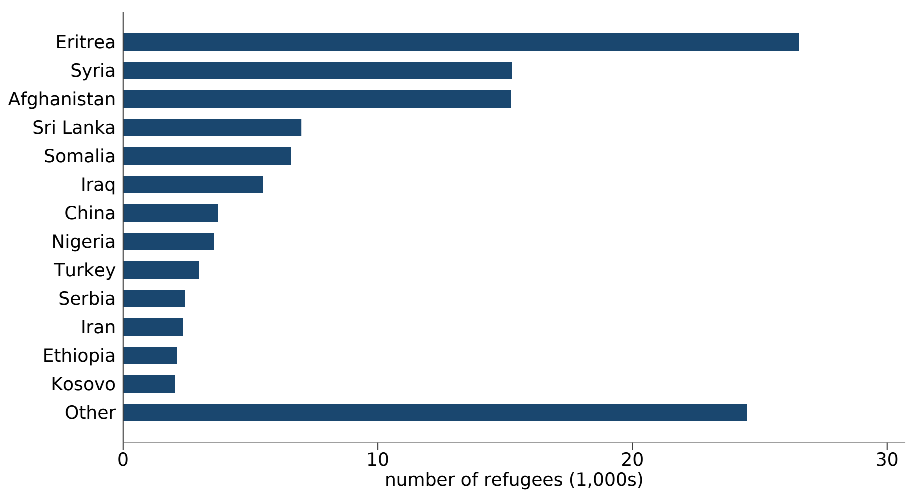
However, while we show that concerns about adverse economic impacts of concentrated immigration are not warranted in our context, there remains an open question as to whether there is a trade off between economic and civic integration. Are larger migrant communities thriving economically, but socially and politically less integrated? Or does economic integration promote social and civic engagement? In ongoing work, we use our data on language spoken at home, residential segregation, intermarriage, female labor force participation, and a novel dataset on all first names of new-born babies over our observation period – all matched to the migration registry – to study how migrant networks affect integration along those dimensions.

1.8 Tables and Figures

Figure 1.1: Refugee numbers and composition



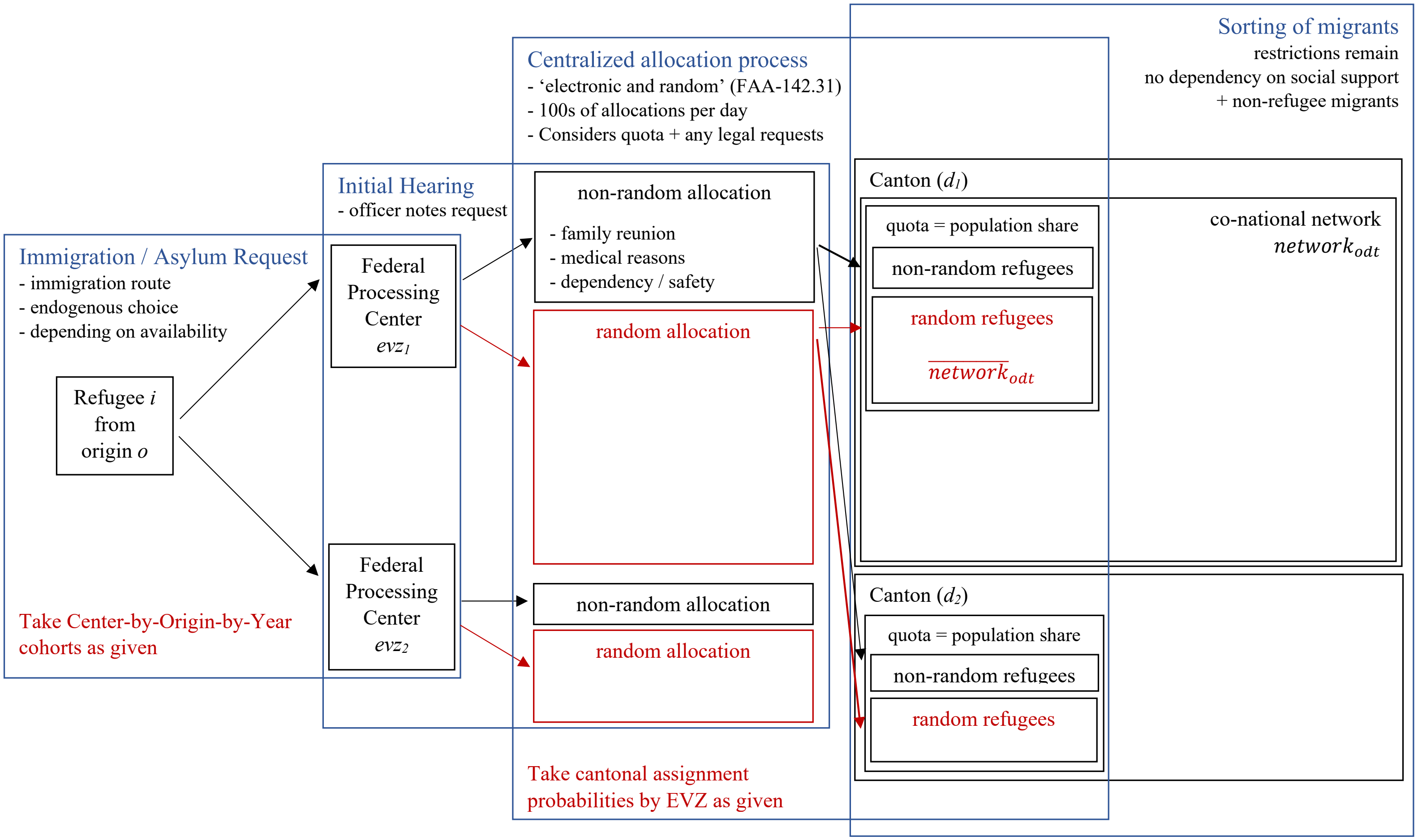
(A) Distribution of asylum applications over time



(B) Main origin nationalities (2008-2017)

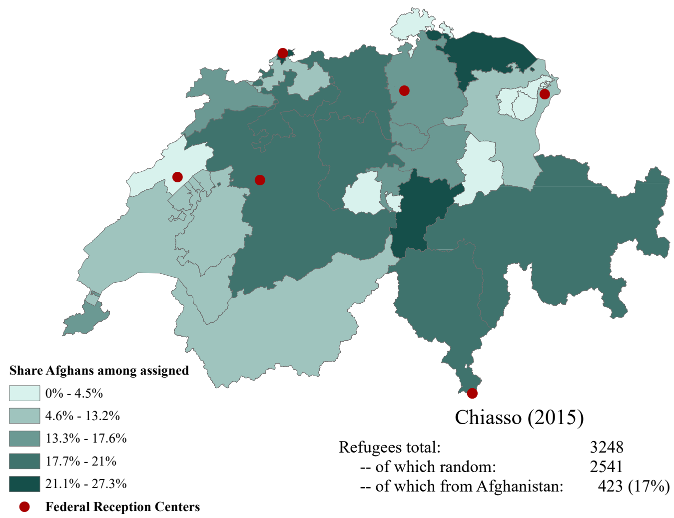
Notes: Panel A shows the distribution of refugee origin nationalities at arrival between 2000 and 2017. We include any refugees present in Switzerland at the end of each year. This is fewer than total asylum applications, since $\approx$ 32% leave Switzerland by the end of each year due to immediate rejections of their asylum requests. Panel B shows counts of the main sending nations between 2008 - 2017, the period on which our main analyses focus.

Figure 1.2: Schematic representation of the cantonal allocation process

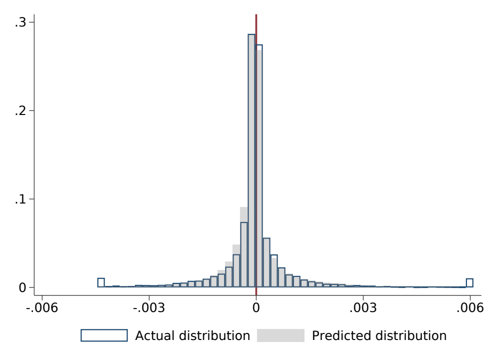


Notes: This figure schematically illustrates the cantonal allocation process of refugees after arrival in Switzerland. Refugee's first arrive at one of several federal reception centers (EVZ), and are then allocated to one of 26 cantons based on each canton's population shares. We highlight in red which aspects of this allocation we consider exogenous, while the other components are taken as given. The area of each box represents the size of each group in the data. Section 1.2 describes this allocation process in detail, and Section 1.3.

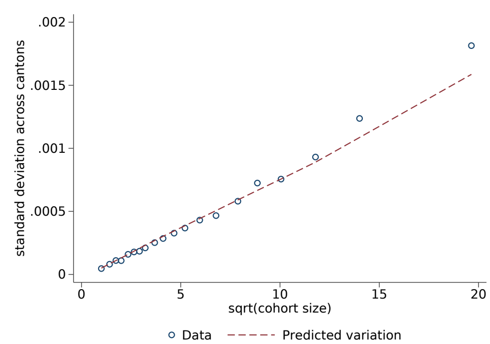
Figure 1.3: Instrument construction and validity



(A) Illustration of quasi-random refugee allocation and instrument construction



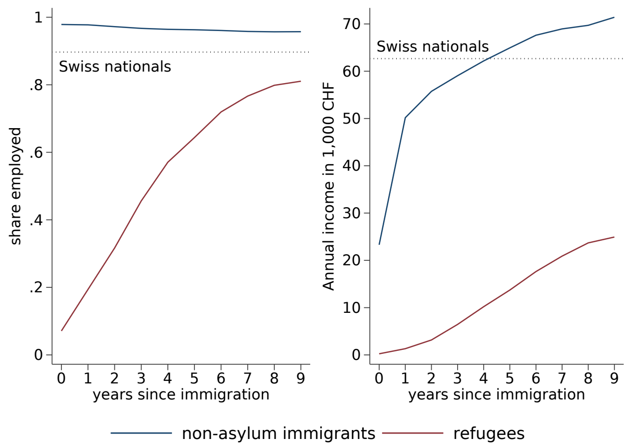
(B) Distribution of the instrument



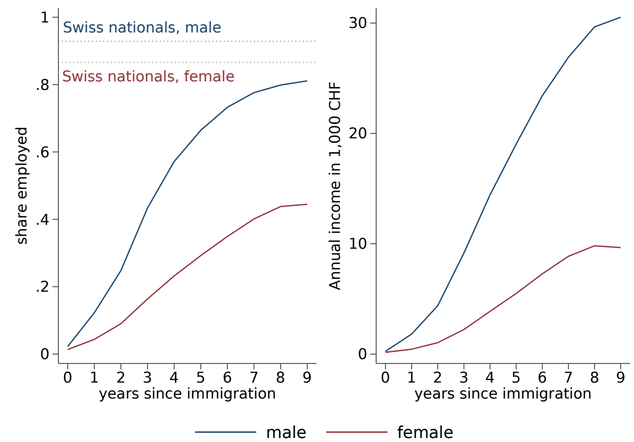
(C) Distribution by cohort size

Notes: Panel A illustrates the cantonal allocation process using the example of Afghan refugees arriving in the federal reception center in Chiasso in 2015. Depicted are Switzerland's 26 cantons, and the permanent federal reception centers. In 2015, Chiasso received 3248 asylum requests, and 2541 of those are random (see Section 1.2). Among those randomly allocated, 423 (17%) were Afghans. For constructing the instrument, we take total allocations to each canton each year as given and permute only within randomly allocated refugees. In expectation, 17% of each canton's randomly assigned refugees should therefore be Afghans. Our instrument takes the difference between actual allocations (represented by the colored shading) and the expected number. Panels B and C plot the distribution of our instrument (in blue) against the expected distribution (in gray/maroon). We obtain expected distributions by simulating the cantonal allocation process – taking nationality-by-reception-center-by-year cohorts, center-by-year-by-canton allocations, any non-random allocations, and family structures of asylum seekers as given. By design, the instrument is mean zero (Panel B). Under our characterization of the allocation mechanism, it follows a Binomial distribution for each cohort, the standard deviation increasing with the square root of the cohort size (Panel C).

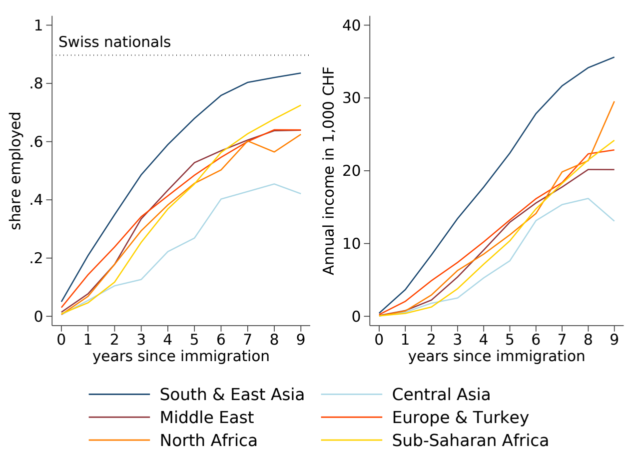
Figure 1.4: Integration trajectories of refugees



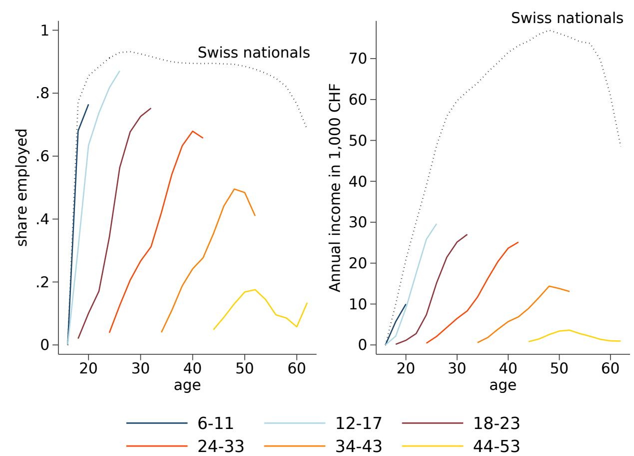
(A) Compared to non-refugee migrants



(B) By gender



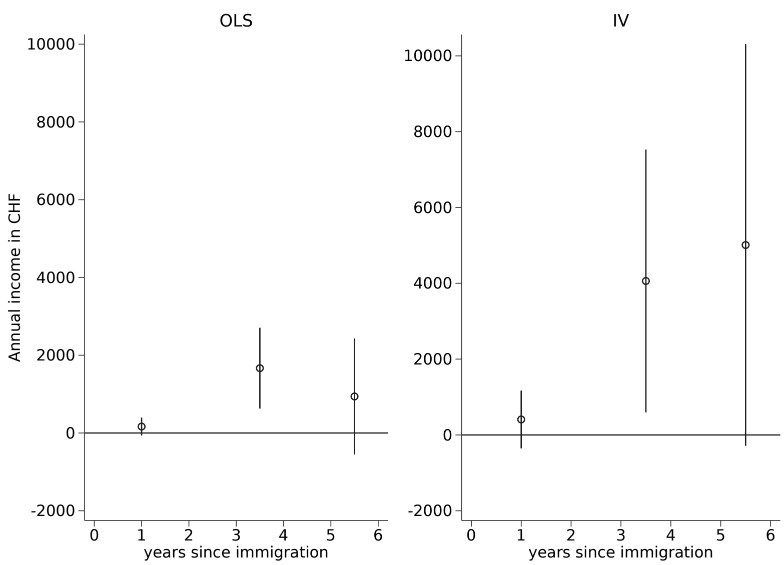
(C) By origin region



(D) By arrival age

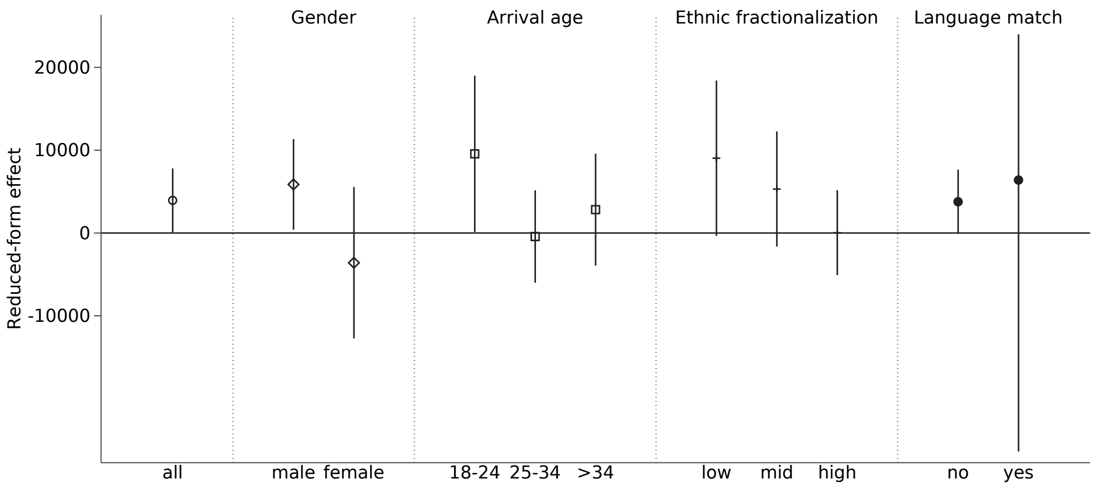
Notes: This figure presents dynamic average employment probabilities and labor income of all refugees arriving between 2009 - 2017, at an age between 19 - 54 (except in Panel D), and conditional on being resident in Switzerland. Averages for non-asylum immigrants and Swiss nationals are also restricted to individuals aged 19-54. Data on employment and labor income are from the central social security registry (see Section 3.2 for details).

Figure 1.5: Dynamic Impact of Networks on Refugee Outcomes



Notes: This figure plots dynamic impacts of the number of co-national network members in the assigned canton on individual refugees' labor market earnings. We include all asylum seekers resident in Switzerland, arriving at age 19 to 54 without any cantonal placement requests, and drop any sending nations with less than 20 refugees over our observation period, or where less than 10% of individuals resident in Switzerland arrived trough the asylum system. Employment and labor income come from the central social security registry. Coefficients are estimated using Equation 1.1, where the instrument is interacted with year-since-arrival dummies. See Figure A.1.3 for reduced form estimates. Standard errors are clustered at the nationality-by-assignment-year level.

Figure 1.6: Heterogeneity in Network Effects



Notes: This figure plots heterogeneity of the reduced form network impacts of the number of co-national network members in the assigned canton in the year prior to arrival on total annual labor income in CHF (roughly 1:1 to USD) 5-6 years after arrival in Switzerland. We include all asylum seekers resident in Switzerland, arriving at age 19 to 54 without any cantonal placement requests, and drop any sending nations with less than 20 refugees over our observation period, or where less than 10% of individuals resident in Switzerland arrived trough the asylum system. Coefficients are estimated using Equation 1.1, separately for each subgroup considered. The instrument is scaled relative to the lagged population, so that coefficients can be interpreted as effects of a quasi-exogenous doubling of the network size (see Section 1.3). Gender and arrival age come from the refugee registry. Ethnic fractionalization is defined as terciles of an index of within-origin-country ethnic fractionalization, obtained from and described in Fearon 2003. Language match is an indicator for when a refugee's origin nationality shares an official language with the assigned canton (e.g. French-speaking West-African nations assigned to the French part of Switzerland). Standard errors are clustered at the nationality-by-assignment-year level.

Table 1.1: Long-run impacts of Networks on Refugee Labor Market Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Employment probability			Annual income (in CHF)			Monthly income (in CHF)		
	(OLS)	(Reduced Form)	(IV)	(OLS)	(Reduced Form)	(IV)	(OLS)	(Reduced Form)	(IV)
ArcSinh(Number of co-nationals)$_{od,\underline{t}-1}$	0.0234		0.153**	939.8		5010.4*	24.14		40.18
	(0.0201)		(0.0761)	(761.7)		(2702.9)	(74.75)		(264.7)
$\sum_{l=1}^{3}$ Inflow of co-nationals$_{od,\underline{t}-l}$		0.121**			3946.0**			30.95	
		(0.0539)			(1967.3)			(203.9)	
		[0.05]			[0.08]			[0.88]	
Observations	8413	8413	8413	8413	8413	8413	4462	4462	4462
Nationality-by-Arrival-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nationality-by-Years-in-CH FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Assignment-Canton FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of origin nations	49	49	49	49	49	49	43	43	43
Number of allocation cantons	26	26	26	26	26	26	26	26	26
Mean of dependent variable	0.537	0.537	0.537	13727.3	13727.3	13727.3	2552.7	2552.7	2552.7
Mean of independent variable	6.557	0.00350	6.557	6.557	0.00350	6.557	6.473	0.00683	6.473
90-10 percentile range of independent variable	0.721	0.122		0.721	0.122		0.727	0.141	
First-Stage coefficient			0.788			0.788			0.770
			(0.132)			(0.132)			(0.143)
First-Stage F-Statistic			35.86			35.86			28.85

Notes: This table presents estimates of the long-run impacts of the number of co-nationals resident at each refugee's assignment canton in the year prior to assignment (Equation 1.1). We include all randomly assigned asylum seekers resident in Switzerland 5-6 years after arrival, arriving at age 19 to 54 without any cantonal placement requests, and drop any origin nations with less than 20 refugees over our observation period, or where less than 10% of individuals resident in Switzerland arrived trough the asylum system. Employment and labor income come from the central social security registry (see Section 3.2). Standard errors (in parentheses) are clustered at the nationality-by-assignment-year level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct level. Square brackets contain Fisher randomization inference p-values based on 999 replications of the allocation process (i.e. the share of t-statistics across all simulations exceeding the realized t-statistic in absolute value), taking nationality-by-year-by-reception center cohorts, exceptions from random allocations, center-by-canton-by-year assignment probabilities and family structures as given (see Section 1.3). Figure A.1.2 plots the distribution of simulated t-statistics.

Table 1.2: Impacts of Network Composition on Long-run Labor Market Outcomes

	(1) Annual income in CHF	(2) Annual income in CHF	(3) Annual income in CHF	(4) Annual income in CHF	(5) Annual income in CHF	(6) Annual income in CHF	(7) Annual income in CHF
ArcSinh(Number of refugee co-nationals)$_{od,\underline{t}-1}$	1381.3***	1806.4***	1800.8***	1690.5***	1542.4***	1413.1**	1597.1***
	(507.0)	(541.4)	(551.3)	(535.3)	(552.3)	(565.1)	(532.0)
Share same sex		7060.8***					6407.3***
		(2024.8)					(1958.6)
Share same age group			2414.7*				235.0
			(1366.1)				(1283.7)
Share with secondary education				-918.6			-367.0
				(1390.8)			(1356.5)
Share employed					3203.8		
					(1942.9)		
Average income of network (in CHF)						0.143***	0.106*
						(0.0539)	(0.0542)
Observations	24840	24691	24652	24648	24691	24691	24648
Controlling for interacted variable(s)		Yes	Yes				Yes
Mean of dependent variable	16673.1	16654.9	16677.6	16680.3	16654.9	16654.9	16680.3
Mean of RHS	5.643	0.563	0.355	0.269	0.248	0.248	
90-10 percentile range of RHS	3.586	0.404	0.373	0.428	0.327	0.327	

Notes: This table presents OLS estimates of the long-run impacts of the size as well as the composition of the network of refugee co-nationals resident in a refugee's assignment canton in the year prior to assignment (Equation 1.1) on total labor income in CHF (roughly 1:1 to USD). We include all randomly assigned asylum seekers arriving between 2008 and 2012, resident in Switzerland 5-6 years after arrival, arriving at age 19 to 54 without any cantonal placement requests, and drop any origin nations with less than 20 refugees over our observation period, or where less than 10% of individuals resident in Switzerland arrived trough the asylum system. For endogenous variables depending on a refugee's own characteristics (e.g. age), we include fixed effects for the categories of that characteristic. Employment and labor income come from the central social security registry (see Section 3.2). Standard errors are clustered at the nationality-by-assignment-year level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 1.3: Impacts of Migrant Networks on Firm Employment and Productivity

	(1)	(2)	(3)	(4)
		Δ_6ArcSinh(No. of Conationals)	$\sum_{l=1}^{5}$ network$_{od,t-l}$	Δ_6ArcSinh(No. of Conationals)
	N	(OLS)	(Reduced Form)	(IV)
Δ_6 **ArcSinh(Employment)**				
Refugee nations	131,224	0.006	0.018**	0.529***
		(0.021)	(0.008)	(0.148)
Other migrants	131,224	-0.003	0.005	0.131
		(0.021)	(0.009)	(0.226)
Swiss nationals	131,224	-0.006	0.004	0.129
		(0.014)	(0.004)	(0.126)
Δ_6 ArcSinh(Employment)	131,224	-0.008	0.008***	0.225***
		(0.016)	(0.002)	(0.070)
Δ_6 ArcSinh(Total wage bill)	127,713	0.019	0.015***	0.442***
		(0.018)	(0.003)	(0.097)
Δ_6 ArcSinh(Monthly wage)	127,713	0.028***	0.006***	0.168***
		(0.009)	(0.001)	(0.058)

Notes: This table presents shift-share estimates of the long-run impacts on employment and wages of employment-weighted refugee inflows into a firm's canton (Equation 1.3). Δ_6 indicates a 6-year long difference. $\sum_{l=1}^{5}$ network$_{od,t-l}$ is the proportional deviation (in ArcSinh) from the expected number of randomly assigned refugees from origin o assigned to canton d over the 5 years prior. Each row represents a regression that includes controls for the baseline *level* of the outcome, a dummy for whether the enterprise had any refugee nation employees in 2011 and includes 2-digit NOGA sector fixed effects as well as canton fixed effects. We include all registered single-location firms in Switzerland active in each year between 2011 - 2017 and with at least 3 employees in 2011. Refugee origin nations are all sending nations with more than 20 refugees over our observation period, and where more than 10% of individuals resident in Switzerland arrived trough the asylum system. Employment is measured from the employer-employee matched enterprise registry, the total wage bill comes from the central social security registry (see Section 3.2). Standard errors come from an equivalent weighted regression transformed to the shock-level (i.e. nationality-by-canton-by-year) to account for correlations across firms with similar initial employment shares, where we cluster at the origin-nationality-level (see Borusyak, Hull, and Jaravel (2021)). The minimum first-stage F-statistic across all rows is 21. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 1.4: Impacts of Migrant Networks on Wages

	(1)	(2)	(3)	(4)
		Δ_6ArcSinh(No. of Conationals)	$\sum_{l=1}^{5}$ network$_{od,t-l}$	Δ_6ArcSinh(No. of Conationals)
N	(OLS)	(Reduced Form)	(IV)	
All employees	127,713	0.028***	0.006***	0.168***
		(0.009)	(0.001)	(0.058)
Refugee nations	15,524	0.062**	-0.013***	-0.301
		(0.024)	(0.004)	(0.187)
Other migrants	66,195	0.025*	-0.002	-0.058
		(0.015)	(0.003)	(0.087)
Swiss nationals	120,349	0.009	0.005*	0.149*
		(0.011)	(0.003)	(0.082)
Stayers	121,499	0.049***	0.006**	0.164***
		(0.010)	(0.002)	(0.058)
Refugee nations	11,999	0.145***	-0.011*	-0.280
		(0.040)	(0.007)	(0.276)
Other migrants	51,975	0.046***	0.005	0.146
		(0.014)	(0.003)	(0.117)
Swiss nationals	111,769	0.029**	0.007**	0.203***
		(0.011)	(0.003)	(0.076)
Swiss nationals: primary education	10,080	-0.008	-0.007	-0.174
		(0.026)	(0.006)	(0.171)
Swiss nationals: secondary education	34,506	-0.006	-0.005	-0.139
		(0.016)	(0.005)	(0.172)
Swiss nationals: tertiary education	24,625	0.010	0.004	0.125
		(0.034)	(0.005)	(0.132)
Stayers: Change in wage percentile within firm	121,499	0.007*	0.002*	0.067**
		(0.004)	(0.001)	(0.027)
Refugee nations	11,999	0.023***	0.002	0.061
		(0.006)	(0.003)	(0.069)
Other migrants	51,975	0.014***	-0.000	-0.013
		(0.004)	(0.001)	(0.034)
Swiss nationals	111,769	-0.007*	0.001	0.043*
		(0.004)	(0.001)	(0.023)

Notes: This table presents shift-share estimates of the long-run impacts on wages of employment-weighted refugee inflows into a firm's canton (Equation 1.3). Δ_6 indicates a 6-year long difference. $\sum_{l=1}^{5}$ network$_{od,t-l}$ is the proportional deviation (in ArcSinh) from the expected number of randomly assigned refugees from origin o assigned to canton d over the 5 years prior. Each row represents a regression that includes controls for the baseline *level* of the outcome, a dummy for whether the enterprise had any refugee nation employees in 2011 and includes 2-digit NOGA sector fixed effects as well as canton fixed effects. We include all registered single-location firms in Switzerland active in each year between 2011 - 2017 and with at least 3 employees in 2011. Refugee origin nations are all sending nations with more than 20 refugees over our observation period, and where more than 10% of individuals resident in Switzerland arrived trough the asylum system. Monthly wages come from the central social security registry, and the within-firm percentile rank of stayer i within firm j is defined as rank$(i)\frac{n^2}{n_j(n_j+1)}$ where n_j is the number of employees in firm j (see Section 3.2). Standard errors come from an equivalent weighted regression transformed to the shock-level (i.e. nationality-by-canton-by-year) to account for correlations across firms with similar initial employment shares, where we cluster at the origin-nationality-level (see Borusyak, Hull, and Jaravel (2021)). First-stage F-statistics range between 5.1 and 24, with an average of 15. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 1.5: Impacts of Living Together on Co-Working

	(1) P(co-worker)	(2) P(co-worker)	(3) P(co-worker)	(4) P(co-worker)	(5) P(co-worker)
1(co-national)	0.00709***				
	(0.000690)				
1(overlapped in cantonal center)	0.00192*	0.0197***	0.0136***	0.0141***	0.00682**
	(0.00101)	(0.00344)	(0.00316)	(0.00337)	0.00250
1(co-national) × 1(overlapped in cantonal center)	0.0178***				
	(0.00308)				
Constant	0.00329***	0.0104***	0.00777***	0.00924***	0.00817***
	(0.000586)	(0.00106)	(0.00129)	(0.00238)	(0.000535)
Observations	103078232	7068486	4116116	3517053	3463459
Individual FE		Yes	Yes	Yes	Yes
Demographic interactions			Yes	Yes	Yes
5-year-arrival-age-group interactions			Yes	Yes	Yes
Request category interactions				Yes	Yes
Municipality interactions					Yes

Notes: This table presents OLS estimates of the impact on the probability of working for the same employer of co-residence in a cantonal residential center in the first few months of arrival (Equation 1.4). Each observation is an ij pair of refugees initially assigned to the same canton, and observed as employed in 2017. We exclude all pairs in the same household / applying for asylum as a family, and all individuals who still live in a refugee center. Column 1 focuses on all pairs of refugees assigned to the same canton, while Columns 2-5 include only pairs sharing the same origin nationality. Demographic interactions include a full set of interactions between all categories of sex, marital status and ethnicity for individuals i and j. Request categories include dummies for the free-text comment in the asylum dossier mentioning any core family members, other family members, any peers, any medical exemptions or conditions, any Swiss national language, any prison sentences, anything related to childbirth, or any disappearances, as well as whether the refugee is a Dublin case and whether the asylum request was rejected (see Section 3.2). Municipality interactions contain 1406-by-1406 interactions for the residence municipalities of i and j in 2017. Standard errors are clustered at the nationality-by-nationality level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Chapter 2

General equilibrium effects of cash transfers: experimental evidence from Kenya

2.1 Introduction

Tracing out the pattern of transactions in an integrated economy, and their contribution to aggregates of interest like overall output or well-being, has long been a central task of economic analysis. For instance, there has been interest in understanding the aggregate impacts of fiscal stimulus and cash infusions for decades (Keynes 1936), and a growing body of empirical evidence from rich countries shows that fiscal multipliers can sometimes be positive and large, based on non-experimental variation generated by policy changes (Chodorow-Reich 2019; Nakamura and Steinsson 2014; Suarez Serrato and Wingender 2016; Auerbach, Gorodnichenko, and Murphy 2020). Until now, however, these issues have not been subjected to experimental examination.

There is also renewed interest in related topics in development economics with the rise of large-scale cash transfer programs, which have now been implemented in scores of low and middle income countries.[1] A large literature on the impacts of these transfers has developed, employing well-identified experimental and quasi-experimental designs. These studies have documented effects on a broad range of behavioral responses among treated households, including consumption, earnings, assets, food security, child growth and schooling, self-reported health, female empowerment, and psychological well-being (Haushofer and Shapiro (2016), Baird, McIntosh, and Ozler (2011), and Bastagli et al. (2016)). Yet there is limited evidence on the aggregate economic impacts or welfare consequences of such policies (for exceptions, see Angelucci and De Giorgi (2009), Cunha, De Giorgi, and Jayachandran (2018), and Filmer et al. (2018)).

The present study was prospectively designed to unite these two disparate literatures by experimentally studying the aggregate impacts of large cash stimulus programs. We designed and carried out a large-scale experiment in rural Kenya that provided one-time

[1] 97% of developing countries in Europe, Latin America and Sub-Saharan Africa have some type of cash transfer program (World Bank 2017).

cash transfers worth roughly USD 1000 (distributed by the NGO GiveDirectly) to over 10,500 poor households in a sample of 653 villages with a population of roughly 300,000. The implied fiscal shock was large, as the cash transfers amounted to over 15% of GDP in treatment villages during the peak 12 months of the program.

Beyond its fiscal scale, at least three aspects of the project represent advances on most existing work. First, we generated substantial spatial variation in the intensity of transfers by deliberately randomizing the allocation of cash transfers not just across households or villages (as is typical), but also across geographic sublocations (groups of 10–15 villages), thereby increasing our ability to detect aggregate impacts. Second, we carried out unusually extensive original data collection, giving us greater visibility into the chain of causal effects linking cash transfers to aggregate outcomes in a complex and interconnected economy. Our household and enterprise censuses of the study area count 65,383 households and 12,095 non-farm enterprises. Within this unusually large sampling frame, we gathered detailed panel (longitudinal) data on household receipt of the transfer; household consumption expenditure patterns (representative for both recipient and non-recipient households); local enterprise production, employment and revenue; labor market conditions; as well as especially high-frequency (monthly) and spatially disaggregated market data on prices. Third, we interpret the results through the lens of a theoretical framework that highlights the links between the individual empirical results, the aggregate transfer multiplier, and welfare in this setting.

Following earlier studies, we first document large direct impacts on households that received transfers, including increases in consumption expenditures and holdings of durable assets eighteen months after the start of transfers. We do not observe meaningful changes in total labor supply among recipient households.

Enterprises in areas that receive more cash transfers also experience meaningful revenue gains, in line with the increases in household expenditure. Interestingly, sales increased without noticeable changes in firm investment behavior (beyond a modest increase in inventories), and did not increase differentially for firms owned by cash recipient households relative to non-recipients. Both patterns suggest a demand- rather than investment-led expansion in economic activity. Increased enterprise revenue in turn translates into moderate increases in wage bills and profits. Methodologically, one important feature of the enterprise (and to a lesser extent household) results is that they are largely driven by the overall intensity of treatment in nearby communities, not solely by the treatment status of the village in which the enterprise is located. This suggests that common study designs which aim to identify spillover effects by clustering treatment at the village level and assuming no spillovers across villages may be mis-specified, at least in densely populated areas such as the one we study.[2]

Despite not receiving transfers, non-recipient households also exhibit large consumption expenditure gains: their annualized consumption expenditure is 13% higher eighteen months after transfers began, an increase roughly comparable to the gains contemporaneously experienced by the recipient households.[3] Increased spending is not financed by dissaving, but

[2]For example, households are located within 2 km of seven other villages on average.

[3]We note below that consumption gains among recipient households are likely to have been larger in the period immediately following transfer receipt.

more likely results in part from the income gains experienced by local firms' owners and workers. Indeed, non-recipients' income gain is driven largely by increases in wage labor earnings, consistent with the fact that enterprise wage bills increase. In a reassuring check, the magnitude of per capita consumption gains among local households lines up roughly with the per capita revenue gains among local firms. On some level this is unsurprising, as increases in local consumption expenditures were spent somewhere; our contribution is to carefully document how such spending spreads locally through a low-income economy.

A further issue is the extent to which transfers affect local prices (as for example Cunha, De Giorgi, and Jayachandran (2018) show in Mexico), and thus the extent to which the effects described above are nominal or real. We study this question through careful monthly measurement of prices for scores of local commodities, consumer goods, and durable goods, as well as prices for inputs like labor and capital. For inputs, we find positive point estimates, but they are economically moderate in magnitude and not always statistically significant. For outputs, we document statistically significant, but economically minimal, local price inflation. Average price inflation is 0.1%, and even during periods with the largest transfers, estimated price effects are less than 1% and precisely estimated across all categories of goods.

We next ask what these effects imply for the aggregate level of economic activity, computing a local transfer multiplier. A standard macroeconomic framework would predict that large multipliers are possible in our rural Kenyan setting: it is a largely closed local economy within a currency union receiving external transfers, with incomplete markets and a large share of hand-to-mouth consumers (Farhi and Werning 2016). Using an expenditure-based approach that takes advantage of our data on the consumption expenditures of representative samples of both recipient and non-recipient households as well as investment by local firms, we estimate a local transfer multiplier of 2.6. A dual income-based approach, relying on distinct and complementary measures of labor and capital income, enterprise profits, and taxes, yields a similar estimate of 2.5. These estimates are broadly in line with what a simple model would imply from households' marginal propensity to consume local value added, which we estimate to be approximately 0.76 over the study period. These results contribute to an active recent empirical literature that estimates multipliers.[4]

A core contribution of this study is thus to exploit a randomized experiment to estimate an important macroeconomic quantity (see Muralidharan and Niehaus (2017) for a related discussion). A notable aspect of our approach is the fact that transfers came from donors outside the study area, rather than being internally tax- or debt-financed; the latter is typically the case in the US programs studied, and may complicate the

[4]Our estimates are somewhat larger than those from a structural simulation, which predicted that the local multiplier from cash transfers in rural Kenya could range from 1.6 to 1.9 (Thome et al. 2013), and are similar in magnitude to non-experimental estimates from a cash transfer program in Mexico (1.5 to 2.6) (Sadoulet, Janvry, and Davis 2001). They are also somewhat larger than recent estimates of the fiscal spending multiplier (which is distinct from the transfer multiplier, since households can save part of the transfer) derived from cross-sectional US policy variation, which often range from 1.5-2.0 (Chodorow-Reich 2019), and from Brazil, which are close to 2 (Corbi, Papaioannou, and Surico 2019). Pennings (2021) focuses on the US transfer multiplier. Kraay (2014) estimates fiscal multiplier estimates less than one when donor lending is used as an instrument for national government spending in developing countries.

interpretation of consumption responses due to contemporaneous tax incidence or Ricardian equivalence issues. The targeting of the transfers to just one region within the larger Kenyan economy also allows us to abstract away from monetary policy and exchange rate responses, simplifying analysis relative to the study of national stimulus policies. A limitation of our approach is that we observe partial data in the months immediately following the transfers, which reduces the precision of some estimates.

Few existing treatments of multipliers also explicitly examine their welfare implications.[5] We interpret the welfare implications of our results using a simple theoretical framework. Transfers directly increase the welfare of those who receive them by \$1 per \$1 received. General equilibrium effects impact welfare through two additional channels: changes in household budget sets (due for example to changes in wages, prices, or firm profits), and changes in peer behaviors that enter directly into own utility (due to externalities or public good provision). The value of budget set expansions depends on what drives them: expansions due to increases in productivity are worth \$1 per \$1, while expansions due to increased factor supply (e.g., labor) come at a partially offsetting opportunity cost. Interpreted through this lens, the results generally suggest that non-recipients were made better off by an expansion in their budget sets driven largely by increased factor productivity, as opposed to factor supply. Externality effects are positive or null, with one possible exception: positive spillovers were large enough that village-level asset inequality increased slightly, which may affect well-being if households have preferences over their relative socioeconomic standing.

The constellation of findings raises an intriguing question about how the economy absorbed such a large shock to aggregate demand, and in particular how it did so without correspondingly large increases in the employment of land (which is in fixed supply), labor, or capital. One plausible, albeit speculative, possibility is that the utilization of these factors was "slack" in at least some enterprises (Lewis 1954). This seems plausible because in the retail and manufacturing sectors, where output responses were concentrated, the typical firm has a single employee (i.e., the proprietor), suggesting that integer constraints may often bind. In addition, many enterprises operate "on demand" in the sense that they produce only when they have customers, and the average non-agricultural enterprise sees just 1.9 customers per hour. In addition to retail, much manufacturing in this setting is on demand; for example, a mill owner waits for customers to bring their grain. The existence of slack may help account for the large multiplier we estimate, as has also recently been argued in US data (Michaillat and Saez 2015; Murphy 2017).

2.2 Study design

Setting: rural western Kenya

The study took place in three contiguous subcounties of Siaya County, a rural area in western Kenya bordering Lake Victoria. The population in Siaya is predominantly Luo, the second largest ethnic group in Kenya, and while rural it is also relatively densely pop-

[5] Recent and notable exceptions include Mankiw and Weinzierl (2011) and Sims and Wolff (2018).

ulated, with 393 people per km^2 compared to a Kenyan average of 88. The main national road running from the port of Mombasa to Nairobi and on to Kampala passes through the study area, likely helping to integrate it into the regional economy.

The NGO GiveDirectly (GD) selected the study area based on its high poverty levels. Within this area GD selected rural (i.e., not peri-urban) villages in which it had not previously worked.[6] This yielded a final sample of 653 villages spread across 84 sublocations (the administrative unit above a village). The mean village consists of 100 households, and at baseline, the average household had 4.3 members, of which 2.2 were children. The average survey respondent was 50 years old and had about 5 years of schooling. 98% of households were engaged in agriculture; at endline, 46% of households in control villages were also engaged in wage work and 49% in self-employment.

Transfers and data collection took place from mid-2014 to early 2017, a period of steady economic growth, relative prosperity, and political stability in Kenya. The World Bank reports annual per capita GDP growth rates ranged between 1.1 to 2.4 percent. All data collection concluded months prior to the August 2017 national election.

Intervention: The GiveDirectly (GD) Cash Transfer Program

GD provides unconditional cash transfers to poor households in low-income countries. For the purpose of this study, to be eligible for transfers, households had to live in homes with thatched roofs, a simple means-test for poverty. In treatment villages, GD enrolled all households that met this criterion ("eligible" households) as classified by their field staff through a village census, and confirmed via two additional visits (see Appendix B.5).

Approximately one-third of all households were eligible. These households received a series of three transfers totaling KES 87,000, or USD 1,871 PPP (USD 1,000 nominal), via the mobile money system M-Pesa, which is widely used in Kenya. (Registering for M-Pesa was a prerequisite for receiving transfers; households without a mobile phone were given the option to purchase one from GD staff with the cost deducted from their transfer.) Households selected the member they wished to receive the transfers. The total transfer is large, corresponding to 75 percent of mean annual household expenditure in recipient households. In aggregate, the transfers were equivalent to approximately 16 percent of annual GDP (based on our data described below) in the treated areas during the peak 12 months of disbursements, and to 24 percent of annual GDP during the full 24 month rollout period. Although small in relation to overall Kenyan GDP in 2015 (<0.1%), locally this is thus a larger relative shock than most government transfer programs, e.g., the ARRA programs studied in the recent US fiscal multiplier literature, see Chodorow-Reich (2019).

Transfers were made in a series of three payments as follows: a token transfer of KES 7,000 (USD 151 PPP) was sent once a majority of eligible households within the village had completed the enrollment process, followed two months later by the first large installment of KES 40,000 (USD 860 PPP). Six months later (and eight months after the to-

[6] The listing of villages was based on the 2009 National Population Census; enumeration areas (which typically correspond to a single village) were treated as villages by GD and this study.

ken transfer), the second and final large installment of KES 40,000 was sent. The median treated household received its token transfer 47 days after being registered for the program; transfers should be interpreted as anticipated during that period to the extent recipients believed GD's promises.[7] The transfers were non-recurring, i.e., no additional financial assistance was provided to recipient households after their final installment, and they were informed of this up front. Households in control villages did not receive transfers.

Experimental design

To identify spillovers both within and across villages, we employed a two-level randomization design (Figure B.1.1, Panel A). First, we randomly assigned sublocations (or in some cases, groups of sublocations) to high or low saturation status, resulting in 33 high- and 35 low-saturation groups. Within high (low) saturation groups we then randomly assigned two-thirds (one-third) of villages to treatment status. We also randomized the order in which treatment was rolled out to treated villages. Within treatment villages, all eligible households received a transfer.[8] This design induces variation in treatment intensity across space due both to the variation in sublocation treatment intensity, and random variation in the location of treated villages within sublocations. Figure B.1.2 illustrates that there is considerable variation both across and within sublocations in the share of neighboring villages treated.[9]

2.3 Data and empirical specifications

We conducted four types of surveys, of households, enterprises, market prices, and local public goods. Results from the public goods surveys are presented primarily in a separate paper (Walker 2018), and discussed briefly here. We filed several pre-analysis plans for this project; for details on the PAPs and where we go beyond these plans, see Appendix B.10.

Household data

We first conducted a baseline household census in all villages, which serves as a sampling frame and classifies household eligibility status. The census was designed to mimic GD's censusing procedure but was conducted by independent (non-GD) enumerators

[7]The precise timing of the first transfer was uncertain, and we believe that recipients may also have perceived the first transfer as less certain than the subsequent ones. However, we do not know of any borrowing against future GD payments, and credit markets are imperfect in our context. In the earlier GiveDirectly study, Haushofer and Shapiro (2016) find evidence for both savings and credit constraints: households that received lump-sum transfers were more likely to own large durable assets at endline than households receiving monthly transfers, even though the total transfer amounts were the same. This suggests that households had trouble borrowing against the promise of the future transfer and saving the early installments. This is also consistent with US evidence finding no anticipation in advance of the receipt of economic stimulus payments (Broda and Parker 2014). We therefore consider all transfers symmetrically in our dynamic regressions and leave analysis of potentially differing effects across installments for future work.

[8]Full details of the randomization are in Appendix B.5.

[9]Figure B.5.1 provides a higher-resolution example for two villages.

across both treatment and control villages for consistency. Throughout this paper, we base our analysis on village membership, household definitions and eligibility as classified by our project data collection field staff. In all, the census identified 65,383 households with a total baseline population of 280,000 people in study villages.

We conducted baseline household surveys within one to two months after the census and before the distribution of any transfers to a village (Figure B.1.1, Panel B).[10] We used census information to sample at random eight eligible and four ineligible households per village to survey. When households contain a married or cohabiting couple, we randomly selected one of the partners as the target survey respondent. Due to time and budget constraints, we sought to complete all baseline household surveys in a single day. If a household on our sampling list was not available on that day, we instead surveyed a randomly-selected replacement household with the same eligibility status. We conducted a total of 7,845 baseline household surveys between September 2014 and August 2015.[11] The survey contained detailed modules on economic activities, asset ownership, psychological well-being, health and nutrition. A large array of baseline characteristics are balanced across treatment and control villages (Table B.6.2, column 2).

Endline household survey data was collected between 9 and 31 months after each household's "experimental start date," meaning the month in which transfers were expected to start in a village assigned to treatment, regardless of their actual treatment timing.[12] The 5th/95th percentiles of timing ranged from 12 to 27 months, and the median survey was conducted 19 months after the experimental start month, or about 11 months after the distribution of the last lump sum transfer (Figure B.1.1, Panel B). This timing implies that some but relatively few households were surveyed in the months immediately following cash transfer receipt, an issue we return to below when estimating the transfer multiplier.

Endline household surveys targeted all households on the initial sampling lists (including those missed at baseline), along with replacement households that were surveyed at baseline. For households that had been surveyed at baseline, we attempted to survey the individual who was the baseline respondent. We conducted a total of 8,239 endline household surveys between May 2016 and June 2017.[13] We achieved high tracking rates at endline, reaching over 90% of eligible and ineligible households in both treatment and control villages, and these rates do not systematically vary by treatment status (Table B.6.1). The only subgroup difference of note is that we are slightly less likely to find ineligible households that were

[10]In a few cases, baseline surveys were conducted before the distribution of transfers but after GD had held meetings in the village informing households that it would be a treatment village.

[11]Of this total, 6,507 households were on the initial sampling list, and 1,338 were randomly-selected replacement households.

[12]The order in which villages were visited by GD and the research team was randomized within subcounties. We calculate the start and end months of when GD started transfers to villages within a subcounty, and then, across these months, evenly assign both treatment and control villages experimental start months based on the random ordering.

[13]This includes 7,016 initially sampled and 1,223 replacement households. Of the initially sampled households, 1,014 had been missed at baseline. The main analysis focuses on the "initially sampled" (which includes those missed at baseline) and "replacement" households; results are similar using only originally sampled households (available upon request).

initially surveyed at baseline in high saturation sublocations (see Appendix B.6 for more information). In addition to the baseline modules, endline surveys collected more detailed data on household expenditures and crop production, additional psychological scales (in particular, related to stress and hope), and female respondents surveyed by a female enumerator were also administered a module on female empowerment and gender-based violence.

Empirics: recipient households

If the general equilibrium effects of transfers were fully contained within administrative units (here, villages and sublocations), then an appropriate specification would be

$$y_{ivs} = \alpha_1 Treat_v + \alpha_2 HighSat_s + \delta_1 y_{ivs,t=0} + \delta_2 M_{ivs} + \varepsilon_{ivs}, \tag{2.1}$$

where y_{ivs} is an outcome of interest for household i in village v in sublocation s.[14] $Treat_v$ is an indicator for residing in a treatment village at baseline, and $HighSat_s$ an indicator for being in a high-saturation sublocation. Here α_1 captures the total average treatment effect for households in treatment versus control villages, including both the direct effect of treatment (for eligible households) and any within-village spillovers; note that our design does not allow us to identify these separately. α_2 is a relatively coarse way to assess cross-village spillovers, as it does not utilize all experimental variation. We include the baseline value of the outcome variable ($y_{ivs,t=0}$), when available, to improve statistical precision.[15] We cluster standard errors at the village level, and weight observations by inverse sampling probabilities to be representative of the population of eligible households.

Overall, we view Equation 2.1 as a useful benchmark but unlikely to capture well the spatial variation in treatment intensity evident in Figure B.1.2. This is because in our study area villages are relatively close to each other; sublocation boundaries are not "hard" in any sense nor reflective of salient ethnic or social divides; and because our data indicate that there is extensive economic interaction in nearby markets regardless of sublocation. To better capture spillovers, we therefore estimate models in which a household's outcomes depend on the amount of money distributed in its own and other geographically proximate villages:

$$y_{iv} = \alpha + \beta Amt_v + \sum_{r=2}^{R} \beta_r Amt_{v,r}^{\neg v} + \delta_1 y_{iv,t=0} + \delta_2 M_{iv} + \varepsilon_{iv}. \tag{2.2}$$

The novel terms here are the amount Amt_v of cash per capita transferred to one's own village v over the entire study, and the amount $Amt_{v,r}^{\neg v}$ of cash per capita transferred to villages other than v in a series of bands with inner radius $r - 2$ km and outer radius r km around the village centroid. We normalize both to be measured as a share of per capita

[14] When we examine individual-level outcomes using Equation (2.1), we define treatment status and eligibility on the basis of the household in which the individual lives.

[15] For observations where the baseline value is missing, we include an indicator variable equal to one denoting a missing value (M_{ivs}), and set the baseline value of the outcome variable equal to its mean.

GDP.[16] The Amt variables depend on both the random assignment of villages to treatment and also on the endogenous share of households in those villages eligible for transfers, so we instrument for them using the own-village treatment indicator $Treat_v$ and the share $s^{e,t}_{\neg v,r}$ of eligible households in each band assigned to treatment. For brevity, we do not report IV first-stage results; however, the minimum first-stage Sanderson-Windmeijer F-statistic is 107.9 across all the cross-sectional specifications (Tables 2.1-2.3) and 88.4 in the multiplier specifications below, minimizing concerns about weak instruments. To account for spatial correlation, we calculate standard errors using a uniform kernel up to 10 km (Conley 2008).[17]

Because we had no a priori knowledge of the relevant distances over which general equilibrium effects might operate, we pre-specified an approach in which we first estimate a series of nested models varying the outer limit R of the spatial bands from 2 km to 20 km in steps of 2 km, and then select the one which minimized the Schwarz Bayesian Information Criterion (BIC). We report estimates of Equation (2.2) using the selected outer limit $\bar{R}$. As it turns out, this algorithm selects only the innermost 0–2 km band for almost all outcomes.[18]

Equation (2.2) correctly identifies the overall effects of the intervention if transfers delivered outside the radius $\bar{R}$ have no effect on i. If not – if, for example, all households were affected to some extent by all transfers in the study area – then the estimated effects are *relative* to these "ambient" effects. The BIC selection procedure determines how reasonable this identifying assumption is by omitting ring $R+2$ from the model if it has little explanatory power for the outcome. In principle, this could be either because variation in treatment intensity at that distance has a precisely estimated but small effect on the outcome (in which case it is a good "control"), or because there simply is not much variation in treatment intensity at that distance (in which case we cannot be sure). Given this, it is important to note that our design generates substantial variation in treatment intensity even at larger distances. Transfers in the 2 km buffer, which the BIC always includes, range from 0 to 25% of GDP, with a 10-90 percentile range of [4%, 14%]. Even in the 4 to 6 km buffer, which the BIC *never* selects for any primary outcome, the 10-90 percentile range remains wide, at [3%, 10%]. This suggests that a subset of our villages can reasonably serve as "pure controls."

[16]We use an expenditure-based measure of GDP that is described in Section 2.5, which we convert to a per-capita measure based on household census data from our study area, and augmented with data from the GiveDirectly census and the 2009 Kenya National Population Census when necessary. Per capita GDP in low saturation control villages is 637 USD PPP (2727 USD PPP per household); see Appendix B.5.

[17]We also conduct Fisher randomization tests for all specifications, where we re-randomize cash transfers across sublocations and villages as well as their roll-out over time as in our experiment and test against the sharp null that effects are zero. Conclusions are robust to this alternative method of inference.

[18]Note that this model selection step introduces some circularity, as we first determine the distances at which effects occur, and then estimate effects at those distances. We check that inference is robust to this model selection, and to alternative approaches more generally. First, we calculate exact p-values using a Fisher permutation test (i.e. randomization inference, see Appendix B.9). Second, we conduct repeated 50-50 splits of the data into training and test sets, using the training data from each split to perform the BIC step and the test set to estimate parameters, and record the proportion of times that the resulting estimates lie within the 95% confidence intervals we report here (Appendix B.9). Third, we estimate effects holding the spatial radius fixed at 2 km, 4 km, and 6 km, respectively, thus eliminating the model selection step (Appendix B.9). Taken together, these results provide reassurance that our methods yield valid inferences.

We also examine in Appendix B.9 how sensitive our main conclusions are to fixing larger maximal radii $\bar{R}$ than the BIC selects, which implies a more conservative definition of the "control group." While there is some variation from outcome to outcome, overall the effects are quite stable as we add the 2-4 km band and fairly stable as we add the 4-6 km band, though as expected standard errors often become much larger. We generally cannot reject that these estimates are different statistically from those estimated using the BIC-optimal bandwidth, giving us greater confidence in the latter. Finally, note that we typically estimate spillovers of the same 'sign' as the direct effects, which suggests that any remaining bias in our estimates likely leads us to understate, rather than overstate, overall effects. All told, we view the problem of estimating spatial effects as unlikely to admit a perfect solution, but believe that our study design and econometric specification allow us to advance meaningfully relative to most existing work.

We estimate Equation (2.2) for all *eligible* households and then use it to obtain estimates of the total effect on *recipient* households, which we report as "Recipient Households" in tables. By "total effect" we mean how the households' outcomes differ from what they would have been in the absence of the intervention. We calculate these by multiplying the estimated coefficients from Equation (2.2) by the average values of the regressors, i.e., $\hat{\beta} \cdot (Amt_v | i \text{ is an eligible recipient}) + \sum_{r=2}^{\bar{R}} \hat{\beta}_r \cdot \left(Amt_{v,r}^{\neg v} | i \text{ is an eligible recipient}\right)$ for all radii bands up to the selected $\bar{R}$. This effect allows for across-village spillovers in addition to direct effects and within-village spillovers.[19] As a benchmark, we also report estimates of α_1 from Equation (2.1), which is the total treatment effect if all spillovers are contained within villages (a common identifying assumption).

Empirics: non-recipient households

We use an analogous approach to estimate total effects on non-recipient households, which include both eligible households in control villages and ineligible households in all villages. Specifically, we estimate

$$y_{iv} = \alpha + \sum_{r=2}^{\bar{R}} \beta_r^1 Amt_{v,r} + \sum_{r=2}^{\bar{R}} \beta_r^2 (Amt_{v,r} \cdot Elig_{iv}) + \gamma Elig_{iv} + y_{iv,t=0} \cdot \delta + \varepsilon_{iv}. \tag{2.3}$$

This specification modifies Equation (2.2) as follows. First, because non-recipient households do not experience direct effects, we no longer separate own-village effects and across-village spillovers: we drop Amt_v and replace $Amt_{v,r}^{\neg v}$ with $Amt_{v,r}$, so that spillovers work entirely through β_r^1 and β_r^2. Second, we include an indicator for eligibility status and its interaction term with amounts to allow for spillovers to differ by eligibility status (recall

[19] Appendix B.6 provides an example of this for outcomes in Table 2.1. We also consider the possibility that effects are non-linear in the per-capita amounts transferred. Figure B.2.1 presents non-linear estimates of equation 2.2 for two key outcomes, total consumption and firm revenue. The relationships appear roughly linear, and we cannot formally reject linearity. We conduct the same test for our 10 pre-specified primary outcomes and eligible / ineligible households separately, and cannot reject linearity at the 10% significance level for any of them.

that eligible households in control villages are non-recipients). As above, we instrument for $Amt_{v,r}$ using the share of eligible households assigned to treatment within the corresponding band, i.e., $s^{e,t}_{v,r}$ and $s^{e,t}_{v,r} \cdot Elig_{iv}$ for each radii band v. When available, we include the baseline value of the outcome variable. We report the average total effect on non-recipients as a population-weighted average of effects for the two groups.[20]

Enterprise data

We employ several complementary sources of data on enterprises. First, we use detailed agricultural and self-employment modules from the household surveys. The agriculture module covers crop-by-crop agricultural production, sales, employment, and input costs; the self-employment module covers revenues, profits, hours worked, and some costs and investments for enterprises run by household members. These data are representative of enterprises owned locally (i.e., by residents of the study area) and allow us to clearly attribute profits to their residual claimants. They do not capture enterprises owned by people living outside the study area, which we capture separately through the enterprise census and surveys.

Specifically, we conducted censuses and surveyed a representative subset of all non-farm enterprises at baseline and endline (see Appendix B.7 for details). The endline census was conducted between November 2016 and April 2017, covering both enterprises identified at baseline and newly established enterprises. This served as the endline survey sampling frame; we randomly sampled up to 5 enterprises per village, stratified by those operating from within and outside of homesteads. Surveys covered revenue, profits, employees, wages, some other costs, and taxes paid. The main endline sample includes 1,673 enterprises operated from within and 1,440 from outside the homestead (both from enterprise surveys), as well as 7,899 agricultural enterprises from the household survey. Enterprise characteristics appear balanced across treatment and control villages at baseline (Table B.7.6).

This integrated approach to household and enterprise surveying allows us to match firms to their owners. We match all agricultural enterprises (as found via household surveys), and 61% of non-agricultural enterprises, for a total of 94% of all enterprises. Based on this match, we estimate that enterprise activity is highly localized, with 92% of total profits and 87% of revenues accruing to owners who live within the village in which the enterprise operates.

Empirics: enterprises

We estimate enterprise-level effects using versions of Equations (2.1) and (2.2), with radii bands selected as above, interacting right hand side variables with enterprise type (Appendix B.7 lists the full specifications). We include village-level means rather than enterprise-level values of the lagged dependent variable given that the enterprise surveys were re-

[20] This is calculated as $s^{e,c}\left(\sum_{r=2}^{\bar{R}}\left(\hat{\beta}^1_r+\hat{\beta}^2_r\right)*(Amt_{v,r}|i \text{ is an eligible non-recipient})\right) + s^i\left(\sum_{r=2}^{\bar{R}}\hat{\beta}^1_r*(Amt_{v,r}|i \text{ is ineligible})\right)$, where $s^{e,c} = 1 - s^i$ is the population share of eligible non-recipient village households among all non-recipient households, and the $\hat{\beta}^1_r$ and $\hat{\beta}^2_r$ terms come from Equation (2.3).

peated cross-sections. We carry out estimation using inverse probability weighting, accounting for enterprise type, except in some cases where we also revenue-weight outcomes. As above, we calculate and report average total effects, weighting effects for the three enterprise types, namely, agricultural enterprises, non-farm enterprises operating within homesteads, and non-farm enterprises operating outside the home; we typically pool data across all enterprise types, except when we do not observe some outcomes for agricultural enterprises. To facilitate comparisons between the household and enterprise results, we also normalize effects as per-household rather than per-enterprise.[21] To examine extensive margin effects, we estimate village-level analogues to this approach with the total number of enterprises censused (per household) as the dependent variable.

Price data

We measure consumer goods prices using monthly surveys of commodity prices in local markets. These surveys were conducted over the course of 2 to 2.5 years in all 61 markets in the study area (and neighboring towns) with at least a weekly market day, for a total of 1,586 market-by-month observations and 321,628 non-missing price observations. We have market price data prior to the disbursement of any local cash transfers for all markets, providing an appropriate baseline for the panel data econometric analysis detailed below, and allowing for the inclusion of market fixed effects. These include market centers located in towns, and so will appropriately reflect the impacts of households (potentially) traveling to towns to spend their transfers. Figure B.1.2 shows the substantial variation in treatment intensity around markets, as well as the heterogeneity in village proximity to markets. The average village had 0.7 markets located within 2 km and 2.3 markets within 4 km, again indicating the rather high density of settlement. Household respondents report an average commuting time to their preferred market of 31 minutes, where over 80% walk to the market.[22]

Market surveys collected prices for 70 relatively homogeneous products, including food (grains, vegetables, fruit, meat), livestock (goats, sheep), hardware (nails, paint), "duka" kiosk store products (non-food and packaged food), and others (e.g., fuel, health items, household items, and farming implements). We collected quotes from three vendors of each product in each market in each month, and use the median for each product-market-month. We then calculate linear log-price indices by weighting prices by household expenditure shares.[23] We also examine effects on subcategories of goods, which include: food items; non-

[21]Specifically, we calculate $\frac{1}{n_{hh}} \sum_g \widehat{\Delta y}_e^g * n_{ent}^g$, where n_{hh} is the total number of households across all control villages (column 2) or treated villages (column 3), $\widehat{\Delta y}_e^g$ is the estimated average effect ($\hat{\beta} * \bar{X}$) for enterprise type g, and n_{ent}^g is the number of enterprises of type g in the control or treated villages.

[22]Enterprises in markets account for 65% of non-agricultural enterprise revenue, based on a 2019 census of enterprises. We did not collect price data as comprehensively from the minority of enterprises located outside of markets and dispersed within villages, both for logistical reasons and because their products tend to be less standardized. That said, estimated impacts on the prices of two common services these enterprises offer, tailoring and maize grinding, are if anything smaller than estimated effects on our main market price index (Table B.8.11).

[23]We use expenditure data from the the Kenya Life Panel Survey (Baird et al. 2016) conducted in 2013-2014 in rural areas of Siaya and neighboring Busia county. We use the KLPS data because we did not collect

food non-durables (such as soap, cooking fat, and firewood); durables (such as iron sheets and jerry cans used for transporting water or fuel); livestock; and temptation goods.[24]

We measure prices of the major factors of production using household survey data on wages, land prices, and interest rates on formal and informal borrowing and lending. Because compositional changes in these inputs may be important, we examine quantity and price effects side by side.

Empirics: prices

We estimate effects on consumer goods prices using both spatial and temporal variation in the amount of cash distributed around each market. In contrast to the household and enterprise data, our repeated measurement of prices, both before and after the start of cash distributions, allows us to estimate equations that include market fixed effects.These absorb any systematic price differences across markets as well as differences in the share of eligible households located around those markets, conditional on which treatment is randomly allocated so that we do not need to instrument for treatment amounts in each buffer. In Appendix B.8, we demonstrate robustness to using an IV approach analogous to that used with household and enterprise data. Specifically, we estimate

$$p_{mt} = \sum_{r=2}^{\bar{R}} \sum_{l=0}^{M} \beta_{rl} Amt_{m(t-l),r} + \alpha_m + \lambda_t + \varepsilon_{mt} \tag{2.4}$$

where p_{mt} is a price outcome for market m in month t. $Amt_{m(t-l),r}$ is the per-household amount transferred within band $r-2$ to r km around market m in month $t-l$, expressed as a fraction of GDP. We exploit our panel setup by conditioning on fixed effects for both markets (α_m) and months (λ_t). The latter account for seasonal differences and other time trends common to all markets. We again account for spatial correlation in calculating standard errors (Conley 2008). We determine both the relevant spatial distance R and the relevant temporal lag M over which price effects persist by minimizing an information criterion conceptually similar to that above, but adapted to account for the fact that the BIC cannot select between non-nested models (such as one with a high R and another with a high M). Specifically, and as pre-specified (Appendix B.10), we first select R while holding M fixed at 3 months by estimating models of the form

$$p_{mt} = \sum_{r=2}^{R} \beta_r \left(Amt_{mt,r} + Amt_{m(t-1),r} + Amt_{m(t-2),r} \right) + \alpha_m + \lambda_t + \varepsilon_{mt} \tag{2.5}$$

where R varies between 2 km and 20 km. We select the value $R = \bar{R}$ that minimizes the Schwarz BIC while imposing weak monotonicity. We then select the number of monthly lags

a full expenditure module at baseline (due to project time and budget constraints) and prefer not to use endline expenditure data which are potentially endogenous. That said, results are nearly unchanged if we use consumption expenditure shares from non-recipient households in our endline survey.

[24]The consumption expenditure measure of temptation goods includes alcohol, tobacco, and gambling. The price index includes the cost of cigarettes.

M by estimating Equation (2.4) with $R = \bar{R}$ and choosing the model that minimizes the Schwarz BIC. This procedure selects only the 0-2 km band (and sometimes the 2-4 km band) around each market and a single temporal effect, implying that we only include contemporaneous transfers in estimating price effects. Appendix B.8 presents results for a specification where we impose $R = 4$ km and $M = 18$ months for robustness, and yields similar results.

Identification in Equation (2.4) comes from the roll-out of treatment across space *and* time, and the project's research design leads to substantial variation in both dimensions. As noted above, there is considerable variation in the total amounts of cash going to the 0–2 km ring around each market. Moreover, the multi-year nature of the market data covers periods both of intensive transfer distribution as well as times when no transfers were going out. As above, we are unable to capture price increases that radiate throughout the whole study area (compared to neighboring counties) over the entire period, but the highly localized nature of the price effects that we do detect suggests that any such effects are unlikely to be large.

We use estimates of Equation (2.4) to calculate two price effects. The implied *average treatment effect* (ATE) is the average price effect across all markets and all months in which any transfers went out to any market in the study area, i.e., during the study period of September 2014 to March 2017. This is simply equal to the estimated coefficients multiplied by the mean of the corresponding regressors of interest. The *average maximum transfer effect* is the average across markets of the estimated effect in the month in which the maximum amount of cash (as a share of GDP) was distributed into the selected radii bands (in other words, out to $\bar{R}$) from the market.

We focus on two sources of heterogeneous price effects. First, we classify goods into those that are more and less tradable, where the former include relatively easily transported, non-perishable items, and the latter include more difficult to transport or perishable items.[25] Second, we classify markets into those with better or worse *market access*. Standard theory in international trade predicts that more integrated markets should be less likely to experience meaningful price changes following a local aggregate demand shock. We examine output price heterogeneity with respect to a commonly used metric of market access[26] (MA_m)

$$MA_m = \sum_d \tau_{md}^{-\theta} N_d \approx \sum_{r=1}^{10} r^{-\theta} N_r \tag{2.6}$$

Geographic distance r is used to proxy for trade costs between origin market m and destination d, i.e., $\tau_{md} = r$. Destinations are 1 km radii bands around each market, with total population N_r in each buffer, and we follow Donaldson and Hornbeck

[25]For instance, more tradable goods include building materials (e.g., timber, cement, nails, iron sheets) and some household goods (soap, firewood, charcoal, batteries, washing powder), while less tradable goods include some food items (e.g., avocado, banana, cabbage, egg, pork, fish) and livestock. These classifications were undertaken based on feedback from Kenyan project staff, but there may, of course, be some ambiguity about specific items. The full pre-specified classification is in Appendix B.8.

[26]Absent data on trade costs, we have to make an assumption about the elasticity of trade costs with respect to distance. We set the elasticity equal to 1 and conclusions are robust to alternative assumptions.

(2016) in setting $\theta = 8$. Within quantiles of this metric, we calculate average and average maximum transfer treatment effects in the manner described above.[27]

We estimate effects on input prices using Equations (2.1), (2.2), and (2.3), as our input price data come from household surveys, and report the corresponding average treatment effects.

Empirics: dynamics

To estimate the multiplier, we extend the cross-sectional analysis by estimating and then integrating effects on components of GDP over time.[28] For a flow variable x (e.g., consumption, investment, etc.), we first estimate the following specification, which is a dynamic extension of previous estimating equations:

$$x_{it,v} = \alpha_t + \sum_{s=0}^{9} \beta_s Amt_{v(t-s)} + \sum_{s=0}^{9} \gamma_s Amt^{\neg v}_{v(t-s),\text{0-2km}} + \varepsilon_{it,v}. \tag{2.7}$$

where $Amt_{v(t-s)}$ is the amount transferred to village v in quarter $t-s$, instrumented by a treatment indicator $Treat_v$ multiplied by the share of total transfers going to village v in quarter $t-s$, and analogously $Amt^{\neg v}_{v(t-s),\text{0-2km}}$ by the share $s^{e,t,\neg v}_{v,\text{0-2km}}$ of eligible households in the 0 to 2 km buffer around v (but not in village v) assigned to treatment multiplied by the share of transfers going to that group in quarter $t-s$. The coefficients in this model are identified by the fact that village treatment status was randomized, and the timing of both cash transfers and survey data collection was rolled out to villages in a randomized order. The main challenge is that the first household surveys started around 9 months after the experimental start date in each village, while enterprise surveys began after about 18 months (see Figure B.1.1). With a few exceptions, recall periods are less than or equal to one month, so we often do not directly observe the initial response in flow variables in the months immediately after the first transfers went out, which is when we might expect to see some of the largest impacts on expenditure. Given that the specification treats each dollar transferred symmetrically, we can still estimate the local response during these early quarters because transfers to recipients rolled out over 8 months. Similarly, we estimate neighborhood effects using the substantial variation in the timing with which nearby villages were treated. However, we tend to obtain less precise estimates of responses in early quarters as they are estimated using less variation in treatment intensity compared to later quarters.

We then integrate dynamic effects on flow variables over time up to 29 months (10 quarters) after treatment. We compute the dynamic profile of treatment effects (or the impulse response function, IRF) using the coefficients estimated above and assuming that the treatment rolled out to recipient households as planned: the timing is a token transfer at time 0,

[27] We also consider an alternative market access metric, namely, road access, defined as the inverse distance from the closest main road (as captured by Open Street Map), see Appendix B.8.

[28] We stated our intention to estimate a multiplier in our pre-analysis plans but did not fully specify the econometric approach for doing so.

a first lump-sum 2 months later, and a second 8 months after the token transfer. We compute this IRF separately for recipient and non-recipient households, and separately for the three categories of enterprise in both treatment and control villages. We then aggregate the quarterly estimates across all villages (using inverse population weights from our household and enterprise census) to compute the study area-wide IRF for each flow component.

We conduct inference on the multiplier estimates this procedure yields using the wild bootstrap clustered by sublocation, the highest unit of randomization (Cameron, Gelbach, and Miller 2008).[29] We focus on two one-sided hypotheses, namely, that the multiplier is less than zero and that it is less than one. We test these hypotheses using our income- and consumption-based multiplier estimates separately, as well as using the average of the two.

2.4 Tracing out the path of spending

We now turn to tracing out the path of spending induced by the cash transfer experimental intervention. We start by documenting effects for recipient households, then for enterprises and untreated households. Monetary units are USD PPP unless otherwise defined (where the transfer was worth USD 1,871 PPP), flow outcomes are annualized, and monetary outcomes are top-coded at the 99th percentile (as pre-specified), unless otherwise noted.[30]

Recipient household effects

The main household expenditure measure is the (annualized) sum of total food consumption in last 7 days, frequent purchases in the last month, and infrequent purchases over the last 12 months.[31] Durables expenditures are the sum of home maintenance, home improvement, and other household durables spending, and the remainder classified as non-durable spending.

As expected, recipient households report significantly higher total expenditure: USD PPP 294 more expenditure than eligible households in control villages (Table 2.1, column 1), an 11.6% increase over the control village in low saturation area mean of USD PPP 2,536. The estimated total treatment effect, including spatial effects, is larger at USD PPP 339, a 13.4% increase (column 2). This pattern between columns 1 and 2 is a first piece of evidence for localized, positive cross-village spillovers, which is repeated across other outcomes.

[29]While this procedure may perform poorly in cases where most units in treated clusters are treated and there are few clusters, here at most two thirds of households in the most intensely treated clusters were treated, and there are 84 clusters, far above the 15-20 that MacKinnon and Webb (2018) deem adequate.

[30]The main measures were pre-specified, though some groupings vary from the PAP to ease exposition.

[31]The survey was quite comprehensive. In addition to food consumption, frequent purchases include airtime and other phone expenses; internet; transport expenses (including petrol); lottery tickets and gambling; clothing and shoes; recreation and entertainment; personal toiletry and grooming; household items, such as cleaning products and candles; firewood, charcoal and kerosene; electricity; and water. Infrequent purchases include house rent/mortgage; home maintenance; home improvements; religious expenses; education expenses; charitable donations; weddings and funerals; medical expenses; household durables, including furniture, lamps, cutlery, pots and pans and other kitchen equipment; and dowry or bride price.

The pattern of expenditure effects by category is broadly consistent with earlier work (Haushofer and Shapiro 2016). Both non-durable and durable spending increase substantially. Food expenditure accounts for a sizable portion of the increase in non-durable expenditure in both columns (42% and 59%, respectively). We can reject meaningful increases in reported spending on temptation goods, consistent with Evans and Popova (2017).[32]

Consistent with increased expenditure on durables, asset stocks also increase (Table 2.1, Panel B). Anecdotally, many recipients withdrew money from M-Pesa immediately and saved via durable assets. The main pre-specified measure of assets includes livestock; transportation (bicycles, motorcycles, and cars); electronics; farm tools; furniture; and other home goods; we add in net household lending to, and borrowing from, both formal and informal sources. This measure of assets increases by USD PPP 183, or 26% of the mean for eligible households in control villages in low saturation sublocations.[33] This measure excludes the values of housing and land, which are harder to measure given thin local markets, but also likely important given existing work shows that households often use GD transfers to spend on housing materials (Haushofer and Shapiro 2016). We separately measure housing value as the respondent's self-reported cost to build a home like theirs, and land value as landholdings multiplied by the household's report of the per-acre cost of land of similar quality (in their village). Estimated housing value increases by USD PPP 477, or 79% of the control mean, and estimated land value increases, though this effect is not statistically significant.

Theoretically, the effect of a large-scale wealth transfer on earnings is ambiguous: it may reduce labor supply through an income effect, but may also enable productive investment or increase labor demand. In the data, recipient households' income from all sources (excluding the GD transfers) does not appear to have decreased: point estimates are positive (USD PPP 79 and 135 in the two main specifications) and the reduced form effect is marginally significant.[34] For labor supply specifically, we do not find that recipient households worked less; if anything, total hours worked by recipient households in agriculture, self-employment and employment increased slightly though not significantly (Table 2.2, Panel A, columns 1 and 2). This is consistent with the studies reviewed by Banerjee et al. (2017), which generally find that cash transfers in low and middle income countries do not reduce labor supply.

Interestingly, we observe little heterogeneity in estimated treatment effects (on assets, expenditure, income, and hours worked) among eligible households across eight pre-specified characteristics (Figure B.2.2), namely, respondent gender, age over 25, marital status, pri-

[32] While there is likely some under-reporting of temptation goods, the fact that the control group mean is non-trivial demonstrates that at least some households feel comfortable reporting such spending. Given our limited expenditure data immediately after transfer receipt, we cannot rule out that temptation good spending increased temporarily at that time.

[33] The mean for eligible households in control villages and low saturation sublocations is USD PPP 716 (with SD 849), less than the overall mean, unsurprisingly since ineligible households are wealthier.

[34] As is common in low-income settings, measured values of consumption are larger than measured household income. Similarly, total measured local area income and firm revenue is lower than expenditures, in part, because measured expenditure includes important categories – including medical and schooling expenses, utilities, rent and mortgage, religious and charitable donations, and dowry, wedding and funeral costs – for which we do not typically measure corresponding revenues in the enterprise data. Expenditure measures may also better capture consumption of own-farm production than the agricultural revenue data.

mary school completion, having a child in the household, an indicator for above median measured psychological well-being, and work status (in self-employment or wage employment).

The effect on net transfers received from other households is also notable: the point estimate is negative but not statistically significant, and we can reject large changes in either direction. This suggests that relatively little of the cash transfer was literally shared with neighbors or social contacts.

Overall, these results highlight that cash transfer recipients substantially increased their expenditure on a broad range of goods. This spending was likely financed primarily by the initial transfers themselves, with possibly some contribution from higher earnings. A large share of this spending likely takes place locally: enterprises report that 88 percent of their customers come from within the same village or sublocation. Below we therefore turn to examining impacts on local enterprises.

Estimating the Marginal Propensity to Consume

The marginal propensity to consume (MPC) sheds light on the inter-temporal decision-making of households, and is an important determinant of the magnitude of a transfer multiplier, as it captures the share of income that is spent—and thus enters the hands of other agents in the economy—rather than being allocated to financial savings or retained in cash (which in our setting might include simply retaining some value on the mobile money platform). The dataset allows us to generate an intuitive estimate of the MPC out of the transfer, obtained conceptually by dividing the total increase in expenditure by the size of the GiveDirectly transfer.[35]

Here we summarize the construction of the MPC in our data; refer to Appendix B.3 for details. An immediate cross-sectional estimate can be obtained by dividing the effect on total household non-durable consumption in Table 2.1 by the size of the transfer among recipient households. However, this underestimates the MPC, as it is based on consumption as captured in the period preceding household survey administration, with a retrospective timeframe of at most 12 months, compared to the full transfer value, which for many households was distributed at least in part more than 12 months ago. It misses any changes to consumption occurring outside this window, particularly in early months when spending may be the highest. We can improve on this by employing the dynamic regression specification (in equation (2.7)), which exploits the fact that the timing of survey data collection, relative to transfer disbursement, varied exogenously across households.

Yet this estimate is also a lower bound. As noted above, a limitation of the data collection is the relative lack of household survey data collection in the months after transfers went out, the period when, anecdotally, a large share of the transfer was spent. We thus augment the analysis by making use of data collected as part of the closely

[35]This measure is comparable to MPC estimates from tax rebates (e.g., Parker et al. 2013). Alternatively, it may be attractive to divide by the transfer size plus any additional non-transfer income generated over the period. Table 2.1 shows that this increase is small and not significant, with a point estimate of $\approx 7\%$ of the transfer value. Thus results do not change substantively if this additional recipient income is included, although its inclusion does reduce the estimated MPC somewhat, see Appendix Table B.3.1

related Haushofer and Shapiro (2016) study of GiveDirectly transfers provided between 2011–2013 in a nearby part of Siaya County (Rarieda subcounty, lying just outside our study area), which gathered information on household consumption immediately after transfers. The MPC of non-durable goods among recipients in the first three quarters following the transfer there was 0.35 (Appendix Table B.3.1), consistent with much spending occurring shortly after transfer receipt. Combined with estimates from our data thereafter, recipients' MPC on non-durables over the 27 months post-transfer is 0.64. This implies that most study households receiving cash transfers were hand-to-mouth consumers, allowing us to soundly reject the permanent income hypothesis in our context.

This estimate still leaves out durable goods expenditure. First note that households report purchasing the vast majority of durables (over 95%) in local shops. These durables may serve as consumption, savings or investment goods. A large share of such purchases in the study sample are consumer durables not primarily intended for productive uses (e.g., radios, furniture). At the same time, formal sector financial savings are limited in rural African settings like ours and much household saving comes in the form of purchases of household durable assets, which necessitates spending on local goods. Thus from the perspective of inter-temporal decision-making, durables are more of a gray area. Yet whether durables are purchased as "savings goods" or "consumption goods", both types of expenditure show up as revenue of local firms and may therefore have similar stimulus effects. Here we rely on the cross-sectional difference in the value of durable assets (including housing) between treatment and control areas among eligible households in our endline data (Table 2.1). Combining this non-durable expenditure yields our best estimate of the overall MPC in the 27 months following transfers, at 0.93.

Because we are interested in estimating the multiplier effect on local economic activity (within the study area), we next refine the MPC estimate by focusing on spending on local value added, excluding spending on intermediates and final goods produced elsewhere. Spending on goods produced in other parts of Kenya (or the world) does not directly contribute to local GDP (although it could generate multiplier effects at larger geographic scales that we cannot readily assess with our data). We thus derive a bound on the share of spending on local value added. This is closely related to the local degree of openness that features prominently in discussions such as Farhi and Werning (2016). We find that most consumption is in fact of locally produced goods, in line with the well-known fact that a large share of household consumption in rural areas consists of locally produced food and other basic necessities (Deaton 2018). In particular, the enterprise data allows us to bound the share of imported intermediate goods sold in the study area, where, recall, over 95% of household shopping occurs.[36] This conservative methodology yields an upper bound of 18% for the expenditure-weighted share of local non-durable expenditure (and 20% for durables) that

[36] As discussed in Appendix B.4, we determine that at most a fraction $1 - \frac{cost_i + profit_i}{revenue_i}$ of the revenue of firm i is spent on intermediate goods; for each firm type, we then generate a revenue-weighted average upper bound for the share of intermediates in its production function. Next, we make assumptions about what share of intermediate goods is likely imported, conservatively erring on the side of assuming a high share; for instance, we assume that all intermediate goods at clothing stores (which spend up to 38% of revenue on intermediate goods) are imported, which is likely to be an upper bound.

may reflect expenditure on imported intermediates, indicating that four fifths of spending is on local value added, and thus that the study area's economy is largely closed.[37]

Combining estimated import shares with our preferred MPC estimate yields a marginal propensity to consume on local value added, which we denote MPC_{Local}, of approximately 0.76 in this context. Of course, this figure is subject to the data and measurement caveats noted above, as well as assumptions on the share of imports, and so should be seen as speculative. Nonetheless, taking this value of 0.76 to a basic static Keynesian model, the transfer multiplier effect on local output would be $\frac{MPC_{Local}}{1-MPC_{Local}} \approx 3.2$. We dynamically estimate the multiplier using all household, enterprise and price data in section 2.5 below.

Enterprise effects

There are large increases in revenue for enterprises in both treatment and control villages (Table 2.3, Panel A). Revenues in treated villages increased by USD PPP 322 per household, a 65% increase, while those in control villages increased by USD PPP 237 (48%). Revenue gains are concentrated in the retail and manufacturing sectors: both treatment and control villages experience statistically significant increases in manufacturing revenue of similar magnitudes – USD PPP 93 and 109, respectively – while treatment villages see larger gains in retail revenue (USD PPP 160 versus USD PPP 82, Appendix Table B.2.2).

Estimated effects on profits are positive, but moderate in magnitude and not significantly different from zero. In fact, profit margins (measured as the ratio of profit to revenues) fell (Table 2.3, Panel A, Row 5). We also see no evidence of firm entry, as one might have expected if enterprises were becoming more profitable (Panel C). Overall, the data indicate that higher revenues were largely absorbed by increased payments to various factors of production. While we do not observe all of these payments, we do see significant increases in the factors that we directly measure, and particularly the wage bill: enterprises in treated (control) villages increase spending on labor by USD PPP 76 (67), a sizable change relative to the mean.

Strikingly, we do *not* see strong evidence of a firm investment response. Estimated increases in fixed capital investment are small, and we can reject large changes (Panel B, Row 2). We do see a modest increase of USD PPP 35 in inventories for enterprises located in treated villages, yet even this appears to be less than proportional to the increase in firm sales; in other words, these enterprises are, if anything, operating leaner business models (Panel B, Row 1). This pattern of results suggests that the expansion in enterprise activity is driven more by the shock to local aggregate demand than by a relaxation of credit constraints that had previously limited investment.

One caveat to this point is that some household assets are difficult to categorize into "productive" assets as opposed to consumer durables. For example, bicycles may be used for personal transportation (i.e., to visit friends), but could also be used as a bicycle taxi

[37] In principle this exercise also depends on migration: money spent elsewhere by migrants appears in our data (as we tracked and surveyed them) but does not contribute towards the share spent on local value added. In practice household migration was uncommon, with 5% of control low-saturation household migrating, and unaffected by treatment (Table B.2.9, Row 1). Estimated treatment effects among non-migrants are also essentially identical to overall average effects (Table B.2.9, Panel 2).

to generate income. We therefore inclusively categorize as "potentially productive" both livestock as well as a number of non-agricultural assets that could potentially be used for income-generating activities (beyond simply renting out the asset).[38] When we do so, overall roughly half of the increase in household asset ownership documented above is in what we believe to be purely non-productive assets, with small gains in productive agricultural assets (e.g., farm tools) and a modest gain for potentially productive assets (Table B.2.1). We also fail to detect any investment response for non-agricultural enterprises owned by recipient households: neither investment nor inventories increase relative to eligible owners in control villages (Table B.2.3, Panel B). Taken together, these patterns are also consistent with the cash transfer program generating only a limited local investment response.

Non-recipient household effects

There are positive and significant expenditure effects for non-recipient households. Column 3 of Table 2.1, Panel A presents results based on Equation (2.3). Notably, the magnitude of these gains (USD PPP 335, p-value < 0.01) are quite similar to those of recipient households (USD PPP 339). The pattern of expenditure increases is also broadly similar to that for recipient households, except that spending on durables does not increase among non-recipient households. One possible reason for the similarity in overall spending impacts is that the timing of effects on recipient and non-recipient households may be different, with recipient households showing impacts earlier than non-recipient households, but effects converging by roughly one year after the final transfer was received. A further potential mechanism is that labor earnings increase differentially: among non-recipients, annual labor income increases by USD 225, while the figure is USD 136 for recipients. For wage earnings, the figures are USD 183 and USD 74, respectively. Thus, the similar impacts on expenditure among recipient and non-recipient households may partly be explained by a lower labor income response among the former. Finally, note that non-recipient households include both eligibles and ineligibles, and, as shown in Table B.2.8, most of the gains accrue to ineligibles. These comparatively wealthier households might be gaining more from business and additional labor income, and may be imperfectly substitutable with eligibles in the labor market. As a result, they may experience a larger increase in wages than recipient and non-recipient eligibles.

How did non-recipients fund these consumption gains? One possibility is that they are dis-saving, perhaps due to social pressure to "keep up with the Joneses", their neighbors who received the transfer. However, this does not appear to be the case: estimated treatment estimates for total assets, housing and land values are all positive, although not all are significant (Table 2.1, Panel B). Nor do we observe a borrowing response for non-recipient households from either formal and informal sources (Table 2.2, Panel C, column 3). A second potential explanation is that expenditure gains reflect inter-household transfers to non-recipient households, as documented in Angelucci and De Giorgi (2009) for Mex-

[38]Potentially productive non-agricultural assets include bicycles, motorcycles, cars, boats, kerosene stoves, sewing machines, electric irons, computers, mobile phones, car batteries, solar panels or systems, and generators. Examples of residual non-productive assets include radio/CD players, kerosene lanterns, beds, mattresses, bednets, tables, sofas, chairs, cupboards, clocks, televisions, and iron sheets.

ico. This also does not seem to be the case, as we find no significant increase in net transfers received by non-recipient households, and the point estimate of USD PPP 8.85 is less than 3 percent of the expenditure gain for non-recipient households; this mirrors the lack of an effect on net transfers among recipient households noted earlier.

Rather, the data suggest that consumption gains are driven by higher earned income: total annualized income increases by USD PPP 225. It is often argued in development economics that survey estimates of consumption are better measured and often substantially larger than estimates of income, particularly for poor households (Deaton 2018). While this is true in our case, we cannot reject that the total effect on income is the same as the effect on consumption expenditure for non-recipient households ($p = 0.23$). Income gains come largely from wage earnings, which increase by USD PPP 183, with a smaller and not significant contribution from profits from owned enterprises. These results are broadly in line with the enterprise results, in which profit increases were modest and marginally significant while the wage bill expanded significantly, by 76 and 67% in treatment and control villages, respectively (Table 2.3, row 4). Higher wage earnings appear more likely to reflect higher wages than increased labor supply, as the point estimate for overall household labor supply is actually somewhat negative (although there does appear to be an increase in respondent hours worked for wages, Table B.2.4). Hourly wages earned by non-recipient household increase meaningfully, although the estimate is only marginally significant (Table 2.2, Panel A).

To sum up the results so far, cash transfer recipient households receive and spend most of the transfer, leading to higher local enterprise revenues. This positive aggregate demand shock, in turn, appears to increase the income of local non-recipient households, leading to higher spending on their part. This pattern provides initial evidence for a positive multiplier effect of the cash transfer program, an issue we return to below.

Effects on output prices

We turn next to effects on consumer goods prices in order to understand the extent to which other monetary impacts are real as opposed to nominal. Overall, we find small, positive and precisely estimated effects on consumer goods prices. For our overall expenditure-weighed log-index of market prices both the ATE and average maximum transfer effect are small and precisely estimated near zero (Table 2.4). The tight standard errors allow us to rule out even relatively small price effects: with 95 percent confidence, the ATE across the study period is below 0.0022 log points, or 0.22 percent. For the average maximum transfer effect across markets, the upper bound of the 95 percent confidence interval is 0.01 log points, or 1 percent. Price effects are also small across almost all product categories. In particular, food prices are in line with the overall price index, and durable prices do not increase meaningfully. To help mitigate concerns that results may be sensitive to the price index weights or product classification, we find that average price inflation is below 1.2% for every product (Figure B.2.4; for alternative specifications and product classifications, see Appendix B.8).

Variation in price responses is generally in line with theoretical predictions. We observe somewhat larger price increases in markets less integrated into the local economy. Columns 3 and 4 split markets into those above and below median market access, with estimated effects typically more positive in more remote markets. Figure B.2.3 further breaks this pattern down by quartile of market access, with lower values reflecting more isolated markets. Panels A-C show a small amount of inflation for less tradable goods only in the most isolated markets, and smaller and less precisely estimated effects for more tradable goods, with less of a clear pattern across market access quartiles. Inflation for less tradable goods in isolated markets nonetheless remains limited, at 0.2-0.3% on average. We also carried out enterprise phone surveys of a subset of enterprise types during the period in which transfers were going out, which collected price data on a limited number of products; inflation for these local manufacturing and services prices is also limited (see Appendix B.8).

These patterns are qualitatively similar to findings in Cunha, De Giorgi, and Jayachandran (2018), who study the price effects of an in-kind food and cash transfer program in Mexico (where the household income shock was similar in magnitude to the Kenya program we study): in-kind transfers there lead to price decreases, while cash transfers lead to price increases, but their estimated effects are small except in remote villages. Filmer et al. (2018) estimate inflation of 5 to 7% for protein-rich foods in the Philippines, with smaller effect for other product categories. Burke, Bergquist, and Miguel (2019) show that a credit intervention impacting the supply of staples also affects local grain market prices in a different Kenyan region. Reconciling these results with ours is a task for future research.

Effects on input prices

We next examine effects on the prices of major factors of production: labor, land and capital. Table 2.2 presents estimated effects on these prices measured in the household survey data. We find some evidence of higher wages. In row 1 of Table 2.2, we examine wages for employees using household survey data.[39] In the reduced form specification, eligible households in treatment villages earn USD PPP 0.1 more per hour, on a base of USD PPP 0.70. This effect is no longer significant, however, when we also estimate across-village spillovers. For non-recipient households, the increase is even more marked at 0.19 USD PPP per hour, and significant at the 10% level. These potentially large wage effects do not seem to be driven by large labor supply responses. In row 2, we calculate the total hours worked by adult household members in agriculture, self-employment and employment, and estimate effects at the household level. Effects are relatively small and not significant. Together with the fact that enterprise wage bills increased, these patterns are strongly suggestive of positive local wage effects (Table 2.3). This in turn suggests that labor markets in this area are fairly localized, at least over the time horizon we study, which is consistent with the fact that we see little evidence of impacts on measures of migration (Table B.2.9). In the longer run, labor may become more mobile, helping to equilibrate any induced wage differentials.

Effects on estimated land prices are positive and economically meaningful (at 9-14%), but not significant (Table 2.2, Panel B). Since our measure of land prices is a somewhat noisy one—formal sales are rare so we use respondents' self-reports of the amount per acre land like

[39] We include all household members that report working for wages, and calculate their hourly wage based on hours worked in the last 7 days and their monthly salary (adjusted to weekly scale).

theirs in the same village would sell for—we also examine land rental prices as a robustness check, which yields data on actual land transactions for a subset of respondents. We do not find significant effects on land rental prices (Table B.2.5). Unsurprisingly, given land should be in relatively fixed supply in the short-run, we find little change in total landholdings among recipient households or those in more heavily treated areas. We also find no effects on total land rentals, nor on the total amount of land used for agriculture (Table B.2.5).

We estimate fairly precise null effects on interest rates and total borrowing (Table 2.2, Panel C), where we measure household borrowing from both formal (e.g., banks, mobile credit services) and informal (moneylenders, family and friends) sources. The loan amount reports total borrowing across sources in the last 12 months, setting those who did not borrow equal to zero. Note that the loan-weighted interest rate is the monthly interest rate on the most recent loan by source, weighted by the total amount of borrowing (by source); we include informal loans without interest, which brings down the average rate.

2.5 The transfer multiplier

We next examine what the household and enterprise responses imply for the aggregate level of economic activity, and specifically for the value of the local multiplier of cash transfers, where 'local' refers to the entire study area. We define this multiplier $\mathbb{M}$ as the cumulative effect of transfers on local real GDP, relative to the total amount T transferred in real terms, over a given time interval:

$$\mathbb{M} = \frac{1}{T}\left(\int_{t=0}^{t=\bar{t}} \Delta GDP_t\right) \tag{2.8}$$

The size of the transfer multiplier is generally thought to depend in part on the policy context in which outlays are made, and in particular on the extent to which (i) monetary policy reacts, and/or (ii) households and firms expect levels of current or future taxation to change. Our setting is unusual in a useful way: because we observe a large one-time fiscal outlay that was made philanthropically, funded from outside of the economy we study, and small relative to the overall Kenyan economy, we can reasonably expect to measure a "pure" external transfer multiplier that should be independent of such effects. This feature generates estimates that can be thought of as a model primitive, and with which estimates from other financing scenarios can be contrasted.

As noted in Section 2.4, an initial calculation of the transfer multiplier as $\frac{MPC_{Local}}{1-MPC_{Local}}$ suggests that it may be substantial, at around 3.2. In this section, we refine this estimate by both accounting as fully as possible for effects on all components of GDP, including spillover effects, and accounting for dynamics. To get at real values, we deflate all monetary outcomes and transfer values to January 2015, linking the overall monthly market price index in the nearest market to each observation (Appendix B.4 presents a nominal version).

Following national accounts definitions, the expenditure-based measure of local GDP is $GDP_t = C_t + I_t + G_t + NX_t$, where C_t is consumption expenditure on non-durables and durables, measured as quarterly consumption plus accumulated assets and housing stock at endline.[40] To avoid potential double-counting, we exclude expenditure on home durables, home improvements and maintenance from the consumption expenditure measure as part of this expenditure may be reflected in an accumulation of assets. In addition, we exclude net lending as well as land values from the asset measure because changes in land values may not be driven purely by investment, and because we think of land supply as being essentially fixed. We exclude local government expenditure, G_t, as Walker (2018) shows that the intervention had a precisely estimated null effect on it.

Since we also measure household and enterprise income, we can construct a dual income-based measure of local GDP as the sum of factor payments and profits: $GDP_t = W_t + R_t + \Pi_t + Tax_t - NFI_t$, where W_t is the total household wage bill, R_t are rental expenses of local enterprises (assuming those are paid to capital owners within our study area), Π_t are enterprise profits, and Tax_t is total enterprise taxes.[41]

For flow variables, we follow Section 2.3 to generate IRFs, which we integrate over time. For two components of GDP, we are instead able to measure impacts on the integral of flows over time by simply measuring impacts on accumulated stocks, simplifying the problem. Specifically, we measure effects on durable consumption expenditure using effects on the stock of endline household durable goods and the value of respondents' home, and effects on inventory investment using effects on current inventory stocks at endline. One drawback is that these figures are likely to under-estimate cumulative spending to the extent that some assets depreciated between the time of purchase and measurement, although over the limited timeframe considered this may be a second-order concern; any such bias would tend to reduce the estimated multiplier. In the graphical presentation, we assume that any effects on these stocks occurred equally across all post-treatment quarters.

Overall, we view the expenditure- and income-based multipliers as two distinct measures of the same underlying concept, each with its own limitations. Reflecting this, below we estimate them jointly and test individual as well as joint hypotheses across the two measures. We discuss limitations and robustness in detail in Section 2.5, including adjustments to account for the fact that we do not directly observe NX_t or NFI_t.

Multiplier estimates

We estimate a sizeable multiplier using both main approaches, in line with the back-of-envelope figure derived above. The estimated expenditure multiplier is 2.53 (Table 2.5, Panel A). 46% of this effect is driven by consumption expenditures. Household asset pur-

[40]Note that by measuring impacts on asset stocks we (correctly) do not count transfers of existing assets between local agents as GDP, since these increase one agent's balance sheet while decreasing another's. Such transactions only introduce bias if they involve a non-local counterparty, as discussed below.

[41]We employ the household rather than enterprise wage bill, as the household survey sample is larger and includes individual-level wage earnings data. We omit land rental income because we do not see any significant evidence of effects on this above. In principle, a third approach to estimating GDP would be to aggregate value added from local enterprises; we do not implement this as we did not collect sufficiently comprehensive data on enterprise expenditures on intermediate inputs.

chases and enterprise investment make up another 32% and 19% respectively, and enterprise inventories are not quantitatively important. While part of the asset response could potentially reflect productive investments by household-operated enterprises, at least 43% of the asset response comes from non-productive assets, across both recipient and non-recipient households (see Table B.2.1). Taking this into account, consumption alone leads to an estimated multiplier of at least 1.5, underscoring the overall point that cash transfers appear to have led to a predominantly demand-side driven increase in local economic activity.

The estimated income-based multiplier is quite similar in magnitude to the expenditure-based multiplier, at 2.28 (Panel B), and we cannot reject that they are the same ($p = 0.88$). This is notable since it is calculated using a completely distinct set of component measures. Of this total effect, we find that 64% reflects increased enterprise profits, 30% increased wages, and a much smaller contribution comes from capital income and taxes taken together. As noted above, the increase in consumption, and the smaller increase in investment, is therefore primarily accounted for by higher profits and wages. Of course, in our context of predominantly single-person firms, "profits" likely reflect some mix of true economic profit along with returns to the owner's capital and labor inputs. Regardless of the exact mix, however, this sum should be appropriate for our goal of calculating the aggregate income-based multiplier.

When we examine the relative contributions of recipient and non-recipient households to both multipliers (as shares of the total household contribution), we find that non-recipient households account for 80% of the household contribution to the expenditure multiplier and 85% of the contribution to the income multiplier, both of which are somewhat higher than their share in the local population of 67%. This suggests that analysis focusing only on recipient households may be missing sizable shares of program effects.

An advantage of this "macro-experimental" approach to estimating the multiplier is the ability to conduct statistical inference. To start, we reject the null of a negative multiplier (with a value less than zero) at the 10% level using either approach (Figure 2.1 and Table 2.5), and reject the null at $p = 0.02$ when testing the joint restriction, and at $p = 0.04$ when testing the average of both multipliers. Since the two measures exploit distinct data, we gain statistical power by examining both measures together. Rejecting a negative multiplier on real GDP is important, ruling out, for example, that prices adjusted immediately to increased spending, netting out any real effects. Testing the null hypothesis of a multiplier less than one has been a central goal of recent research on the fiscal multiplier, since it would imply a crowd out of private spending, but this value does not have the same interpretation for transfer multiplier estimates like ours where all spending is private. Nonetheless, rejecting a transfer multiplier value less than one constitutes a conservative test for the existence of positive output spillovers for non-recipients, holding even in the extreme case of $MPC = 1$.[42] Using the expenditure or income-based approach alone, the p-value on this test is 0.14 and 0.23 respectively. The average of the

[42] A less conservative test of the no-spillovers hypothesis would be to test if the multiplier estimate is less than the MPC of recipients (which is strictly less than 1). Because this MPC is itself somewhat imprecisely estimated in our data (see section 2.4), this approach does not necessarily increase power.

two multipliers is 2.40 ($SE = 1.38$, $p = 0.15$), and the hypothesis that the multipliers are jointly less than one is rejected at the marginally significant $p = 0.07$ level.

Figure 2.1 presents these results graphically, breaking up the aggregate multipliers into quarters after transfers went out. Panel A presents the expenditure-based multiplier. The increase in GDP is fairly stable over time; in fact, we cannot reject that the expenditure response is constant across all quarters (p-value of 0.73). It increases slightly up to a peak after 9 months (when the second lump-sum transfer has been received), and then slowly declines. Interestingly, we reject a null effect as late as two years after the transfer, suggesting that the true multiplier (out to an infinite time horizon) could be larger still and that our estimates are likely to be lower bounds. The less precisely estimated effects (with larger confidence intervals) during the first three quarters afters transfers go out are visually apparent. The income multiplier, on the other hand, visually appears to fluctuate more over time (Panel B): it is marked by a strong early response in profits, while wages appear to take longer to rise. Yet as with the expenditure measure, we cannot reject equality of all quarterly coefficients (p-value of 0.76).

These estimates are somewhat larger than the higher end of recent fiscal multiplier estimates in the context of public spending in the United States (Chodorow-Reich 2019; Nakamura and Steinsson 2014), where they tend to range from 1.5 to 2.0. As noted above, the magnitudes of transfer versus fiscal spending multipliers are not directly comparable. The differences between our results and existing estimates may also reflect the relative levels of economic development and other structural differences between the Kenyan and US economies (such as the degree of openness of local economies, the share of hand-to-mouth consumers and the existence of financial savings opportunities), differences in data and measurement, as well as any effects on (or expectations of effects on) either monetary policy or future taxes in the US, the latter being response effects that this study's experimental design usefully allows us to avoid.

Alternative assumptions

We also consider several alternative multiplier estimates that treat prices, exports/imports, and the first three quarters of data post-transfer in different ways. A first alternative presents the multiplier in nominal rather than real terms: the nominal expenditure (income) multiplier is 2.66 (2.55), see Appendix Table B.4.6. Given our quantitatively small price effects, the differences between these and the real estimates presented in Table 2.5 are mainly due to the moderate degree of overall price inflation during the study period.

The expenditure- and income-based measures of GDP we generate are based on unusually rich underlying data, but each has potential limitations. In particular, each may misattribute transactions between agents located in the study area and counterparties located outside. In the expenditure case, the main concern is that we do not directly observe net exports (NX_t). Imports show up as expenditure but are not local GDP, while exports do not show up in expenditure but are part of local GDP. To the extent that cash transfers decrease (increase) net exports from the study area, our expenditure multiplier would overstate (understate) the multiplier. Intuitively, we might expect net exports to fall following a large external income transfer: since many local firms are retail establishments, imports of intermediate goods (including packaged consumer goods ready for sale) would likely increase. This suggests that the expenditure-based approach might be upwardly biased.[43] Note that transactions between agents *within* our study area are correctly accounted for: for example, if study village A imports goods from study village B then the value of these goods should be included in local GDP as they are produced within our study area. Of course, increases in net imports could in part reflect increases in economic activity outside of the study area due to the cash transfers, which our concept of the local multiplier does not capture but which are a part of the broader impact of the intervention.

As a robustness check to gauge the magnitude of the potential bias in the expenditure-based measure due to imports of intermediate inputs, we first assign each component in the non-durable and durable expenditure measures to enterprise types at which the good is most likely to be purchased (using revenue shares of different enterprise types, where appropriate). As noted in Section 2.4, this conservative methodology yields an upper bound of 20% of local spending that may reflect expenditure on imported intermediate goods. If imports scale linearly with expenditure, this suggests a transfer multiplier of at least 2.05 on local expenditure alone (see Appendix B.4).

In the income case, potential bias could arise if there are changes in net wage income (NFI_t) earned outside the study area, since this is not considered part of local GDP. This bias seems unlikely to be quantitatively important in our setting: 83% of all non-farm employees are family labor (and therefore presumably overwhelmingly local), and among individuals employed for a wage, only 6% report an employment contact address outside the study area. To the extent some bias remains, we would expect it to be negative (towards zero), if net labor income earned outside the study area decreases in response to higher local business revenue, employment and wages. This suggests that the income-based approach may yield a lower bound on the multiplier. Consistent with this bounding logic, the estimated income multiplier is somewhat smaller in magnitude than the expenditure multiplier.

Next, we examine how alternative estimates of effects over the first three quarters affect the overall multiplier estimate. A conservative approach excludes effects on GDP during the first three quarters after transfers arrive, which may be statistically attractive in a mean squared error sense since it yields more precise, if surely somewhat downward biased, estimates. Under this assumption, the expenditure (income) multiplier estimate is 2.04 (1.45), which is smaller than the preferred estimates in Table 2.5, as expected, and estimates attain greater statistical significance (see Appendix Table B.4.4). An arguably more realistic method utilizes the household consumption data from Haushofer and Shapiro (2016) for the first three quarters post-transfer, as we did in the construction of the MPC discussed above. This yields a larger estimated expenditure multiplier of 3.09, with the increase due to a greater contribution from recipients' non-durable consumption (see Ap-

[43] Note that direct imports by households themselves are unlikely to increase because on average only 10% of households report ever shopping at a market outside our study area, and overall the impacts we see on household spending and local enterprise revenue are fairly similar, suggesting that consumer spending was quite localized. Similarly, non-farm businesses report only 5% of customers coming from outside the study area, and that share does not change significantly in response to treatment.

pendix Table B.4.4). When this refinement is combined with our preferred assumptions on input shares, the expenditure multiplier estimate is 2.48 (Appendix Table B.4.5). Though it relies on additional assumptions, some readers may prefer this estimate since it addresses several limitations in measurement of the main expenditure multiplier estimate.

2.6 Welfare implications

Transfer multiplier estimates have typically been used for positive economic analysis, to predict how fiscal policy will affect output. Yet since output is not social welfare, how fiscal policy affects welfare is a distinct issue. Classic derivations of fiscal multipliers from accounting relationships such as the "Keynesian cross" could not deliver parallel statements about welfare as they were not grounded in models of individual preferences. While recent papers have focused primarily on estimation (Ramey 2019), a few have examined the relationship between fiscal multipliers and welfare in the context of micro-founded models, emphasizing that multipliers need not be sufficient statistics for welfare or (consequently) for optimal policy (Sims and Wolff 2018). In fact, Mankiw and Weinzierl (2011) construct examples in which the interventions with the largest multipliers have the *least* impact on social welfare. The program evaluation literature on cash transfers, meanwhile, has largely focused on estimating behavioral responses without exploring what these mean for welfare.

Here we examine the broad channels through which transfers could affect household welfare, and how these relate to the transfer multiplier. Let indirect utility function $v_i(T_i, T)$ define the utility achieved by household i when it receives a (possibly zero) transfer T_i while other eligible households in the area receive transfers of T each. We are interested in characterizing how T affects the quantity T_i^* defined by $v_i(T_i^*, 0) = v_i(T_i, T)$, in other words, the transfer that would make household i indifferent between receiving T_i^* on the one hand, and experiencing the intervention we study on the other. Notice that if there were no general equilibrium effects, in the sense that v_i did not depend on T, then we would simply have $T_i^* = T_i$, i.e., the tautology that the value of receiving a dollar is a dollar.

We think of v_i as the value of some generic underlying optimization problem

$$v_i(T_i, T) = \max_{x_i} u_i(x_i, x_{-i}(T)) \text{ s.t. } x_i \in X(T_i, T) \tag{2.9}$$

Here u_i represents preferences over variables x_i which the household chooses from a set X, as well as variables x_{-i} chosen by others. This formulation delineates two ways in which T can affect the utility of household i. First, it may change market outcomes that determine the choice set X – for example, prices or income from various sources. We therefore need to interpret the impacts on output that generate the transfer multiplier through this lens. Second, it may change behaviors $x_{-i}(T)$ that affect i's well-being without appearing in the transfer multiplier (e.g., through externalities).

Market outcomes

An increase in (real) output must reflect some combination of (i) an increase in the *employment* of factors of production and (ii) an increase in their aggregate *productivity.* While the latter represents an unambiguous welfare gain, the former comes at an opportunity cost – the value of foregone leisure, for example, in the case of labor inputs, or of foregone present consumption in the case of capital inputs. Appendix B.11 provides a formal illustration of the mapping from household welfare to aggregate output, emphasizing this point. The discussion also illustrates how household welfare differs from household expenditure, which is often used in the program evaluation literature as a proxy for well-being. Specifically, expenditure does *not* take into account the opportunity cost of supplying labor (or other inputs), and over any finite time interval incorrectly interprets dis-saving as a welfare gain.

A key question is thus the extent to which the output response we observe can be explained by increases in the supply of scarce factors of production. In the data, we find fairly limited evidence of increases in the employment of either land, labor, or capital. Land is in relatively fixed supply; agricultural households do not report owning or renting more of it (Table B.2.5) and we would not expect it to be a limiting factor in the sectors in which the output expansion is concentrated (namely, retail and manufacturing). Total household labor supply does not change significantly (Table 2.2), though we do see a net shift out of self-employment and into wage employment (Table B.2.4, Panel A), with the latter increasing by 1.9 hours per person per week on average across recipients and non-recipients. These estimates are not statistically different from zero, however, and even under generous assumptions can explain only around a 5% increase in real output, well below the observed response.[44]

As for capital, the non-agricultural enterprises that increased their output did not increase investment in fixed capital (Table 2.3, row 7) and, while increasing inventories somewhat, actually decreased them slightly in proportion to sales (Table 2.3, row 6). Moreover, if investment were driving output increases then we would expect to see these increases concentrated in enterprises owned by recipients, who gained access to a new source of capital, but if anything we find the opposite (Table B.2.3). Overall, the limited factor supply response suggests that the bulk of the output response we estimate must be attributable to productivity gains, and should thus be valued at roughly \$1 per \$1 in welfare terms. (We discuss productivity further in Section 2.7.)

The distribution of benefits also matters for welfare to the extent we value more highly expansions in the budget sets of poorer households. While transfers were targeted to relatively poor households, we have seen that large spillovers accrued to their somewhat richer neighbors. Indeed, we find no significant reductions in village-level Gini coefficients for consumption expenditure or wealth in treatment villages, and a small and significant ($p < 0.05$) *increase* for wealth in control villages (Table B.2.7). We also reject in most cases the null that observed effects on Gini coefficients are equal to the counterfactual changes we might have expected had there been *no* spillovers. Overall, the patterns underscore the

[44]Specifically, an increase of 1.9 hours per person is a 8.1% increase in wage labor hours. Assuming a Cobb-Douglas production function with a labor share of 2/3, and no productive value of time given up from self-employment, this implies a 5.4% increase in real output.

large spillover gains for non-recipient households: wealthier non-recipients benefit along with recipients, on some dimensions so much that inequality may slightly increase.

Distributional effects could also work through prices; while the overall price level changed only slightly, changes in relative prices could transfer value between net buyers and net sellers of goods and services. However, as for the overall index, effects are muted across all individual goods prices that we measured, with nearly all changes within a -1% to +1% range, indicating that any redistributive effects via price changes are likely to be very small (Figure B.2.4).

Non-market outcomes and externalities

We measured several outcomes that do not enter into our multiplier calculation but that arguably influence well-being or proxy for it, and may thus capture externalities either between or within households ($x_{-i}(T)$ in Equation 2.9). Specifically, we examine indices for psychological well-being, health status, food security, education, female empowerment, and security from crime. Each index is the inverse-covariance-weighted sum of component z-scores signed so that positive values indicate better outcomes.[45] The index for psychological well-being can be interpreted as a measure of overall well-being. The next four indices arguably capture intra-household externalities, while security from crime is an inter-household externality.

For recipient households, we find positive and significant reduced-form effects for four of the six indices: psychological well-being, food security, education and security. Estimates are close to zero and not significant for health and female empowerment.[46] Total effects including spillovers are similar for all but the security index. For non-recipient households, on the other hand, we find no significant effects except for a 0.1 SD increase in the education index ($p < 0.10$). We do not find evidence of adverse spillover effects for non-recipient households on any index, with point estimates positive for all but the security index, which is indistinguishable from zero (-0.02 SD, SE 0.07). Village public good provision was also unchanged (Walker 2018).

Overall, this pattern of findings suggests that the most important welfare effects were market-mediated, though of course there may be other external effects we did not measure. A possible exception is the impacts on inequality noted above: to the extent that households care about comparisons with neighbors, these may constitute a form of "psychic externality."

2.7 Discussion: utilization of productive capacity

The results raise the question of which features of the local economy enabled it to respond elastically to a large aggregate demand shock. While fully addressing this is beyond the scope of the present project, we outline what can be said given available data.

[45]The first five of these were pre-specified as primary outcomes; the components of the security index were pre-specified as part of a family of outcomes, though combining them into an index was not. Details on index construction and results for components are in Appendix B.2, and PAP details in Appendix B.10.

[46]The latter (non)-result contrasts with Haushofer and Shapiro (2016) who found increases in female empowerment and reductions in domestic violence among households receiving a similar transfer.

Any explanation of these patterns must apply to the retail and manufacturing sectors specifically, as it is here rather than in agriculture or services that output gains are concentrated (Table B.2.2). Moreover, it cannot rely on an increase in the *employment* of factors of production, since we find little evidence of this (Section 2.6). It must instead reflect an increase in the *utilization* of factors employed, as well as in the throughput of intermediate goods. This notion is consistent with our observation (during fieldwork) of the retail and manufacturing enterprises in the area, which typically involve some degree of "on-demand" production. A retail establishment, for example, requires premises and an employee to "mind the shop," but once these are in place the volume of goods it sells depends largely on consumer demand. Similarly, many small-scale manufacturing enterprises require equipment and staff to be in place but then produce only when customers arrive. In fact, about 60% of manufacturing revenue accrues to just two enterprise types, grain mills and welding shops, both of which largely operate in this way.

These examples suggest retail and manufacturing sectors in which there are important inputs whose costs are fixed over the relevant ranges – e.g., a building, milling machinery, or hiring an employee – and whose utilization thus depends on demand. While we did not measure capacity utilization directly, some indirect evidence suggests the existence of meaningful slack. The average non-agricultural enterprise saw just 1.9 customers per hour, in between which other inputs (i.e., employee time, fixed capital, inventories, etc.) may sit idle. For labor inputs in particular, 69% of non-agricultural enterprises have just a single employee, which suggests that (due to integer constraints) the labor input is essentially fixed over the relevant range.[47]

Given this structure of production, we would expect the revenue from additional sales to be paid out to the suppliers of intermediate goods, the suppliers of elastic factors of production (whose marginal product increased as they became better utilized), and to enterprise owners to the extent they can extract economic profits. We do not directly measure purchases of intermediates, but upper bounds for the expenditure-weighted share of intermediate inputs in total sales are sizeable, at 57% in the retail and 18% in the manufacturing sector (see Appendix B.4). We also see an increase in wage bills, which accounts for 26% of increased revenue (Table 2.3). Estimated effects on profits, meanwhile, are positive but modest and not statistically significant (and may in any case be better interpreted as returns to the owners' capital or labor which, as usual, are difficult to distinguish from true economic profits).

While suggestive, this interpretation of the supply side response to a demand shock is consistent with other recent findings. In Uganda, Bassi et al. (2019) find that employees in on-demand manufacturing (e.g., welding, furniture-making) spend about 25% of time "waiting for customers" or "eating and resting." More broadly, it relates to the old idea in development economics that it might be possible to expand production without notable price inflation due to the availability of slack capacity. Classic arguments focused on "surplus labor" due to artificially high wages (Lewis 1954), while here both labor and capital

[47]Note that while we do not observe a large response in reported labor supply, we are not measuring utilization of labor capacity at the intensive margin; e.g., we do not distinguish between the time a shopkeeper waits for customers or serves them.

appear to have been underutilized due to limited flexibility to scale their employment to match demand. Local mechanisms to address this through better coordination, such as periodic markets, do so imperfectly, leaving some degree of residual excess capacity.[48]

A deeper question is whether specific market failures contribute to slack capacity in steady-state in rural Kenya. Here we speculate about possibilities and directions worthy of future investigation. The most immediate explanation revolves around the small scale of local market activity; for instance, a single grain mill typically serves each village. While the capacity provided by the standard grinding machine and the worker staffing it may generate positive profit for the owner, this capacity may also exceed average local demand, implying excess steady-state capacity which could be engaged following a demand shock. The small scale of local markets is itself likely to reflect the poor road quality and high transport costs that characterize rural Africa (Foster and Briceno-Garmendia 2010). The same logic suggests that multipliers could be smaller in cities due to their greater population density and better transportation infrastructure, not to mention the fact that rural economies are often relatively closed, with large shares of consumption coming in the form of locally produced food and other basic necessities.

Contracting frictions and institutions may also affect local market structure and capacity. For instance, Bassi et al. (2019) document a pattern of small industrial clusters in neighboring Uganda, in which a dozen small carpentry firms producing nearly identical products may co-exist in the same area. Each of these separately owned firms has one or at most a few employees, and they are characterized by the slack labor capacity noted above. Consolidation into fewer, larger firms – each better utilizing workers' time and any installed machinery, and run by a more capable manager – could conceivably reduce slack and free up labor to shift to alternate activities. Further research on the legal, financial and output market frictions that prevent horizontal integration of this kind would be useful.[49]

2.8 Conclusion

A large-scale cash transfer program in rural Kenya led to sharp increases in the consumption expenditures of recipient households, and extensive broader effects on the local economy, including large revenue gains for local firms (that line up in magnitude with household consumption gains), as well as similar increases in consumption expenditures for non-recipient and recipient households approximately a year and a half after the initial transfers. Firms do not meaningfully increase investment, and there was minimal local price inflation, with precisely estimated effects of less than 1% on average across a wide range of goods. Two independent calculations of the local transfer multiplier using consumption data and income data yield estimates of approximately 2.5, and reject the hypothesis that the multipliers are less than or equal to 1 with 90% confidence. Several

[48] A growing literature also finds evidence of excess capacity in rich countries, especially in periods of recession (Murphy 2017; Michaillat and Saez 2015; Chodorow-Reich 2019).

[49] Another fruitful direction for further investigation is whether cash transfers triggered a productivity-enhancing re-allocation of factors of production across sectors and firms (Hsieh and Klenow 2009).

suggestive patterns are consistent with the existence of "slack" on the production side in our context, which may partially account for the large estimated multiplier.

Concerns that cash transfer programs like the one we study could have adverse consequences for non-recipients were not borne out in our setting. Firm revenues and non-recipient households' consumption expenditures rise substantially in areas receiving large cash transfers; there is little price inflation; overall economic inequality does not increase meaningfully in treated areas; nor are there negative effects in terms of domestic violence, health, education, psychological well-being, and local public goods. Instead, the positive spillovers we find suggest that RCTs of cash transfer programs that simply compare outcomes in treatment versus control villages may understate true overall impacts by ignoring the general equilibrium effects that we capture (along the lines that Miguel and Kremer (2004) argue in the context of a health program).

This study is among the first to exploit randomized controlled trial methods to directly estimate macroeconomic parameters and more broadly capture large-scale aggregate effects of a development program. The multiplier effects that we focus on here have been the subject of intense interest since at least the seminal work of Keynes (1936). Our approach thus provides a novel counter-example to the well-known critique that RCT methods are not well-suited to studying the "big" questions in development economics (Bardhan 2005; Easterly 2006; Deaton 2010). We demonstrate that there need not always be a trade-off between a study's rigor and its relevance: economics research can increasingly achieve both (Muralidharan and Niehaus 2017; Burke, Bergquist, and Miguel 2019).

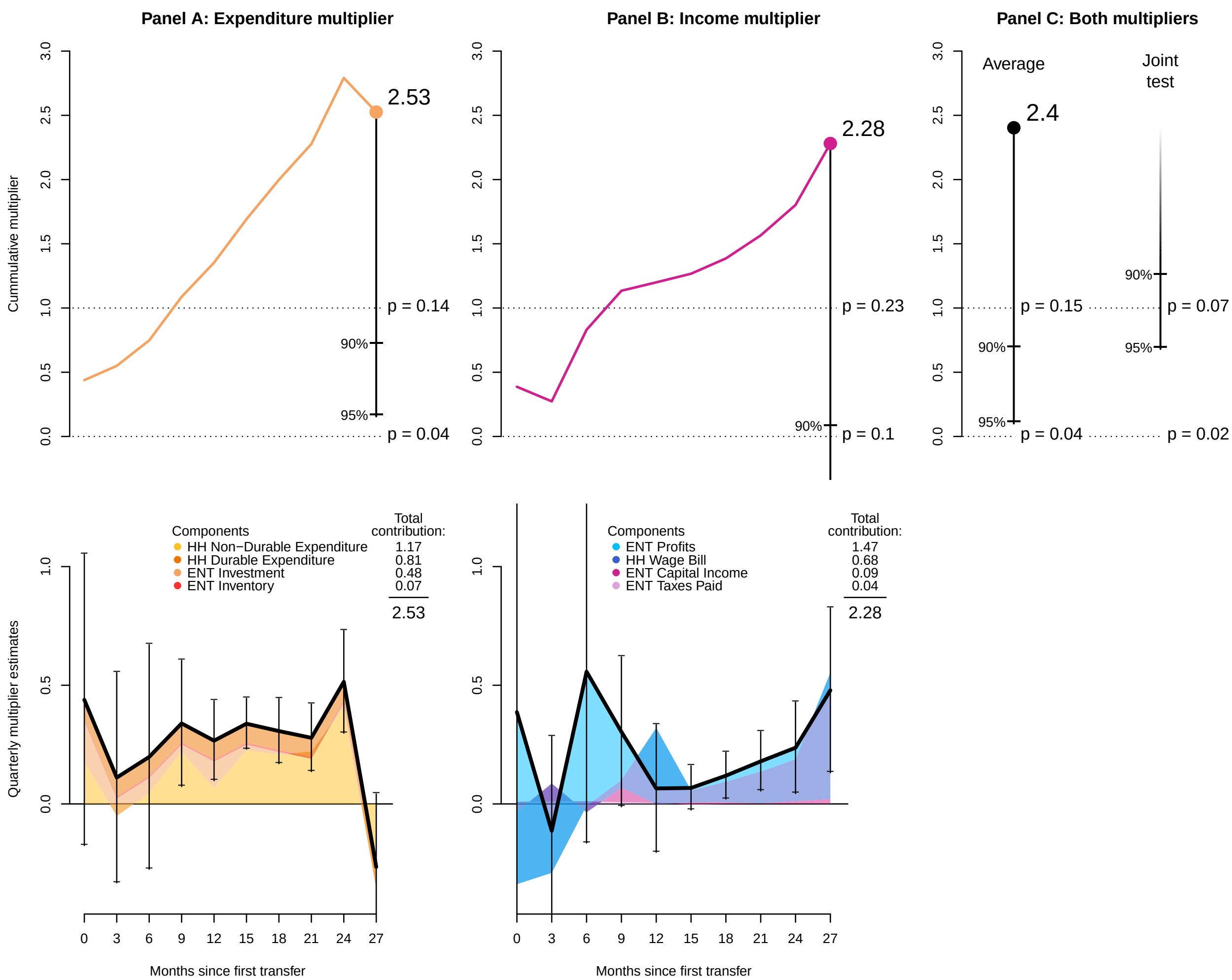
The extent to which the multiplier results apply to other settings merits further discussion. They are likely particularly relevant for rural areas of low and middle income economies that share structural and institutional features with Kenya, including many other African settings. One open question is the extent to which targeting of particular types of households, and the distribution of spending propensities across households, affect the multiplier: for example, spillover effects might have been more muted if the program had also targeted transfers to some better-off households with lower marginal propensities to spend on local goods than the poor rural households we study. A second issue is how the multiplier may vary over the business cycle. It is noteworthy that we estimate a large multiplier during a period when the Kenyan economy was experiencing steady economic growth, rather than a recession; this suggests that any under-utilization of supply side capacity is not simply temporary or cyclical in rural Kenya, but may be more persistent.[50] All that said, the results do not necessarily imply that "helicopter drops" of money as part of a scaled-up national cash transfer program would yield similar results, taking into account potential differences across rural and urban locations. The source of funding would also matter: simply printing money, for example, would likely have different inflationary consequences than financing via foreign contributions, as we study here (and is common for many social protection schemes).

Looking ahead, a traditional perspective in the case of an open economy with complete markets is that the economy should eventually revert to its previous steady-state after a

[50]Recent work argues that there may be a related phenomenon of steady-state "liquidity traps" or "secular stagnation" in advanced economies (e.g., Rachel and Summers 2019, Mian, Straub, and Sufi, forthcoming).

local aggregate demand shock like the one we study ends, with only transient effects on consumption and prices (Farhi and Werning 2016). However, other theoretical perspectives from international trade, economic geography, and development (e.g., Marshall 1890, Rosenstein-Rodan 1943, Murphy, Shleifer, and Vishny 1989, Krugman 1991), as well as the liquidity traps literature, suggest there could be persistent local effects of a temporary cash infusion, due to agglomeration effects, increasing returns, changes in income inequality, market structure and firm specialization, and even shifts in the social networks of traders and suppliers. Temporary cash transfers and other forms of assistance have also been shown to have effects on long-run human capital accumulation and earnings (Bouguen et al. 2019; Baird et al. 2016). An evaluation of long-run patterns of economic activity, firm dynamics, migration, and household living standards in the sample communities would provide a valuable experimental test of these theories.

Figure 2.1: Transfer multiplier over time



Notes: Panel A shows the cumulative expenditure multiplier over the first 29 months after start of the transfers in the top panel, and the corresponding quarterly impulse response function (IRF) in the bottom panel. The integral under this IRF yields our overall point estimate of 2.53. Colored areas below the IRF represent the different components of expenditure and the adjacent table indicates their total (over time) contribution. Darker shading indicates cases where a component turns negative in a given quarter, leading some areas to overlap. Brackets around the quarterly IRF point estimates indicate ±1SE confidence intervals obtained from 2000 wild bootstrap replications. Whiskers below the overall point estimate indicate one-sided confidence intervals from the same bootstrap procedure, with p-values corresponding to tests of the one-sided hypotheses $H_0 : \mathbb{M} < 0$ and $H_0 : \mathbb{M} < 1$ presented at the horizontal lines at 0 and 1 respectively. Panel B repeats the same exercise for the income multiplier. Panel C presents results from aggregating the two estimators either by averaging them (left-hand side) or testing the joint null that both are less than the indicated critical values (right-hand side). In each case whiskers indicate one-sided confidence intervals obtained via the bootstrap as above.

Table 2.1: Expenditures, Savings and Income

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	$\mathbb{1}$(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Panel A: Expenditure				
Household expenditure, annualized	293.59 ***	338.57 ***	334.77 ***	2,536.01
	(60.11)	(109.38)	(123.20)	(1,933.51)
Non-durable expenditure, annualized	187.65 ***	227.20 **	317.62 ***	2,470.69
	(58.59)	(99.63)	(119.76)	(1,877.23)
Food expenditure, annualized	72.04*	133.84 **	133.30 **	1,578.05
	(36.96)	(63.99)	(58.56)	(1,072.00)
Temptation goods expenditure, annualized	6.55	5.91	-0.68	37.07
	(5.79)	(8.82)	(6.50)	(123.54)
Durable expenditure, annualized	95.09***	109.01 ***	8.44	59.41
	(12.64)	(20.24)	(12.50)	(230.83)
Panel B: Assets				
Assets (non-land, non-house), net borrowing	178.78 ***	183.38 ***	133.06 *	1,131.66
	(24.66)	(44.26)	(78.33)	(1,419.70)
Housing value	376.92 ***	477.29 ***	80.65	2,032.11
	(26.37)	(38.80)	(215.81)	(5,028.27)
Land value	51.28	158.47	544.85	5,030.03
	(186.22)	(260.91)	(459.57)	(6,604.66)
Panel C: Household balance sheet				
Household income, annualized	79.43*	135.70	224.96 ***	1,023.36
	(43.80)	(92.10)	(85.98)	(1,634.02)
Net value of household transfers received, annualized	-1.68	-7.43	8.85	130.08
	(6.81)	(13.06)	(19.11)	(263.65)
Tax paid, annualized	1.94	-0.09	1.68	16.92
	(1.28)	(2.02)	(2.02)	(36.50)
Profits (ag & non-ag), annualized	26.24	35.85	36.37	485.56
	(23.67)	(47.66)	(44.88)	(786.92)
Wage earnings, annualized	42.43	73.66	182.63 ***	494.95
	(32.23)	(60.82)	(65.53)	(1,231.12)

Notes: Column 1 reports the coefficient on an indicator for treatment village from a regression using data from eligible households (as classified by the GE census team), and includes an indicator for saturation status of the sublocation (Equation 2.1). Column 2 reports the total effect on treated households (eligible recipients) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a household's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the household (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer), as in Equation 2.2. For this analysis, the sample is restricted to eligible households, including between 5,372 and 5,424 observations. Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households (5,448 to 5,509 observations), coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer), as in Equation 2.3. The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. Standard errors are clustered at the sublocation in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.2: Input Prices and Quantities

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	$\mathbb{1}$(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Panel A: Labor				
Hourly wage earned by employees	0.10***	0.04	0.19*	0.70
	(0.03)	(0.04)	(0.10)	(0.89)
Household total hours worked, last 7 days	2.44	1.41	-4.69	63.19
	(1.71)	(3.69)	(3.17)	(54.12)
Panel B: Land				
Land price per acre	168.02	366.46	557.44	3,952.48
	(201.18)	(290.85)	(412.34)	(3,147.29)
Acres of land owned	-0.19	-0.10	0.08	1.42
	(0.14)	(0.09)	(0.10)	(2.37)
Panel C: Capital				
Loan-weighted interest rate, monthly	-0.01	0.01	-0.01	0.06
	(0.01)	(0.01)	(0.01)	(0.07)
Total loan amount	5.53	3.12	6.12	80.57
	(4.95)	(8.34)	(13.23)	(204.28)

Notes: Column 1 reports the coefficient on an indicator for treatment village from a regression using data from eligible households (as classified by the GE census team), and includes an indicator for saturation status of the sublocation (Equation 2.1). Column 2 reports the total effect on treated households (eligible recipients) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a household's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the household (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer), as in Equation 2.2. For this analysis, the sample is restricted to eligible households, including between 2,828 and 5,423 observations for variables at the household level, and 2,832 observations at the individual level for wages. Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households, coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer), as in Equation 2.3. We have between 2,781 to 5,509 observations at the household level and 2,391 wage observations at the individual level. The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. In addition, prices are quantity-weighted. That is, wages are weighted by the number of hours worked, land prices by the number of acres purchased, and interest rates by size of each loan. Standard errors are clustered at the sublocation in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.3: Enterprise Outcomes

	(1)	(2)	(3)	(4)
	Treatment Villages		**Control Villages**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation weighted mean (SD)
Panel A: All enterprises				
Enterprise profits, annualized	-2.27	55.77	35.08	156.79
	(21.42)	(36.73)	(37.36)	(292.84)
Enterprise revenue, annualized	-29.61	322.16 **	237.16 **	494.45
	(102.74)	(138.17)	(112.72)	(1,223.07)
Enterprise costs, annualized	-13.32	89.35**	73.08	117.22
	(28.63)	(38.51)	(46.77)	(263.46)
Enterprise wagebill, annualized	-15.90	75.99**	66.57*	97.35
	(25.49)	(30.64)	(35.86)	(237.01)
Enterprise profit margin	0.01	-0.11*	-0.12**	0.33
	(0.02)	(0.06)	(0.05)	(0.30)
Panel B: Non-agricultural enterprises				
Enterprise inventory	11.02	34.69***	16.90	50.41
	(9.14)	(13.39)	(10.66)	(131.86)
Enterprise investment, annualized	4.00	13.58	6.82	46.57
	(7.05)	(13.10)	(7.96)	(167.44)
Panel C: Village-level				
Number of enterprises	0.01	0.02	0.01	1.12
	(0.01)	(0.01)	(0.01)	(0.14)

Notes: Column 1 reports the coefficient on an indicator for treatment village, and includes an indicator for saturation status of the sublocation. Column 2 reports the total effect on enterprises in treatment villages (own-village effect plus across-village spillover) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a enterprise's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the enterprise (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer). Column 3 reports the total effect on enterprises in control villages (across-village spillover only). For each Column, we stack 3 separate regressions for own-farm enterprises, non-agricultural enterprises operated within the household, and non-agricultural enterprises operated outside the household, due to our independent sampling across these enterprise categories (Equations B.1 and B.2). We have between 9,997 and 10,254 observations for all enterprises, and 2,389 to 2,398 for variables we collect for non-ag enterprises only, and 653 villages. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across all enterprise categories). Each regression is weighted by inverse sampling weights and contains village-level baseline averages of the outcome variable by enterprise category when available. For monetary values, we convert effects to a per-household level by multiplying the average effect per enterprise in each enterprise category by the number of enterprises in that category, dividing by the number of households in our study area, and summing over all enterprise categories. For the number of enterprises, we run regressions at the village level, where the outcome is the number of enterprises per household in each category, we weight by the number of households in each village and sum up over all enterprise categories. For the profit margin, we weight the effects across all enterprise categories by their share in the economy, and across each enterprise by revenue, so that our estimate represents the effect on the revenue-weighted average enterprise in the economy. Standard errors are clustered at the sublocation in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.4: Output Prices

		(1)	(2)	(3)	(4)
		Overall Effects		**ATE by market access**	
		ATE	Average maximum effect (AME)	below median	above median
All goods		0.0010*	0.0042	0.0017*	0.0007
		(0.0006)	(0.0031)	(0.0009)	(0.0007)
By tradability	More tradable	0.0014	0.0062	0.0023	0.0021
		(0.0015)	(0.0082)	(0.0023)	(0.0018)
	Less tradable	0.0009	0.0034	0.0015	0.0001
		(0.0006)	(0.0032)	(0.0011)	(0.0008)
By sector	Food items	0.0009	0.0036	0.0016	0.0002
		(0.0006)	(0.0033)	(0.0012)	(0.0008)
	Non-durables	0.0014	0.0061	0.0026	0.0019
		(0.0017)	(0.0089)	(0.0026)	(0.0019)
	Durables	0.0019*	0.0070	-0.0009	0.0034**
		(0.0011)	(0.0061)	(0.0011)	(0.0016)
	Livestock	-0.0008	-0.0027	-0.0008*	-0.0017
		(0.0010)	(0.0052)	(0.0004)	(0.0020)
	Temptation goods	-0.0011	-0.0112	-0.0008	-0.0003
		(0.0026)	(0.0143)	(0.0036)	(0.0035)

Notes: Each row represents a regression of the logarithm of a price index on the "optimal" number of lags and distance buffers of per capita Give Directly transfers in each buffer. Price indices are based on 321,628 non-missing price quotes for 70 commodities and products. For each product, we take the logarithm of the median price quote in a market-month, and create our market price indices as an expenditure weighted average of these median price quotes across all goods in that market-month. Regressions include a panel of 1,734 market-by-month observations. The number of radii bands and lags is chosen sequentially by minimizing the BIC, as pre-specified, for the overall price index, which selects a 4km radius; subcomponents use this value as well. Regressions include a full set of market and month fixed effects. Column 1 reports the implied ATE, calculated by evaluating the "optimal" regression specification at the average level of treatment intensity between September 2014 and March 2017, the time during which transfers went out. Column 2 reports the average maximum effect, calculated at the average across all markets of the month in which the largest per capita transfers went into a market's neighborhood (up to the largest buffer selected by the algorithm). Columns 3 and 4 break down the ATE by market access, defined as $MA_m = \sum_{r=0}^{10} r^{-\theta} N_r$, where $\theta = 8$ and N_r is the population in in the $r-2$ to r km buffer around each market. Standard errors (in parentheses) are as in and we allow for spatial correlation up to 10km and autocorrelation up to 12 months. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2.5: Transfer Multiplier Estimates

	(1) $\mathbb{M}$ Estimate	(2) H_0: $\mathbb{M} < 0$ p-value	(3) H_0: $\mathbb{M} < 1$ p-value
Panel A: Expenditure multiplier	2.53 (1.42)	0.04**	0.14
Household non-durable expenditure	1.17 (1.32)	0.19	
Household durable expenditure	0.81 (0.05)	0.00***	
Enterprise investment	0.48 (0.42)	0.13	
Enterprise inventory	0.07 (0.03)	0.02**	
Panel B: Income multiplier	2.28 (1.73)	0.10*	0.23
Enterprise profits	1.47 (1.28)	0.13	
Household wage bill	0.68 (1.15)	0.27	
Enterprise capital income	0.09 (0.17)	0.31	
Enterprise taxes paid	0.04 (0.03)	0.08*	
Panel C: Expenditure and income multipliers			
Average of both multipliers	2.40 (1.38)	0.04**	0.15
Joint test of both multipliers		0.02**	0.07*

Notes: Results are from the joint estimation of expenditure and income multipliers. Column 1 reports point estimates of both multipliers and their respective components. Each component is estimated individually and the multiplier is obtained by aggregating components as described in the main text. Effects of the cash infusion on flow variables (non-durable consumption, investment, wages, profits, capital income, and taxes) are obtained by dynamically estimating effect sizes over 29 months after the first transfer and computing the integral under this curve (Equation 2.7). Effects on remaining stock variables are the estimated the total endline treatment effects (Equations 2.2, 2.3 and B.2). Transfer amounts and outcome variables are deflated to January 2015 using the overall consumer price index in the geographically closest market. Standard errors are computed from 2,000 replications of a wild clustered bootstrap, which re-allocates within-sublocation Rademacher-perturbed residuals from the main population regressions to fitted outcome values to create perturbed samples. Columns 2 and 3 conduct one-sided tests of each multiplier estimate $\mathbb{M}$ against 0 and 1 respectively, using the bootstrapped distributions of $\mathbb{M}$. Panel C conducts two tests regarding both multipliers. The first row computes the average of both estimates and conducts tests on this average using the same bootstrap procedure. The last row reports p-values from joint tests of both multipliers against the same nulls. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Chapter 3

Balancing Work and Childcare: Evidence from COVID-19 School Closures and Reopenings in Kenya

3.1 Introduction

The availability and cost of childcare have been shown to significantly affect adult labor supply in high-income countries, particularly for women. But there is less evidence on this relationship in low- and middle-income countries (LMICs), particularly in Sub-Saharan Africa (Halim, Perova, and Reynolds 2021). Yet, a historical perspective highlights the important role of women's labor supply in economic development (Boserup, Tan, and Toulmin 2013). Understanding how childcare and adult labor supply interact is therefore crucial in these settings.

Sub-Saharan African countries differ from high-income countries in many ways relevant to this question. Households have more children but also more adults on average (UN 2020). Formal early childhood care availability is increasing, but from a low base and there are concerns around quality and cost (Samman et al. 2016). Female labor participation is high but concentrated in informal activities (ILO 2017). Family farm or non-farm enterprise work is widespread, and may be more accommodating of childcare needs than wage employment. Critically, older children play an important role in household productive activities (Kielland and Tovo 2006), including sibling childcare (Jakiela et al. 2020), meaning they are not just childcare recipients within the household. It is not clear *a priori* how these differences would affect the relationship between childcare needs and labor participation.

An important factor influencing household childcare needs is the availability of low- or no-cost schooling. In 2020, countries around the world closed schools in response to the COVID-19 pandemic. This paper leverages school closure policies in Kenya as exogenous shocks to provide empirical estimates of the impact of childcare responsibilities on adult labor supply in an LMIC setting.

Kenya closed all schools nationwide after its first COVID-19 cases in March 2020, partially reopened schools for specific grades in October 2020, and fully reopened for all grades in January 2021. Household childcare needs increase during school closures creating trade-offs for adults' time allocation across childcare, work in different sectors, and other activities. We exploit quasi-random variation in when children enrolled in different grades were eligible to return to school and use data from the nationally-representative panel Kenyan Rapid Response Phone Survey to implement a difference-in-differences analysis comparing changes in labor supply after the October partial reopening for adults in households with children in grades 4 or 8—eligible to return (99% did)—against those with children in adjacent grades.

Weekly work hours increase by 3.6 (22%) after the partial reopening for adults with a child eligible to return to school, driven by a 26% increase in household agriculture hours. Agricultural households drive average impacts and poorer households— based on an index of housing and assets—increase work hours by more than wealthy ones but not significantly so.

Surprisingly, the impacts are not significantly different by sex, contrasting with evidence on pandemic labor supply changes from high-income contexts (Alon et al. 2021; Amuedo-Dorantes et al. 2020; Collins et al. 2021; Hansen, Sabia, and Schaller 2022; Heggeness 2020) and expectations based on women's role as primary caregivers in most Kenyan households. One reason for the lack of difference is that in our sample, both sexes contribute to childcare and increased childcare hours during school closures. A second reason is that in this setting, school-age children are both *receivers* of childcare and *contributors* to household productive activities, including childcare to siblings and household agriculture. We find that women's labor supply responds relatively more to changes in childcare burdens while men respond more to changes in child agricultural labor, leading to offsetting impacts.

Effects of the partial reopening vary with household composition, consistent with differences in how the partial reopening affects treated households' childcare burdens due to economies of scale in childcare and the important role of older siblings as care providers. Increases in work hours are driven by households without below-school-age children, where the student's return to school decreases adults'—and particularly women's—childcare burdens. A second mechanism for the impacts we observe is participation of school-age-children in household agriculture. Child agriculture hours decrease in treated households by approximately one-quarter of the total increase in adult agriculture hours, suggesting some of this increase substituted for child labor during Kenya's main harvest season.

This paper explores a new dimension of the relationship between childcare and labor supply (e.g., Browning 1992; Connelly 1992; Ribar 1992) by considering how formal childcare for school-age children (through schooling) affects households through changes in both childcare burdens and availability of child labor. The current literature largely studies childcare for below-school-age children and treats children solely as childcare recipients, while focusing on settings dominated by wage employment. These characteristics do not generalize to many LMIC contexts. Among studies of childcare and labor supply in African LMICs (Bjorvatn et al. 2021; Clark et al. 2019; Delecourt and Fitzpatrick 2021; Heath 2017; Lokshin, Glinskaya, and Garcia 2000; Martinez, Naudeau, and Pereira 2012; Quisumbing, Hallman, and Ruel 2007), causal identification is limited, only two include rural areas, and none consider the role of children as household labor providers. This paper estimates causal impacts of a change in childcare needs using a natural experiment with a nationally-representative sample of households in an African LMIC with most engaged in household farm and non-farm enterprise rather than wage work. Analyzing a shock affecting formal care provision for school-age children further allows us to shed light on the role of child household labor in the relationship between childcare needs and adult labor supply.

We also contribute to understanding labor impacts of pandemics and pandemic-related policies. Many studies have analyzed the gendered effects of the COVID-19 pandemic on childcare and employment (see e.g., Alon et al. 2021; Amuedo-Dorantes et al. 2020; Collins et al. 2021; Del Boca et al. 2020; Furman, Kearney, and Powell 2021; Grantham et al. 2021; Hansen, Sabia, and Schaller 2022; Heggeness 2020; Prados and Zamarro 2021; Zamarro and Prados 2021). Though there is descriptive evidence from COVID-19 in India (Chauhan 2020; Deshpande 2020) and South Africa (Casale and Posel 2020) and from Ebola in Sierra Leone and Liberia (Wenham et al. 2020), and one causally-identified study on COVID-19 in Shaanxi province, China (Ma, Sun, and Xue 2020), causal estimates of impacts of changes to household childcare during a pandemic on adults' labor supply in LMICs are currently lacking.

A back-of-the-envelope calculation indicates that pandemic school closures decreased work hours across Kenya by 2.1 billion in 2020—at the average hourly earnings in the data a cost of USD 3 billion (3.1% of 2019 GDP). More generally, we demonstrate that reducing household childcare burdens can broadly increase adults' labor participation in an African LMIC context.

3.2 Context and Data

This section summarizes Kenyan COVID-19 school closure policies, the data we use to analyze their impacts on employment, and information on childcare arrangements.

Context

Formal education in Kenya begins around age 6 and is compulsory for the first nine years. Pre-primary education has also become broadly available.Public education is free, but school-related costs such as materials, meals, and examinations are typically in the range of 25-75 USD per year for primary schools (Zuilkowski et al. 2018) and 100-500 USD for secondary schools (Bonds 2021). Kenya's academic year consists of three terms from January to October.

Schools in Kenya closed on 16 March 2020 as part of a broad set of national restrictions to reduce risk of disease transmission after the first reported COVID-19 cases.[1] The rest of academic Term 1 was cancelled. National top-down changes in school closure policies represent exogenous shocks to households, unrelated to local economic or health conditions.

On 15 September the Ministry of Education released guidelines for safe reopening of schools, but the specifics remained uncertain until 6 October when the Ministry announced

[1] Figure C.1.1 shows a timeline of school closures and reopenings, other key pandemic-related policy changes, and weekly confirmed COVID-19 cases in Kenya, along with the timing of data collection.

that students in grades 8 and 12—those sitting national exams—along with students in grade 4 should return to school on 12 October for Term 2 of 2020. This announcement was presented in the media as "a shocking move that caught parents and candidates off guard" (The Star 2020). Students in grades 4, 8, and 12 returned for Term 3 from January-March 2021 while other students returned for Term 2; their Term 3 was shifted to May-July 2021. Terms and breaks for the 2021-2023 academic calendars were shortened to allow a gradual return to the pre-pandemic term schedule in time for the 2024 academic year.

We focus on the partial school reopening for several reasons. First, unlike initial school closures, the partial reopening did not coincide with other pandemic-related policies. Second, we exploit discontinuities in eligibility to return by grade to isolate the effect of the shock. Further, because households vary in whether the students eligible for the partial reopening are net suppliers or recipients of childcare—depending on the presence of younger siblings—this shock sheds light on the importance of sibling-provided childcare.

Data

Data come from the Kenya COVID-19 Rapid Response Phone Survey (RRPS) panel, collected by the World Bank in collaboration with the Kenya National Bureau of Statistics and the University of California at Berkeley (Pape 2021).[2] The main sample ($\sim$ 80%) is drawn from the nationally-representative Kenya Integrated Household Budget Survey conducted in 2015-2016, and is supplemented by random digit dialing. The sample is intended to be representative of the population of Kenya using cell phones—80% of households nationally own a mobile phone, and these have better socioeconomic conditions on average than households that do not (Pape et al. 2021). We use data from four survey rounds covering May 2020-March 2021, along with recall data for February 2020.

The outcomes of interest are measures of labor supply.[3] The extensive margin is measured by participation in the last 7 days in three activities: employed/wage labor, household non-farm enterprise, and household agriculture. The intensive margin is captured using hours of work by activity; individuals not working in a given activity are coded as working 0 hours. The survey also captures total child hours spent in household agriculture.

Information on what grades children were enrolled in prior to the initial school closures allows us to identify households affected by the partial reopening. Nearly 99% of eligible students are reported to have returned to school.[4] We define 'treatment' households as those with children enrolled in grades 4 or 8 prior to the pandemic (eligible for the partial reopening)[5] while 'control' households have children in grades 3, 5, 6, 7, or 9, but not in grade 4 or 8. We separate 'mixed' households with children

[2] See Appendix D for more detail.

[3] We use the term labor 'supply' to refer to equilibrium outcomes, acknowledging that individuals may have been willing to supply additional labor but faced limited demand.

[4] A survey of 3,000 grade 8 students in Busia County, Kenya similarly shows that 97% reported back to school after the partial reopening (Bonds 2021). Across all grades, 97% of previously enrolled students in the RRPS returned to school after the full reopening in January 2021.

[5] Few households report any children in grade 12; we test robustness of our results to including them.

in both grade groups from 'treatment' households as they might experience different effects when not all children in the relevant grade range return to school. The main analysis sample includes 335 treatment, 361 mixed, and 948 control households.

Finally, the data include questions on household childcare arrangements, including respondent childcare hours.[6]

Childcare

Over 98% of children ages 6-17 in the RRPS are reported to have been enrolled in school in February 2020. After the March closures these children required care and supervision during the working day, representing a large unexpected shock to household childcare needs. Children primarily stayed at home with a parent during the closures (Figure C.1.2), including situations where parents were simultaneously working. Almost no households report their children spending time with childcare providers outside the home or with a maid/-domestic helper at home. Adults with schoolchildren at home will have faced trade-offs in their allocation of time across childcare, work in different sectors, and other activities given a limited time budget to accommodate increased childcare burdens.

Figure 3.1 Panel A presents how hours of childcare from different providers (excluding schools) vary with the number of household children, using data from after schools fully reopened. Non-household members provide very little childcare on average—86% of households report 0 hours of care from non-household members in the last 7 days. While formal childcare availability has been increasing in Kenya (particularly in urban areas), affordability remains a challenge for most households (Clark et al. 2021; Murungi 2013). Other adults besides the parents are present in 37% of households with children, and on average provide around 10 hours per week of childcare.

Older siblings also play an important role. In households with at least 2 children 55% of children provided childcare to siblings in the last 7 days, for 15-20 hours on average in total, demonstrating how some older siblings may be net providers rather than recipients of household childcare. Sibling childcare hours may have been higher during school closures as school-age children were home, but we only measure sibling childcare after the full reopening.

[Figure 3.1]

Respondents provide 30-35 hours per week of childcare. Figure 3.1 Panel B shows that while female respondents provide around 10-15 hours more than men, men still contribute around 25 hours on average. This contrasts with the image of fathers in African countries primarily providing economic support and little childcare, but is consistent with recent evidence (Clark, Cotton, and Marteleto 2015; Kah 2012). While the gender gap increased during school closures—women's childcare hours increased by 13.4 on average compared to 9.8 for men—the burden increased significantly for both sexes. This pattern is similar to findings for changes in domestic work during the pandemic in India (Deshpande 2020),

[6] The survey does not distinguish between time actively spent caring for a child and time spent on other activities while responsible for a child. We topcode reported childcare hours at 20 hours per day.

South Africa (Casale and Posel 2020), and many higher-income countries (see e.g., Andrew et al. 2020; Del Boca et al. 2020; Farré et al. 2020; İlkkaracan and Memiş 2021). After schools fully reopened, respondent childcare hours returned to slightly below pre-pandemic levels.

There are significant economies of scale in childcare hours in Kenya: respondent childcare hours increase very little after the first child, with total childcare hours likely determined by the child that requires the most care. Sibling childcare provision may also contribute to these economies of scale. The importance of sibling childcare suggests that a student returning to school might increase rather than decrease parents' childcare burden, in situations where they were net childcare providers during school closures.

3.3 Empirical Approach

We identify the effect of partial school reopenings through a difference-in-differences analysis comparing outcomes before and after the reopening between households with and without eligible children. We estimate regressions of the form

$$y_{iht} = \alpha + \beta_1 \cdot Post_t \times Treat_h + \beta_2 \cdot Post_t + \mu_h + County_h \times \tau_t + X_{iht} + \epsilon_{iht} \tag{3.1}$$

y_{iht} are outcomes for adult (age 18-64) i in household h at time t. $Post_t$ is an indicator for observations after the partial reopening on 12 October 2020. We include observations from May-November, omitting data from after schools fully reopened. $Treat_h$ indicates whether all household children in grades 3-9 were eligible to return to school (treatment), none were eligible (control), or some were eligible and others not (mixed). Household fixed effects μ_h absorb time invariant characteristics of households which may affect labor supply outcomes. County-by-month fixed effects control for common shocks affecting households across locations and over time. Finally, X_{iht} is a vector of controls, including individual sex, age, and household head status, number of adults, young children (age 0-4), and school-age children (5-17) in the household, and household dummies for engagement in agriculture and in enterprise. We cluster standard errors at the household level.

We exploit quasi-random discontinuities in which households are affected by the partial reopening by restricting our control group to households with children in grades adjacent to those eligible to return to school. Identification is based on the argument that unobserved factors that could affect outcomes are continuous around the thresholds of children being in adjacent grades.

Respondent and household characteristics are similar for treatment and control households during the school closures period (Table C.1.1). Mixed households look different in terms of household composition by construction as they must have one additional child on average. We focus our analysis on the comparison between control and treatment households; differences in household composition may affect estimated impacts of the partial reopening for mixed households.

Mean work hours trend almost identically for adults in treatment and control households from February to early October 2020. Differences emerge following the partial reopening but are eliminated after schools fully reopen, when all households become 'treated' (Figure C.1.3). Figure 3.2 shows further evidence of parallel trends in labor supply while schools were closed, for both women and men. There are no significant differences for treatment adults in the periods when schools were fully closed, and there is no evidence of anticipation effects in the period from September to 11 October.

[Figure 3.2]

Our main analyses pool women and men as both contribute to childcare and increased childcare hours during school closures, though we also test for different impacts by sex, as well as by particular household characteristics.

3.4 Results

Table 3.1 presents results for the impacts of partial reopening on labor supply by activity. Fifty-nine percent of adults ages 18-64 among control households were working during the school closures period, primarily in household agriculture. Mean work hours of 16.4 reflect that many workers were not working 'full-time.' Labor supply does not change in control households after schools partly reopen. This indicates that general labor conditions were not changing, consistent with no major simultaneous pandemic policy changes.

[Table 3.1]

We find no effects of treatment on the extensive margin of labor supply but a large impact on the intensive margin. Work hours in the last 7 days increase by 3.6 (22.1%) relative to adults in control households, driven by a 26.0% increase in household agriculture hours. Greater impacts on household agricultural hours than in wage work or in work participation are not surprising given that we estimate short-term impacts in the weeks following the partial reopening. Wage work is dependent on employers so may be constrained, while household agriculture is more flexible. Increased agricultural work may also be a response to reduced child labor.

Non-significant impacts on household agriculture engagement indicate that hours increase primarily among those already working in agriculture. Household agriculture engagement was affected less than other work activities by school closures and other pandemic restrictions. Adults may have been more likely to pause their engagement in household enterprise—more exposed to infections and pandemic restrictions as well as potentially more challenging to combine with childcare—and slower to resume these activities.

'Mixed' households with children eligible to return to school as well as children in adjacent grades do not change labor supply following the partial reopening. This is not surprising given what we observe about economies of scale in childcare hours in Kenya: one child returning while another of a similar age stays home is unlikely to meaningfully change adult childcare burdens.

Impacts across treatment households are driven more by households with a grade 8 student than those with a grade 4 student (Table C.1.2). Though grade 4 children likely require more care than grade 8 children, they also likely contribute less to household agriculture and are more likely to have young siblings, mechanisms we explore in section 3.4. Impacts on work hours are smaller if we expand our treatment definition to include grade 12 students also eligible to return to school. This result is not surprising as grade 12 students are net providers of household childcare.

We conduct a variety of robustness tests (Table C.1.3). Results are unchanged when using individual rather than household fixed effects, when focusing on sub-samples of adults more likely to be parent caregivers or engaged in work, and when defining *Post* by the date the potential reopening was announced.

Heterogeneity

We test for heterogeneity in impacts by estimating Equation 3.1 and fully interacting a characteristic Z with all right-hand side variables other than the household fixed effects, focusing on total working hours (Table C.1.4).Figure 3.3 displays estimated effects from regressions for sub-samples with particular characteristics.

Impacts of schools reopening on work hours are not significantly different for women (54% of the sample) relative to men. This contrasts with evidence from high-income countries, which consistently report larger effects of the pandemic on mothers' labor supply relative to fathers' and other women's, pointing to school closures as an important mechanism (e.g., Alon et al. (2021) and Collins et al. (2021)). But this result aligns with the data on childcare hours in Kenya: responsibilities prior to the pandemic are less gendered than expected and both women's and men's hours increase by over one-third during school closures. The childcare shock thus affects both parents' labor supply, though the similar increases for women and men also partly reflect different offsetting mechanisms, which we discuss in section 3.4.

[Figure 3.3]

The impact of partial reopening on work hours is over four times as large for adults in agricultural households (61%, defined as households with any agricultural activity), consistent with effects on total hours driven by household agriculture. Adults in non-agricultural households do not significantly increase work hours, likely reflecting constraints on increasing wage or enterprise labor supply in the short term.

We observe no differences in impacts between urban (46%) and rural households. The definition of 'urban' in the data includes many peri-urban areas; over 35% of household classified as urban are engaged in agriculture. The difference remains insignificant when restricting urban households to those in counties with the largest cities in Kenya, though even there 25% of households engage in agriculture. The sample may not be representative of all rural households: RRPS households must have a mobile phone and are on average better off than the population (Pape et al. 2021). This may further blur the urban/rural distinction in these data. Low take-up of formal childcare services and low wage employment in the sample may also limit urban/rural heterogeneity.

Adults in poor households increase work hours by more than those in wealthy households (measuring wealth with an index based on housing and asset ownership), though the difference is not significant. Below-mean-wealth households in the analysis sample are around 30% more likely to engage in agriculture than wealthier households and may have had fewer resources to absorb increased childcare burdens during school closures and thus been more affected by the reopening.

Adults in households with children aged 0-4 (41%) likely faced *increased* childcare burdens when an older sibling returned to school due to reduced sibling childcare. In households without young children, returning students would more likely be net childcare recipients. Indeed, we observe that the increase in work hours is 3 times larger and only significant in households with no children ages 0-4, consistent with differences in the nature of the childcare shock.

Changes in work hours do not differ by whether the household has more than two adults (45%), perhaps due to two competing mechanisms. Households with more adults may have spread out increased childcare burdens during school closures more than households with 1-2 adults, but would also better absorb any reduction in sibling-provided childcare after the reopening.

Different types of work may also be more or less affected by a childcare shock (Table C.1.6). Increased work hours are driven by adults engaged in household agriculture before the pandemic, consistent with overall impacts driven by agriculture hours. Works hours only increase after the partial reopening for treated adults that were working at some point during the school closures, reflecting how impacts are concentrated on the intensive margin of labor supply. Treated adults working for a wage or in household enterprise during school closures do not increase hours after the partial reopening, which again may reflect constraints in increasing hours in these activities in the weeks we observe after reopening.

Mechanisms

Adults must allocate their limited time across work in different sectors, childcare, other activities, and leisure. Changes in household childcare burdens after the partial school reopening may therefore affect adults' time allocation. As a proxy for household childcare burdens, we directly test for impacts of the partial school reopening on childcare hours using Equation 3.1, though childcare data are only available for respondents in these survey rounds so we cannot fully capture household-level changes. Table 3.2 column 1 shows that respondent childcare hours in the last 7 days do not change significantly for treatment respondents on average after the partial school reopening. This is not surprising as most households have multiple children and we observe significant economies of scale in household childcare hours. Point estimates are negative for treated households with a grade 4 student and positive and larger for the return a grade 8 student (Table C.1.2), consistent with older children more likely being net providers of childcare on average. The average impact of treatment on childcare is negative but not significantly different for female respondents.[7]

[Table 3.2]

Given what we observe about the role of siblings and economies of scale in childcare in the sample (Figure 3.1), children in grades 4 and 8 (typically around age 9 and 13) are likely net providers of childcare when home from school in households with younger children. Their return to school would thus *increase* the childcare burden on adults. In households without younger children, the return to school of these students would instead decrease that burden. Columns 3-6 of Table 3.2 support this, showing increased childcare hours for treated adults with young children and decreases for those without.[8] Differences are more pronounced by the presence of below-school-age children (ages 0-4) than by the presence of younger children more generally (ages 0-8), consistent with much higher childcare needs of very young children.

Positive labor supply impacts of the partial school reopening driven by households without below-school-age children (Figure 3.3) are consistent with these different childcare effects. Adults increase work hours when the partial reopening constitutes a positive childcare shock but not when it is negative (the student returning to school is a net childcare provider). These results highlight the importance of sibling-provided childcare and the role of childcare in explaining labor impacts of the reopening.

The results indicate that childcare may be particularly gendered only for care of the youngest children. Female respondents in treatment households with children ages 0-4 increase childcare by 39 hours more than those without after the partial reopening, a much larger difference than for men. Positive labor supply impacts of treatment among households without below-school-age children are driven more by women, suggesting the childcare mechanism is more important for women. Limited reductions in childcare hours on average may help explain the lack of significant difference in average impacts of the partial school reopening by sex.

School reopenings may also affect adult labor supply through reduced child labor, particularly as the timing coincides with the main harvest season for most of Kenya. In the 39% of agricultural households with children in grades 3 through 9 reporting some child agricultural labor, children worked an average of 18.1 total hours per week (28.6% of the household total) during the school closures period.

The partial school reopening reduces treatment households' child agricultural labor by 1.48 hours (Table 3.2 columns 7-8), indicating students do not contribute as much after returning to school. Reductions are larger in households without younger children to help make up for the lost child labor. For a two-parent household, the reduction in child labor represents 23.9% of the increase in adult agricultural hours. Substitution for reduced child agricultural labor could thus explain part but not all of the impacts on adult work hours.

[7] The positive point estimate on *Post* for women may be due to increased burden of childcare on women during the harvest period.

[8] Control adults—particularly women—with young children also increase childcare hours following reopening. We suspect this reflects a general household reallocation of childcare responsibilities toward women during the harvest period which coincides with the partial school reopening.

The impact of partial reopening on agricultural hours is larger for men than for women (Table C.1.5), indicating men are more responsive to the change in child agricultural labor. As women are more responsive to the change in childcare burdens, these offsetting mechanisms could lead to similar impacts of the reopening by sex. Child labor could also contribute to the different impacts by household wealth: poor households engaged in agriculture are 20.8% more likely to report child labor than more wealthy households.

Finally, though there were no additional fees incurred when schools reopened adults in treatment households may have also increased labor supply to generate income to pay for materials, meals, and extra lessons. Such costs would be higher for students in grade 12 than in grades 4 or 8. Estimated impacts of the partial reopening are smaller when including households with grade 12 children in the sample (Table C.1.2). Given expected differences in school-related costs, net childcare burden, and child agricultural labor by student grade, smaller increases in parent work hours for a grade 12 student returning to school relative to a younger student suggest childcare burdens are the main mechanism.

3.5 Discussion

The partial school reopening affected a subset of children older than those with the greatest childcare needs, whereas initial closures affected all school-age children. Reducing labor supply is also likely easier than increasing it. Impacts of the partial reopening should thus provide a conservative estimate of the contribution of school closures to initial pandemic labor participation decreases in Kenya.

Labor force participation in the last 7 days across RRPS respondents ages 18-64 fell from 76% in February 2020 before the pandemic to 59% in May-July, and average working hours fell from 23.9 to 16.9.[9] For respondents ages 18-64 in our analysis sample households, average weekly hours fell from 30.4 in February to 19.3 in May-July. Adult work hours in the last 7 days increased by 3.6 after the partial school reopening, corresponding to 32.4% of the pandemic reduction in work hours in this sample.

Increasing work hours for adults in households with school-age children (66.4% of households) during the school closure period by our estimate of the amount they increased due to the partial reopening—to approximate the counterfactual with no closures—reduces the drop in average weekly work hours from February to May-July among all adults nationally from 7.0 to 4.8 hours. We therefore estimate that school closures account for (at least) 30% of the pandemic decrease in work hours.

Across Kenya's labor force of 23.7 million (ILO 2021), a reduction of 2.2 work hours per week over the period of the school closures adds up to over 2.1 billion hours, or USD 2.96 billion at the average hourly income observed in our sample—3.1% of Kenya's 2019 GDP. This is a simplified back-of-the-envelope calculation but provides a likely conservative rough estimate of the magnitude of the labor supply impact of

[9]We focus on survey respondents as February 2020 data are limited for non-respondents. We apply household survey weights for all analyses of national changes in labor supply.

Kenya's school closures. A better understanding of their labor supply impacts may inform discussion of school closures as a potential pandemic policy response.

Although the shock we analyze takes place in the context of a global pandemic, the results will continue to have relevance as COVID-19 is unfortunately unlikely to be completely overcome in the immediate future. For example, after fully reopening schools in January 2021, Kenya closed them again in late March after a spike in COVID-19 cases before reopening again in mid-May. Further, although some pandemic-related restrictions were still in effect at the time schools partly reopened in Kenya in October 2020, many had been relaxed, so these estimated impacts of a childcare shock may generalize to similar settings with ongoing COVID-19 caseloads and related government health policies—potentially the new normal moving forward.

3.6 Conclusion

We present nationally-representative results for the impacts of childcare on labor supply in an LMIC setting, using pandemic-related school closure policy changes in Kenya as exogenous childcare shocks. Having a child eligible to return increases adult work hours in the weeks after schools partially reopen, suggesting childcare burdens constrain labor supply in this context.

Unlike studies of pandemic school closures in high-income countries, impacts are not concentrated primarily among women. The role of school-age children as both recipients of childcare as well as providers of household childcare and agricultural labor creates offsetting effects: women benefit relatively more from the reduction in the net childcare burden when children return to school, while men pick up a larger share of the lost child agricultural labor. Some studies of changes in childcare availability or cost in Africa similarly report significant impacts for men as well as women (Bjorvatn et al. 2021; Martinez, Naudeau, and Pereira 2012), but most focus exclusively on women, and none consider the role of children's household labor. Considering how childcare burdens are allocated across *all* household members is critical for understanding the intra-household distributional impacts of childcare shocks in this setting.

Our study generates three main policy-relevant takeaways. First, parents in Kenya appear to have limited options for dealing with increased childcare burdens beyond reducing work hours or combining work and childcare. This is despite many households having additional adults, many parents being engaged in potentially more flexible household farm work, and adults working just 24 hours a week on average (less than 'full-time') before the pandemic.

Second, both cost and availability of childcare may be important constraints. Older siblings are an important source of (unpaid) childcare, and the results suggest that the partial reopening increases childcare burdens for parents with younger children when sibling caregivers return to school. This indicates that households lack alternative childcare options or that they cost more than adults could earn by working instead of caring for children themselves. RRPS households report almost no childcare provision by non-household members, and several studies point to high costs as a main constraint to using formal childcare centres in Kenya and advocate for public subsidies to facilitate access (Clark et al. 2021; Murungi 2013). Policies aiming to increase childcare availability may therefore be less effective if they are not complemented by policies to reduce cost.

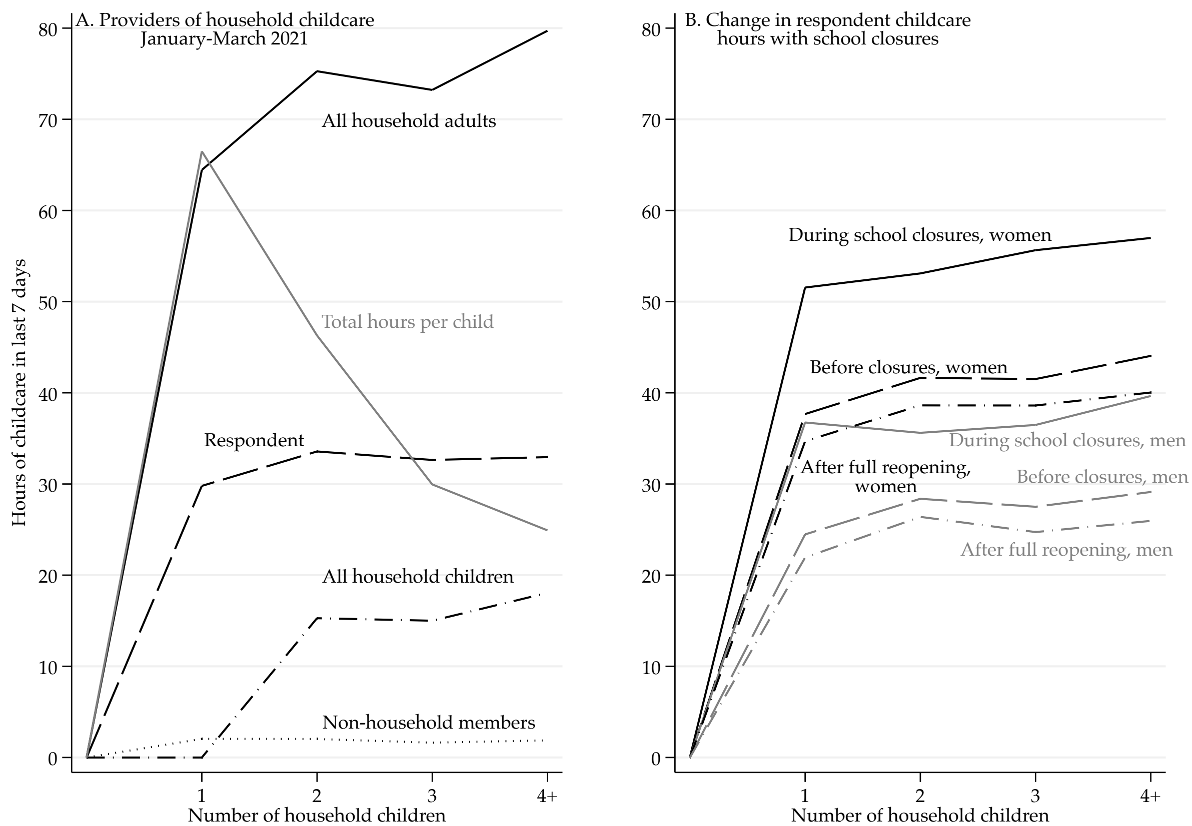
Third, the timing of when children are in school affects some households through child agricultural labor. The 2020 school closures disrupted academic calendars with implications for the timing of school terms and breaks relative to the agricultural cycle over 2021-2023 and thereafter. This will affect whether children are in school during labor-intensive agricultural periods in Kenya. Given the important role of children in agricultural production for many households, future work could consider how these changes affect children's school attendance and household production decisions.

The results also have relevance to other policies affecting household childcare. If we expect that childcare needs decrease with child age, we would expect the estimated impacts on labor supply to be lower bounds on the impact of policies giving households access to free full-day childcare for young children during the working week (as schools implicitly provide to students). Clark et al. (2019) show that subsidies for childcare centres increase labor supply for women in an informal settlement in Nairobi. Our results indicate such policies could have positive effects outside urban settings and also for men. Women could particularly benefit from policies that allow them to maintain labor market attachment while children are very young. Older children might also benefit from reduced need to care for younger siblings.

Finally, the large magnitude of the labor supply effects of the partial school reopening highlights another point that is often underappreciated: universal primary school (now close to reality across the globe), pre-school, and other forms of childcare may play a substantial role in increasing adult labor supply and promoting economic growth, and are a key component of development.

3.7 Figures and Tables

Figure 3.1: Count of children and childcare hours in the last 7 days, by provider of care and school closure status

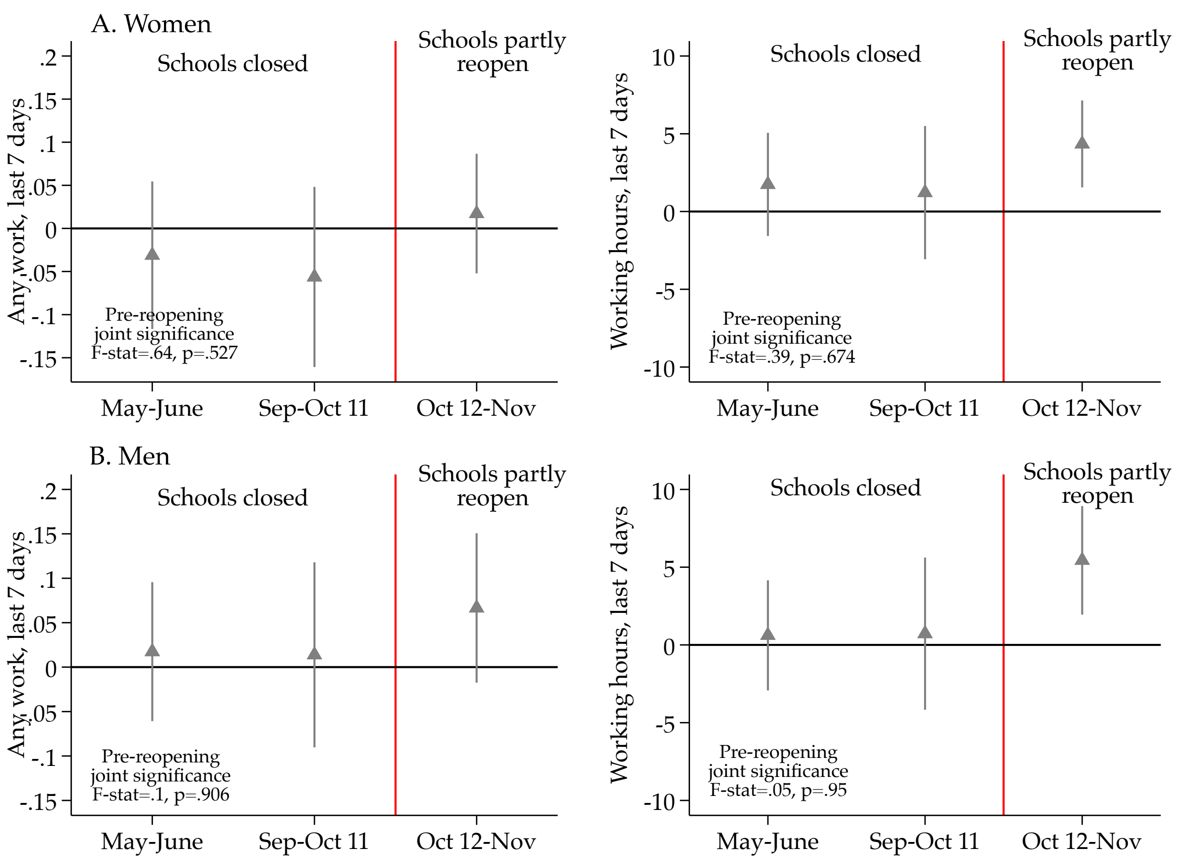


The figures show mean childcare hours in the last 7 days by number of household children (ages 0-17).

Panel A presents data from RRPS round 4 (January-March 2021) which asks about childcare hours for each household adult, for all children in total, and for all non-household members in total. Previous rounds only ask about childcare hours for the respondent. The hours for 'all household adults' include the respondent's hours. Total hours per child is the sum of all childcare hours divided by the number of children.

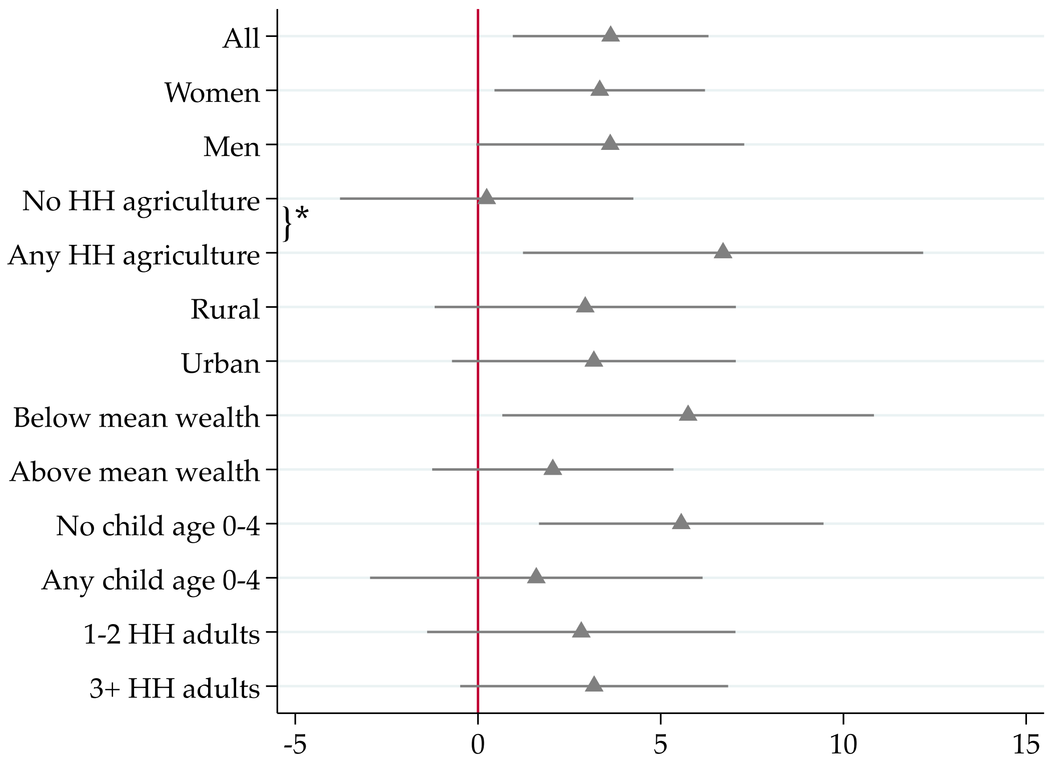
Panel B presents data for female (black) and male (gray) *respondent* childcare hours before school closures (dashed lines, recall data from RRPS round 4 about January-early March 2020), during school closures (solid lines, RRPS rounds 1-3 covering May-November 2020), and after schools fully reopened (dash-dot lines, RRPS round 4 covering January-March 2021). Data on childcare hours before and during the school closures period for other care providers are not available.

Figure 3.2: Impact of treatment on labor participation in the last 7 days, by time period



The figures show estimated coefficients and 95% confidence intervals for the interaction between *Treat* and time period from Equation 3.1 for treatment households, where *Post* is replaced with time period dummies, separately for women (panel A) and men (panel B). Outcomes are any work engagement (left) and total work hours (right) in the 7 days prior to the interview. Treatment households have a child enrolled in grades 4 or 8, and control households have a child enrolled in grades 3, 5, 6, 7, or 9. We do not show coefficients for mixed households with children in both grade groups. The reference period is July-August, while schools were closed and before the partial reopening was announced. The red bars indicate the timing of Kenya's partial school reopening.

Figure 3.3: Heterogeneity in impacts of partial school reopening on adult work hours



The figure summarizes estimated coefficients and 95% confidence intervals for the effect of $Post * Treat$ from Equation 3.1 for sub-samples with specified characteristics. Only coefficients for treatment households are shown. The outcome is total work hours in the 7 days prior to the interview. Results are reported in Table C.1.4. Brackets indicate significant differences between pairs of characteristics. Data include observations from May-November 2020. Household characteristics are from the first time they are observed. Wealth is measured by an index based on housing and asset ownership.

Table 3.1: Impacts of partial school reopening on adult labor supply

	N	Control Mean (SD)	Post (SE)	Post x Treat (SE)	Post x Mixed (SE)
Engaged in any work in last 7 days	8538	0.587	-0.003	0.041	0.030
		(0.492)	(0.031)	(0.026)	(0.026)
Engaged in wage employment in last 7 days	8538	0.062	-0.006	0.011	-0.006
		(0.241)	(0.018)	(0.013)	(0.013)
Engaged in HH agriculture in last 7 days	8538	0.510	0.015	0.037	0.006
		(0.500)	(0.027)	(0.027)	(0.023)
Engaged in HH non-ag enterprise in last 7 days	8538	0.072	-0.007	0.015	0.019
		(0.259)	(0.019)	(0.016)	(0.015)
Total work hours, last 7 days	8538	16.434	0.074	3.630***	-0.569
		(20.027)	(1.806)	(1.365)	(1.413)
Wage hours, last 7 days	8538	1.986	0.053	0.395	-0.367
		(9.374)	(0.731)	(0.565)	(0.582)
Ag hours, last 7 days	8538	11.895	0.575	3.090***	-0.564
		(15.403)	(1.362)	(1.091)	(1.125)
Enterprise hours, last 7 days	8538	2.434	-0.178	0.281	0.036
		(10.023)	(0.681)	(0.625)	(0.630)

This table presents estimates of Equation 3.1 for individual labor supply. Individuals not working in a given sector are coded as working 0 hours. From left to right, the columns show the dependent variable, number of observations, the control mean prior to the partial reopening, and the impacts of being in the partial reopening period for control households (Post), treatment households (Post x Treat), and mixed households (Post x Mixed). Control households have a child in grades 3, 5, 6, 7, or 9, treatment households have a child in grades 4 or 8, and mixed households have both. 'Post' is a dummy for being observed on or after the partial school reopening on October 12. Regressions include household and county by month fixed effects and controls for household and individual characteristics. Standard errors are clustered at the household level. Data include observations for adults age 18-64 from May to November 2020. Significant treatment impacts on total and agricultural work hours are robust to multiple testing adjustment using FDR q-values. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2: Impacts of partial school reopening on respondent childcare hours and child agricultural labor

	Respondent Childcare Hours						Child Ag. Hours	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	0.253	6.414	-6.786	-10.657	1.995	5.817	0.829	0.744
	(6.443)	(9.764)	(9.798)	(13.736)	(12.157)	(17.687)	(1.084)	(2.862)
Post × Treat	0.461	-0.575	-5.972	-11.648	-3.170	-6.997	-1.480**	-2.635
	(5.032)	(6.922)	(6.960)	(10.601)	(9.311)	(16.262)	(0.739)	(1.758)
Post × Any child age 0-4			17.426	49.104**				
			(13.770)	(23.043)				
Post × Treat × Any child age 0-4			8.394	39.520**				
			(11.204)	(19.482)				
Post × Any child age 0-8					0.670	5.050		-0.636
					(14.686)	(23.237)		(3.125)
Post × Treat × Any child age 0-8					5.729	9.185		1.715
					(11.563)	(18.999)		(1.992)
Observations	3073	1722	3073	1722	3073	1722	3077	3077
Control Mean	52.743	59.905	52.743	59.905	52.743	59.905	3.848	3.848
Adult Sex	Both	Women	Both	Women	Both	Women	-	-

This table presents estimates of Equation 3.1 for respondent childcare hours (columns 1-6) and total household child agriculture hours (columns 7-8). Dependent variables are defined over the last 7 days. Childcare hours are not measured for household adults besides the respondent in these survey rounds. Households not engaged in agriculture are coded as having 0 child agriculture hours. Observations include data from May to November 2020, and include treatment households with children in grades 4 or 8 (indicated by 'Treat'), control households with children in an adjacent grade, and 'mixed' households with both (results not shown). 'Post' is a dummy for being observed on or after the partial school reopening on October 12. Regressions include household and county by month fixed effects, and additional household and individual controls. SEs clustered at household level.

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

Bibliography

Abdulkadiroğlu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. 2017. “Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation.” *Econometrica* 85 (5): 1373–1432.

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. “Shift-Share Designs: Theory and Inference*.” *The Quarterly Journal of Economics* 134, no. 4 (August): 1949–2010.

Ahani, Narges, Tommy Andersson, Alessandro Martinello, Alexander Teytelboym, and Andrew C. Trapp. 2021. “Placement Optimization in Refugee Resettlement.” *Operations Research* 69 (5): 1468–1486.

Allport, Gordon W. 1954. *The Nature of Prejudice.* Cambridge, Mass.: Addison-Wesley Pub. Co.

Alon, Titan, Sena Coskun, Matthias Doepke, David Koll, and Michele Tertilt. 2021. *From Mancession to Shecession: Women's Employment in Regular and Pandemic Recessions.* National Bureau of Economic Research WP 28632.

Amuedo-Dorantes, Catalina, Miriam Marcén, Marina Morales, and Almudena Sevilla. 2020. *COVID-19 School Closures and Parental Labor Supply in the United States.* IZA Discussion Papers No. 13827.

Andrew, Alison, Sarah Cattan, Monica Costa Dias, Christine Farquharson, Lucy Kraftman, Sonya Krutikova, Angus Phimister, and Almudena Sevilla. 2020. *How are mothers and fathers balancing work and family under lockdown?* Institute for Fiscal Studies Briefing Note 290.

Angelucci, Manuela, and Giacomo De Giorgi. 2009. “Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption?” *Am. Econ. Rev.* 99 (1): 486–508.

Auer, Daniel. 2018. “Language roulette–the effect of random placement on refugees' labour market integration.” *Journal of Ethnic and Migration Studies* 44 (3): 341–362.

Auer, Daniel, and Johannes S Kunz. 2021. *Communication Barriers and Infant Health: Intergenerational Effects of Randomly Allocating Refugees Across Language Regions.* Technical report. Monash University, SoDa Laboratories.

Auerbach, Alan J, Yuriy Gorodnichenko, and Daniel Murphy. 2020. “Local Fiscal Multipliers and Fiscal Spillovers in the United States.” *IMF Economic Review,* 195–229.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. 2016. “Worms at Work: Long-run Impacts of a Child Health Investment.” *QJE* 131, no. 4 (July): 1637–80.

Baird, Sarah, Craig McIntosh, and Berk Ozler. 2011. “Cash or Condition? Evidence from a Cash Transfer Experiment.” *QJE* 126 (4): 1709–53.

Banerjee, Abhijit, Rema Hanna, Gabriel E. Kreindler, and Benjamin A. Olken. 2017. "Debunking the Stereotype of the Lazy Welfare Recipient: Evidence from Cash Transfer Programs." *World Bank Research Observer* 32 (2): 155–184.

Bansak, Kirk, Jeremy Ferwerda, Jens Hainmueller, Andrea Dillon, Dominik Hangartner, Duncan Lawrence, and Jeremy Weinstein. 2018. "Improving refugee integration through data-driven algorithmic assignment." *Science* 359 (6373): 325–329.

Bardhan, Pranab. 2005. "Theory or Empirics in Development Economics." *Economic and Political Weekly* 40 (40).

Bartel, Ann P. 1989. "Where Do the New U.S. Immigrants Live?" *Journal of Labor Economics* 7 (4): 371–91.

Barwick, Panle Jia, Yanyan Liu, Eleonora Patacchini, and Qi Wu. 2019. *Information, Mobile Communication, and Referral Effects.* Working Paper, Working Paper Series 25873. National Bureau of Economic Research, May. https://doi.org/10.3386/w25873.

Bassi, Vittorio, Raffaela Muoio, Tommaso Porzio, Ritwika Sen, and Esau Tugume. 2019. *Achieving Scale Collectively.* Technical report.

Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, and T. Schmidt with L. Pellerano. 2016. *Cash transfers: what does the evidence say?* Technical report. Overseas Development Institute.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy* 116 (6): 1150–1196.

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg, and Maisy Wong. 2019. "Unity in Diversity? How Intergroup Contact Can Foster Nation Building." *American Economic Review* 109, no. 11 (November): 3978–4025.

Beaman, Lori A. 2011. "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S." *The Review of Economic Studies* 79, no. 1 (August): 128–161.

Biavaschi, Costanza, Corrado Giulietti, and Yves Zenou. 2021. *Social Networks and (Political) Assimilation in the Age of Mass Migration.* Unpublished Manuscript.

Bjorvatn, Kjetil, Denise Ferris, Selim Gulesci, Arne Nasgowitz, Vincent Somville, and Lore Vandewalle. 2021. "Childcare and cash grants for labor supply and wellbeing: experimental evidence from Uganda." Presented at the NBER Economics of Caregiving Conference.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent dirichlet allocation." *the Journal of machine Learning research* 3:993–1022.

Blumenstock, Joshua, Guanghua Chi, and Xu Tan. 2021. *Migration and the Value of Social Networks.* Unpublished Manuscript.

Bonds, Stephanie. 2021. *Information, student-parent communication, and secondary school choice: Experimental evidence from Kenya.* Working Paper, University of California at Berkeley.

Borjas, George J. 2003. "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market." *The Quarterly Journal of Economics* 118 (4): 1335–1374.

Borusyak, Kirill, and Peter Hull. 2020. *Non-Random Exposure to Exogenous Shocks: Theory and Applications.* Working Paper, Working Paper Series 27845. National Bureau of Economic Research, September. https://doi.org/10.3386/w27845.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2021. "Quasi-Experimental Shift-Share Research Designs." Rdab030, *The Review of Economic Studies* (June).

Boserup, Ester, Su Fei Tan, and Camilla Toulmin. 2013. *Woman's role in economic development.* Routledge.

Bouguen, A., Y. Huang, M. Kremer, and E. Miguel. 2019. "Using RCTs to Estimate Long-Run Impacts in Development Economics." *Annual Rev. of Econ.* 11:523–61.

Breiman, Leo. 2001. "Random forests." *Machine learning* 45 (1): 5–32.

Brell, Courtney, Christian Dustmann, and Ian Preston. 2020. "The Labor Market Integration of Refugee Migrants in High-Income Countries." *Journal of Economic Perspectives* 34, no. 1 (February): 94–121.

Broda, Christian, and Jonathan A. Parker. 2014. "The Economic Stimulus Payments of 2008 and the Aggregate Demand for Consumption." *J. of Monetary Econ.* 68 (Fall): S20–36.

Browning, Martin. 1992. "Children and household economic behavior." *Journal of Economic Literature* 30 (3): 1434–1475.

Burke, Marshall, Lauren Falcao Bergquist, and E. Miguel. 2019. "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets." *QJE* 134, no. 2 (December): 785–842.

Caldwell, Sydnee, and Nikolaj Harmon. 2019. *Outside Options, Bargaining, and Wages: Evidence from Coworker Networks.* Unpublished Manuscript.

Calvó-Armengol, Antoni, and Matthew O. Jackson. 2004. "The Effects of Social Networks on Employment and Inequality." *American Economic Review* 94, no. 3 (June): 426–454.

Cameron, A.C., J.B. Gelbach, and D.L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Econ. and Stat.* 90 (3): 414–27.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *ILR Review* 43 (2): 245–257.

———. 2009. "Immigration and Inequality." *American Economic Review* 99, no. 2 (May): 1–21.

Casale, Daniela, and Dorrit Posel. 2020. "Gender and the early effects of the COVID-19 crisis in the paid and unpaid economies in South Africa." *NIDS-CRAM Policy Paper* 18.

Chandrasekhar, Arun G., Melanie Morten, and Alessandra Peter. 2020. *Network-based Hiring: Local Benefits; Global Costs.* Unpublished Manuscript.

Chauhan, Priyanshi. 2020. "Gendering COVID-19: Impact of the Pandemic on Women's Burden of Unpaid Work in India." *Gender Issues,* 1–25.

Chodorow-Reich, Gabriel. 2019. "Geographic Cross-Sectional Fiscal Spending Multipliers: What Have We Learned?" *AEJ: Economic Policy* 11, no. 2 (May): 1–34.

Clark, Shelley, Cassandra Cotton, and Leticia J Marteleto. 2015. "Family ties and young fathers' engagement in Cape Town, South Africa." *Journal of Marriage and Family* 77 (2): 575–589.

Clark, Shelley, Midanna De Almada, Caroline W Kabiru, Stella Muthuri, and Milka Wanjohi. 2021. "Balancing paid work and child care in a slum of Nairobi, Kenya: the case for centre-based child care." *Journal of Family Studies* 27 (1): 93–111.

Clark, Shelley, Caroline W Kabiru, Sonia Laszlo, and Stella Muthuri. 2019. "The impact of childcare on poor urban women's economic empowerment in Africa." *Demography* 56 (4): 1247–1272.

Collins, Caitlyn, Liana Christin Landivar, Leah Ruppanner, and William J Scarborough. 2021. "COVID-19 and the gender gap in work hours." *Gender, Work & Organization* 28:101–112.

Conley, Timothy G. 2008. "Spatial Econometrics." In *The New Palgrave Dictionary of Economics,* Second Edition, edited by Steven N. Durlauf and Lawrence E. Blume, 7:741–47. Houndsmills: Palgrave Macmillan.

Connelly, Rachel. 1992. "The effect of child care costs on married women's labor force participation." *The review of Economics and Statistics,* 83–90.

Corbi, Raphael, Elias Papaioannou, and Paolo Surico. 2019. "Regional Transfer Multipliers." *Review of Economic Studies* 86:1901–1934.

Cunha, Jesse M, Giacomo De Giorgi, and Seema Jayachandran. 2018. "The Price Effects of Cash Versus In-Kind Transfers." *Review of Economic Studies* 86, no. 1 (April): 240–81.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica* 77 (4): 1003–1045.

Damm, Anna. 2009. "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence." *Journal of Labor Economics* 27 (2): 281–314.

Danzer, Alexander M., and Firat Yaman. 2013. "Do Ethnic Enclaves Impede Immigrants' Integration? Evidence from a Quasi-experimental Social-interaction Approach." *Review of International Economics* 21 (2): 311–325.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48, no. 2 (June): 424–55.

Deaton, Angus S. 2018. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* New York: World Bank Group.

Del Boca, Daniela, Noemi Oggero, Paola Profeta, and Mariacristina Rossi. 2020. "Women's and men's work, housework and childcare, before and during COVID-19." *Review of Economics of the Household* 18 (4): 1001–1017.

Delecourt, Solène, and Anne Fitzpatrick. 2021. "Childcare Matters: Female Business Owners and the Baby-Profit Gap." *Management Science.*

Deshpande, Ashwini. 2020. *The Covid-19 Pandemic and Lockdown: First Order Effects on Gender Gaps in Employment and Domestic Time Use in India.* GLO Discussion Paper No. 607.

Division, United Nations Population. 2020. *Database on Household Size and Composition 2019.* United Nations.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A 'Market Access' Approach." *QJE* 131 (2): 799–858.

Dustmann, Christian, and Albrecht Glitz. 2015. "How Do Industries and Firms Respond to Changes in Local Labor Supply?" *Journal of Labor Economics* 33 (3): 711–750.

Dustmann, Christian, Albrecht Glitz, Uta Schönberg, and Herbert Brücker. 2015. "Referral-based Job Search Networks." *The Review of Economic Studies* 83, no. 2 (October): 514–546.

Dustmann, Christian, Uta Schönberg, and Jan Stuhler. 2016. "The Impact of Immigration: Why Do Studies Reach Such Different Results?" *Journal of Economic Perspectives* 30, no. 4 (November): 31–56.

Dustmann, Christian, Kristine Vasiljeva, and Anna Piil Damm. 2018. "Refugee Migration and Electoral Outcomes." *The Review of Economic Studies* 86, no. 5 (September): 2035–2091.

Easterly, William. 2006. *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done so Much Ill and so Little Good.* Penguin Books.

Edin, Per-Anders, Peter Fredriksson, and Olof Åslund. 2003. "Ethnic Enclaves and the Economic Success of Immigrants—Evidence from a Natural Experiment." *The Quarterly Journal of Economics* 118 (1): 329–357.

Evans, David K., and Anna Popova. 2017. "Cash Transfers and Temptation Goods." *Economic Development and Cultural Change* 65 (2): 189–221.

FAA-142.31. 1998. *Federal Act on Asylum.*

Farhi, Emmanuel, and Ivan Werning. 2016. "Fiscal Multipliers: Liquidity Traps and Currency Unions." In *Handbook of Macroeconomics,* edited by J.B. Taylor and H. Uhlig, 2:2417–92. Amsterdam: Elsevier. https://doi.org/10.1016/bs.hesmac.2016.06.006.

Farré, Lídia, Yarine Fawaz, Libertad González, and Jennifer Graves. 2020. *How the COVID-19 lockdown affected gender inequality in paid and unpaid work in Spain.* IZA Discussion Paper No. 13434.

Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth* 8 (2): 195–222.

Filmer, Deon, Jed Friedman, Eeshani Kandpal, and Junko Onishi. 2018. "Cash Transfers, Food Prices, and Nutrition Impacts on Nonbeneficiary Children." March.

Fisher, R. A. 1936. "Design of Experiments." *BMJ* 1 (3923): 554–554.

Foged, Mette, and Giovanni Peri. 2016. "Immigrants' Effect on Native Workers: New Analysis on Longitudinal Data." *American Economic Journal: Applied Economics* 8, no. 2 (April): 1–34.

Foster, Vivien, and Cecilia Briceno-Garmendia. 2010. *Africa's Infrastructure: A Time for Transformation.* Washington DC: World Bank Publications.

Furman, Jason, Melissa Schettini Kearney, and Wilson Powell. 2021. *The Role of Childcare Challenges in the US Jobs Market Recovery During the COVID-19 Pandemic.* National Bureau of Economic Research WP 28934.

Gee, Laura K., Jason Jones, and Moira Burke. 2017. "Social Networks and Labor Markets: How Strong Ties Relate to Job Finding on Facebook's Social Network." *Journal of Labor Economics* 35 (2): 485–518.

Giulietti, Corrado, Jackline Wahba, and Yves Zenou. 2018. "Strong versus weak ties in migration." *European Economic Review* 104:111–137.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360–1380.

Grantham, Kate, Leva Rouhani, Neelanjan Gupta, Martha Melesse, Diva Dhar, Soumya K Mehta, and Kanika J Kingra. 2021. *Evidence review of the global childcare crisis and the road for post-Covid-19 recovery and resilience.* Technical report. International Development Research Centre.

Halim, Daniel, Elizaveta Perova, and Sarah Reynolds. 2021. *Childcare and Mothers' Labor Market Outcomes in Lower-and Middle-Income Countries.*

Hansen, Benjamin, Joseph J Sabia, and Jessamyn Schaller. 2022. *Schools, Job Flexibility, and Married Women's Labor Supply: Evidence from the COVID-19 Pandemic.* National Bureau of Economic Research WP 29660.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *QJE* 131 (4): 1973–2042.

Heath, Rachel. 2017. "Fertility at work: Children and women's labor market outcomes in urban Ghana." *Journal of Development Economics* 126:190–214.

Heggeness, Misty L. 2020. "Estimating the immediate impact of the COVID-19 shock on parental attachment to the labor market and the double bind of mothers." *Review of Economics of the Household* 18 (4): 1053–1078.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *QJE* 124 (4): 1403–1448.

İlkkaracan, İpek, and Emel Memiş. 2021. "Transformations in the Gender Gaps in Paid and Unpaid Work During the COVID-19 Pandemic: Findings from Turkey." *Feminist Economics* 27 (1-2): 288–309.

ILO. 2017. *World employment social outlook: Trends for women.* Technical report. International Labour Organization.

———. 2021. *ILOSTAT database.* International Labour Organization.

Jakiela, Pamela, Owen Ozier, Lia Fernald, and Heather Knauer. 2020. *Big Sisters.* World Bank WP 9454.

Kah, Henry Kam. 2012. "Husbands in wives' shoes: Changing social roles in child care among Cameroon's urban residents." *Africa Development* 37 (3): 101–114.

Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money.* London: Macmillan.

Kielland, Anne, and Maurizia C Tovo. 2006. *Children at work: Child labor practices in Africa.* Lynne Rienner Publishers Boulder, CO.

Kraay, Aart. 2014. "Government Spending Multipliers in Developing Countries: Evidence from Lending by Official Creditors." *AEJ: Macroeconomics* 6 (4): 170–208.

Kramarz, Francis, and Oskar Nordström Skans. 2014. "When Strong Ties are Strong: Networks and Youth Labour Market Entry." *The Review of Economic Studies* 81, no. 3 (January): 1164–1200.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo. 2013. "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment*." *The Quarterly Journal of Economics* 128, no. 3 (June): 1123–1167.

Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99 (3): 483–499.

Lazear, Edward P. 1999. "Culture and Language." *Journal of Political Economy* 107 (S6): S95–S126.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labour." *The Manchester School* 22 (2): 139–191.

Lokshin, Michael M, Elena Glinskaya, and Marito Garcia. 2000. *The effect of early childhood development programs on women's labor force participation and older children's schooling in Kenya.* The World Bank.

Lowe, Matt. 2021. "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." *American Economic Review* 111, no. 6 (June): 1807–44.

Ma, Sen, Zhengyun Sun, and Hao Xue. 2020. *Childcare Needs and Parents' Labor Supply: Evidence from the COVID-19 Lockdown.* Working Paper, Harvard University.

MacKinnon, James G., and Matthew D. Webb. 2018. "The wild bootstrap for few (treated) clusters." *Econometrics Journal* 21 (2): 114–35.

Mankiw, N. Gregory, and Matthew Weinzierl. 2011. "An Exploration of Optimal Stabilization Policy." *Brookings Papers on Economic Activity* 42 (1): 209–272.

Manning, A., and Michael Amior. 2021. *Monopsony and the Wage Effects of Migration.* CEP Working Paper 1690.

Marshall, Alfred. 1890. *The Principles of Economics.* McMaster University.

Martén, Linna, Jens Hainmueller, and Dominik Hangartner. 2019. "Ethnic networks can foster the economic integration of refugees." *Proceedings of the National Academy of Sciences* 116 (33): 16280–16285.

Martinez, Sebastian, Sophie Naudeau, and Vitor Pereira. 2012. *The promise of preschool in Africa: A randomized impact evaluation of early childhood development in rural Mozambique.* Technical report.

Mian, Atif, Ludwig Straub, and Amir Sufi. Forthcoming. "Indebted Demand." *QJE.*

Michaillat, Pascal, and Emmanuel Saez. 2015. "Aggregate Demand, Idle Time, and Unemployment." *QJE* 130 (2): 507–569.

Miguel, Edward, and Michael Kremer. 2004. "Worms: identifying impacts on education and health in the presence of treatment externalities." *Econometrica* 72 (1): 159–217.

Mitaritonna, Cristina, Gianluca Orefice, and Giovanni Peri. 2016. *Immigrants and Firms' Outcomes: Evidence from France.* Working Paper, Working Paper Series 22852. National Bureau of Economic Research, November. https://doi.org/10.3386/w22852.

Mousa, Salma. 2020. "Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq." *Science* 369 (6505): 866–870.

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U. S. Labor Market*." *The Quarterly Journal of Economics* 118, no. 2 (May): 549–599.

Muralidharan, Karthik, and Paul Niehaus. 2017. "Experimentation at Scale." *Journal of Econ. Perspectives* 31, no. 4 (November): 103–24.

Murphy, Daniel. 2017. "Excess capacity in a fixed-cost economy." *European Economic Review* 91 (C): 245–60.

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny. 1989. "Industrialization and the Big Push." *Journal of Political Economy* 97 (5): 1003–1026.

Murungi, Catherine Gakii. 2013. "Reasons for low enrolments in early childhood education in Kenya: The parental perspective." *International Journal of Education and Research* 1 (5): 1–10.

Musterd, Sako. 2005. "Social and Ethnic Segregation in Europe: Levels, Causes, and Effects." *Journal of Urban Affairs* 27 (3): 331–348.

Nakamura, Emi, and Jón Steinsson. 2014. "Fiscal Stimulus in a Monetary Union: Evidence from US Regions." *American Economic Review* 104, no. 3 (March): 753–92.

Ottaviano, Gianmarco I. P., and Giovanni Peri. 2012. "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association* 10, no. 1 (February): 152–197.

Pallais, Amanda. 2014. "Inefficient Hiring in Entry-Level Labor Markets." Reprinted in Learning in Labor Markets (2017) edited by Michael Waldman. *American Economic Review* 104 (11): 3565–3599.

Pallais, Amanda, and Emily Glassberg Sands. 2016. "Why the Referential Treatment? Evidence from Field Experiments on Referrals." *Journal of Political Economy* 124 (6): 1793–1828.

Pape, Utz Johann. 2021. *Kenya COVID-19 Rapid Response Phone Survey Households 2020-2021, Panel.* The World Bank, November.

Pape, Utz Johann, Javier Baraibar Molina, Antonia Johanna Sophie Delius, Caleb Leseine Gitau, and Laura Abril Rios Rivera. 2021. *Socio-Economic Impacts of COVID-19 in Kenya on Households: Rapid Response Phone Survey Round 1.* The World Bank, January.

Parker, J.A., N.S. Souleles, D.S. Johnson, and R. McClelland. 2013. "Consumer Spending and the Economic Stimulus Payments of 2008." *AER* 103 (6): 2530–53.

Pennings, Steven. 2021. "Cross-Region Transfer Multipliers in a Monetary Union: Evidence from Social Security and Stimulus Payments." *AER* 111, no. 5 (May): 1689–1719.

Peri, Giovanni. 2012. "The Effect Of Immigration On Productivity: Evidence From U.S. States." *The Review of Economics and Statistics* 94, no. 1 (February): 348–358.

Prados, Maria J, and Gema Zamarro. 2021. *School re-openings, childcare arrangements, and labor outcomes during Covid-19.* Working Paper, University of Southern California.

Quisumbing, Agnes R, Kelly Hallman, and Marie T Ruel. 2007. "Maquiladoras and market mamas: Women's work and childcare in Guatemala City and Accra." *The Journal of Development Studies* 43 (3): 420–455.

Rachel, Lukasz, and Lawrence H. Summers. 2019. "On falling neutral real rates, fiscal policy, and the risk of secular stagnation." *Brookings Papers on Economic Activity* BPEA Conference Drafts.

Ramey, Valerie A. 2019. "Ten Years after the Financial Crisis: What Have We Learned from the Renaissance in Fiscal Research?" *Journal of Econ. Perspectives* 33, no. 2 (May): 89–114.

Ribar, David C. 1992. "Child care and the labor supply of married women: Reduced form evidence." *Journal of Human Resources,* 134–165.

Rosenstein-Rodan, Paul N. 1943. "Problems of Industrialisation of Eastern and South-eastern Europe." *Economic Journal* 53:202–11.

Sadoulet, E., A. de Janvry, and Benjamin Davis. 2001. "Cash Transfer Programs with Income Multipliers: PROCAMPO in Mexico." *World Development* (6): 1043–56.

Sale, Laura. 2021. *The social determinants of refugee integration and optimal dispersal: a study in Denmark.* Unpublished Manuscript.

Samman, Emma, Elizabeth Presler-Marshall, Nicola Jones, Maria Stavropoulou, and John Wallace. 2016. *Women's work: Mothers, children and the global childcare crisis. Report.* Technical report. Overseas Development Institute (ODI).

Sims, Eric, and Jonathan Wolff. 2018. "The Output and Welfare Effects of Government Spending Shocks Over the Business Cycle." *International Econ. Rev.* 59 (3): 1403–35.

State Secretariat for Migration. 2015. "Handbuch Asyl und Rueckkehr." *Bern: Swiss State Secretary for Migration.*

Suarez Serrato, Juan Carlos, and Philippe Wingender. 2016. *Estimating Local Fiscal Multipliers.* Unpublished.

The Star. 2020. *Short notice: Rush against time for parents and candidates as CS orders phased resumption of classes next week,* October.

Thome, Karen, Mateusz Filipski, Justin Kagin, J. Edward Taylor, and Benjamin Davis. 2013. "Agricultural spillover effects of cash transfers: What does LEWIE have to say?" *American Journal of Agricultural Economics* 95 (5): 1338–1344.

Walker, Michael. 2018. "Informal Taxation Responses to Cash Transfers: Experimental Evidence from Kenya." July.

Wenham, Clare, Julia Smith, Sara E Davies, Huiyun Feng, Karen A Grépin, Sophie Harman, Asha Herten-Crabb, and Rosemary Morgan. 2020. *Women are most affected by pandemics—lessons from past outbreaks.* Nature Publishing Group, Comment.

Witte, Mark. 2021. *Why do Workers Make Job Referrals.* Unpublished Manuscript.

World Bank. 2017. *Closing the Gap: The State of Social Safety Nets.* Washington, D.C., April.

Zamarro, Gema, and María J Prados. 2021. "Gender differences in couples' division of childcare, work and mental health during COVID-19." *Review of Economics of the Household* 19 (1): 11–40.

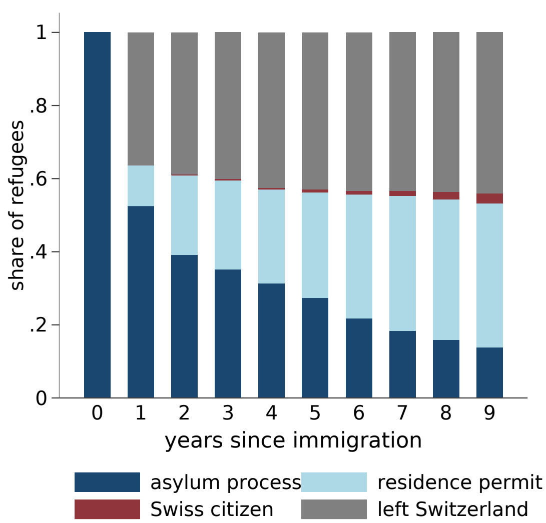
Zuilkowski, Stephanie Simmons, Benjamin Piper, Salome Ong'ele, and Onesmus Kiminza. 2018. "Parents, quality, and school choice: Why parents in Nairobi choose low-cost private schools over public schools in Kenya's free primary education era." *Oxford Review of Education* 44 (2): 258–274.

Appendix A

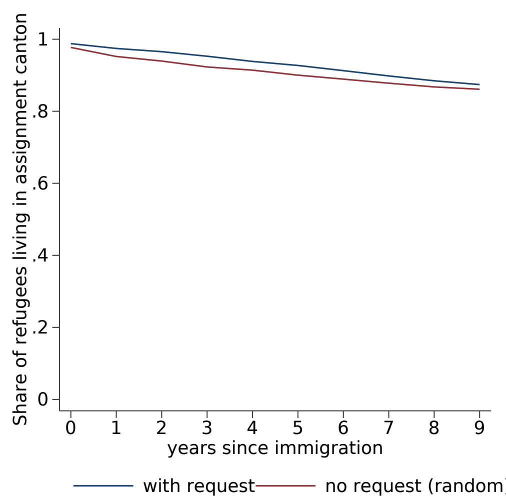
Supplementary Appendix – Chapter 1

A.1 Supporting Figures and Tables

Figure A.1.1: Dynamics of refugee legal status and residence



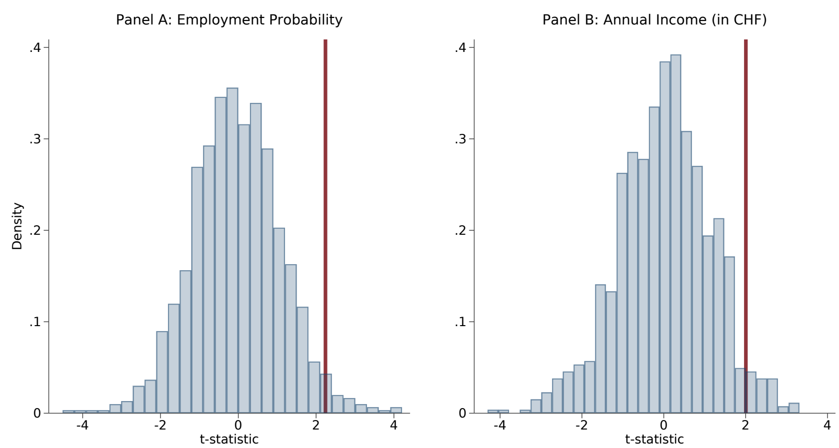
(A) Residence status over time



(B) Residence within assigned canton

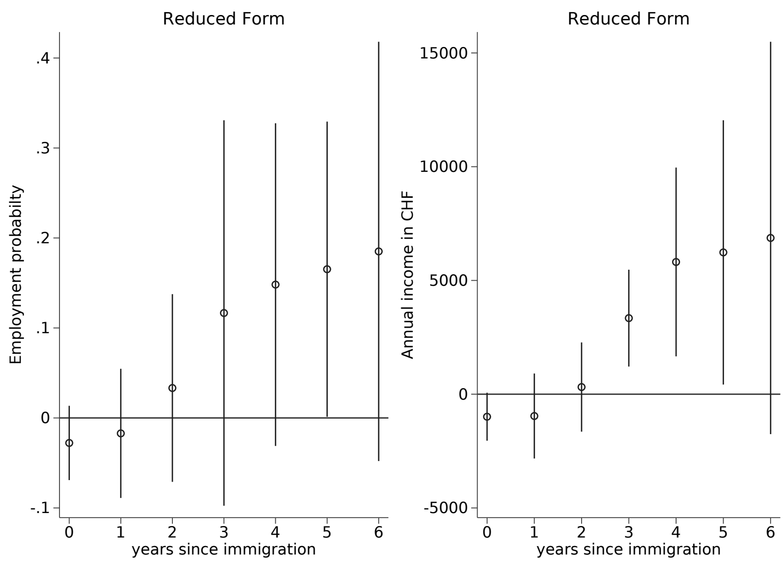
Notes: Panel A shows status of refugees based on their residence permit captured in the full annual census registry data. Panel B plots the probability of each refugee still living in their assigned canton (conditional on being in Switzerland), for refugees with a canton allocation requests and those without (see Section 1.2).

Figure A.1.2: Randomization Inference of Network Effects on Refugees



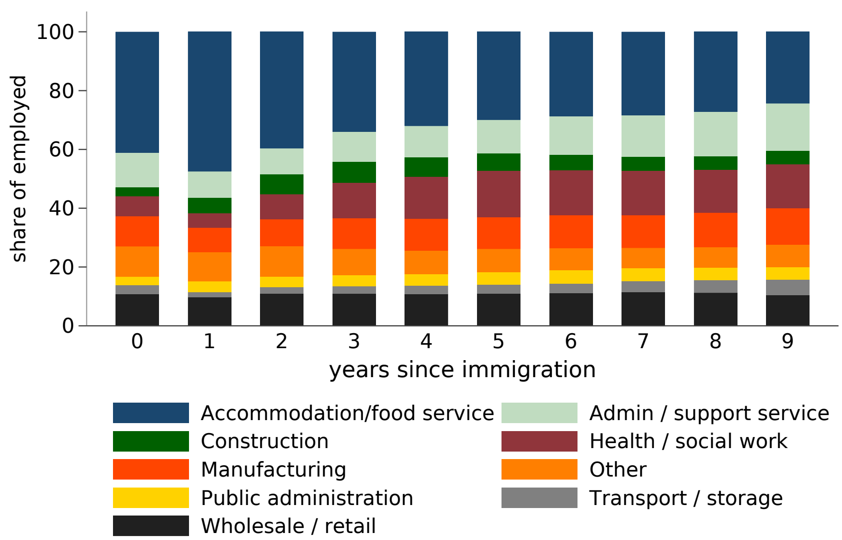
Notes: This figure plots the t-statistics obtained from estimating the reduced form of Equation 1.1 across 999 simulated exogenous refugee assignments. For each assignment, we re-assign refugees, taking exceptions from random allocations, origin-by-reception-center-by-year cohorts as well as reception-center-by-year assignment probabilities as given. Each family of refugees is assigned to the same canton. For each iteration, we also re-calculate the values of the instrument based on the simulated re-shuffling of assignments. The red line indicates the realized t-statistic in our main specification.

Figure A.1.3: Dynamic Impact of Networks on Refugee Outcomes



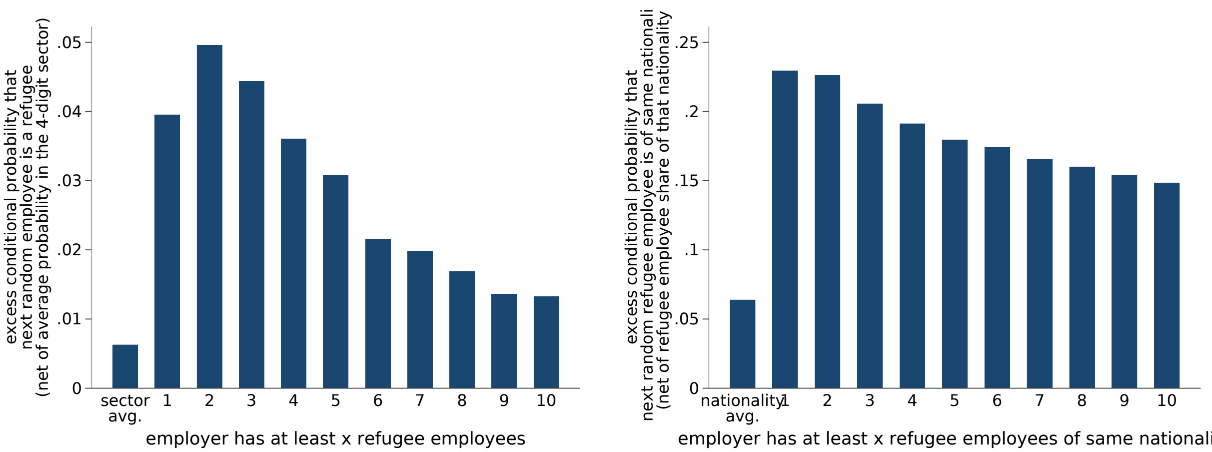
Notes: This figure plots dynamic reduced form impacts of the number of co-national network members in the assigned canton on individual refugees' labor market trajectories. We include all asylum seekers resident in Switzerland, arriving at age 19 to 54 without any cantonal placement requests, and drop any sending nations with less than 20 refugees over our observation period, or where less than 10% of individuals resident in Switzerland arrived trough the asylum system. Employment and labor income come from the central social security registry. Coefficients are estimated using Equation 1.1, where the instrument is interacted with year-since-arrival dummies. The instrument is scaled relative to the lagged population, so that coefficients can be interpreted as effects of a quasi-exogenous doubling of the network size (see Section 1.3). Standard errors are clustered at the nationality-by-assignment-year level.

Figure A.1.4: Refugee employment by sector over time



Notes: This figure plots the employment sector of refugees arriving between 2008 - 2017, conditional on employmnt. Sector information comes from the employer-employee-matched enterprise registry.

Figure A.1.5: Sorting of refugees into firms within sectors



(A) Refugee sorting into firms

(B) Sorting of nationalities within refugees

Notes: Panel A shows the conditional probability that a randomly chosen employee of an employer is a refugee in 2017, net of the average share of refugees employed by employers in the same 4-digit NOGA sector. Each bar conditions on the first x employees being refugees, i.e. the x's bar shows the probability that an employee chosen at random among all $n-x$ remaining employees of the same employer is also a refugee (net of the sectoral employment share). Panel B focuses only on employers hiring any refugees, and plots the probability that a randomly chosen individual among the remaining $n-x$ employees is of the same nationality, conditional on the first x employees sharing a nationality.

Table A.1.1: Balance table for characteristics of main refugee estimation sample

	(1)	(2)	(3)	(4)	(5)	(6)
	Randomly allocated refugees			**Non-randomly allocated refugees**		
	Mean	N	Lag of ArcSinh(Population) in assigned canton	Mean	N	Lag of ArcSinh(Population) in assigned canton
Demographics						
1(female)	.421	5629	0.003	.542	3535	0.039
			(0.013)			(0.025)
Arrival age	30	5629	-0.098	30.6	3535	0.432
			(0.331)			(0.356)
Majority religion in origin country	.671	4680	-0.017	.649	2907	-0.005
			(0.015)			(0.023)
Majority ethnicity in origin country	.564	4680	-0.038	.528	2907	-0.018
			(0.026)			(0.026)
Average education	1.37	297	-0.101	1.38	231	-0.145
			(0.180)			(0.108)
Completed secondary education	.269	297	-0.035	.29	231	-0.136*
			(0.115)			(0.077)
Asylum process						
Asylum process duration (day	620	4515	14.290	548	2748	24.737
			(15.240)			(21.757)
Temporarily accepted	.551	4680	0.002	.481	2907	0.032
			(0.016)			(0.021)
Rejected	.545	4680	-0.008	.452	2907	0.039
			(0.016)			(0.029)
Dublin case	.232	5629	-0.029*	.13	3535	-0.013
			(0.016)			(0.015)
Number of family member applicants	2.35	4680	0.080	3.64	2907	-0.087
			(0.056)			(0.113)
Accompanied by children	.421	4680	0.002	.658	2907	-0.018
			(0.019)			(0.028)
Unaccompanied minor	.00171	4680	0.000	0	2907	0.000
			(0.002)			(0.000)
Allocation requests						
Core family mentioned	.0615	5629	0.003	.574	3535	0.019
			(0.008)			(0.026)
Extended family mentioned	.0483	5629	-0.006	.177	3535	-0.016
			(0.008)			(0.021)
Peers or friends mentioned	.00497	5629	0.002	.028	3535	0.011
			(0.002)			(0.009)
Medical reason mentioned	.0103	5629	0.004	.0331	3535	0.018**
			(0.003)			(0.009)
Childbirth mentioned	.00195	5629	-0.001	.019	3535	-0.007
			(0.002)			(0.008)
Swiss language mentioned	.022	5629	0.006	.0209	3535	-0.008
			(0.005)			(0.007)

Notes: Each row represents a regression of a baseline characteristic of each refugee in our main estimation sample (see Table 1.1) on the number of co-nationals in the assigned canton. Regressions include nationality-by-arrival-year fixed effects and canton-of-assignment fixed effects. (Equation 1.1). Columns 1-3 are all refugees without a cantonal placement request, 4-6 those with such a request (see Section 1.2). Standard errors are clustered at the nationality-by-assignment-year level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.1.2: Balance table for baseline network characteristics

	(1) Mean	(2) Lag 5: ArcSinh(No. of co-nationals) in assigned canton	(3) Instrument: co-nationals assigned in the last 3 years
Demographics			
Lag 5: ArcSinh(No. of co-nationals)			-0.228
			(0.172)
Lag 5: Share of co-nationals - arrived as refugees	.519	-0.023***	-0.005
		(0.008)	(0.031)
Lag 5: Share of co-nationals - female	.398	0.029***	-0.045
		(0.005)	(0.028)
Lag 5: Share of co-nationals - age 18-30	.308	-0.026***	0.047
		(0.007)	(0.038)
Lag 5: Share of co-nationals - age 31-44	.33	-0.004	-0.013
		(0.004)	(0.028)
Lag 5: Share of co-nationals - age 45-62	.124	-0.002	0.014
		(0.003)	(0.014)
Lag 5: Share of co-nationals -arrived less than 3 years ago	.436	-0.025***	0.018
		(0.007)	(0.038)
Economics			
Lag 5: Share of co-nationals - employed	.367	0.017**	-0.051
		(0.008)	(0.036)
Lag 5: Average income of co-nationals	13.6	0.524	-2.045
		(0.360)	(1.423)
Lag 5: Share of co-nationals with secondary eduacation	.306	-0.001	0.043
		(0.006)	(0.043)

Notes: Each row represents a regression with a characteristic of the population / network of residents from a refugee origin country 5 years prior as the dependent variable. The RHS variable is the size of that network 5 years prior in Column 2, and the deviation from expected inflows (i.e. the instrument for the network size) over the past 3 years. We leave 1 year buffer between instrumented inflows and baseline network characteristics to allow for potential correlation (e.g. within families arriving together) of assignment around the end of of a year. Each regression includes a nationality-by-year fixed effect and a canton-fixed effect. We drop any origin nations with less than 20 refugees over our observation period, or where less than 10% of individuals resident in Switzerland arrived trough the asylum system. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.1.3: Balance table for baseline firm characteristics (SSIV)

	(1)	(2)	(3)	(4)
	N	Mean	Δ_6ArcSinh(No. of Conationals)	$\sum_{l=1}^{5}$ network$_{od,t-l}$
ArcSinh(Employment) in 2011	131,224	2.71	-0.037	-0.013^{*}
			(0.029)	(0.007)
Migrant employment share in 2011	131,224	.243	0.023^{***}	0.007
			(0.008)	(0.006)
Share of employees with tertiary education in 2011	93,029	.29	0.063^{***}	-0.004^{**}
			(0.009)	(0.002)
Share of refugee nation employees with tertiary education in 2011	10,713	.0998	0.133^{***}	-0.002
			(0.019)	(0.010)
ArcSinh(Wage bill) in 2011	129,854	13.4	-0.044	-0.016
			(0.037)	(0.014)
ArcSinh(Average monthly wage) in 2011	129,854	9.03	-0.005	-0.007
			(0.015)	(0.007)
Wage premium of Swiss Nationals in 2011	78,845	.267	-0.034	-0.013
			(0.025)	(0.009)

Notes: Each row represents a regression with a baseline characteristic of each firm in our main estimation sample (see Table 1.3) as the dependent variable. In column 3, the RHS variable is 6-year proportional change between 2017 and 2011 (in ArcSinh) of the cantonal population of each refugee sending nation, weighted by the baseline employment share as a share of all refugee nation employment (see Equation 1.3). The RHS variable in column 4 is our instrumental variable, the cumulative proportional deviation (in ArcSinh) from expected inflows of each refugee nation, again aggregated using initial employment share for each firm. Each regression includes a dummy for whether the enterprise had any refugee nation employees in 2011, 2-digit NOGA sector fixed effects as well as canton fixed effects. Refugee sending nations are those with at least 20 refugees over our observation period, and where at least 10% of individuals resident in Switzerland arrived trough the asylum system. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.1.4: Robustness of main network effects to varying included fixed effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Annual income in CHF					
$\sum_{l=1}^{3}$ Inflow of co-nationals$_{od,\underline{t}-l}$	7958.0*	7176.9*	4553.3**	7302.7*	7349.3*	3946.0**
	(4395.8)	(3898.9)	(1964.3)	(4381.3)	(4404.7)	(1967.3)
Observations	8440	8447	8440	8433	8413	8413
Nationality FE	Yes		Yes			
Year FE		Yes	Yes			
Nationality-by-Arrival-Year FE				Yes	Yes	Yes
Nationality-by-Years-in-CH FE					Yes	Yes
Assignment-Canton FE			Yes			Yes
Assignment-Canton-by-Year FE						
Number of origin nations	56	63	56	55	49	49
Number of allocation cantons	26	26	26	26	26	26
Mean of dependent variable	13745.4	13749.7	13745.4	13736.3	13727.3	13727.3
Mean of independent variable	0.00358	0.00348	0.00358	0.00360	0.00350	0.00350

Notes: This table presents estimates of the long-run impacts of the number of co-nationals resident at each refugee's assignment canton in the year prior to assignment (Equation 1.1). We include all asylum seekers resident in Switzerland 5-6 years after arrival, arriving at age 19 to 54 without any cantonal placement requests, and drop any origin nations with less than 20 refugees over our observation period, or where less than 10% of individuals resident in Switzerland arrived trough the asylum system. Standard errors are clustered at the nationality-by-assignment-year level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.1.5: Long-run impacts of Networks on Refugee Residence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	1(Employed in Switzerland)			1(Resident in Switzerland)			1(Resident in assigned canton)		
	(OLS)	(Reduced Form)	(IV)	(OLS)	(Reduced Form)	(IV)	(OLS)	(Reduced Form)	(IV)
ArcSinh(Number of co-nationals)$_{od,\underline{t}-1}$	0.0307***		0.0751*	0.00246		0.0435	0.0369**		-0.0790
	(0.0102)		(0.0442)	(0.00740)		(0.0356)	(0.0168)		(0.0605)
$\sum_{l=1}^{3}$ Inflow of co-nationals$_{od,\underline{t}-l}$		0.0508			0.0294			-0.0622	
		(0.0346)			(0.0240)			(0.0430)	
Observations	18337	18337	18337	18337	18337	18337	8413	8413	8413
Nationality-by-Arrival-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nationality-by-Years-in-CH FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Assignment-Canton FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of origin nations	63	63	63	63	63	63	49	49	49
Number of allocation cantons	26	26	26	26	26	26	26	26	26
Mean of dependent variable	0.247	0.247	0.247	0.460	0.460	0.460	0.877	0.877	0.877
Mean of independent variable	6.105	-0.000726	6.105	6.105	-0.000726	6.105	6.557	0.00350	6.557
90-10 percentile range of independent variable	1.191	0.164		1.191	0.164		0.721	0.122	
First-Stage F-Statistic			19.28			19.28			35.86

Notes: This table presents estimates of the long-run impacts of the number of co-nationals resident at each refugee's assignment canton in the year prior to assignment (Equation 1.1). We include all asylum seekers resident in Switzerland 5-6 years after arrival, arriving at age 19 to 54 without any cantonal placement requests, and drop any origin nations with less than 20 refugees over our observation period, or where less than 10% of individuals resident in Switzerland arrived trough the asylum system. Residence indicators come from the annual census registry data (see Section 3.2), and the indicator for living in the initial assignment canton is conditional on still being in Switzerland. Standard errors are clustered at the nationality-by-assignment-year level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

A.2 Random allocation

Caseworkers

In this section, we provide more detail on the distribution of refugees to different allocation officers (caseworker), as well as the potential discretion of these allocation officers. Between 2011 and 2017, 24 caseworker allocate a total of 179,553 refugees. As discussed in the main text, caseworker never interact with refugees, nor observe any additional information than what is observed in the limited data shared with them, and which we have access to.[1]

Roughly half of all refugees are allocated by a single allocation caseworker (CW1); the second largest share (approx. 33%) by CW2, and 22 other CWs allocated the remaining. This is mostly due to the fact that caseworkers work across different time spans and with differing intensity. Figures and , show that there are no specialisations among caseworkers: all are allocating similar shares of refugees from each reception center and origin country, refugee characteristics are balanced across caseworkers, and all caseworkers achieve a similar distribution across cantons. Moreover, Auer and Kunz 2021 provides additional evidence that allocations to cantons are strongly consistent with the target share that is set nationally based on population size. This is consistent with caseworkers being as good as randomly assigned to refugees.

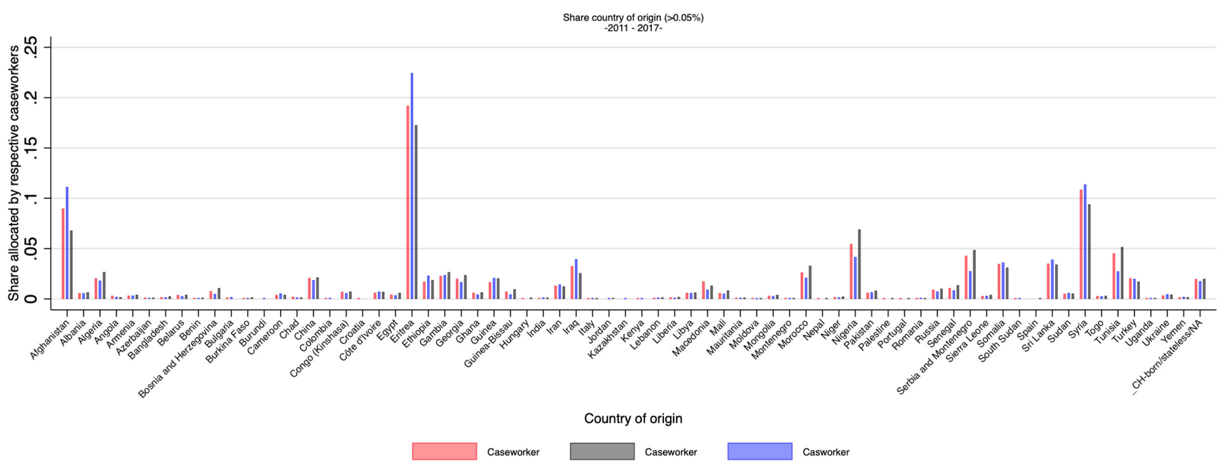


Figure A.2.1: Shares Country of origin within Caseworkers

Note: The Figure displays within caseworker raw share of countries distributed across all years. Two main caseworkers allocate most refugees, remaining caseworkers are subsumed in other caseworker category, the figure does not show which caseworker is which due to potential privacy concerns.
Source: ZEMIS 2011-2017, own calculations.

[1]In the following, we perturbate the allocation officer identifiers in the following and show only shares and conditional regression results to respect their privacy.

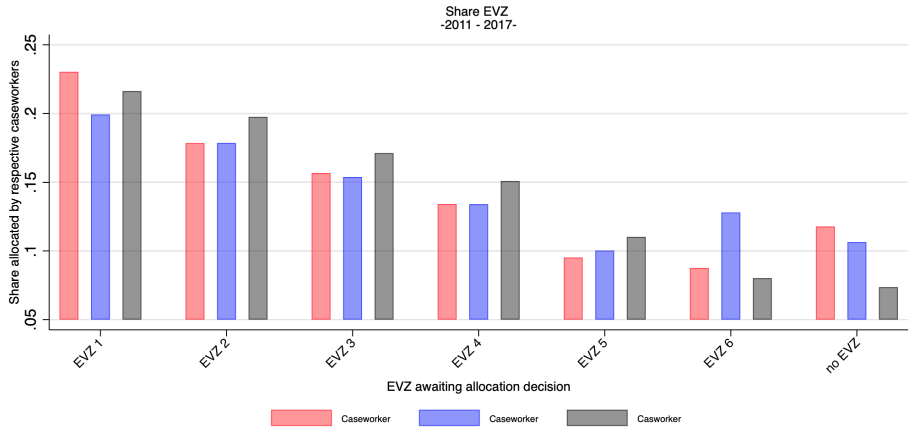


Figure A.2.2: Shares EVZ within Caseworkers

Note: The Figure displays within caseworker raw share of refugees that are located in the 5 largest EVZs, 6 is subsuming all other EVZ (permanent and non-permanent), and a category for not located in any EVZ, distributed across all years. Two main caseworkers allocate most refugees, remaining caseworkers are subsumed in other caseworker category, the figure does not show which caseworker is which due to potential privacy concerns.
Source: ZEMIS 2011-2017, own calculations.

Table A.2.1 further presents descriptive statistics of the caseworkers allocation populations. Overall, they appear to be balanced, with the exception that not all caseworkers are similarly active across years. CW3 is slightly more likely to allocate refugees that arrived later (2014.71 versus 2013.76 for CW1), most likely driven by work schedules. Since we control for year-of-arrival in all our regressions, this should not bias our results. Note that we randomize which coworker is CW1, CW2, and CW3 in each row to protect their identity.

Allocation officers often allocate several hundred refugees a day. The daily average (conditional on observing any allocations on that day) is around 100. The female share is highly balanced, as is the age at arrival of the refugees. There is slightly more variability in the size of the family which is likely due to chance. However, relative to the overall mean, deviations are very small. Corroborating evidence from Figure A.2.2, all officers allocate refugees from different reception centers in similar proportions. Moreover, there is also little variation in the proportion assigned to each canton.

Table A.2.1: Caseworker descriptive statistics

	All (1)	CW1 (2)	CW2 (3)	CW3 (4)
Year (Min)	2011	2011	2011	2011
Year (Max)	2017	2017	2017	2017
Year (Av)	2014.081	2013.764	2013.803	2014.710
Refugee population characteristics				
Av. # allocated per day	107.479	83.204	117.181	107.967
	(0.134)	(0.297)	(0.255)	(0.174)
Share female refugees	0.299	0.300	0.298	0.279
	(0.001)	(0.002)	(0.002)	(0.003)
Age at arrival (in years)	21.919	21.527	23.314	21.978
	(0.031)	(0.056)	(0.084)	(0.045)
Share w. placement request	0.288	0.292	0.258	0.296
	(0.001)	(0.002)	(0.003)	(0.002)
Share w. placement request, based on family	0.179	0.174	0.176	0.182
	(0.001)	(0.002)	(0.002)	(0.001)
Share from main French speaking country	0.266	0.287	0.250	0.269
	(0.001)	(0.003)	(0.002)	(0.001)
Share from official French speaking country	0.064	0.073	0.061	0.063
	(0.001)	(0.002)	(0.001)	(0.001)
Av. arrival family size[a]	2.792	2.677	2.726	3.105
	(0.007)	(0.018)	(0.009)	(0.013)
Largest EVZ	0.181	0.179	0.198	0.178
	(0.001)	(0.002)	(0.003)	(0.001)
2nd largest EVZ	0.158	0.154	0.171	0.157
	(0.001)	(0.002)	(0.002)	(0.001)
Allocation characteristics				
Share sent to Zurich (target 0.17)[a]	0.161	0.162	0.171	0.154
	(0.001)	(0.002)	(0.002)	(0.001)
Log refugee enclave size[c]	5.325	5.191	5.244	5.575
	(0.004)	(0.005)	(0.010)	(0.007)

Note: Mean and standard error in parentheses, based on regressing variable indicated in row on an indicator. [a]pseudo family size based on arrived together, allocated together, same request, etc. [b]same origin, same destination, allocated before index refugee (leaving out family members). [c]excludes 0s and reduces reliance on outliers. *Source*: BFS provided Refugee Registry 2011-2017.

A.3 Variance Decomposition of Destination Effects

In this section, we decompose the variance in long-run labor income of refugees into its various components. We are interested in the share of the overall variance in refugee outcomes that is due to *causal* differences in assigned destinations, and the networks initially located at that destination. For comparison, we also add the contribution to the overall income variance of sex-by-arrival-age fixed effects, since we know these demographics are highly predictive of labor market integration (see Figure 1.4). We take two approaches, one based on adjusted R2, and another based on (bias-corrected) plug-in estimates of variance components.

We start with our main sample of refugees arriving between age 19 and 54 that were exogenously assigned between 2011 and 2017, and whom we observe 5 to 6 years after arrival. Exogenous assignment to cantons allows us to estimate a *causal* effect of each destination canton for each origin nationality.

$$y_{iodt} = \alpha_{od} + \delta_{o\underline{t}} + \varepsilon_{iodt}$$

We include nationality-by-year arrival cohort fixed effects ($\delta_{o\underline{t}}$), and decompose the variability of long-run outcomes only *within* cohorts. The total variance of labor income 5 to 6 years after arrival *within* cohorts is given by $V(\alpha_{od} + \varepsilon_{iodt})$. The part of the variance in overall outcomes attributable to the causal effect of cantonal assignment is $V(\alpha_{od})$. We take two approaches to correcting for sampling error of this variance component. First, we report the adjusted R2. Second, we use a plug-in estimator that corrects for the sampling bias.[2]. The estimated variance of the origin-by-destination fixed effects is:

$$\widehat{V(\alpha_{od})} = \frac{1}{N}\sum_i (\hat{\alpha}_{o(i)d(i)} - \bar{\alpha})^2 - \text{se}(\hat{\alpha}_{o(i)d(i)})^2$$

For comparison of location effects with individual refugee demographics, we follow an analogous approach, replacing α_{od} with a set of sex-by-arrival-age fixed effects (where arrival-age is defined in 5-year bins).

We next want to compare the overall effect of a destination with the estimated causal effects of networks in that destination upon arrival. Our predicted network effect for individual i randomly assigned to destination d at time $\underline{t}$ is:

$$\widehat{\text{network-effect}}_{iod\underline{t}} = \text{network}_{od,\underline{t}-1} \cdot \hat{\beta}_{IV}$$

where β is estimated using Equation 1.1 and our IV strategy (Section 1.3). Analogous to the approach above, the variance component of network effects is then estimated as:

$$\hat{V}(\text{network-effect}_{iodt}) = \frac{1}{N}\sum_i \left(\hat{\beta}_{IV} - \text{se}(\hat{\beta}_{IV})\right)^2 \left(\text{network}_{o(i)d(i),\underline{t}(i)-1} - \bar{\text{network}}\right)^2$$

[2] In ongoing work, we plan to refine the estimates of variance components using the leave-out estimator proposed by Kline, Saggio, and Sølvsten 2020. Since the number of our origin-by-destination fixed effects is relatively small compared to AKM-type applications where the number of fixed effects grows with the number of individuals, the resulting bias is likely small in our case

We define $\overline{\text{network}}$ in two ways: first, we take the average across cantons within origin-by-arrival-year cohorts; and second, we additionally net out a canton-of-assignment-fixed effect. The former accounts for all the variation in network effects across destinations within cohorts but implies an extrapolation of the LATE from the IV across a wide range of network sizes. The latter more conservatively accounts only for the relative differences in networks within a canton, and is closer to the variation driving our IV estimates. It also controls for the total cantonal population and size.

Table A.3.1: Variance Decomposition of Long-run Refugee Outcomes

	(1)	(2)	(3)	(4)
	Adj. R2	Adj. R2 relative to orig-by-dest FE	$\hat{V}$	$\hat{V}$ rel. to orig-by-dest FE
Total	1		100	
Origin-by-destination FE	.07	1	3	1
Network effect	.10	.83	7	2.09
Network effect (net of assignment-canton-FE)	.01	.09	1	.23
Sex-by-arrival-age FE	.11		11	

Notes: Column (1) presents the adjusted R2 of each model, as a share of the overall variation of labor income 5-6 years after arrival within origin-by-nationality arrival cohorts. We include only randomly assigned refugees, and origin-by-destination fixed effects are therefore interpreted as causal. Column (2) contains the share of the explained sum of squares of origin-by-destination fixed effects, i.e. $\frac{N-1}{N-k}\frac{ESS_{FE}}{ESS_{network}}$. Column (3) shows the bias-adjusted variance estimator of each component as described above, and Column (4) shows the ratio of this estimated variance component to the estimated variance of origin-by-destination fixed effects.

The causal effects of the assigned destination explain between 3% (Variance component estimate) and 7% (adjusted R2) of the overall variation in individual refugee outcomes (Row 2, Table A.3.1). This is roughly half as much as sex and arrival age combined, which explain 11% of income variability 5 to 6 years after arrival. Given how large gender and age gaps are in terms of labor market integration (see Figure 1.4), this is a substantial effect.

Estimated network effects (without netting out average differences across cantons for all nationalities) explain between 7 and 10% of the overall variation in outcomes – again a substantial effect. Network effects are between 0.8 and 2.1 times as consequential as the destination overall (Row 3). Note that it is possible for network effects to explain more of the overall variance in outcomes compared to destination effects, migrant networks are negatively selected into locations. But it is also possible that these large effects are due to sampling variation, or the fact that we extrapolate a LATE from IV across the entire observed distribution of network sizes[3]. Our preferred (and more conservative) approach therefore nets out average variation in network size across cantons. Using this measure, *relative* differences in network sizes across cantons explain between 9 and 23% of the total variation due to causal effects of randomly assigned cantons.

[3]In ongoing work, we plan to follow Kline, Saggio, and Sølvsten 2020 to conduct inference on variance components

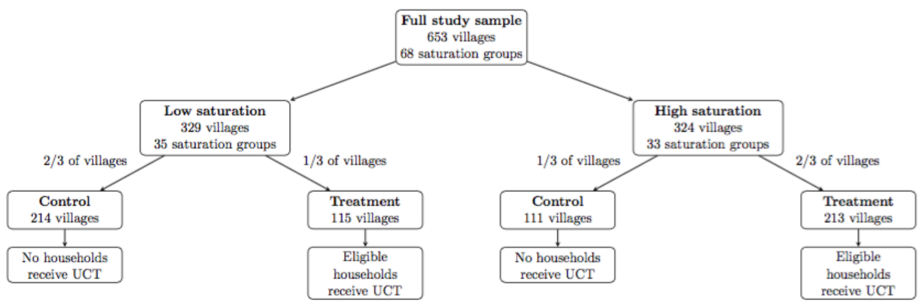
Appendix B

Supplementary Appendix – Chapter 2

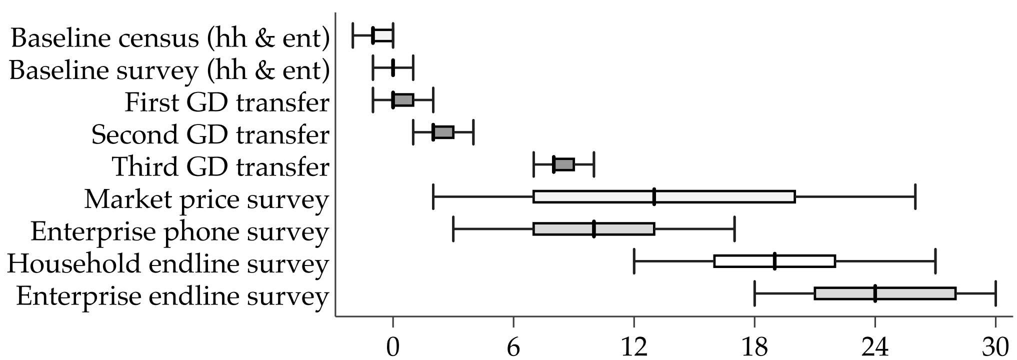
B.1 Study Timeline and Study Area

Figure B.1.1: Study design and timeline

(A) Randomization

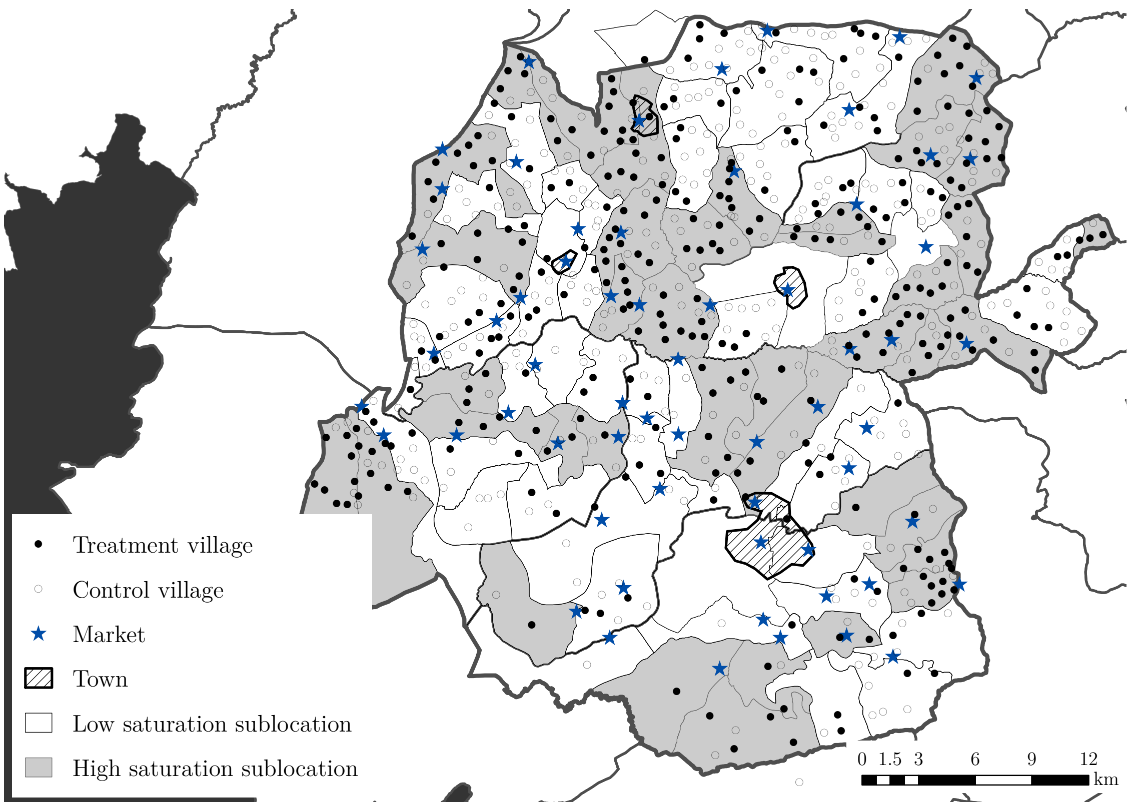


(B) Timing relative to experimental start



Notes: Panel A illustrates the two-level randomized controlled trial experimental design. 653 villages were grouped into 68 saturation groups based on the sublocation (the administrative unit directly above the village level) in which they are located. Saturation groups were then randomly assigned to either high or low saturation status. In the 33 high saturation groups, two-thirds of villages were assigned to treatment status, while in the 35 low saturation groups, one-third of villages are assigned to treatment status. In the 328 treatment villages, all eligible households received an unconditional cash transfer, while no households within control villages received a transfer. Panel B plots the 5th, 25th, 50th, 75th and 95th percentiles of study activities. Timing is reported relative to the anticipated start of activities in each village (the "experimental start"). The experimental start for a village is calculated based on the random ordering of treatment and control villages that both GD and research team field enumerators worked in, as well as GD's mean monthly pace of enrolling villages in the subcounty in which the village is located. As markets were not assigned to treatment, we use the first date transfers were distributed within the subcounty in which the market is located. The value of the first GD transfer is USD 151 PPP, while the second and third are both USD 860 PPP.

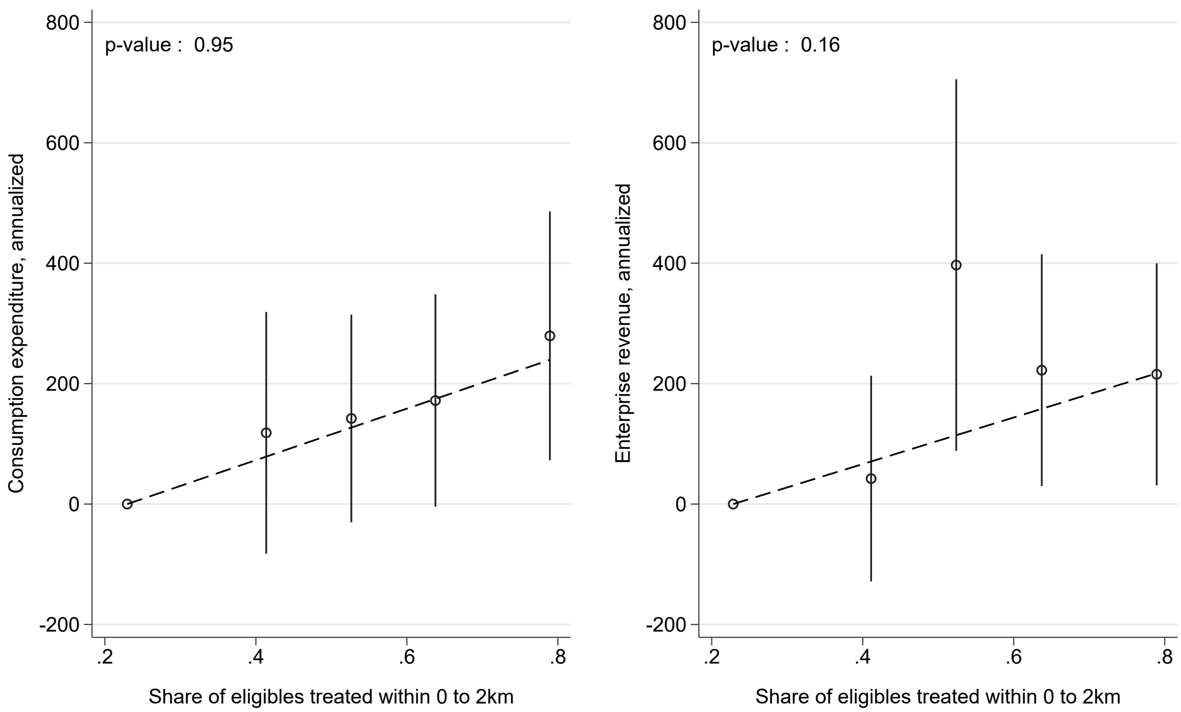
Figure B.1.2: Study area



Notes: This figure plots study villages, sublocation boundaries, and weekly markets in the study area in Siaya County, Kenya. Control villages are denoted by hollow circles, treatment villages are denoted by solid circles, and blue stars indicate the locations of markets. High saturation sublocations are shaded in gray, while low saturation sublocations are those in white. Town boundaries are shaded with diagonal lines.

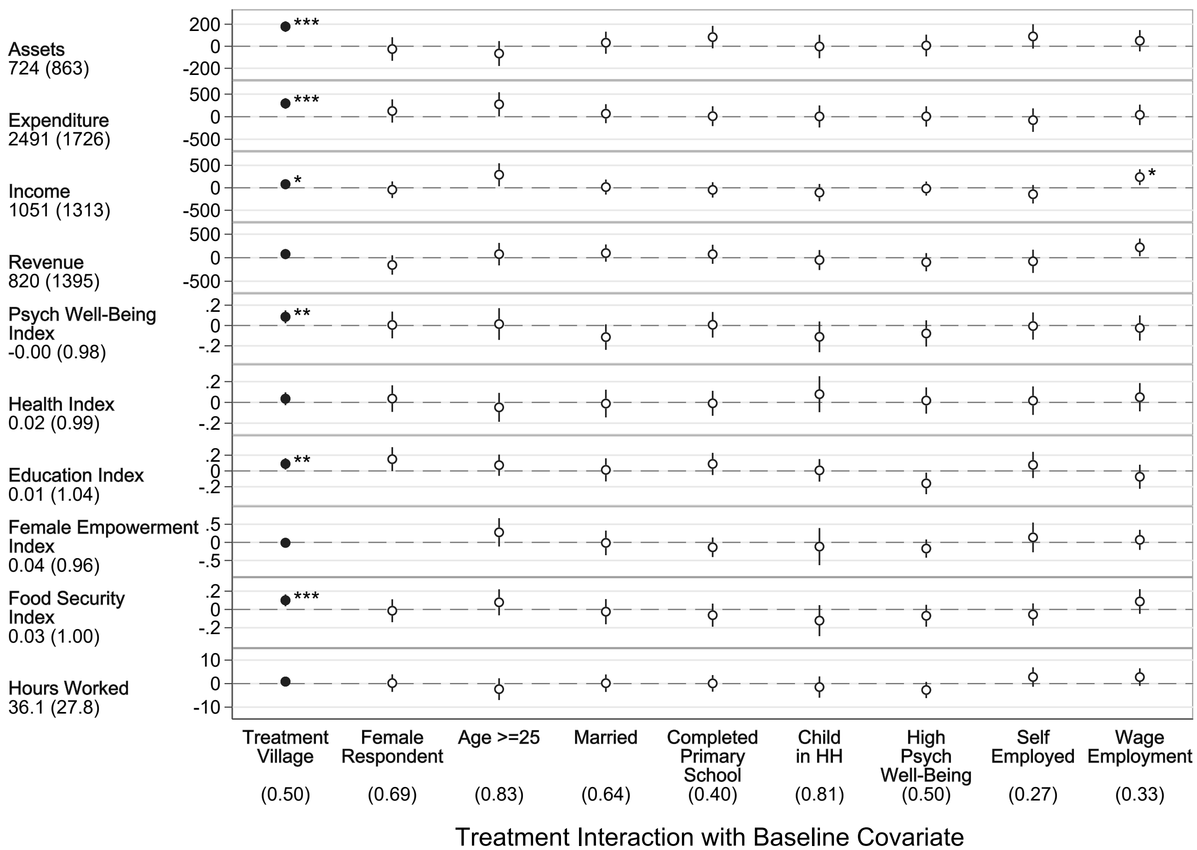
B.2 Supporting figures & tables

Figure B.2.1: Non-linear Spillover Estimates



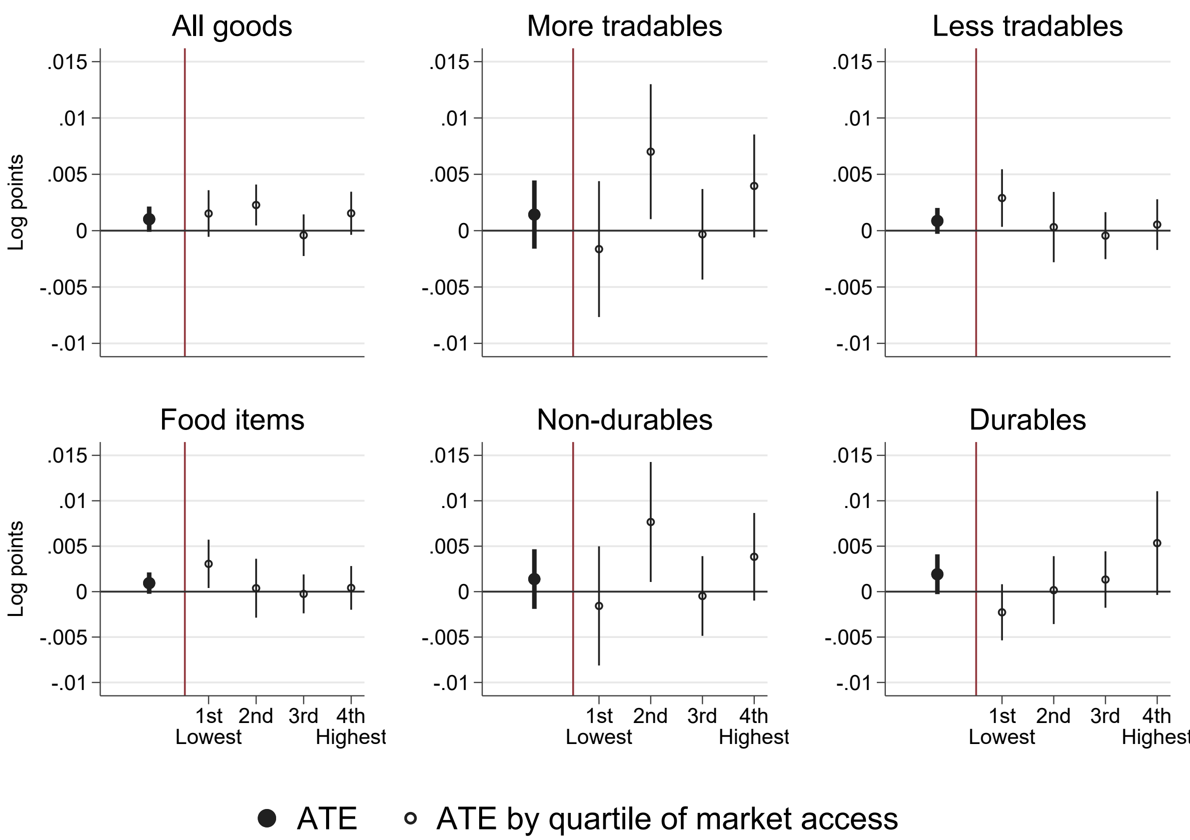
Notes: Each panel represents a reduced-form regression of household consumption and enterprise revenue on 5 quintile bins of the share of eligibles assigned to treatment 2km around each household / enterprise. Effects are relative to the lowest quantile. For consumption, we control for direct effects by including eligibility and a dummy for treatment status of each household. For enterprises, we include interactions with dummies for 3 enterprise types (within homestead, outside homestead, and own-farm). We then sum and weight coefficients to obtain total revenue effects per household in our study area. We control for baseline revenue at the village-by-enterprise type level and use inverse sampling weights. 95% confidence intervals are obtained using Conley (2008). Dashed lines start at zero, the slope coming from the same regression, with quantile bins of treatment intensity replaced by a linear term (and weighted across enterprise types as above). We cannot formally reject that our estimated non-linear regression is linear, i.e. that $\frac{\beta_2-\beta_1}{\Delta X_2} = ... = \frac{\beta_n-beta_{n-1}}{\Delta X_n}$. The p-values of this test are 0.95 and 0.11 for consumption and revenue respectively. We did the same test for all 10 pre-specified primary outcomes and treated / untreated households separately; we cannot reject linearity at the 10% level for any of them.

Figure B.2.2: Little heterogeneity in pre-specified primary outcomes



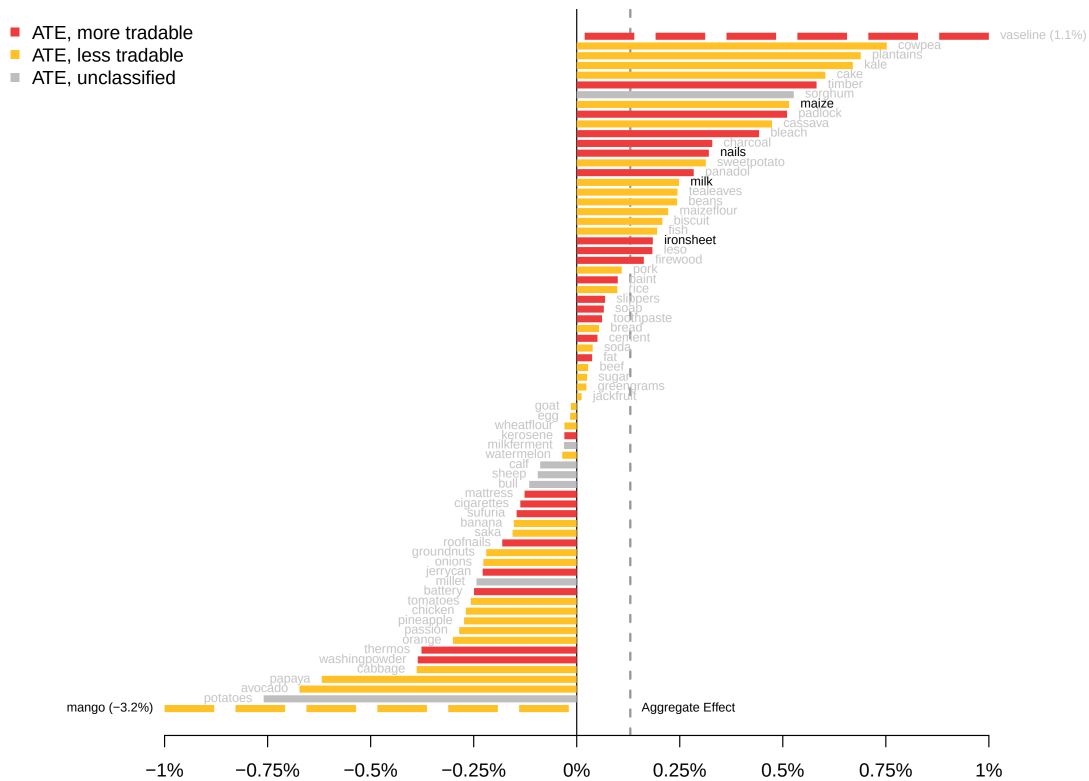
Notes: This figure presents estimates for treatment effect heterogeneity for eligible households in pre-specified primary outcomes along 8 pre-specified dimensions of heterogeneity (Haushofer et al. 2017a). Each plotted coefficient is from a separate regression. Each row represents a separate primary outcome; the mean (SD) for eligible households in control, low saturation villages is reported below the outcome label. The first column (Treatment Village) plots estimated effects for the coefficient on an indicator for being in a treatment village from Equation (2.1), where the sample is restricted to eligible households. Columns 2 through 8 plot the coefficient on the interaction term of the listed baseline covariate with the treatment village indicator; this interaction term and baseline covariate are added to Equation (2.1). Values in parentheses on the x-axis denote the mean of the baseline covariate. Standard errors are clustered at the village level. Reported significance levels correspond to FDR q-values, calculated following Benjamini, Krieger, and Yekutieli (2006). * denotes significance at 10 pct., ** denotes significance at 5 pct., and *** denotes significance at 1 pct. level.

Figure B.2.3: Output price effects by market access



Notes: Each panel represents a regression of the logarithm of a price index on the "optimal" number of lags and distance buffers of per capita GiveDirectly transfers in each buffer, as calculated for the overall price index. The number of radii bands and lags is chosen sequentially by minimizing the BIC, as pre-specified, for the overall price index. Regressions include a full set of market and month fixed effects. We report the implied ATE, calculated by evaluating the "optimal" regression specification at the average level of treatment intensity between September 2014 and March 2017, the time during which transfers went out. Bold markers correspond to the ATE across all markets. Hollow markers break down this average by quartiles of market access (with low market access referring to more remote markets), defined as $MA_m = \sum_{r=1}^{10} r^{-\theta} N_r$, where $\theta = 8$ and N_r is the population in in the $r-1$ to r km buffer around each market. Bars represent 95% confidence intervals based on standard errors as in Conley (2008), where we allow for spatial correlation up to 10km and autocorrelation up to 12 months.

Figure B.2.4: Output price effects at the product level



Notes: Each bar represents a regression of the logarithm of a median price index for each good, using a 4km distance buffer and no lags (the "optimal" number of lags and distance buffers of per capita GiveDirectly transfers for the overall price index). The number of radii bands and lags is chosen sequentially by minimizing the BIC, as pre-specified. Regressions include a full set of market and month fixed effects. Colors denote our classification into more tradable vs. less tradable goods. For each good, we report the implied ATE, calculated by evaluating the regression specification at the average level of treatment intensity between September 2014 and March 2017, the time during which transfers went out. Bold product names indicate significance at the 95% level.

Table B.2.1: Household Assets by Productivity Status

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	$\mathbb{1}$(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Assets (non-land, non-house)	174.49 ***	175.62 ***	151.53 *	1,205.22
	(25.51)	(46.95)	(82.92)	(1,459.67)
Productive Agricultural Assets	4.26***	4.16**	-0.37	32.50
	(0.93)	(1.96)	(2.47)	(38.93)
Potentially Productive Assets	90.03***	52.80	36.46	700.16
	(25.85)	(49.31)	(65.84)	(1,025.10)
Livestock Assets	50.60***	44.81	-6.88	461.88
	(17.03)	(27.90)	(35.77)	(723.23)
Non-Ag Assets	37.10***	24.64	25.71	218.90
	(10.43)	(22.85)	(23.15)	(423.88)
Non-Productive Assets	79.00***	92.71***	52.49*	449.32
	(9.32)	(14.28)	(29.60)	(468.53)

Notes: This table presents results on household asset ownership based on classifications of assets by productivity status. Productive agricultural assets include agricultural tools. Potentially productive assets include livestock and non-agricultural assets, made up of the following: bicycle, motorcycle, car, boat, kerosene stove, sewing machine electric iron, computer, mobile phone, car battery, solar (panels or system), and generators. Non-productive assets include: radio/cd player, kerosene lantern, bed, mattress, bednet, table, sofa, chair, cupboards, clock, television, iron sheets. Column 1 reports the coefficient on an indicator for treatment village from a regression using data from eligible households (as classified by the GE census team), and includes an indicator for saturation status of the sublocation (Equation 2.1). Column 2 reports the total effect on treated households (eligible recipients) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a household's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the household (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer), as in Equation 2.2). For this analysis, the sample is restricted to eligible households, including 5,420 observations. Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households, coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer), as in Equation 2.3). We have 5,505 observations. The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC for total assets (Row 1). Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. Standard errors are clustered at the village level in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.2.2: Enterprise revenue effects by sector

	(1)	(2)	(3)	(4)
	Treatment Villages		**Control Villages**	
	$\mathbb{1}$(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation weighted mean (SD)
Retail revenue	65.46	160.21 **	81.50*	235.98
	(41.84)	(68.09)	(43.38)	(414.95)
Manufacturing revenue	-49.59	92.74**	108.51	81.19
	(73.46)	(46.42)	(70.22)	(177.10)
Services revenue	-77.25 *	7.20	43.37	115.09
	(40.75)	(46.57)	(31.35)	(175.76)
Agriculture revenue	3.11**	5.51***	2.15*	37.91
	(1.27)	(1.43)	(1.29)	(46.39)

Notes: Column 1 reports the coefficient on an indicator for treatment village, and includes an indicator for saturation status of the sublocation (Equation B.1). Column 2 reports the total effect on enterprises in treatment villages (own-village effect plus across-village spillover) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a enterprise's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the enterprise (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer). Column 3 reports the total effect on enterprises in control villages (across-village spillover only). For non-agricultural sectors (retail, services and manufacturing), we stack 2 separate regressions for non-agricultural enterprises operated within the household, and non-agricultural enterprises operated outside the household, due to our independent sampling across these enterprise categories (as in Equation B.2). We have 1,300 observations for retail enterprises, 576 for manufacturing, 400 for services and 7,896 for agriculture. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across all enterprise categories). Each regression is weighted by inverse sampling weights and contains village-level baseline averages of the outcome variable by enterprise category when available. We convert effects to a per-household level by multiplying the average effect per enterprise in each enterprise category by the number of enterprises in that category, dividing by the number of households in our study area, and summing over all enterprise categories. Standard errors are clustered at the village level in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.2.3: Enterprise outcomes by owner eligibility

	(1)	(2)	(3)	(4)
	Recipient Owners		**Non-Recipient Owners**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation weighted mean (SD)
Panel A: All enterprises				
Enterprise profits, annualized	6.78	18.51*	43.55***	156.79
	(7.39)	(11.08)	(14.44)	(292.84)
Enterprise revenue, annualized	51.79**	100.98	171.83 ***	494.45
	(22.82)	(86.46)	(42.78)	(1,223.07)
Enterprise costs, annualized	24.04**	28.11	37.27**	117.22
	(9.41)	(17.39)	(17.18)	(263.46)
Enterprise wagebill, annualized	21.13**	27.71	36.93**	97.35
	(8.69)	(17.48)	(17.10)	(237.01)
Enterprise profit margin	-0.05**	-0.05	-0.01	0.33
	(0.02)	(0.05)	(0.04)	(0.30)
Panel B: Non-agricultural enterprises				
Enterprise inventory	2.88	7.74	5.58	50.41
	(2.79)	(7.47)	(3.91)	(131.86)
Enterprise investment, annualized	-5.15	-15.61	5.49	46.57
	(5.34)	(15.75)	(8.36)	(167.44)

Notes: Column 1 reports the coefficient on an indicator for treatment village, and includes an indicator for saturation status of the sublocation, among matched eligible enterprise owners (Equation B.1). Column 2 reports the total effect on enterprises with a treated owner relative to eligible owners in control villages (own-village effect plus across-village spillover) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a enterprise's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the enterprise (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer). We have between 5,424 and 5,555 owner-matched observations for all enterprises, and 415 for non-ag outcomes. Column 3 reports the total effect on enterprises with untreated owners (spillover only), where we have between 6,584 to 6,739 observations for all enterprises, and 1,454 to 1,459 for non-ag outcomes. For each column, we stack 3 separate regressions for own-farm enterprises, non-agricultural enterprises operated within the household, and non-agricultural enterprises operated outside the household, due to our independent sampling across these enterprise categories (Equation B.2. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across all enterprise categories). Each regression is weighted by inverse sampling weights and contains village-level baseline averages of the outcome variable by enterprise category when available. We convert effects to a per-household level by multiplying the average effect in each enterprise category by the number of enterprises in that category, dividing by the number of households in our study area, and summing over all enterprise categories. Standard errors are clustered at the village level in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.2.4: Input prices and quantities: additional labor supply outcomes

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Household hours worked on own farm	2.07*	0.97	-6.26**	35.32
	(1.15)	(2.30)	(2.61)	(38.79)
Individual hours worked in self-employment	1.80	4.23**	-1.38	26.82
	(1.14)	(1.96)	(1.76)	(23.53)
Individual hours employed last week	0.52	-1.37	2.51	23.60
	(0.98)	(2.32)	(2.67)	(25.95)
Individual hours employed last week in agriculture	-1.53***	-2.28***	0.33	6.00
	(0.56)	(0.75)	(1.11)	(12.78)
Individual hours employed last week not in agriculture	1.67	0.62	1.93	17.08
	(1.03)	(2.31)	(2.65)	(26.40)
Hourly wage earned by employees	0.10***	0.04	0.19*	0.70
	(0.03)	(0.04)	(0.10)	(0.89)
Hourly wage earned by employees in agriculture	0.15**	0.21**	-0.06	0.67
	(0.06)	(0.08)	(0.13)	(0.67)
Hourly wage earned by employees not in agriculture	0.04	0.08	0.20	1.09
	(0.08)	(0.10)	(0.23)	(1.45)

Notes: Column 1 reports the coefficient on an indicator for treatment village from a regression using data from eligible households (as classified by the GE census team), and includes an indicator for saturation status of the sublocation (Equation 2.1). Column 2 reports the total effect on treated households (eligible recipients) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a household's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the household (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer), as in Equation 2.2. For this analysis, the sample is restricted to eligible households, including between 5,420 observations at the household level, and between 1,201 and 4,085 observations for individual-level outcomes. Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households, coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer), as in Equation 2.3. We have between 5,505 household observations, and between 1,019 and 3,486 individuals. The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. In addition, prices are quantity-weighted. That is, wages are weighted by the number of hours worked. Standard errors are clustered at the sublocation in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.2.5: Input prices and quantities: additional land outcomes

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Acres of land owned	-0.19	-0.10	0.08	1.42
	(0.14)	(0.09)	(0.10)	(2.37)
Acres of land rented out	-0.04	-0.05	0.06	0.93
	(0.11)	(0.21)	(0.18)	(0.91)
Acres of land rented in	0.03	0.04	0.08	0.70
	(0.03)	(0.06)	(0.07)	(0.64)
Acres of land used for crops	0.03	-0.03	0.09	0.96
	(0.02)	(0.04)	(0.06)	(1.18)
Land price per acre	168.02	366.46	557.44	3,952.48
	(201.18)	(290.85)	(412.34)	(3,147.29)
Monthly land rental price per acre	-0.05	-0.02	1.80	9.71
	(0.56)	(0.96)	(1.41)	(8.33)
Total ag land rental costs	6.97***	8.99*	10.14	51.76
	(2.47)	(5.21)	(9.39)	(39.67)

Notes: Column 1 reports the coefficient on an indicator for treatment village from a regression using data from eligible households (as classified by the GE census team), and includes an indicator for saturation status of the sublocation (Equation 2.1). Column 2 reports the total effect on treated households (eligible recipients) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a household's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the household (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer), as in Equation 2.2. For this analysis, the sample is restricted to eligible households, including between 352 and 5,418 observations (indicating land markets are often thin). Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households, coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer), as in Equation 2.3. We have between 348 and 5,505 observations. The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. In addition, prices are quantity-weighted. That is, land prices and rental rates are weighted by land size. Standard errors are clustered at the sublocation in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.2.6: Non-market Outcomes and Externalities

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	$\mathbb{1}$(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Psychological well-being index	0.09***	0.12*	0.08	0.01
	(0.03)	(0.07)	(0.06)	(1.01)
Health index	0.03	0.06	0.01	0.03
	(0.03)	(0.06)	(0.05)	(1.01)
Food security index	0.10***	0.05	0.08	0.01
	(0.03)	(0.07)	(0.06)	(1.00)
Children food security	0.13***	0.17**	0.09	-0.04
	(0.04)	(0.08)	(0.09)	(1.12)
Education index	0.09**	0.09*	0.10*	0.01
	(0.04)	(0.05)	(0.06)	(1.02)
Female empowerment index	-0.01	0.08	0.09	0.05
	(0.07)	(0.14)	(0.15)	(0.94)
Security index	0.11***	-0.02	-0.02	0.03
	(0.04)	(0.07)	(0.07)	(0.96)

Notes: Outcome indices in each row are calculated as weighted, standardized indices of multiple survey questions, as described in detail in Appendix B.6. Column 1 reports the coefficient on an indicator for treatment village from a regression using data from eligible households (as classified by the GE census team), and includes an indicator for saturation status of the sublocation (Equation 2.1). Column 2 reports the total effect on treated households (eligible recipients) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a household's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the household (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer), as in Equation 2.2. For this analysis, the sample is restricted to eligible households, including between 4,121 and and 5,423 observations (and a subset of 1,118 for female empowerment). Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households, coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer), as in Equation 2.3. We have between 4,048 and 5,509 observations (and a subset of 978 for female empowerment). The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in Columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. Standard errors are clustered at the sublocation in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.2.7: Inequality

	(1)	(2)	(3)	(4)
	Treatment Villages		**Control Villages**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation weighted mean (SD)
Panel A: Expenditure				
Gini coefficient	0.7	0.8	0.2	32.3
	(0.7)	(1.3)	(1.1)	(7.8)
Counterfactual Gini coefficient	-1.1 *	-2.1	0	32.3
	(0.7)	(1.3)		(7.8)
P-value: effect = counterfactual effect	p=0.08	p=0.05	p=0.84	
Panel B: Assets				
Gini coefficient	-1.1	2.2	2.8**	45.4
	(0.9)	(1.6)	(1.4)	(10.1)
Counterfactual Gini coefficient	-7.6 ***	-6.7 ***	0	45.8
	(0.8)	(0.5)		(10.7)
P-value: effect = counterfactual effect	p=0.00	p=0.00	p=0.04	

Notes: This table reports results on village level inequality as measured by Gini coefficients (0-100). Panel A presents expenditure-based Gini coefficients and Panel B presents assets-based Gini coefficients. For each panel, the first row presents results on actual Gini coefficients measured from our data. The second row estimates the same specifications as the first row, but using counterfactual Gini coefficients assuming that only recipient households gained from the cash transfers, and untreated households experienced no spillovers. We construct a hypothetical consumption expenditure and assets distribution from its baseline distribution (for assets) or by imputing a baseline distribution based on endline non-missing values in control and low-saturation villages (for expenditure). We add in the associated gain, assuming recipients spend 66% of the transfer on consumption, and 34% on assets, following the relative magnitude of the point estimates on expenditure and assets in Table 2.1. This is also in line with our preferred dynamic MPC estimates, where we find recipients spent 93% of the transfer in the first 29 months, 63% on non-durables and 30% on durable assets (see Appendix 2.4 for details). The p-value reported in the third row tests if the actual effect (Row 1) equals the counterfactual effect (Row 2). Gini estimates and effect estimates are weighted by inverse sampling probabilities and village size. Standard errors are clustered at the village level in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.2.8: Expenditures, Savings and Income: Extended version

	Recipient households		Non-recipient households			
	(1)	(2)	(3)	(4)	(5)	(6)
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control Eligibles	Ineligibles	Control, low-saturation mean (SD)
Panel A: Expenditure						
Household expenditure, annualized	293.59 ***	338.57 ***	334.77 ***	21.03	411.55 ***	2,536.01
	(60.11)	(109.38)	(123.20)	(83.76)	(147.81)	(1,933.51)
Non-durable expenditure, annualized	187.65 ***	227.20 **	317.62 ***	24.68	389.31 ***	2,470.69
	(58.59)	(99.63)	(119.76)	(79.05)	(144.86)	(1,877.23)
Food expenditure, annualized	72.04*	133.84 **	133.30 **	10.59	163.33 **	1,578.05
	(36.96)	(63.99)	(58.56)	(50.09)	(71.26)	(1,072.00)
Temptation goods expenditure, annualized	6.55	5.91	-0.68	10.65	-3.46	37.07
	(5.79)	(8.82)	(6.50)	(8.02)	(7.80)	(123.54)
Durable expenditure, annualized	95.09***	109.01 ***	8.44	5.69	9.12	59.41
	(12.64)	(20.24)	(12.50)	(16.83)	(15.00)	(230.83)
Panel B: Assets						
Assets (non-land, non-house), net borrowing	178.78 ***	183.38 ***	133.06 *	-12.25	168.63 *	1,131.66
	(24.66)	(44.26)	(78.33)	(39.93)	(98.04)	(1,419.70)
Housing value	376.92 ***	477.29 ***	80.65	26.90	93.80	2,032.11
	(26.37)	(38.80)	(215.81)	(37.33)	(268.31)	(5,028.27)
Land value	51.28	158.47	544.85	192.35	631.12	5,030.03
	(186.22)	(260.91)	(459.57)	(291.51)	(545.93)	(6,604.66)
Panel C: Household balance sheet						
Household income, annualized	79.43*	135.70	224.96 ***	83.37	259.61 **	1,023.36
	(43.80)	(92.10)	(85.98)	(58.32)	(105.27)	(1,634.02)
Net value of household transfers received, annualized	-1.68	-7.43	8.85	-6.84	12.69	130.08
	(6.81)	(13.06)	(19.11)	(10.27)	(23.18)	(263.65)
Tax paid, annualized	1.94	-0.09	1.68	-0.92	2.31	16.92
	(1.28)	(2.02)	(2.02)	(1.65)	(2.39)	(36.50)
Profits (ag & non-ag), annualized	26.24	35.85	36.37	-1.74	45.70	485.56
	(23.67)	(47.66)	(44.88)	(36.54)	(55.63)	(786.92)
Wage earnings, annualized	42.43	73.66	182.63 ***	90.01**	205.30 **	494.95
	(32.23)	(60.82)	(65.53)	(39.13)	(80.22)	(1,231.12)

Notes: See Table 2.1 for a description of Columns 1 to 3 and 6. Columns 4 and 5 break out the total effects from Column 3 separately for eligible households in control villages and ineligible households (in both treatment and control villages), respectively. Standard errors are clustered at the village level in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 - 5. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.2.9: Expenditures, savings and income results excluding respondents that migrated

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Panel 1: Full Sample				
Respondent migrated	0.01	-0.03	-0.00	0.05
	(0.01)	(0.02)	(0.01)	(0.22)
Net change in household members since baseline	0.03	0.02	-0.18**	-0.10
	(0.04)	(0.08)	(0.08)	(1.30)
Household size	0.02	0.03	-0.10	4.05
	(0.05)	(0.09)	(0.08)	(2.35)
Panel 2: Non-Migrant Sample				
Panel 2.A: Expenditure				
Hou	312.43 ***	376.89 ***	325.86 ***	2,511.75
	(61.11)	(113.62)	(120.61)	(1,926.63)
Non-durable expenditure, annualized	200.77 ***	261.03 **	307.48 ***	2,445.60
	(59.11)	(101.60)	(117.66)	(1,868.88)
Food expenditure, annualized	81.11**	152.28 **	124.06 **	1,572.87
	(38.00)	(67.15)	(60.61)	(1,069.55)
Temptation goods expenditure, annualized	4.50	1.91	-0.61	37.91
	(6.15)	(9.51)	(6.83)	(125.53)
Durable expenditure, annualized	102.07 ***	113.36 ***	8.56	60.03
	(13.19)	(20.84)	(12.63)	(231.69)
Panel 2.B: Assets				
Assets (non-land, non-house), net borrowing	175.10 ***	173.60 ***	136.72	1,145.54
	(25.28)	(50.55)	(84.10)	(1,414.55)
Housing value	403.73 ***	473.12 ***	44.72	2,096.91
	(27.55)	(39.64)	(216.19)	(5,132.21)
Land value	51.69	87.92	525.14	5,141.36
	(193.43)	(279.70)	(464.96)	(6,685.90)
Panel 2.C: Household balance sheet				
Household income, annualized	39.63	84.96	197.25 **	992.84
	(43.31)	(95.51)	(87.03)	(1,600.14)
Net value of household transfers received, annualized	0.69	-10.87	10.58	135.84
	(7.03)	(14.04)	(20.71)	(266.48)
Tax paid, annualized	1.58	-0.95	1.42	16.65
	(1.32)	(2.28)	(2.21)	(35.72)
Profits (ag & non-ag), annualized	13.45	-3.07	14.64	488.97
	(23.56)	(52.34)	(41.43)	(786.27)
Wage earnings, annualized	16.03	67.17	175.24 ***	460.98
	(31.57)	(59.78)	(67.66)	(1,185.04)

Notes: Panel 1 presents estimates of migration impacts on 3 indicators of migration: Whether the respondent themselves migrated out of the study area, the net change in household members since baseline, and the endline household size. Panel 2 reports results from Table 2.1 for respondents that have not migrated, where migration is defined as living in another administrative sublocation for over 4 months. See Table 2.1 for a descriptions of Columns 1-4. In Panel A, we have between 5,403 and 5,422 observations for columns 1-2 and 5,489 and 5,508 for column 3. In Panels B and C, we have 4,982 to 5,024 observations in columns 1-2 and 5,170 to 5,220 observations in column 3. Standard errors are clustered at the village level in Column 1, and calculated following Conley (2008) using a uniform kernel out to 10 km in Columns 2 - 5. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

B.3 Estimating the marginal propensity to consume and spend locally

This appendix section provides details on the marginal propensity to consume (MPC) estimates reported in section 2.4.

We focus on recipients' propensity to spend as a result of the transfer, which is directly relevant for the first-round spending impacts in the local economy. To the extent that recipient households generate additional earned income as the result of the transfer, and also spend out of this income, the main marginal propensity to consume estimate may be an overestimate. Below, we therefore also present recipient expenditure effects relative to the transfer amount received plus any additional income generated as a result of the transfer. (We are also able to obtain an estimate of the marginal propensity to consume among transfer non-recipients, by taking the ratio of spending impacts relative to income effects over the same time period. In fact, the estimates for non-recipients are quantitatively similar to those estimated among cash transfer recipients. Since income is likely to be imperfectly measured relative to expenditure in this context, see Deaton and Zaidi (2002), and because first-round spending impacts are particularly important, we focus on the MPC among transfer recipients.)

In rural African settings like ours, formal sector financial savings (e.g., in bank accounts) or cash savings are limited. Only 11% of households in our study area report having a bank account at endline. In ongoing work in a similar Kenyan context, total savings in mobile money, cash and bank accounts amounted to roughly 100 USD PPP in the control group, a small share of total assets. The effect on total savings of a 1000 USD PPP transfer (which is roughly half the size of the transfer in our study) after 14 months was only roughly 25 USD PPP, or 2.5% of the transfer. Instead, most household saving comes in the form of purchases of relatively liquid durable assets such as livestock or even housing materials. In what follows, we separately present recipient spending on durable assets and non-durable consumption goods. From an intertemporal decision-making perspective, the latter represents pure "consumption", while the former is likely have both a "consumption" and a "savings" component.

Whether they are "consumed" or "saved", expenditures on both durables and non-durables are predominantly local: over 95% of respondents report shopping locally for both types of goods. In a context where financial savings options are limited, high marginal propensities to spend — which as noted above, is not necessarily the same as to consume — should not be unexpected. From the perspective of quantifying the transfer multiplier, it is this marginal propensity to *spend* that matters, as spending on both "consumption goods" and "savings goods" show up as revenue for local firms, and have a similar stimulus effect on the local economy. Our main estimate of the MPC (MPC total) therefore includes both components.

Importantly, recipient expenditures only enter the local economy, and thus generate a local multiplier, if they occur locally and contribute to the income of another local agent. We call the measure of this type of expenditure the marginal propensity to spend locally (MPC local). Since the vast majority of individuals in the study sample work locally and firms are overwhelmingly locally owned (as noted in the main text), we expect nearly all factor payments to remain in the local economy. The main reason why local revenue might not end up as local income is the importing of intermediate goods. In Appendix Section B.4, we calculate that up to 19% of non-durable consumption and 20% of durable purchases indirectly reflect imports of intermediate goods from outside the study area. Our preferred measure of the MPC local adjusts the overall marginal propensity to spend (MPC total) to account for such imports, leading the MPC local to be smaller in magnitude than MPC total.

Table B.3.1: Estimates of recipients' marginal propensity to consume

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Transfer					**Transfer + Income Gains**	
	MPC non-durables		**MPC durables**	**MPC total**	**MPC local**	**MPC total**	**MPC local**
	q1-q3	q4-q10					
Our data only	-0.21	0.29	0.30	0.38	0.30	0.34	0.27
	(0.14)	(0.06)	(0.04)	(0.18)	(0.15)	(0.15)	(0.12)
Rarieda data q1-3, our data q4-10	0.35	0.29	0.30	0.93	0.76	0.84	0.68
	(0.12)	(0.06)	(0.04)	(0.11)	(0.09)	(0.10)	(0.09)

Notes: This table presents estimates of recipients' marginal propensity to consume. Columns 1 and 2 report total effects on non-durable expenditure over the first 3 and the next 7 quarters after the transfer respectively relative to the average treatment amount received, and estimated dynamically according to Equation 2.7. Column 3 adds the effect on accumulated durable assets (including house value) for recipients at endline, estimated using Equation 2.2. Column 4 sums up Columns 1-3, presenting our main marginal propensity to spend (MPC total). Column 5 adjusts Column 4 by accounting for an estimated 20% of durables and 18% of non-durables expenditure that reflects imports of intermediates, yielding the marginal propensity to spend locally (MPC local). See Appendix B.4 for details. Columns 6 and 7 present these effects relative to the average transfer amount plus the average additional income recipients' generated over the 27 months after the transfer, again estimated using Equation 2.2. The first row estimates the MPC using only data from this study. The second row estimates the first 3 quarters of the non-durable expenditure effect in Column 1 using midline expenditure data from households in neighboring Rarieda county that received similar transfers as part of Haushofer and Shapiro (2016), which has the advantage of capturing the immediate expenditure response to the transfer. It is estimated analogously using Equation 2.7, but excluding spillover terms. Standard errors (in parentheses) come from 2000 iterations of a wild-bootstrap, clustered at the sublocation level for our data, and the village level for Rarieda data.

Row 1 in Table B.3.1 presents MPC estimates using data only from this study. We estimate a marginal propensity to spend on non-durables of 0.08 over the first 10 quarters after the initial transfer, and 0.30 on on durables. Combined, this yields a marginal propensity to spend of 0.38. We are thus able to directly account for 38% of the transfer. Adjusted for imported intermediates, the MPC local is 0.30.

As noted in the main text, this estimate faces the important limitation that the endline data collection started about 9 months after transfers in a village went out (see Figure B.1.1B). Recall periods for non-durable consumption range from a week to a month, making it hard to convincingly estimate *direct* spending effects for recipients on non-durables in the initial months after the transfer. Regarding measures from our data, we show in Figure 2.1 that the observed variation in transfer amounts received in the 3 quarters prior to our surveys is limited, leading to imprecise estimates (that are also small or even negative). This limitation is less relevant for the estimation of across-village spillovers because transfers to surrounding villages may have randomly gone out earlier or later relative to the timing of the survey, thus providing ample variation to estimate early spillover effects over a short time horizon. Note that estimates of durables expenditure effects do not face the same problem as non-durable consumption, since we measure these as the difference in the stock of durables at endline (between treatment and control), rather than as an integral of flows.

To overcome this limitation, the second row of Table B.3.1 brings in additional evidence using data collected as part of a closely related project in the neighboring sub-county of Rarieda that collected more detailed data on household spending in the months immediately after receipt of similar GiveDirectly transfers (Haushofer and Shapiro 2016). Endline surveys were conducted an average of 9 months after transfers. In addition, a random subset of this sample was surveyed in each of the first 6 months after transfers went out. Here, we use the combined midline and endline data for households which were surveyed in both rounds to estimate the short-run impacts of transfers on recipient spending[1]. The setting of the Rarieda study is remarkably similar to the one studied in this paper: same implementing partner, same eligibility criteria, similar geography and economic structure and only 3 years between them. However, there are two key features that differed and warrant discussion.

First, the Rarieda study randomized treatment among eligibles *within* villages, while in our study, all eligibles within a village are treated. Moreover, the Rarieda study design did incorporate geographic density of treatment across villages. Thus, the Rarieda data allows us to obtain only estimates of the *direct* impact of cash transfers on recipient spending, not including within-village spillovers or across-village spillovers. We expect the bias from excluding spillovers to be small for the initial non-durable spending impacts on recipients. Table 2.1 shows that across-village spillover effects for recipient non-durable spending are small 18 months after transfers (compare columns (1) and (2)). As we expect spillovers to increase over time, as money begins circulating, they are likely to be even smaller in the initial months. Moreover, Haushofer and Shapiro (2016) show that within-village spillovers in their setting were small and not statistically significant over the first 9 months.

Second, average transfer amounts in Rarieda were only about half the size of transfers in our study – recipients randomly received either 404 USD PPP or 1525 USD PPP – and transfers were randomly either sent as a lump-sum or monthly installments over 9 months. In our study, transfers were sent in 3 instalments over 8 months, a schedule that lies somewhere inbetween the two Rarieda transfer schedules. For estimation, we assume that recipient spending effects are linear in transfer amounts, and do not vary with the scheduling of transfers. Haushofer and Shapiro (2016) show that although initial spending impacts increase slightly less than linearly with the transfer amount, there is also a larger increase in early purchases of large, expensive items in the lump-sum arm. While the former may lead estimates from Rarieda to be overstated compared to our larger transfers, the latter may lead to a bias in the opposite direction.

Although we cannot exactly estimate the potential bias resulting from differences in study design, we can test whether estimated impacts on recipients' spending path are comparable between Rarieda and our data at a time horizon where we have sufficient data in both studies. The p-value for the hypothesis that the impact of cash on recipient non-durable

[1] Note that Haushofer and Shapiro (2016) focus solely on endline data.

spending 4-5 quarters after transfer are the same in our data and in Rarieda is $p = 0.31$. Together with the considerations above, we view these two studies as broadly comparable.

In our preferred estimate of the marginal propensity to consume, we therefore estimate the non-durable spending impact for recipients in the first 3 quarters from Rarieda data, and use our own data thereafter. Specifically, we estimate the dynamic impact of transfers on recipient spending according to Equation 2.7 as we do for our data, but excluding spillover terms as discussed above. We deflate monetary values using the overall Kenyan CPI for Rarieda, and our own market price indices for the GE data. Using per-dollar coefficient estimates from Rarieda data, we then simulate the initial spending impact from transfers sent according to the schedule in our study, i.e., 3 transfers totalling USD 1,871 PPP (USD 1,000 nominal) over 8 months based on the Rarieda coefficients.

Column 1 shows that initial direct spending impacts on non-durable goods in Rarieda were indeed far higher than what we estimate in our data, at 0.35. Combined with our data on non-durable expenditure in the quarters thereafter, we estimate that recipients' spend 64% of the transfer on non-durables over the first 10 quarters. Adding in durable expenditure yields our preferred estimate of the marginal propensity to spend (MPC total) of 0.93. This indicates that we are close to accounting for the entire transfer amount being spent, and highlights that the study population can be characterized as largely hand-to-mouth consumers. Even when we account for increased income generated by recipients over the same period in Column 6, the estimate of the total marginal propensity to spend remains very high, at 0.84. This is again in line with the observation that savings in formal financial products or even in cash are unlikely to be substantial in this context.

The preferred estimate of the marginal propensity spend locally, which accounts for imports of intermediate goods is presented in Column 5, and yields an estimated MPC local of 0.76. An alternative estimate that accounts for any additional income generated (among transfer recipients) is similar, at 0.68 (Column 7). These calculations illustrate that a large share of transfer is spent by recipient households within our study period, and roughly three quarters re-enters the local economy and ends up as income of another local agent. In a simple static Keynesian framework, an MPC local in the range of 0.68 to 0.76 implies a local economy transfer multiplier $\frac{MPC}{1-MPC}$ between 2.1 to 3.2

B.4 Transfer multiplier - robustness

This section conducts three main robustness checks regarding the multiplier analysis. In the first subsection, we attempt to account for transactions between agents in our study area and those located outside it. Using a combination of household and enterprise data, and conservative assumptions on import shares by type of enterprise, we provide an upper bound on the share of the expenditure multiplier that may reflect increased imports from outside the study area. The second subsection makes alternative assumptions about the expenditure effects in the initial months after transfers, which as noted in the main text are noisily estimated in our data because the average endline survey took place 18 months after the first transfers were received. Third, we present estimates in nominal terms (rather than real terms).

Accounting for imports of intermediate goods

As described in Section 2.5, the main expenditure multiplier incorrectly includes imports which are not part of local value added. There are many reasons to believe that any resulting bias is relatively small. From household shopping patterns, we know that only 10% of households report ever shopping at a market outside our study area. Non-farm businesses report only 5% of customers coming from outside the study area. In addition, the estimated effects on household consumption and enterprise revenue are fairly similar, suggesting that consumer spending was quite localized and direct imports by households are relatively small. The main concern is therefore imported intermediate goods.

To gauge whether this bias is quantitatively important, we first assign each component of our non-durable expenditure and durable asset measures to one of 48 enterprise types where it is most likely to be purchased. When there are multiple possible types of enterprises, we use overall revenue shares of different enterprise types to distribute expenditure between them. Reassuringly, this correspondence implies expenditure shares by enterprise type that match their revenue shares from the enterprise survey fairly well (correlation coefficient of 0.62). For each enterprise type, we then obtain an upper-bound for the share of intermediate inputs in overall value added as: $1 - \frac{1}{N}\sum_i w_i \frac{cost_i + profit_i}{revenue_i}$ (where we first winsorize at the 1% and 99% cut-offs, then average across enterprises of each type using revenue weights, and cap at 0 and 1), and where i denotes a firm and N is the total number of firms of that type, and w_i the revenue weight of firm i (re-scaled to sum to 1). This is clearly an upper bound, since the enterprise survey cost measure only contains selected components of firm costs.

Next, we make assumptions based on an understanding of the local context about what share of intermediate inputs is imported from outside the study area. In doing so, we try to err on the side of an import share that is too high. The total share of imports in consumption expenditure and assets is then calculated as the expenditure-weighted share of imports of intermediate goods for each expenditure and asset category. For the exact correspondence between each consumption good or asset and enterprise types, consult Tables B.4.1 and B.4.2.

Using this methodology, the upper bound estimate of the share of imports in non-durable consumption goods is 18%, and for assets, the figure is 20%. This shows that imports of intermediate goods may be non-negligible, but that a large majority of spend-

Table B.4.1: Non-durable expenditure: Intermediate input and import shares

Item	Bought atenterprise type	(1) Expenditure share (data)	(2) Intermediate input share (data)	(3) Intermediate import share (assumed)	(4) Overall import share
Cereals	Cereals	5.9%	60 %	50 %	30 %
	Posho mill	5.9%	26 %	0 %	0 %
	Small retail	2.9%	65 %	75 %	49 %
Roots and tubers	Food stall / Raw food and fruits vendor	2.6%	44 %	25 %	11 %
Pulses	Food stall / Raw food and fruits vendor	3.7%	44 %	25 %	11 %
Vegetables	Food stall / Raw food and fruits vendor	8.6%	44 %	25 %	11 %
Fruits	Food stall / Raw food and fruits vendor	2.9%	44 %	25 %	11 %
Meat	Butcher	4.2%	58 %	0 %	0 %
	Livestock / Animal (Products) / Poultry Sale	0.5%	20 %	50 %	10 %
Fish	Fish Sale / Mongering	6.0%	41 %	0 %	0 %
Dairy and eggs	Food stall / Raw food and fruits vendor	4.6%	44 %	25 %	11 %
Other animal products	Livestock / Animal (Products) / Poultry Sale	0.5%	20 %	50 %	10 %
Cooking fat	Small retail	3.7%	65 %	75 %	49 %
Sugar products	Jaggery	2.6%	54 %	0 %	0 %
	Small retail	2.6%	65 %	75 %	49 %
Jam, honey, sweets, candies	Small retail	0.2%	65 %	75 %	49 %
Tea, coffee	Small retail	1.5%	65 %	75 %	49 %
Salt, pepper, condiments, etc.	Small retail	0.7%	65 %	75 %	49 %
Food eaten outside the house	Food stand / Prepared food vendor	0.8%	56 %	25 %	14 %
	Restaurant	0.6%	48 %	50 %	24 %
Alcohol, tobacco	Bar	0.2%	41 %	100 %	41 %
	Homemade alcohol / liquor	1.0%	52 %	0 %	0 %
	Small retail	0.5%	65 %	75 %	49 %
Other foods	Small retail	0.3%	65 %	75 %	49 %
Clothing and shoes	Clothes / Mtumba / Boutique	1.0%	37 %	100 %	37 %
	Tailor	1.8%	18 %	100 %	18 %
Personal items	Barber shop	0.8%	0 %	100 %	0 %
	Beauty shop / Salon	0.2%	12 %	100 %	12 %
	Photo studio	0.0%	0 %	100 %	0 %
	Small retail	1.0%	65 %	75 %	49 %
Household items	Small retail	2.3%	65 %	75 %	49 %
Transport, travel	Guesthouse/ Hotel	0.5%	18 %	75 %	14 %
	Petrol station	2.3%	86 %	100 %	86 %
	Piki driver	1.9%	26 %	100 %	26 %
Airtime and phone expenses	M-Pesa	2.7%	54 %	100 %	54 %
Internet	Cyber café	0.1%	18 %	100 %	18 %
Firewood, charcoal, kerosene	Charcoal sale / burning	1.6%	16 %	0 %	0 %
	Kerosene	0.1%	36 %	100 %	36 %
	Timber / Firewood	0.1%	45 %	50 %	22 %
Electricity	Local	0.3%		0 %	0 %
Water	Local	0.3%		0 %	0 %
Recreation	Bookshop	0.0%	21 %	100 %	21 %
	Small retail	0.1%	65 %	75 %	49 %
	Video Room/Football hall	0.0%	57 %	100 %	57 %
Lottery tickets and gambling	Small retail	0.1%	65 %	75 %	49 %
Religious expenses	Local	0.6%		0 %	0 %
Weddings, funerals	Local	1.0%		0 %	0 %
Charitable expenses	Local	0.1%		0 %	0 %
House rent / mortgage	Local	0.5%		0 %	0 %
School expenses	Local	10.7 %		0 %	0 %
Medical expenses	Chemist	2.3%	27 %	100 %	27 %
Other expenses	Local	4.2%		0 %	0 %
Total	100.0%				18%

Notes: Each row corresponds to an item in the expenditure module of our household surveys. We match each expenditure item to the enterprise type (from our enterprise census) at which it was most likely purchased (using revenue shares where possible to distribute expenditure where a good may be purchased in multiple enterprise types). Column 1 contains expenditure shares (sample-weighted across all households). Column 2 reports the upper bound of the share of intermediates in value added for each enterprise type, calculated from our enterprise surveys as $1 - \frac{1}{N}\sum_i w_i \frac{cost_i + profit_i}{revenue_i}$ (where we first winsorize at the 1% and 99% cut-offs, then average across enterprises of each type using revenue weights w_i, and cap at 0 and 1). Column 3 contains our upper-bound assumptions about the share of intermediate goods for each enterprise type that is imported, and Column 4 displays the implied import share for this item (multiplying Columns 2 and 3). The last row is expenditure-weighted across all categories.

Table B.4.2: Durable assets: Intermediate input and import shares

Item	Bought at enterprise type	(1) Asset share (data)	(2) Intermediate input share (data)	(3) Intermediate import share (assumed)	(4) Overall import share
Bicycle	Bicycle repair / mechanic shop	0.5%	0 %	100 %	0 %
	Import	0.5%		100 %	100 %
Motorcycle	Motorcycle Repair / Shop	0.5%	45 %	100 %	45 %
	Import	1.9%		100 %	100 %
Car	Import	2.5%		100 %	100 %
Boat	Import	0.0%		100 %	100 %
Bed	Carpenter	2.0%	10 %	75 %	7 %
Chair	Carpenter	1.1%	10 %	75 %	7 %
Table	Carpenter	1.3%	10 %	75 %	7 %
Cupboard	Carpenter	1.5%	10 %	75 %	7 %
Sofa	Carpenter	4.2%	10 %	75 %	7 %
Mattress	Import	1.8%		100 %	100 %
Bednet	Hardware store	0.1%	41 %	100 %	41 %
Solar energy system	Electric accesory/repair	0.3%	6 %	100 %	6 %
	Import	1.0%		100 %	100 %
Generator	Hardware store	0.1%	41 %	100 %	41 %
Car battery	Hardware store	0.2%	41 %	100 %	41 %
Kerosene	Kerosene	0.1%	36 %	100 %	36 %
Lantern	Hardware store	0.2%	41 %	100 %	41 %
Clock	Electric accesory/repair	0.1%	6 %	100 %	6 %
Radio	Electric accesory/repair	0.6%	6 %	100 %	6 %
Sewing machine	Electric accesory/repair	0.4%	6 %	100 %	6 %
Electric Iron	Electric accesory/repair	0.0%	6 %	100 %	6 %
Mobile phone	Electric accesory/repair	0.7%	6 %	100 %	6 %
	Import	0.7%		100 %	100 %
Television	Electric accesory/repair	0.7%	6 %	100 %	6 %
Computer	Electric accesory/repair	0.0%	6 %	100 %	6 %
	Import	0.0%		100 %	100 %
Cattle	Livestock / Animal (Products) / Poultry Sale	11.4 %	20 %	50 %	10 %
Pig	Livestock / Animal (Products) / Poultry Sale	0.3%	20 %	50 %	10 %
Sheep	Livestock / Animal (Products) / Poultry Sale	0.6%	20 %	50 %	10 %
Goat	Livestock / Animal (Products) / Poultry Sale	0.6%	20 %	50 %	10 %
Chicken	Livestock / Animal (Products) / Poultry Sale	1.4%	20 %	50 %	10 %
Other birds	Livestock / Animal (Products) / Poultry Sale	0.1%	20 %	50 %	10 %
Farm tools	Hardware store	0.6%	41 %	100 %	41 %
Ox plow	Hardware store	0.1%	41 %	100 %	41 %
Wheel barrow	Hardware store	0.3%	41 %	100 %	41 %
Hand cart	Hardware store	0.0%	41 %	100 %	41 %
Iron sheets	Hardware store	0.4%	41 %	100 %	41 %
House value (maintenance, improvement)	Welding / metalwork	12.3 %	0 %	100 %	0 %
	Carpenter	12.3 %	10 %	75 %	7 %
	Hardware store	18.4 %	41 %	100 %	41 %
	Local	18.4 %		0 %	0 %
Total	100.0%				20%

Notes: Each row corresponds to an item in the asset module of our household surveys. We match each asset to the enterprise type (from our enterprise census) at which it was most likely purchased (using revenue shares where possible to distribute assets where a good may be purchased in multiple enterprise types). Column 1 contains asset shares (sample-weighted across all households). Column 2 reports the upper bound of the share of intermediates in value added for each enterprise type, calculated from our enterprise surveys as $1 - \frac{1}{N}\sum_i w_i \frac{cost_i + profit_i}{revenue_i}$ (where we first winsorize at the 1% and 99% cut-offs, then average across enterprises of each type using revenue weights w_i, and cap at 0 and 1). Column 3 contains our upper-bound assumptions about the share of intermediate goods for each enterprise type that is imported, and Column 4 displays the implied import share for this item (multiplying Columns 2 and 3). The last row is asset-share-weighted across all categories.

ing still reflects local economic activity (and recall that these figures are upper bounds). To get a sense of how this impacts the multiplier estimate, we further assume that (i) all inventories are in the form of intermediate goods rather than final goods (leading us to err on the side of overstating their import share, at 62%), (ii) the import share of enterprise investment is the same as that of household assets (in our context, household and enterprise assets are often comparable or even shared), and (iii) imports scale linearly with expenditure. We then compute the share of the expenditure-based multiplier that is spent locally (see Table B.4.3). Even under the set of conservative assumptions discussed above, the transfer multiplier for local expenditure remains similar at 2.01.

Table B.4.3: Transfer Multiplier Estimates: Adjusting for Imported Intermediates

	(1) $\mathbb{M}$ Estimate	(2) Share imported	(3) Import adjusted
Panel A: Expenditure multiplier	2.53	0.20	2.01
Household non-durable expenditure	1.17	0.18	0.96
Household durable expenditure	0.81	0.20	0.65
Enterprise investment	0.48	0.20	0.38
Enterprise inventory	0.07	0.62	0.03

Notes: Results from the joint estimation of the expenditure multiplier (as in Table 2.5). Column 1 reports our main point estimates of the expenditure multiplier components. Each component is estimated individually and the multiplier is obtained by aggregating components as described in the main text. Column 2 presents our upper-range estimates of the share of imports captures for each expenditure component, and Column 3 presents the import-share adjusted estimate of the multiplier on local expenditure only.

Alternative assumptions for initial expenditure responses

Figure B.1.1 illustrates the timing of endline household surveys and enterprise surveys, and the substantial time lag between when the first transfers were scheduled in each village and survey administration (with time lags of 9 and 18 months, respectively). This limitation implies that treatment effects on expenditures in the three quarters post-transfer are quite imprecisely estimated, as discussed in detail in Appendix Section B.3.

We conduct two additional robustness tests to partially address these concerns. First, from the perspective of minimizing mean squared error, it may be preferable to exclude the noisily estimated initial quarters for all components of each multiplier that rely on flow values. This almost certainly leads to a substantial downward bias, since all early spending, profits and investment are excluded, but may improve statistical precision. The results of this exercise are presented in Panel A of Table B.4.4. As expected, the estimated expenditure and income multipliers are both lower compared to the main specifications, with the average of both multipliers falling to 1.75. The standard standard error on this estimate also declines substantially, by more than half. When testing both multipliers jointly, we reject a multiplier smaller than one with a p-value of 0.04.

Second, we utilize data from a closely related project in a neighboring county Rarieda that collected more detailed data on recipient household spending in the months immediately after they received similar GiveDirectly transfers a few years prior to this experiment (Haushofer and Shapiro 2016). While this project did not collect data on ineligible households, its data complements our data precisely where we think the timing of surveys and transfers imposes the most significant limitation for us, namely for estimating the direct impacts of transfers on *recipients* in the initial period post-transfer. In this exercise, we replace the noisily estimated consumption impacts among recipient households in the first 3 quarters post-transfer with estimates from the Rarieda data. Specifically, we estimate the same equation 2.7 as we do for our data, but exclude across-village spillover terms (see Appendix B.3 for more details). For all other components, and for responses among non-recipients, the inputs into the multiplier estimate are unchanged.

Panel B of Table B.4.4 shows that augmenting the spending impact estimates with the data from Haushofer and Shapiro (2016) leads to a larger expenditure multiplier estimate of 3.09 (that is also slightly more precisely estimated than our main estimate). When testing both multipliers jointly, we reject a multiplier smaller than one with a p-value of 0.04. In Table B.4.5 we take this augmented estimate of the expenditure multiplier, and additionally adjust for imported intermediates using the same methodology as in B.4. Combining these adjustments, the expenditure multiplier is estimated to be 2.48.

The nominal transfer multiplier

The main multiplier estimate is based on real GDP, in which transfer amounts and all outcome measures are deflated to January 2015 US Dollars using the overall consumer price index in the geographically closest market to each household or enterprise (see Section 2.3 for a description of the price data). Table B.4.6 presents the same exercise in nominal terms. Since we estimate small treatment effects on prices, the difference between the real and nominal measures is mainly driven by overall inflation in the study area. As shown in Figure B.8.2, prices in the study area rose by about 10% per year on average. Roughly in line with this, the nominal multiplier over the first two years after transfers went out is roughly 5% larger than the real multiplier (2.66 versus 2.53) on the expenditure side, and approximately 12% larger (2.55 versus 2.28) on the income side.

Table B.4.4: Transfer Multiplier: Alternative Assumptions for the Initial Spending Impact

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Main estimate	**Alternative Specification I: Setting initial 3 quarters = 0**			**Alternative Specification II Initial 3 quarters from Haushofer & Shapiro (2016)**		
		$\mathbb{M}$ Estimate	H_0: $\mathbb{M} < 0$ *p*-value	H_0: $\mathbb{M} < 1$ *p*-value	$\mathbb{M}$ Estimate	H_0: $\mathbb{M} < 0$ *p*-value	H_0: $\mathbb{M} < 1$ *p*-value
Panel A: Expenditure multiplier	2.53	2.04	0.00***	0.06*	3.09	0.01**	0.06*
	(1.42)	(0.67)			(1.38)		
Household non-durable expenditure	1.17	0.99	0.06*		1.73	0.08*	
	(1.32)	(0.63)			(1.25)		
Household durable expenditure	0.81	0.81	0.00***		0.81	0.00***	
	(0.05)	(0.05)			(0.05)		
Enterprise investment	0.48	0.17	0.06*		0.48	0.15	
	(0.42)	(0.11)			(0.44)		
Enterprise inventory	0.07	0.07	0.02**		0.07	0.02**	
	(0.03)	(0.03)			(0.03)		
Panel B: Income multiplier	2.28	1.45	0.01***	0.23	2.28	0.12	0.24
	(1.73)	(0.65)			(1.76)		
Enterprise profits	1.47	0.00	0.48		1.47	0.14	
	(1.28)	(0.35)			(1.28)		
Household wage bill	0.68	1.34	0.01***		0.68	0.28	
	(1.15)	(0.54)			(1.15)		
Enterprise capital income	0.09	0.10	0.05*		0.09	0.32	
	(0.17)	(0.06)			(0.17)		
Enterprise taxes paid	0.04	0.01	0.03**		0.04	0.08*	
	(0.03)	(0.01)			(0.03)		
Panel C: Expenditure and income multipliers							
Average of both multipliers	2.40	1.75	0.00***	0.09*	2.69	0.03**	0.12
	(1.38)	(0.58)			(1.39)		
Joint test of both multipliers			0.00***	0.04**		0.01***	0.04**

Notes: Results from the joint estimation of expenditure and income multipliers. Column 1 reports our main point estimates of both multipliers and their respective components from Table 2.5. Columns 2 - 4 repeat this exercise, imposing that the impact of each dynamically estimated flow component is zero in the first 3 quarters after the transfer. Columns 5 - 7 estimate the initial 3 quarters of the impact on non-durable consumption expenditure for recipients using data from a related project that collected more detailed data for recipient expenditure in the initial months after the transfer (Haushofer and Shapiro 2016). All other components remain the same as in our main specification. Transfer amounts and outcome variables are deflated to January 2015 using the overall consumer price index in the geographically closest market. Standard errors are computed by 2,000 replications of a clustered wild clustered bootstrap. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.4.5: Transfer Multiplier Estimates: Adding initial Quarters from Haushofer and Shapiro (2016) and Adjusting for Imported Intermediates

	(1) $\mathbb{M}$ Estimate	(2) Share imported	(3) Import adjusted
Panel A: Expenditure multiplier	3.09	0.20	2.48
Household non-durable expenditure	1.73	0.18	1.42
Household durable expenditure	0.81	0.20	0.65
Enterprise investment	0.48	0.20	0.38
Enterprise inventory	0.07	0.59	0.03

Notes: Results from the joint estimation of the expenditure multiplier, using data from Haushofer and Shapiro (2016) for the expenditure response of recipients in the first 3 quarters (as in Table B.4.4). Column 1 reports our main point estimates of the expenditure multiplier components. Each component is estimated individually and the multiplier is obtained by aggregating components as described in the main text. Column 2 presents our upper-range estimates of the share of imports captures for each expenditure component, and Column 3 presents the import-share adjusted estimate of the multiplier on local expenditure only.

Table B.4.6: Nominal Transfer Multiplier

	(1) $\mathbb{M}$ Estimate	(2) H_0: $\mathbb{M} < 0$ p-value	(3) H_0: $\mathbb{M} < 1$ p-value
Panel A: Expenditure multiplier	2.66	0.04**	0.12
	(1.48)		
Household non-durable expenditure	1.22	0.18	
	(1.37)		
Household durable expenditure	0.89	0.00***	
	(0.06)		
Enterprise investment	0.47	0.15	
	(0.43)		
Enterprise inventory	0.08	0.02**	
	(0.04)		
Panel B: Income multiplier	2.55	0.08*	0.19
	(1.80)		
Enterprise profits	1.47	0.13	
	(1.30)		
Household wage bill	0.94	0.22	
	(1.17)		
Enterprise capital income	0.10	0.29	
	(0.18)		
Enterprise taxes paid	0.04	0.07*	
	(0.03)		
Panel C: Expenditure and income multipliers			
Average of both multipliers	2.60	0.04**	0.12
	(1.44)		
Joint test of both multipliers		0.01**	0.06*

Notes: This table is analogous to Table 2.5 (see table notes for detail). The only difference is that here, monetary values are nominal, whereas in Table 2.5 transfer amounts and outcome variables are deflated to January 2015 using the overall consumer price index in the geographically closest market. Standard errors and test statistics are computed from 2,000 replications of a wild clustered bootstrap. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

B.5 Details on study design and intervention

Cash transfer intervention

The NGO GiveDirectly (GD) provides cash transfers to poor households, and for the purposes of this study, enrolled households with grass-thatched roofs in villages assigned to treatment. GD worked on a rolling basis across villages in the study area. The village order was randomized. GD's enrollment process in treatment villages consisted of the following 6 steps:

1. Village meeting (*baraza*): Before beginning work in a village, GD held a meeting for all households in the village to inform residents that GD would be working in their village and explain the program and GD as an organization. To prevent gaming, the eligibility criteria were not disclosed.
2. Census: GD staff then conducted a household census of the village, collecting names of household members, contact information, and information about housing materials. The information on housing materials was used to determine program eligibility.
3. Registration: Households identified as eligible based on the household census were visited by GD's registration team. During these visits, GD staff confirmed the eligibility of the household, informed the household of their eligibility, and registered the household for the program. Households could select the member that they wished to be registered for the program. This visit was the point at which households learned they would be receiving transfers, as well as the amount of the transfers, the transfer schedule, and the fact that the transfer was unconditional.[2] Households were instructed and coached on how to register for M-Pesa, which was a prerequisite for being able to receive transfers. Households that did not have a mobile phone were given the option to purchase one from GD staff, the cost of which was deducted from the transfer amount.
4. Back-check: All registered households were back-checked to confirm eligibility in advance of transfers being sent. Importantly, the census, registration, and back-check teams consisted of separate staff members; this fact, and the multiple eligibility confirmations, were security measures to prevent gaming by households and field staff.
5. Transfers: Transfers were made in a series of three payments via M-Pesa, according to the following schedule: (i) A token transfer of KES 7,000 (USD 151 PPP) was sent once a majority of eligible households within the village had completed their backchecks, to ensure that the system was working properly, to ensure that the system was working properly. (ii) Two months after the token transfer, a first large installment of KES 40,000 (USD 860 PPP) was sent. (iii) Six months later (eight months after the token transfer), a second and final large installment of KES 40,000 was sent. If households elected to receive a mobile phone from GD, this cost (KES 1600 or USD 34 PPP) was subtracted from the second large installment. Transfers were typically sent at a single time per month (usually around the 15th) to all households scheduled to receive transfers.

[2]To emphasize the unconditional nature of the transfer, households were provided a brochure that listed a large number of potential uses of the transfer.

6. Follow-up: After transfers were sent, GD staff followed up by phone with transfer recipients to ensure that transfers were received. In addition, recipients could contact a GD helpline with questions. If GD staff learned that household conflicts had arisen as a result of the transfers, transfers were occasionally delayed while these problems were worked out.

Randomization details

Villages were randomly assigned to treatment status following the two-level randomization design described in Figure B.1.1A. The randomization was conducted in two batches as GD expanded its operations, with the first batch covering villages in Alego subcounty, and the second batch covering villages in Ugunja and Ugenya subcounties.

In Alego, we compiled a list of rural villages eligible for GD expansion. We then grouped sublocations into 23 saturation groups, ensuring that each saturation group was formed from contiguous sublocations, had at least three study villages, and (where possible) the number of study villages was a multiple of three (given that either one-third or two-thirds of villages are assigned to treatment within each sublocation). In 11 sublocations, we declared the sublocation itself as the saturation group. The remaining 13 saturation groups were formed by combining contiguous sublocations into saturation groups. In this manner, the 39 sublocations in Alego were allocated to 23 saturation groups, which were later randomized into high- and low-saturation status.

GD had worked in 193 villages in Alego prior to the start of this study. To account for previous participation in GD's program, we stratified assignment of high and low saturation by the level of previous exposure to the GD program within the saturation group, measured as the share of villages covered by a previous GD campaign, splitting the exposure level at the median.

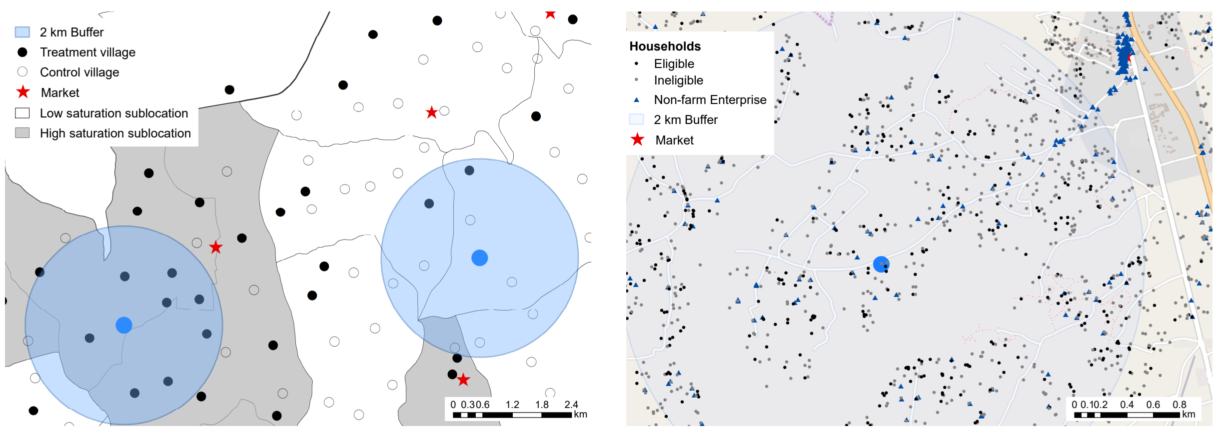
We then randomly assigned villages to three groups, and randomly assigned these groups to either a) treatment, b) treatment in high saturation, control in low saturation, and c) control. In Alego, 12 saturation groups were assigned to high saturation status, covering 98 villages (65 treatment, 33 control), and 11 saturation groups were assigned to low saturation status, covering 105 villages (37 treatment, 68 control). Across these 203 villages, a total of 7,891 households were classified as eligible by the GE census team (37 percent of households), with 3,950 of these households in treatment villages. We randomly generated an order for GD program expansion by first randomly ordering the saturation groups, and then villages within saturation groups.

The second randomization batch included villages in Ugunja and Ugenya subcounties. GD had not previously worked in any villages in these subcounties, so we did not stratify on any variables related to previous exposure for these villages. These subcounties had a larger number of villages per sublocation than Alego on average; as a result, we randomized saturation at the sublocation level. We assigned villages to one of three groups as above, pooled the "residual" villages that were not a multiple of 3, and randomly assigned one third of these to treatment, one third to treatment in high saturation sublocations and control in low saturation sublocations, and one third to control. GD worked first in Ugunja and then Ugenya. Across Ugunja and Ugenya, 115 sublocations covering 227 villages (148 treatment, 79 control) were assigned to high saturation status, while 79 sublocations covering 224 villages (78 treatment, 146 control) were assigned to low saturation status. These 450 villages had a total of 13,846 households classified as eligible by the GE census team (31 percent), with 7,105 of these households in treatment villages. We generated a random order within these subcounties by first ordering locations (the administrative unit above the sublocation), then sublocations within the location, then villages within the sublocation.

Illustrating spatial variation in treatment

Figure B.5.1: Spatial variation of data and treatment

(A) Comparing 2 villages (B) Households and enterprises



Notes: This figure provides an example of the spatial variation that we use to identify spillover effects. Both panels provide zoomed-in views on a selection of villages from Figure B.1.2. Panel A illustrates variation in the density of treatment villages around 2 treated villages. It plots village centers for treatment (filled circles) and control (open circles) villages, as well as a 2 km radius around the village center. While both villages themselves are not treated, the share of treated villages around them varies considerably. Panel B zooms in on one of these villages and plots eligible and ineligible households, as well as non-farm enterprises. Market centers are plotted as red stars.

We construct the amount per capita GDP in each buffer around a village or market i ($Amt_{it,r}$) as the per capita transfers in each buffer r, divided by per capita GDP. We obtain amount transferred into each buffer r at time t from the GPS location of recipients, as well as information from GiveDirectly (GD) on transfers sent to each recipient. Per capita GDP is calculated as the population-weighted average, across all households in the study area, of our expenditure-based measure of GDP (see Section 2.5). To convert stock values into annual flows, we assume a 10% depreciation rate. This yields an average per capita GDP of 676 USD PPP (or 2,897 USD PPP per household). For $Amt^{\neg v}_{vt,r}$, we exclude households in buffer r but located in the same village v.

The population in each radius band around each market or village is determined using the GPS location of each household in our baseline household census data. Each household is then multiplied by the average number of people per household from the baseline household survey. This provides a population measure for each village in our study sample. To account for villages not included in our sample, but within radii bands of study markets or villages, we take two approaches. First, in villages that were not part of our sample but where GD had worked previously, we use household GPS locations provided to us by GD. For areas which were neither in our sample nor had been visited by GD previously, we calculate the population by uniformly distributing the sublocation population from the 2009 Kenyan census, net of the population in study area or GD census villages, over the area of the sublocation that was not already covered by a village in our study or a village where GD had worked previously. Village areas are defined as convex hulls around GPS coordinates of all village households. 2009 Kenyan census numbers are inflated by the overall average population growth rate in Kenya between 2009 and 2014.

B.6 Household data appendix

Construction of index outcomes

Our index variables are constructed from the following components:

1. Psychological well-being index: Weighted, standardized average of depression (10 question CES-D scale), happiness, life satisfaction, and perceived stress (PSS-4), appropriately signed so that positive values represent better psychological well-being.
2. Health index: weighted, standardized average of self-reported health (on a scale of 1 to 5), an index of indicators for common health indicators, and an indicator for whether the respondent has experienced a major health problem since the date of baseline surveys, appropriately signed so that positive values represent better health.
3. Food security index: weighted, standardized index of the number of days a) adults and b) children i) skipped or cut meals, ii) went to bed hungry, iii) went entire days without food out of the last 7 days, appropriately signed so that higher values represent better food security. The Children food security index is made up of the child-related food security questions.
4. Education index: weighted, standardized average of total education expenditure and proportion of school-aged children in school, appropriately signed so that higher values represent better education outcomes.
5. Female empowerment index: weighted, standardized average of a violence index and attitudes index, appropriately signed so that positive values reflect more female empowerment/less domestic violence. The violence index is calculated as from the frequency of physical, emotional as sexual violence over the last 6 months. The attitudes index is calculated from an index of male-oriented attitudes and an index on the justifiability of domestic violence.
6. Security index: a weighted, standardized index of the number of times victimized by i) theft or ii) assault, arson or witchcraft in the last 12 months, an indicator for experiencing but not reporting a crime, and an indicator for reporting to be worried about crime or safety in the neighborhood.

Tracking and attrition

We achieved high tracking rates at endline, reaching over 90 percent of both treatment and control households. To assess levels of attrition, and whether attrition at endline is affected by treatment status and hence might confound our results, we estimate Equation (2.1) using as an outcome an indicator r_{hvs} for whether household h in village v in sublocation s is observed at endline, and do this separately for eligible and ineligible households. We investigate whether this indicator of non-attrition varies with treatment status in Table B.6.1.

We observe high tracking rates of 90.3 and 90.8 in the two types of households, respectively, in low-saturation control villages. These rates are very similar in other villages and sublocations: We observe broadly insignificant treatment coefficients in both tables, suggesting that attrition does not systematically vary with treatment status. This result is robust to defining r_{hvs} as an indicator for being reached at both baseline and endline (Column 2). It is also robust to restricting the sample to only households reached at endline (Panel B) or only households surveyed at baseline (Panel C). The one significant coefficient is for ineligible households in high-saturation sublocations: these are significantly less likely to be reached twice (Panel A, Column 4).

Table B.6.1: Household survey tracking and attrition

	(1)	(2)	(3)	(4)
	Eligible		**Ineligible**	
	Surveyed at endline	Surveyed at baseline and endline	Surveyed at endline	Surveyed at baseline and endline
Panel A: All households targeted at endline				
Treatment Village	0.004	-0.000	0.011	0.013
	(0.008)	(0.013)	(0.011)	(0.015)
High Saturation Sublocation	0.002	-0.014	-0.014	-0.035**
	(0.008)	(0.012)	(0.011)	(0.016)
Control, Low Sat Mean (SD)	0.892	0.797	0.901	0.800
	(0.311)	(0.403)	(0.299)	(0.400)
Observations	6,039	6,039	3,111	3,111
Panel B: Among households surveyed at endline				
Treatment Village		-0.005		0.004
		(0.011)		(0.014)
High Saturation Sublocation		-0.017		-0.025*
		(0.011)		(0.014)
Control, Low Sat Mean (SD)		0.894		0.889
		(0.309)		(0.315)
Observations		5,423		2,816
Panel C: Among households surveyed at baseline				
Treatment Village	-0.005	-0.005	0.011	0.011
	(0.008)	(0.008)	(0.011)	(0.011)
High Saturation Sublocation	0.003	0.003	-0.019*	-0.019*
	(0.008)	(0.008)	(0.011)	(0.011)
Control, Low Sat Mean (SD)	0.916	0.916	0.929	0.929
	(0.278)	(0.278)	(0.256)	(0.256)
Observations	5,185	5,185	2,648	2,648

Notes: This table reports tracking and attrition rates for households, by classification as eligible or ineligible to receive GD transfers by GE project field staff. Each Column represents a regression of an indicator for being surveyed at endline, or at both baseline and endline on an indicator for being in a treatment village, and and an indicator for the saturation status of the sublocation. Panel A includes all households that were targeted for endline surveys. Panel B looks at households that completed endline surveys, and serves as our main analysis sample. Panel C looks at households that completed baseline surveys, and provides information on households that attrited from baseline to endline. Standard errors are clustered at the village level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Baseline balance

We re-estimate our main specifications using baseline demographic and outcome data, following the format of Table 2.1.[3] Panel A presents household demographics, while Panel B presents outcomes from Tables 2.1 and 2.2 for which we have baseline data (we did not collect consumption expenditure data at baseline). We are generally balanced across a wide range of variables. In our main specifications, we include baseline values of the outcome variable as a control when available to improve statistical precision.

Household weights

We weight household-level analyses with inverse sampling probability weights to ensure results are representative of the full population. In each village, we have baseline census data that provides the total number of households, classified by transfer eligibility status (based on research team reports). We targeted 8 eligible households and 4 ineligible households for surveys at baseline, and at endline targeted households surveyed at baseline, as well as those targeted and missed at baseline. The number of eligible households varies across villages; we thus weight households surveyed at endline by the inverse of the share of eligible households surveyed within the village. We do the same for ineligible households.

For hourly earnings, land prices and household interest rates, we interact these household level weights with the number of hours worked, acres of land owned, and total loan amounts, respectively, to make price effects interpretable as unit price effects.

[3] We pre-specified a different set of balance checks that did not incorporate spatial variation; these are available in Egger et al. (2020). These checks also show the experiment is well-balanced.

Table B.6.2: Household balance

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	1(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Panel A: Respondent demographics				
Female	0.02	0.02	-0.01	0.75
	(0.02)	(0.03)	(0.02)	(0.43)
Respondent aged 25 or older	0.00	0.01	-0.01	0.92
	(0.01)	(0.02)	(0.01)	(0.27)
Is married	0.02	0.02	0.01	0.50
	(0.02)	(0.03)	(0.04)	(0.50)
Completed primary school	0.02	0.02	0.05*	0.33
	(0.02)	(0.03)	(0.03)	(0.47)
Has child	0.01	0.02	0.04*	0.73
	(0.01)	(0.02)	(0.02)	(0.44)
Self-employed	-0.01	-0.01	0.00	0.28
	(0.02)	(0.02)	(0.03)	(0.45)
Employed in wage work	-0.02	-0.01	0.01	0.25
	(0.02)	(0.05)	(0.03)	(0.43)
Panel B: Household assets				
Assets (non-land, non-house), net borrowi	3.25	-19.41	-40.08	1,021.51
	(23.01)	(35.93)	(100.45)	(1,392.92)
Housing value	1.58	-12.43	-10.60	1,595.66
	(7.63)	(13.21)	(344.97)	(4,236.82)
Land value	-220.35	-239.49	-132.66	4,386.95
	(164.13)	(349.61)	(438.85)	(5,873.34)
Panel C: Household cash flow				
Household non-ag income, annualized	-4.83	25.58	-13.79	196.61
	(15.79)	(32.83)	(29.16)	(460.56)
Self-employment profits, annualized	1.83	9.23	-2.25	89.08
	(7.42)	(13.54)	(19.00)	(287.39)
Wage earnings, annualized	-10.16	5.32	-5.40	95.96
	(12.64)	(25.99)	(13.52)	(305.39)
Tax paid, annualized	2.00*	3.38**	3.29	16.35
	(1.20)	(1.72)	(2.45)	(44.72)
Panel C: Input Prices				
Land price per acre	-58.12	201.74	274.14	3,303.33
	(94.69)	(171.17)	(263.14)	(2,984.00)
Acres of land owned	36.39	73.30	-0.32**	1.36
	(36.43)	(74.18)	(0.16)	(2.39)
Total loan amount	1.61	5.83	-3.19	54.08
	(3.12)	(4.65)	(12.19)	(162.29)

Notes: This table presents regression specifications from Table 2.1 using baseline demographic and outcome variables. We did not collect consumption expenditure data at baseline. We have 4,674 to 4,768 observations for columns 1 and 2 (3,962 for land price) and 4,696 to 4,831 observations for column 3 (4,201 for land price). * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Constructing average effects from coefficient estimates

Table B.6.3 presents the coefficient estimates underlying our reported average effects shown in Table 2.1. These average effects are constructed using the average values of each of the regressors included in the selected specification, denoted $\bar{X}$ and presented in the bottom row. For instance, the recipient household total effect for household expenditure (Table 2.1, row 1, column 2) is found by multiplying the coefficient on the amount going into households' own village Amt_v (row 1, column 2) by the mean amount (relative to village GDP) going into own village $Amt_v|i$ is an eligible recipient (last row, column 2) and adding the coefficient on the amount going to other villages within 0-2km $Amt_{v,2}^{\neg v}$ (row 1, column 3) times the mean amount going into this radii band for treated villages $Amt_{v,2}^{\neg v}$ (last row, column 3). We proceed in the same manner for the other tables.

Table B.6.3: Coefficient estimates for Expenditures, Savings and Income

	Recipient households			Non-recipient households		
	(1)	(2)	(3)	(4)	(5)	(6)
	𝟙(Treat village) Reduced form	Amt Own Village IV	Amt Other Villages 0-2km IV	Amount, Control Eligibles 0-2km IV	Amount, Ineligibles 0-2km IV	Control, low-saturation mean (SD)
Panel A: Expenditure						
Household expenditure, annualized	296.67 ***	1,248.05 ***	296.86	194.07	4,723.85 ***	2,537.82
	(59.75)	(240.70)	(1,070.65)	(1,111.62)	(1,696.62)	(1,938.31)
Non-durable expenditure, annualized	189.24 ***	812.65 ***	240.60	251.37	4,472.33 ***	2,472.60
	(58.23)	(223.18)	(977.19)	(1,052.20)	(1,663.59)	(1,881.73)
Food expenditure, annualized	73.29**	380.13 ***	436.39	123.43	1,884.12 **	1,579.19
	(36.81)	(135.03)	(613.36)	(667.01)	(819.74)	(1,073.91)
Temptation goods expenditure, annualized	6.10	23.61	4.49	150.16	-40.07	37.09
	(5.72)	(23.24)	(96.96)	(107.64)	(89.54)	(123.47)
Durable expenditure, annualized	96.55***	396.11 ***	141.73	69.54	102.21	59.21
	(12.66)	(60.98)	(207.04)	(224.15)	(171.20)	(230.66)
Panel B: Assets						
Assets (non-land, non-house), net borrowing	178.10 ***	716.26 ***	80.21	-185.89	1,879.49 *	1,136.61
	(24.55)	(93.48)	(398.26)	(531.93)	(1,117.90)	(1,423.95)
Housing value	377.63 ***	1,563.60 ***	1,064.65 ***	339.50	1,102.27	2,044.23
	(26.12)	(94.98)	(387.54)	(500.73)	(3,077.40)	(5,038.31)
Land value	49.94	351.14	677.21	2,354.67	7,167.92	5,041.59
	(186.74)	(507.69)	(3,091.03)	(3,894.74)	(6,211.05)	(6,614.97)
Panel C: Household balance sheet						
Household income, annualized	79.27*	361.97 *	551.85	1,146.13	2,965.25 **	1,026.71
	(43.55)	(194.61)	(769.06)	(773.28)	(1,231.82)	(1,643.68)
Net value of household transfers received, annualized	-1.57	-5.91	-64.66	-86.14	141.78	130.67
	(6.77)	(27.54)	(124.17)	(136.24)	(264.70)	(264.40)
Tax paid, annualized	1.87	7.53	-23.00	-12.72	27.16	16.95
	(1.28)	(4.78)	(20.55)	(22.08)	(27.48)	(36.55)
Profits (ag & non-ag), annualized	25.67	127.33	79.45	-2.79	524.79	485.91
	(23.54)	(101.00)	(407.04)	(486.58)	(639.27)	(787.17)
Wage earnings, annualized	42.61	210.45 *	236.21	1,191.26 **	2,356.19 **	498.29
	(32.23)	(125.91)	(587.42)	(522.76)	(945.28)	(1,243.65)
$\bar{X}$		0.25	0.09	0.07	0.09	

Notes: This table reports the coefficient estimates that underlie the average effects reported in Table 2.1, see corresponding table note for more details. $\bar{Amt}$ reports the average of each RHS variable for the sample studied (recipients or non-recipients), which we multiply with the coefficient to get the average effects reported.

B.7 Enterprise data appendix

Enterprise census and survey details

We conducted a baseline enterprise census in each village on the same day as the baseline household census. The household census included a question on whether the household was running an enterprise from their homestead or from a fixed kiosk/shop. The enterprise census targeted enterprises operating outside of homesteads. We then returned to survey enterprises operating outside of the homestead and open on the day of our visit, coincident with baseline household surveys. In villages with over 20 enterprises operating outside of homesteads, e.g., those that overlapped a market center, we randomly selected 20 enterprises to survey.

Our endline enterprise census sought to re-identify all enterprises operating from within or outside homesteads, both those identified at baseline and any new enterprises. In order to maintain a representative sample, we randomly sampled up to 2 enterprises operating from within homesteads and up to 3 outside of homesteads to be surveyed, including those in market centers in villages containing a market.

Enterprise surveys cover profits, revenues, and a subset of costs (including the wage bill), and at endline collected information on inventories and investment. We measure (annualized) revenues and profits for non-agricultural enterprises directly by asking respondents about these quantities with a one month recall period (de Mel, McKenzie, and Woodruff 2009). We calculate costs as the sum of the employee wage bill, rent and security costs; this is not a comprehensive measure of all costs, and hence we do not expect the revenue measure to equal our measure of profits plus measured costs. In particular, we do not directly measure expenditure on intermediate inputs such as materials or supplies.

Information on agricultural enterprises comes from our household surveys. Baseline household surveys did not include sufficient detail to construct measures of agricultural revenue and profit, so we only use endline measures for these outcomes. For agricultural enterprises, total revenue is calculated as the sum of crop output (measured at the crop level) plus the value of pastoral and poultry output sold, and the value of the household's own consumption of pastoral and poultry output. When crop output was reported in non-monetary units, we convert these to monetary values using the 2016 mean of the median crop output price measured in the market price surveys in the household's sub-county. Agricultural costs are the wage bill, all agricultural inputs (e.g., seed and fertilizer), and land rental costs. We then calculate agricultural profits as total agricultural revenue minus agricultural costs.

Enterprise specifications

We estimate the following equations for enterprises:

$$y_{ivs} = \alpha_1 Treat_v \cdot X_{ivs} + \alpha_2 HighSat_s \cdot X_{ivs} + X_{ivs}\gamma + \delta_1 y_{ivs,t=0} \cdot X_{ivs} + \delta_2 M_{ivs} \cdot X_{ivs} + \varepsilon_{ivs}, \tag{B.1}$$

$$y_{iv} = \beta Amt_v \cdot X_{ivs} + \sum_{r=2}^{R} \beta_r Amt_{v,r}^{\neg v} \cdot X_{ivs} + X_{ivs}\gamma + \delta_1 \bar{y}_{iv,t=0} \cdot X_{ivs} + \delta_2 M_{iv} \cdot X_{ivs} + \varepsilon_{iv}. \tag{B.2}$$

Here, y_{ivs} is an outcome for enterprise i in village v (and sublocation s), $X_{iv(s)}$ is a vector of indicators for enterprise type (agricultural, non-agricultural operating outside the homestead, non-agricultural operating from the homestead), and other terms are defined as in Section 2.3. We interact our treatment indicator and transfer amount variables with this vector of enterprise types, effectively estimating a stacked version of Equations 2.1 and 2.2. This allows treatment effects and controls to vary flexibly across enterprise type. Table B.7.4 reports the share of enterprises by sector weighted by count and by revenue. Since enterprise surveys were conducted as repeated cross-section rather than a panel, we control for the village-level baseline mean of the outcome variable where available in our main specification. Results are similar if we omit this control (Table B.7.5).

We use our endline enterprise census data to construct weights that are representative of the full population of enterprises. In particular, we weight enterprises by the inverse of the share of surveyed enterprises of a particular type (agricultural, operating from homesteads, operating outside homesteads) within each village. For hourly wages, we interact these enterprise-level weights with the total hours worked to make wage effects interpretable as the average effect per hour worked.

Tracking, balance and attrition

Our enterprise samples are repeated cross-sections, so we do not report attrition rates between baseline and endline. We do check baseline balance for enterprises, taking the same approach as in Table 2.3 but using baseline values for outcomes that are available. (We did not collect enterprise investment or inventories, nor do we have revenue and profit measures for non-agricultural enterprises at baseline.) The baseline sample generally appears balanced; there are no statistically significant differences at the 5% level (Table B.7.6).

Matching enterprise owners

Through our integrated approach to enterprise and household censusing, we are able to match all agricultural enterprises (as found via household surveys), and 61% of non-agricultural enterprises, for a total of 94% of all enterprises. To match non-agricultural enterprises to the households that own them we apply both automatic and manual procedures to our detailed name, phone number and GPS data. As we relied heavily on the reported operating location, we excluded enterprise census data without this information. The proportion of matched enterprises are relatively evenly split by treatment status for both eligible and ineligible households: 52% of matched eligible enterprise owners and 51% of matched ineligible owners are in treatment villages.

Patterns with respect to the eligibility status of the owner are generally sensible: 28% of non-agricultural enterprises are owned by an eligible household, slightly below their share in the population (33%), and enterprises owned by ineligible (and thus on average somewhat richer) households have 10% higher profits and 21% higher revenues on average than those owned by eligibles.

Table B.7.4: Composition of enterprises by sector

	Overall		**Non-Ag**	
Sector	Count Share	Revenue Share	Count Share	Revenue Share
Retail	0.33	0.49	0.54	0.52
Manufacturing	0.15	0.23	0.24	0.24
Services	0.13	0.23	0.21	0.24
Agriculture	0.40	0.06		

Notes: This table describes enterprise shares by sector, both in terms of counts and shares of total revenue. Data on counts comes from the endline enterprise census (for non-agricultural enterprises) and the baseline household census (for agricultural enterprises). Data on revenue shares for the non-agricultural sectors comes from endline enterprise surveys, while data on agricultural revenue shares comes from endline household surveys.

Table B.7.5: Enterprise outcomes without baseline controls

	(1)	(2)	(3)	(4)
	Treatment Villages		**Control Villages**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation weighted mean (SD)
Panel A: All enterprises				
Enterprise profits, annualized	-12.27	48.57	36.04	156.79
	(28.06)	(155.40)	(131.66)	(292.84)
Enterprise revenue, annualized	-99.47	340.24 **	289.53 **	494.45
	(128.29)	(146.36)	(138.41)	(1,223.07)
Enterprise costs, annualized	-15.18	95.11**	73.79	117.22
	(34.39)	(41.02)	(50.59)	(263.46)
Enterprise wagebill, annualized	-17.57	83.83**	72.97*	97.35
	(30.39)	(33.36)	(40.48)	(237.01)
Enterprise profit margin	—	—	—	—
Panel B: Non-agricultural enterprises				
Enterprise inventory	—	—	—	—
Enterprise investment, annualized	—	—	—	—
Panel C: Village-level				
Number of enterprises	0.01	0.01	-0.00	1.12
	(0.01)	(0.02)	(0.02)	(0.14)

Notes: This table replicates Table 2.3 but without village level baseline control variables. We omit outcomes for which baseline controls were not available in the original table, as results for those outcomes are unaffected. See notes to Table 2.3 for further details. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.7.6: Enterprise Balance

	(1)	(2)	(3)	(4)
	Treatment Villages		**Control Villages**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation weighted mean (SD)
Panel A: Non-agricultural enterprises				
Enterprise profits, annualize	-10.31	8.21	13.74	238.33
	(15.92)	(21.83)	(21.54)	(393.18)
Enterprise revenue, annualized	-93.00	85.33	130.63	1,010.90
	(84.39)	(103.77)	(109.85)	(2,370.59)
Panel B: All enterprises				
Enterprise costs, annualized	2.73	11.36	4.76	37.72
	(5.14)	(8.65)	(7.78)	(107.56)
Enterprise wagebill, annualized	2.11	8.54	5.29	36.10
	(5.03)	(7.33)	(5.79)	(106.31)
Panel C: Village-level				
Number of enterprises	-0.00	-0.00	-0.01	1.07
	(0.01)	(0.02)	(0.02)	(0.14)

Notes: This table presents regression specifications from Table 2.3 using corresponding baseline enterprise outcomes where available. We did not collect enterprise inventories and investment data at baseline. We also exclude baseline agricultural revenues and profits, as these were not collected in the same manner as at endline. We have between 4,125 and 4,193 observations in Panel A, 9,245 to 9,264 in Panel B, and 653 in Panel C. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct.

B.8 Price data appendix

Categorizing market survey products

Our market surveys included questions about 84 commodities. As outlined in our pre-analysis plan, we excluded products that are not present in at least 5 percent of market-month observations; this affects 11 products (bicycle, bull (grade), calf (grade), donkey, duck, piglet, turkey, goat (meat), lamb, milk powder, and mosquito net). Three products (Waterguard, fertilizer, and improved maize seeds) do not have consumption expenditure analogues in the KLPS-3 surveys that we use to construct expenditure weights, so these are also excluded, leaving a final list of 70 products.

Table B.8.7 presents the classification of the products we use in our price analysis into more and less tradable categories, and the subcategories (food, livestock, (non-food) non-durables, durables and temptation goods) shown in Table 2.4.

Table B.8.7: List of market products by category

Less tradable (locally produced)			**More tradable**		
Food		Livestock	Non-Food Non-Durables	Durables	Temptation Goods
Cassava	Papaya	Bull (local)	Bar soap	1 Iron sheet (32 gauge)	Cigarettes
Irish potato	Pineapple	Calf (local)	Toothpaste	Cement	
Maize	Water Melon	Chicken (hen)	Vaseline/lotion	Large Padlock	
Millet	Jackfruit	Goat	Washing powder	Nails (3 inch)	
Plantains	Passion Fruit	Sheep	Bleach	Roofing Nails	
Rice	Beef		Panadol/aspirin	Timber (2x2)	
Sorghum	Fish (Tilapia)		Cooking fat	Water Paint	
Sweet potato	Pork		Batteries (3-volt)	20L Jerry can	
Beans	Eggs		Firewood	Thermos flask	
Cabbage	Milk (Fresh)		Kerosene	3 1/2 X 6 Mattress	
Cowpea leaves	Biscuits		Charcoal		
Green grams	Bread		Leso		
Groundnuts	Cake		Small sufuria		
Kales	Maize flour		Slippers		
Onions	Wheat flour				
Saka (Local Vegetable)	Milk (Fermented)				
Tomatoes	Soda				
Avocado	Sugar				
Banana-sweet	Tea				
Mango					
Orange					

Notes: This table presents the classification of the 70 products used in our analysis of output prices. The classification follows our midline pre-analysis plan (Appendix B.10). The market survey collected information on 85 products. As outlined in our pre-analysis plan, we exclude any product that, at the market-product level, is missing for more than 95% of cases, a total of 11 products. We also drop three products that do not match items in our expenditure share data.

Price analyses robustness checks

Alternative definition of market access

Our main specifications separates price effects by market access as defined in Donaldson and Hornbeck 2016: $MA_m = \sum_{r=0}^{10} r^{-\theta} N_r$, where $\theta = 8$ and N_r is the population in in the $r-2$ to r km buffer around each market. Here we present alternative results based on a definition of market access as the inverse distance from the closest 'main' road, where we define a main road as any road in Open Street Maps classified as motorway, trunk, primary, secondary or tertiary road (excluding residential streets, tracks, paths, and unclassified roads). While price effects were concentrated in low-market-access areas using our main population-density-based market access measure, they seem to be fairly similarly small when splitting by road access. In a context where most people walk to their nearest market, this may not be surprising. However, we cannot reject that results are the same as our main results.

Table B.8.8: Output Prices using distance to main road as market access measure

		(1)	(2)	(3)	(4)
		Overall Effects		**ATE by road access**	
		ATE	Average maximum effect (AME)	below median	above median
All goods		0.0010*	0.0042	0.0010	0.0011
		(0.0006)	(0.0031)	(0.0008)	(0.0008)
By tradability	More tradable	0.0014	0.0062	0.0006	0.0021
		(0.0015)	(0.0082)	(0.0021)	(0.0021)
	Less tradable	0.0009	0.0034	0.0012	0.0007
		(0.0006)	(0.0032)	(0.0009)	(0.0009)
By sector	Food items	0.0009	0.0036	0.0014	0.0007
		(0.0006)	(0.0033)	(0.0009)	(0.0010)
	Non-durables	0.0014	0.0061	0.0005	0.0020
		(0.0017)	(0.0089)	(0.0023)	(0.0022)
	Durables	0.0019*	0.0070	0.0012	0.0031
		(0.0011)	(0.0061)	(0.0013)	(0.0019)
	Livestock	-0.0008	-0.0027	-0.0023*	0.0012
		(0.0010)	(0.0052)	(0.0013)	(0.0013)
	Temptation goods	-0.0011	-0.0112	-0.0035	0.0022
		(0.0026)	(0.0143)	(0.0036)	(0.0041)

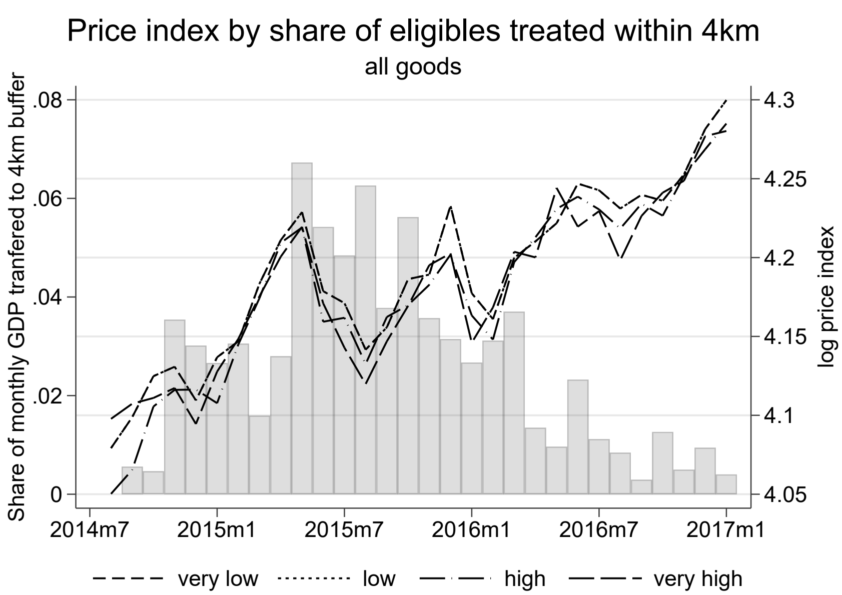
Notes: This table replicates Table 2.4. See notes for details. The only difference is the definition of market access of each market in Columns 3 and 4: It is defined as the inverse distance from the closest main road, classified by Open Street Map as motorway, trunk, primary, secondary or tertiary road (excluding residential, tracks, paths, and unclassified roads): $MA_m = \frac{1}{distance_m}$. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Spatial and temporal analysis of price effects

Our main analysis in Section 2.4 follows our pre-specified algorithm, which selects the number of lags and distance buffers by minimizing the Schwarz BIC. While we allow for up to 18 months lags, and 20km spatial dependence, the algorithm selects a specification that includes only contemporaneous transfers up to 4km. In this section, we present three pieces of additional exploratory analysis that serve as robustness checks on our primary pre-specified results and explore the spatial and temporal dimensions of price effects in turn.

First, we show our price data in raw form: Figure B.8.2 shows that prices in more vs. less exposed markets as measured by the share of eligible households within 4km that were assigned to treatment evolved very similarly over the course of the study period and afterwards. We can visually reject large differences in the evolution of prices in response to treatment.

Figure B.8.2: Price index by treatment intensity



Notes: The figure shows the log price index across all goods in more vs. less exposed markets as measured by quartiles of the the share of eligible households within 4km that were assigned to treatment. Bars represent average transfer amounts relative to monthly GDP going to the 4km buffer across all 61 markets in each month.

Second, we estimate Equation (2.4) for a range of outer radii R from 2km to 6km while fixing the number of temporal lags at the (BIC selected) value of 0 months. This allows us to test whether our algorithm is indeed picking up the relevant spatial horizon, or whether we might be missing part of the effect. Table B.8.9 shows that price effects are robust to including additional radii bands. For none of the price indices can we reject that adding an additional buffer on top of that selected by our pre-specified algorithm leads to significantly different average price effects as those in the main specification.

Table B.8.9: Robustness to fixing alternative radii bands: Output Prices

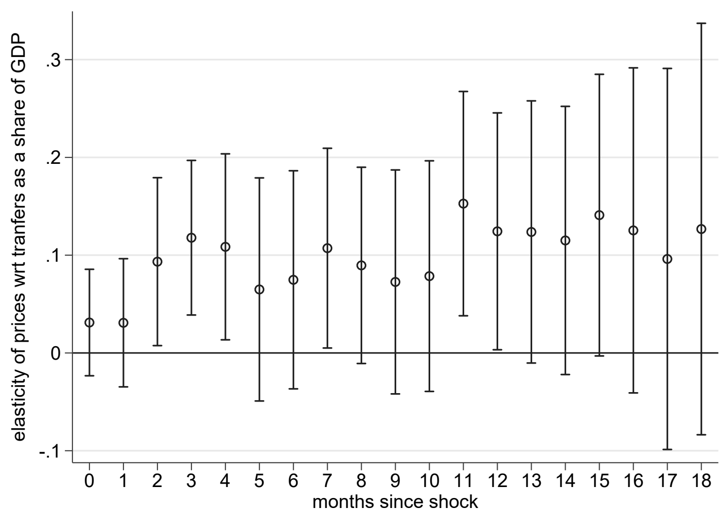
		Overall Effects			
		(1) ATE Optimal Radius	(2) ATE $\bar{R} = 2$	(3) ATE $\bar{R} = 4$	(4) ATE $\bar{R} = 6$
All goods		0.0010*	0.0001	0.0010*	0.0014*
		(0.0006)	(0.0004)	(0.0006)	(0.0008)
By tradability	More tradable	0.0014	0.0003	0.0014	0.0021
		(0.0015)	(0.0010)	(0.0015)	(0.0020)
	Less tradable	0.0009	0.0001	0.0009	0.0012
		(0.0006)	(0.0004)	(0.0006)	(0.0008)
By sector	Food items	0.0009	0.0001	0.0009	0.0012
		(0.0006)	(0.0005)	(0.0006)	(0.0009)
	Non-durables	0.0014	0.0003	0.0014	0.0021
		(0.0017)	(0.0011)	(0.0017)	(0.0021)
	Durables	0.0019*	-0.0000	0.0019*	0.0027
		(0.0011)	(0.0008)	(0.0011)	(0.0016)
	Livestock	-0.0008	0.0001	-0.0008	-0.0011
		(0.0010)	(0.0006)	(0.0010)	(0.0011)
	Temptation goods	-0.0011	-0.0022	-0.0011	0.0002
		(0.0026)	(0.0019)	(0.0026)	(0.0034)

Notes: This table replicates Column 1 of Table 2.4, and then estimates the same ATE based on specifications where the maximum radius is imposed to be at $R \in [2\text{km}, 6\text{km}]$. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Third, we do the analogous exercise temporally, estimating Equation (2.4) for lag structures of up to $M = 18$ months while fixing the maximum radius at $R = 4$km. We then calculate the cumulative effect of a shock of 100% of monthly GDP in each month up to L months. Figure B.8.3 shows that prices adjust rapidly. We cannot reject that lags of treatment beyond 3 months have no additional effect on prices, and the elasticity of prices with respect to transfers as a share of GDP stabilizes around 0.1.

The economic implication is that any part of the price response we miss using our pre-specified algorithm to select a temporal horizon does not make a meaningful difference quantitatively. In the most intense 12 month period, 4.5% of annual GDP was transferred to the 0 - 4km buffer around the average market (see Figure B.8.2). With an elasticity of 0.1, this implies a price effect of 0.4% in the average market over the most intense period of transfers (nearly identical to the 0.4% we arrive at using our pre-specified algorithm).

Figure B.8.3: Cumulative price effects



Notes: The figure is based on estimating Equation 2.5, where we impose a maximum lag or price effects up to $M = 18$ months and a maximum spatial radius of $R = 4$km. We then calculate the cumulative effect of a shock of 100% of monthly GDP in each month over L months on the overall logarithmic price index ($= \sum_{r=2km}^{R=4km} \sum_{l=0}^{L} \hat{\beta}_{rl}$.). Confidence intervals are as in Conley (2008) and we allow for spatial correlation up to 10km and autocorrelation up to 12 months.

IV specification for market price effects

In our preferred estimates we identify effects on market prices using a different method than the IV strategy we use for identifying effects on households and firms (i.e. specifications 2.2, 2.3 and B.2). This is because, unlike for households and firms, we have rich panel data on market prices including prices both before and after treatment onset. This lets us identify effects using a difference-in-differences design (as described in section 2.3) and leveraging the random roll-out of transfers into different buffers around each market over time. Specifically, conditional on market fixed effects, which control for the share of eligible households around each market as well as other time-invariant characteristics of markets, treatment roll-out is exogenous.

For comparability, however, we can also estimate price effects using the same IV strategy we use for firm and household outcomes. Concretely, we run the same pre-specified radii and buffer selection algorithm (specifications 2.4 and 2.5) based on the BIC for each price effect. But instead of including market fixed effects and $Amt_{m(t-l),r}$, we drop market fixed effects, and instrument $Amt_{m(t-l),r}$ with share $s^{e,t}_{m,r}$ of eligible households assigned to treatment in the buffer r around market m multiplied by the share of transfers going to that buffer in month $t - l$. Note that this approach is analogous to our dynamic IV specification 2.7 we use to estimate the impulse response functions for flow variables underlying our multiplier estimates. After selecting radii $\bar{R}$ and lags L to be included for each price index, we run:

$$p_{mt} = \sum_{r=2}^{\bar{R}} \sum_{l=0}^{L} \beta_{rl} Amt_{m(t-l),r} + \lambda_t + \varepsilon_{mt} \tag{B.3}$$

Table B.8.10 reports average and average maximum effects as in Table 2.4 resulting from this strategy. Effects are broadly in line with those in our main specification. Although none of the effects are statistically significant when using the IV strategy, we are still able to reject large effects on prices.

Enterprise price analyses

In addition to prices collected as part of our market price surveys, we also collected some price data as part of our enterprise surveys. We make use of enterprise price data collected via seven rounds of phone surveys of enterprises between August 2015 and June 2016. These surveys were conducted with four types of enterprises: small retailers, hardware stores, maize grinders, and tailors. We focus on prices for services provided by the latter two, as hardware and retail prices are well-covered by our market price data. To ensure consistent quality, unit size and availability we collected prices for a small number of services these enterprises commonly provide. In particular, we focus on the price of grinding 1kg of maize at a posho mill, and for patching a small hole at a tailor shop.

Phone surveys overlapped with an intense period of treatment rollout. During those 11 months the share of overall transfers sent went from 52% to 92%, and the variation in transfers was substantial, both across space and time: The 10-90 percentile range of per capita GDP transferred within 2km of a village over the period is $[0.1\%, 7.8\%]$, and the average village experienced 1.5% of GDP more inflows in the most intense month compared to the least intense month.

We analyze these prices analogously to our market prices, running the following specification:

$$p_{evt} = \sum_{r}^{\bar{R}} \sum_{l=0}^{M} \beta_{rl} Amt_{v(t-l),r} + \alpha_v + \lambda_t + \varepsilon_{evt} \tag{B.4}$$

where p_{evt} is the logarithm of the price from enterprise e in village v in month t, α_v are village fixed effects, γ_t are month fixed effects. We select the included radii bands $\bar{R}$ and the number of treatment lags M using the same pre-specified algorithm as for market prices. Table B.8.11 reports the average treatment effect across the intervention period (ATE) as well as the average maximum effect across villages (AME) from the optimal specification, and investigates heterogeneity by market access (see Section 2.3 for details on the methodology).

We find limited effects on these two selected services, with magnitudes in the range of product-specific effects for our our market price measures. Tailoring prices rise by 0.02% on average, and 0.1% in the month of most intense transfer, though those coefficients are not statistically significant. As with market prices, the effects are concentrated in more remote areas. Maize grinding prices fall, if anything, but the estimated effects are not statistically significantly different from zero.

Table B.8.10: Output Prices - IV Specification

		(1)	(2)	(3)	(4)
		Overall Effects		**ATE by market access (in %)**	
		ATE	Average maximum effect (AME)	below median	above median
All goods		0.0017	0.0079	0.0007	0.0021
		(0.0012)	(0.0063)	(0.0013)	(0.0017)
By tradability	More tradable	0.0020	0.0082	0.0018	0.0026
		(0.0020)	(0.0103)	(0.0028)	(0.0022)
	Less tradable	0.0016	0.0079	0.0003	0.0020
		(0.0014)	(0.0075)	(0.0016)	(0.0022)
By sector	Food items	0.0017	0.0084	0.0003	0.0022
		(0.0015)	(0.0078)	(0.0016)	(0.0023)
	Non-durables	0.0019	0.0080	0.0020	0.0024
		(0.0021)	(0.0111)	(0.0031)	(0.0024)
	Durables	0.0029*	0.0110	-0.0005	0.0043*
		(0.0017)	(0.0103)	(0.0019)	(0.0025)
	Livestock	-0.0008	-0.0046	0.0009	-0.0028
		(0.0014)	(0.0077)	(0.0010)	(0.0023)
	Temptation goods	-0.0026	-0.0190	-0.0046	-0.0017
		(0.0028)	(0.0156)	(0.0041)	(0.0039)

Notes: Each row represents an IV regression of the logarithm of a price index on the "optimal" number of lags and distance buffers of per capita Give Directly transfers in each buffer (Equation B.3). Price indices are based on 321,628 non-missing price quotes for 70 commodities and products. For each product, we take the logarithm of the median price quote in a market-month, and create our market price indices as an expenditure weighted average of these median price quotes across all goods in that market-month. Regressions include a panel of 1,734 market-by-month observations. The number of radii bands and lags is chosen sequentially by minimizing the BIC, as pre-specified, for the overall price index, which selects 2km. Treatment amounts for each buffer-month are instrumented by the share of eligible households assigned to treatment in that buffer, multiplied by the share of all transfers in that buffer going out in that month. Regressions include a full set of month fixed effects. Column 1 reports the implied ATE, calculated by evaluating the "optimal" regression specification at the average level of treatment intensity between September 2014 and March 2017, the time during which transfers went out. Column 2 reports the average maximum effect, calculated at the average across all markets of the month in which the largest per capita transfers went into a market's neighborhood (up to the largest buffer selected by the algorithm). Columns 3 and 4 break down the ATE by market access, defined as $MA_m = \sum_{r=1}^{10} r^{-\theta} N_r$, where $\theta = 8$ and N_r is the population in in the $r-2$ to r km buffer around each market. Standard errors (in parentheses) are as in Conley (2008) and we allow for spatial correlation up to 10km and autocorrelation up to 12 months. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.8.11: Local manufacturing and services prices

	(1)	(2)	(3)	(4)
	Overall Effects		**ATE by market access**	
	ATE	Average maximum effect (AME)	below median	above median
Tailor, patch small hole	-0.0414***	-0.0387	-0.0423**	-0.0330**
	(0.0154)	(0.0329)	(0.0197)	(0.0153)
Posho mill, grind 1kg of maize	-0.0114***	-0.0316***	-0.0149***	-0.0105***
	(0.0025)	(0.0069)	(0.0043)	(0.0037)

Notes: Each row represents a regression of the logarithm of a price on the “optimal” number of lags and distance buffers of per capita GiveDirectly transfers in each buffer (EquationB.4). We include 2,347 monthly price observations for tailors (simple patch), and 4,577 observations from posho mills (grinding 1kg of maize) collected between Aug 15 - Jun 16, around the time of peak transfer intensity. The number of radii bands and lags is chosen sequentially by minimizing the BIC, as pre-specified. Regressions include a full set of market and month fixed effects. Column 1 reports the implied ATE, calculated by evaluating the “optimal” regression specification at the average level of treatment intensity between September 2014 and March 2017, the time during which transfers went out. Column 2 reports the average maximum effect, calculated at the average across all markets of the month in which the largest per capita transfers went into a market's neighborhood (up to the largest buffer selected by the algorithm). Columns 3 and 4 break down the ATE by market access, defined as $MA_m = \sum_{r=1}^{10} r^{-\theta} N_r$, where $\theta = 8$ and N_r is the population in in the $r-2$ to r km buffer around each market. Regressions are weighted by inverse sampling weights. Standard errors (in parentheses) are as in Conley (2008) and we allow for spatial correlation up to 10km and autocorrelation up to 3 months. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

B.9 Robustness to alternative spatial modelling approaches

In this section we examine the robustness of our statistical inferences and overall conclusions to several alternative ways of dealing with the issue of radius selection in our spatial models.

Fixed radii

We first examine results holding the spatial radius fixed at 2km (the maximal radius our BIC procedure usually selects) as well as at 4km and 6km. Generally speaking we expect to capture additional spillovers using the larger radii that we might miss at the shorter 2km radius, but at the cost of precision. Tables B.9.1, B.9.2, and B.9.3 mirror our main Tables 2.1, 2.2 and 2.3 but using this approach. In each table Column 1 reproduces our main estimate while Columns 2, 3 & 4 report estimates using fixed 2km, 4km and 6km radii, all for recipient households (or in Table B.9.3 for treated villages). Columns 5-8 then repeat this exercise for non-recipient households and control villages.

At a fixed 2km radius our results are (not surprisingly) similar if not identical to those from our default specification, but with the advantage that inference post-model selection is less of a concern. At higher radii the point estimates are generally similar to (or in some cases larger than) our benchmark estimates, though as expected the precision of our estimates decreases at higher radii. In almost all cases we cannot reject that the fixed-radius estimates are the same as our benchmark estimates.

Split-sample estimation

We next examine robustness to selecting a radius and estimating coefficients using different splits of the data. Specifically, we select 200 random 50-50 splits of our data, stratified by village treatment assignment and (for households) eligibility status, into training and estimation samples. For each split we use the training sample to select a radius and the estimation sample to estimate parameters. We repeat this exercise, using the estimation sample as the training sample and vice-versa. We record the proportion of splits in which we calculate the same optimal radius band as when using our full dataset; we take the mean of the two point estimates and report the proportion of cases in which the resulting estimate effect lies within the 95% confidence interval reported in the paper.

Tables B.9.4, B.9.5, and B.9.6 presents results for the outcomes found in Tables 2.1, 2.2 and 2.3, respectively. For Tables B.9.4 and B.9.5, columns 1 and 2 reproduce the estimates and radii selection for recipient households. Column 3 reports the fraction of these splits that produce estimates for non-recipients falling into the 95%-CI of the initial estimate, and Column 4 the proportion that select the same radius as when using the full dataset. Columns 5-8 do the same for non-recipient households. Note that as in producing our main estimates we do not separately estimate an optimal radius for subcomponents of larger totals or indices. For enterprises, TableB.9.6 columns 1 and 3 reproduce the main estimates for treatment and control villages, respective. Columns 2 and 4 report the share of mean estimates falling within the 95%-CI of the initial estimate. As we use a common radius for treatment and control villages, column 5 reports the selected radius, and column 6 reports the share selecting the same radius.

Overall we see congruence between the full data and the subsamples regarding the optimal radius over which to estimate effects, with most agreement rates in the 90%s. We also see good coverage, with 95% or more of the mean replicate point estimates falling within our original 95% confidence interval in most cases.

Heterogeneous radii

We next examine whether our BIC algorithm selects different maximal radii for different geographic sub-groups of villages. Specifically, we (i) allow the BIC to select a different radius for markets with above versus below median market access, and then (ii) allow the BIC to select a different radius for each of the three sub-counties in which our study is set. Tables B.9.7, B.9.8 and B.9.9 report the maximal radius selected in each case, with the full sample radius selected for comparison. For enterprise results, optimal radii bands were selected only once for each outcome across treatment and control villages, as the enterprises were not direct recipients of the cash transfers. In the great majority of cases we end up selecting the same radius (which in almost every case is 2km). Specifically, out of 190 radii selected (15 outcomes * 2 treatment status groups * 5 geographic subgroups for households + 8 outcomes * 5 geographic subgroups for enterprises) we select a different radius than in the corresponding pooled approach 10 times, or 5.3% of the total. Overall we conclude that, while there are surely are differences in the relevant radii or more generally the relevant "catchment areas" for different units, our data do not reveal systematic differences.

Randomization Inference

Finally, we examine the sensitivity of our conclusions to randomization inference. This approach sidesteps concerns about model selection; we simply interpret the coefficients we obtain from the entire model selection and estimation procedure as a statistic whose distribution should be invariant to reassignments of treatment and control status under the null of no treatment effects for any unit. Specifically, we generate 500 replicates in each of which we re-assign each village and household's treatment status using the same algorithm with which actual treatment was assigned, recalculate the our derived spatial exposure measures using these assignments, and then re-estimate total effects.

Tables B.9.10, B.9.11 and B.9.12 report results for the outcomes in Tables 2.1, 2.2 and 2.3, respectively. Table B.9.13 does the same but also simulating the randomized rollout of the transfer program in order to conduct randomization inference for output price outcomes in Table 2.4. Randomization inference yields very similar substantive conclusions to our main analysis, rejecting the null of no treatment effects for almost exactly the same outcomes as our main tests reject the null of no average effect.

Table B.9.1: Robustness to fixing alternative radii bands: Expenditures, Savings and Income

	Recipient households				Non-recipient households				
	(1) Total Effect IV Optimal Radius	(2) Total Effect IV $\bar{R} = 2$	(3) Total Effect IV $\bar{R} = 4$	(4) Total Effect IV $\bar{R} = 6$	(5) Total Effect IV Optimal Radius	(6) Total Effect IV $\bar{R} = 2$	(7) Total Effect IV $\bar{R} = 4$	(8) Total Effect IV $\bar{R} = 6$	(9) Control, low-saturation mean (SD)
Panel A: Expenditure									
Household expenditure, annualized	338.57 ***	338.57 ***	517.77 ***	640.22 *	334.77 ***	334.77 ***	452.77 **	-29.04	2,536.01
	(110.07)	(110.07)	(168.12)	(386.23)	(120.36)	(120.36)	(208.75)	(445.10)	(1,933.51)
Non-durable expenditure, annualized	227.20 **	227.20 **	461.10 ***	472.53	317.62 ***	317.62 ***	422.19 **	-120.07	2,470.69
	(98.83)	(98.83)	(158.17)	(379.85)	(117.29)	(117.29)	(201.68)	(431.20)	(1,877.23)
Food expenditure, annualized	133.84 **	133.84 **	323.88 ***	218.26	133.30 **	133.30 **	152.05	-79.54	1,578.05
	(64.01)	(64.01)	(90.54)	(196.81)	(61.62)	(61.62)	(104.21)	(212.30)	(1,072.00)
Temptation goods expenditure, annualized	5.91	5.91	14.95	28.37	-0.68	-0.68	8.31	-17.10	37.07
	(9.20)	(9.20)	(12.93)	(26.82)	(6.60)	(6.60)	(11.98)	(21.57)	(123.54)
Durable expenditure, annualized	109.01 ***	109.01 ***	59.78***	147.72 ***	8.44	8.44	21.52	104.00 **	59.41
	(19.93)	(19.93)	(22.87)	(38.78)	(12.32)	(12.32)	(15.91)	(40.43)	(230.83)
Panel B: Assets									
Assets (non-land, non-house), net borrowing	183.38 ***	183.38 ***	205.21 ***	290.28 **	133.06	133.06	246.04 *	125.17	1,131.66
	(45.59)	(45.59)	(62.84)	(141.12)	(81.79)	(81.79)	(149.40)	(260.24)	(1,419.70)
Housing value	477.29 ***	477.29 ***	477.45 ***	631.26 ***	80.65	80.65	558.21	-389.59	2,032.11
	(37.66)	(37.66)	(49.91)	(117.76)	(205.76)	(205.76)	(492.54)	(840.52)	(5,028.27)
Land value	158.47	158.47	496.77	206.65	544.85	544.85	810.44	-320.21	5,030.03
	(239.20)	(239.20)	(353.56)	(651.37)	(456.42)	(456.42)	(666.42)	(1,380.27)	(6,604.66)

Table B.9.1: Robustness to fixing alternative radii bands: Expenditures, Savings and Income (continued)

	Recipient households				**Non-recipient households**				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Total Effect IV Optimal Radius	Total Effect IV $\bar{R} = 2$	Total Effect IV $\bar{R} = 4$	Total Effect IV $\bar{R} = 6$	Total Effect IV Optimal Radius	Total Effect IV $\bar{R} = 2$	Total Effect IV $\bar{R} = 4$	Total Effect IV $\bar{R} = 6$	Control, low-saturation mean (SD)
Panel C: Household balance sheet									
Household income, annualized	135.70	135.70	223.93	382.17	224.96 ***	224.96 ***	308.76 **	174.25	1,023.36
	(94.80)	(94.80)	(156.96)	(308.93)	(84.57)	(84.57)	(151.26)	(280.17)	(1,634.02)
Net value of household transfers received, annualized	-7.43	-7.43	-6.50	-21.58	8.85	8.85	57.38*	7.67	130.08
	(12.58)	(12.58)	(14.81)	(30.57)	(20.11)	(20.11)	(31.80)	(60.42)	(263.65)
Tax paid, annualized	-0.09	-0.09	0.19	3.59	1.68	1.68	2.81	0.99	16.92
	(2.24)	(2.24)	(3.02)	(4.92)	(2.15)	(2.15)	(3.97)	(6.80)	(36.50)
Profits (ag & non-ag), annualized	35.85	35.85	83.63	149.22	36.37	36.37	70.41	92.50	485.56
	(50.17)	(50.17)	(87.40)	(148.49)	(41.55)	(41.55)	(71.04)	(132.44)	(786.92)
Wage earnings, annualized	73.66	73.66	96.72	201.72	182.63 ***	182.63 ***	180.62	-11.58	494.95
	(63.70)	(63.70)	(99.14)	(198.47)	(66.77)	(66.77)	(126.57)	(246.51)	(1,231.12)

Notes: Columns 1 and 5 replicate columns 2 and 3 from Table 2.1, selecting the number of radii bands included using our pre-specified algorithm as described in Section 2.3. The optimal radius selected is 2km for all outcomes. Columns 2-4 estimate the Total Effect (IV) for treated households, imposing a maximum radius $\bar{R}$ of 2, 4 and 6km respectively. Similarly, Columns 6-8 replicate Column 5, while including more spatial buffers. Standard errors are calculated following Conley (2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.9.2: Robustness to fixing alternative radii bands: Input Prices and Quantities

	Recipient households				**Non-recipient households**				
	(1) Total Effect IV Optimal Radius	(2) Total Effect IV $\bar{R} = 2$	(3) Total Effect IV $\bar{R} = 4$	(4) Total Effect IV $\bar{R} = 6$	(5) Total Effect IV Optimal Radius	(6) Total Effect IV $\bar{R} = 2$	(7) Total Effect IV $\bar{R} = 4$	(8) Total Effect IV $\bar{R} = 6$	(9) Control, low-saturation mean (SD)
Panel A: Labor									
Hourly wage earned by employees	0.04	0.04	0.04	0.14	0.19*	0.19*	0.02	-0.11	0.70
	(0.06)	(0.06)	(0.08)	(0.14)	(0.11)	(0.11)	(0.18)	(0.31)	(0.89)
Household total hours worked, last 7 days	1.41	1.41	5.77	10.99	-4.69	-4.69	-3.13	-4.71	63.19
	(3.83)	(3.83)	(5.35)	(10.23)	(3.09)	(3.09)	(5.32)	(11.76)	(54.12)
Panel B: Land									
Land price per acre	366.46	366.46	696.58	806.35	557.44	557.44	558.25	-232.42	3,952.48
	(293.20)	(293.20)	(537.25)	(1,090.38)	(393.60)	(393.60)	(856.97)	(1,677.59)	(3,147.29)
Acres of land owned	-0.10	-0.10	0.02	0.10	0.08	0.08	0.12	0.04	1.42
	(0.09)	(0.09)	(0.16)	(0.55)	(0.11)	(0.11)	(0.16)	(0.39)	(2.37)
Panel C: Capital									
Loan-weighted interest rate, monthly	0.01	0.01	0.04**	0.02	-0.01	-0.01	-0.02	0.00	0.06
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.04)	(0.07)
Total loan amount	3.12	3.12	20.52	59.30**	6.12	6.12	29.06	35.04	80.57
	(9.07)	(9.07)	(12.88)	(25.10)	(12.71)	(12.71)	(20.02)	(35.27)	(204.28)

Notes: Columns 1 and 5 replicate Columns 2 and 3 from Table 2.2, selecting the number of radii bands included using our pre-specified algorithm as described in Section 2.3. The optimal radius selected is 2km for all outcomes for both recipients and non-recipientsand groups. Columns 2, 3 and 4 estimate the Total Effect (IV) for treated households, imposing a maximum radius R of 2, 4 and 6km respectively. Similarly, Columns 6-8 replicate Column 5, while including more spatial buffers. Standard errors are calculated following Conley (2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.9.3: Robustness to fixing alternative radii bands: Enterprise Outcomes

	Treatment Villages				Control Villages				
	(1) Total Effect IV Optimal Radius	(2) Total Effect IV $\bar{R} = 2$	(3) Total Effect IV $\bar{R} = 4$	(4) Total Effect IV $\bar{R} = 6$	(5) Total Effect IV Optimal Radius	(6) Total Effect IV $\bar{R} = 2$	(7) Total Effect IV $\bar{R} = 4$	(8) Total Effect IV $\bar{R} = 6$	(9) Control, low-saturation mean (SD)
Panel A: All enterprises									
Enterprise profits, annualized	55.77	55.77	118.91 *	111.12	35.08	35.08	93.35	85.57	156.79
	(36.73)	(36.73)	(69.27)	(135.12)	(37.36)	(37.36)	(67.06)	(124.24)	(292.84)
Enterprise revenue, annualized	322.16 **	322.16 **	644.21 ***	659.36	237.16 **	237.16 **	532.19 ***	542.03	494.45
	(138.17)	(138.17)	(230.99)	(420.80)	(112.72)	(112.72)	(203.31)	(373.79)	(1,223.07)
Enterprise costs, annualized	89.35**	89.35**	75.50	-88.60	73.08	73.08	59.68	-97.38	117.22
	(38.51)	(38.51)	(59.79)	(125.46)	(46.77)	(46.77)	(64.27)	(120.78)	(263.46)
Enterprise wagebill, annualized	75.99**	75.99**	60.31	-58.69	66.57*	66.57*	54.14	-50.04	97.35
	(30.64)	(30.64)	(49.57)	(108.97)	(35.86)	(35.86)	(48.66)	(91.57)	(237.01)
Enterprise profit margin	-0.11*	-0.07**	-0.11*	-0.09	-0.12**	-0.07***	-0.12**	-0.09	0.33
	(0.06)	(0.03)	(0.06)	(0.12)	(0.05)	(0.02)	(0.05)	(0.12)	(0.30)
Panel B: Non-agricultural enterprises									
Enterprise inventory	34.69***	34.69***	33.41*	-41.10	16.90	16.90	16.09	-49.22	50.41
	(13.39)	(13.39)	(17.75)	(44.75)	(10.66)	(10.66)	(13.86)	(38.27)	(131.86)
Enterprise investment, annualized	13.58	13.58	9.47	-22.42	6.82	6.82	3.22	-24.51	46.57
	(13.10)	(13.10)	(22.93)	(42.65)	(7.96)	(7.96)	(16.33)	(35.84)	(167.44)
Panel C: Village-level									
Number of enterprises	0.02	0.02	0.02	-0.04	0.01	0.01	0.01	-0.05	1.12
	(0.01)	(0.01)	(0.02)	(0.04)	(0.01)	(0.01)	(0.02)	(0.04)	(0.14)

Notes: Columns 1 and 5 replicate Columns 2 and 3 from Table 2.3, selecting the number of radii bands included using our pre-specified algorithm as described in Section 2.3. The optimal radius selected is 2km for all outcomes and groups. Columns 2, 3 and 4 estimate the Total Effect (IV) for enterprises in treated villages, imposing a maximum radius R of 2, 4 and 6km respectively. Similarly, Columns 5 and 6 replicate Column 4, while including more spatial buffers. Standard errors are calculated following Conley (2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.9.4: BIC split sample approach for household expenditure, savings and income outcomes

	Recipient Households				Non-Recipient Households			
	Main Estimate		200 Split Sets		Main Estimate		200 Split Sets	
	(1) Total Effect (IV)	(2) Selected Radius	(3) Share in 95% CI	(4) Share selecting same radius	(5) Total Effect (IV)	(6) Selected Radius	(7) Share in 95% CI	(8) Share selecting same radius
Panel A: Expenditure								
Household expenditure, annualized	338.57 ***	2 km	100 %	100 %	334.77 ***	2 km	100 %	100 %
	(110.07)				(120.36)			
Non-durable expenditure, annualized	227.20 **	2 km	100 %	100 %	317.62 ***	2 km	100 %	100 %
	(98.83)				(117.29)			
Food expenditure, annualized	133.84 **	2 km	100 %	100 %	133.30 **	2 km	100 %	100 %
	(64.01)				(61.62)			
Temptation goods expenditure, annualized	5.91	2 km	100 %	100 %	-0.68	2 km	100 %	100 %
	(9.20)				(6.60)			
Durable expenditure, annualized	109.01 ***	2 km	100 %	100 %	8.44	2 km	100 %	100 %
	(19.93)				(12.32)			
Panel B: Assets								
Assets (non-land, non-house), net borrowing	183.38 ***	2 km	100 %	100 %	133.06	2 km	100 %	100 %
	(45.59)				(81.79)			
Housing value	477.29 ***	2 km	100 %	91 %	80.65	2 km	100 %	99 %
	(37.66)				(205.76)			
Land value	158.47	2 km	100 %	100 %	544.85	2 km	100 %	100 %
	(239.20)				(456.42)			
Panel C: Household balance sheet								
Household income, annualized	135.70	2 km	100 %	100 %	224.96 ***	2 km	99 %	99 %
	(94.80)				(84.57)			
Net value of household transfers received, annualized	-7.43	2 km	100 %	100 %	8.85	2 km	100 %	100 %
	(12.58)				(20.11)			
Tax paid, annualized	-0.09	2 km	100 %	100 %	1.68	2 km	100 %	100 %
	(2.24)				(2.15)			
Profits (ag & non-ag), annualized	35.85	2 km	100 %	100 %	36.37	2 km	100 %	100 %
	(50.17)				(41.55)			
Wage earnings, annualized	73.66	2 km	100 %	100 %	182.63 ***	2 km	100 %	92 %
	(63.70)				(66.77)			

Notes: Columns 1 and 2 reproduce the total effect estimates and radii selection for recipient households in Table 2.1. Columns 3 and 4 show the fraction of 200 random 50-50 splits of the sample for which the mean point estimates (across the two estimation samples within each split) for recipients falling into the 95%-CI of the initial estimate (Column 3) and select the same radius as the initial selection (Column 4). Columns 5 - 8 do the same for non-recipient households. See Table 2.1 for more details on variable construction and regression specification. Standard errors in Columns 1 and 5 are calculated following Conley (2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.9.5: BIC split sample approach for input prices and quantities

	Recipient Households				Non-Recipient Households			
	Main Estimate		200 Split Sets		Main Estimate		200 Split Sets	
	(1) Total Effect (IV)	(2) Selected Radius	(3) Share in 95% CI	(4) Share selecting same radius	(5) Total Effect (IV)	(6) Selected Radius	(7) Share in 95% CI	(8) Share selecting same radius
Panel A: Labor								
Hourly wage earned by employees	0.04	2 km	100 %	100 %	0.19*	2 km	100 %	91 %
	(0.06)				(0.11)			
Household total hours worked, last 7 days	1.41	2 km	100 %	100 %	-4.69	2 km	100 %	100 %
	(3.83)				(3.09)			
Panel B: Land								
Land price per acre	366.46	2 km	100 %	99 %	557.44	2 km	100 %	95 %
	(293.20)				(393.60)			
Acres of land owned	-0.10	2 km	100 %	99 %	0.08	2 km	100 %	100 %
	(0.09)				(0.11)			
Panel C: Capital								
Loan-weighted interest rate, monthly	0.01	2 km	99 %	100 %	-0.01	2 km	100 %	97 %
	(0.01)				(0.01)			
Total loan amount	3.12	2 km	100 %	100 %	6.12	2 km	100 %	100 %
	(9.07)				(12.71)			

Notes: Columns 1 and 2 reproduce the total effect estimates and radii selection for recipient households in Table 2.2. Columns 3 and 4 show the fraction of 200 random 50-50 splits of the sample for which the mean point estimates (across the two estimation samples within each split) for recipients falling into the 95%-CI of the initial estimate (Column 3) and select the same radius as the initial selection (Column 4). Columns 5 - 8 do the same for non-recipient households. See Table 2.2 for more details on variable construction and regression specification. Standard errors in Columns 1 and 5 are calculated following Conley (2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.9.6: BIC split sample approach for enterprise outcomes

	Treatment Villages		Control Villages		Radii Selected	
	(1) Total Effect (IV)	(2) Share in 95% CI	(3) Total Effect (IV)	(4) Share in 95% CI	(5) Main Estimate	(6) Share selecting same radius
Panel A: All enterprises						
Enterprise profits, annualized	55.77 (36.73)	2 km	100 %	92 %	35.08 (37.36)	2 km
Enterprise revenue, annualized	322.16 ** (138.17)	2 km	100 %	97 %	237.16 ** (112.72)	2 km
Enterprise costs, annualized	89.35** (38.51)	2 km	90 %	86 %	73.08 (46.77)	2 km
Enterprise wagebill, annualized	75.99** (30.64)	2 km	89 %	82 %	66.57* (35.86)	2 km
Enterprise profit margin	-0.11* (0.06)	4 km	97 %	21 %	-0.12** (0.05)	4 km
Panel B: Non-agricultural enterprises						
Enterprise inventory	34.69** (13.39)	2 km	100 %	100 %	16.90 (10.66)	2 km
Enterprise investment, annualized	13.58 (13.10)	2 km	100 %	100 %	6.82 (7.96)	2 km
Panel C: Village-level						
Number of enterprises	0.02 (0.01)	2 km	89 %	89 %	0.01 (0.01)	2 km

Notes: Columns 1 and 3 reproduce the total effect estimates for enterprises located in treatment and control villages from Table 2.3. Column 5 reports the radii selection (which is done across all enterprises jointly, since enterprises are not direct recipients of cash transfers). Columns 2 and 4 show the fraction of 200 random 50-50 splits of the sample for which the mean point estimates (across the two estimation samples within each split) produce estimates falling into the 95%-CI of the initial estimates in Columns 1 and 3 respectively. Column 6 shows the share of these splits where the algorighm selects the same radius as the initial selection in Column 5. See Table 2.3 for more details on variable construction and regression specification. Standard errors in Columns 1 and 3 are calculated following Conley (2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table B.9.7: Maximum Radius Chosen by the BIC Algorithm (in km), expenditure, saving and income outcomes

	Recipient Households						**Non-recipient Households**					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Full Sample	low market access	high market access	Alego	Ugunja	Ukwala	Full Sample	low market access	high market access	Alego	Ugunja	Ukwala
Panel A: Expenditure												
Household expenditure, annualized	2	2	2	2	2	2	2	2	2	2	2	2
Panel B: Assets												
Assets (non-land, non-house), net borrowing	2	2	2	2	2	2	2	2	2	2	2	2
Housing value	2	2	2	2	**4**	**4**	2	2	2	2	2	2
Land value	2	2	2	2	2	2	2	2	2	2	2	2
Panel C: Household balance sheet												
Household income, annualized	2	2	2	2	2	2	2	2	2	2	2	2
Net value of household transfers received, annualized	2	2	2	2	2	2	2	2	2	2	2	2
Tax paid, annualized	2	2	2	2	2	2	2	2	2	2	2	2
Profits (ag & non-ag), annualized	2	2	2	2	2	2	2	2	2	2	2	2
Wage earnings, annualized	2	2	2	2	2	2	2	2	2	2	2	2

Notes: This table reports the maximum radius selected minimizing a Baysian Information Criterion within different sub-samples of our data for outcomes in Table 2.1. Columns 1-6 report the radius selected when estimating effects on recipient households and when using the full sample (Column 1), when restricting the sample to villages with higher and lower than median market access (Columns 2 and 3), and to households in the three sub-counties Alego, Ugunja and Ukwala (Columns 4 - 6). Columns 7-12 report the same for non-recipient households.

Table B.9.8: Maximum Radius Chosen by the BIC Algorithm (in km), input prices and quantities

	Recipient Households						**Non-recipient Households**					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Full Sample	low market access	high market access	Alego	Ugunja	Ukwala	Full Sample	low market access	high market access	Alego	Ugunja	Ukwala
Panel A: Labor												
Hourly wage earned by employees	2	2	2	2	**4**	2	2	2	2	2	2	2
Household total hours worked, last 7 days	2	2	2	2	2	2	2	2	2	2	2	2
Panel B: Land												
Land price per acre	2	2	2	2	**16**	2	2	2	2	2	2	2
Acres of land owned	2	2	2	2	2	2	2	2	2	2	2	2
Panel C: Capital												
Loan-weighted interest rate, monthly	2	2	2	2	2	2	2	2	2	2	2	**4**
Total loan amount	2	2	2	2	2	2	2	2	2	2	2	2

Notes: This table reports the maximum radius selected minimizing a Baysian Information Criterion within different sub-samples of our data for outcomes in Table 2.2. Columns 1-6 report the radius selected when estimating effects on recipient households and when using the full sample (Column 1), when restricting the sample to villages with higher and lower than median market access (Columns 2 and 3), and to households in the three sub-counties Alego, Ugunja and Ukwala (Columns 4 - 6). Columns 7-12 report the same for non-recipient households.

Table B.9.9: Maximum Radius Chosen by the BIC Algorithm (in km), enterprise outcomes

	Market Access	**Subcounty**				
	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample	low market access	high market access	Alego	Ugunja	Ukwala
Panel A: All enterprises						
Enterprise profits, annualized	2	**4**	2	2	**20**	2
Enterprise revenue, annualized	2	**4**	2	2	2	2
Enterprise costs, annualized	2	**8**	2	2	2	2
Enterprise wagebill, annualized	2	2	2	2	2	2
Enterprise profit margin	4	**2**	4	**2**	**16**	4
Panel B: Non-agricultural enterprises						
Enterprise inventory	2	**8**	2	2	2	2
Enterprise investment, annualized	2	2	2	2	2	2
Panel C: Village-level						
Number of enterprises	2	2	**6**	2	2	2

Notes: This table reports the maximal radius selected by minimizing a Baysian Information Criterion within different sub-samples of our data for enterprise outcomes in Table 2.3. We report the radius selected when using the full sample (Column 1), when restricting the sample to villages with higher and lower than median market access (Columns 2 and 3), and to enterprises in the three sub-counties Alego, Ugunja and Ukwala (Columns 4 - 6).

Table B.9.10: Randomization inference for expenditure, savings and income outcomes

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	Total Effect IV	Spatial RI p-value	Total Effect IV	Spatial RI p-value
Panel A: Expenditure				
Household expenditure, annualized	338.57 ***	[0.00]***	334.77 ***	[0.02]**
	(110.07)		(120.36)	
Non-durable expenditure, annualized	227.20 **	[0.03]**	317.62 ***	[0.02]**
	(98.83)		(117.29)	
Food expenditure, annualized	133.84 **	[0.07]*	133.30 **	[0.08]*
	(64.01)		(61.62)	
Temptation goods expenditure, annualized	5.91	[0.58]	-0.68	[0.91]
	(9.20)		(6.60)	
Durable expenditure, annualized	109.01 ***	[0.00]***	8.44	[0.43]
	(19.93)		(12.32)	
Panel B: Assets				
Assets (non-land, non-house), net borrowing	183.38 ***	[0.00]***	133.06	[0.23]
	(45.59)		(81.79)	
Housing value	477.29 ***	[0.00]***	80.65	[0.78]
	(37.66)		(205.76)	
Land value	158.47	[0.60]	544.85	[0.31]
	(239.20)		(456.42)	
Panel C: Household balance sheet				
Household income, annualized	135.70	[0.17]	224.96 ***	[0.08]*
	(94.80)		(84.57)	
Net value of household transfers received, annualized	-7.43	[0.50]	8.85	[0.58]
	(12.58)		(20.11)	
Tax paid, annualized	-0.09	[0.97]	1.68	[0.42]
	(2.24)		(2.15)	
Profits (ag & non-ag), annualized	35.85	[0.40]	36.37	[0.54]
	(50.17)		(41.55)	
Wage earnings, annualized	73.66	[0.36]	182.63 ***	[0.06]*
	(63.70)		(66.77)	

Notes: This table presents randomization inference results for outcomes in Table 2.1. Column 1 reproduces the total effect for recipient households (Column 2 in Table 2.1), and Column 3 reproduces the total effect for non-recipient households (Column 3 in Table 2.1). Columns 2 and 4 report randomization inference p-values from 500 repetitions (i.e. the share of replications where the estimated total effect is larger than the main estimate), where the cross-sectional distribution of cash distributed in radii bands across villages was randomly shuffled.

Table B.9.11: Randomization inference for input prices and quantities

	(1)	(2)	(3)	(4)
	Recipient Households		**Non-recipient Households**	
	Total Effect IV	Spatial RI p-value	Total Effect IV	Spatial RI p-value
Labor				
Hourly wage earned by employees	0.04	[0.56]	0.19*	[0.21]
	(0.06)		(0.11)	
Household total hours worked, last 7 days	1.41	[0.73]	-4.69	[0.15]
	(3.83)		(3.09)	
Land				
Land price per acre	366.46	[0.34]	557.44	[0.15]
	(293.20)		(393.60)	
Acres of land owned	-0.10	[0.64]	0.08	[0.56]
	(0.09)		(0.11)	
Capital				
Loan-weighted interest rate, monthly	0.01	[0.62]	-0.01	[0.62]
	(0.01)		(0.01)	
Total loan amount	3.12	[0.78]	6.12	[0.69]
	(9.07)		(12.71)	

Notes: This table presents randomization inference results for outcomes in Table 2.2. Column 1 reproduces the total effect for recipient households (Column 2 in Table 2.2), and Column 3 reproduces the total effect for non-recipient households (Column 3 in Table 2.2). Columns 2 and 4 report randomization inference p-values from 500 repetitions (i.e. the share of replications where the estimated total effect is larger than the main estimate), where the cross-sectional distribution of cash distributed in radii bands across villages was randomly shuffled.

Table B.9.12: Randomization inference for enterprise outcomes

	(1)	(2)	(3)	(4)
	Treated Villages		**Control Villages**	
	Total Effect IV	Spatial RI p-value	Total Effect IV	Spatial RI p-value
Panel A: All enterprises				
Enterprise profits, annualized	55.77	[0.27]	35.08	[0.41]
	(36.73)		(37.36)	
Enterprise revenue, annualized	322.16 **	[0.09]*	237.16 **	[0.08]*
	(138.17)		(112.72)	
Enterprise costs, annualized	89.35**	[0.06]*	73.08	[0.04]**
	(38.51)		(46.77)	
Enterprise wagebill, annualized	75.99**	[0.08]*	66.57*	[0.03]**
	(30.64)		(35.86)	
Enterprise profit margin	-0.11*	[0.06]*	-0.12**	[0.02]**
	(0.06)		(0.05)	
Panel B: Non-agricultural enterprises				
Enterprise inventory	34.69***	[0.07]*	16.90	[0.21]
	(13.39)		(10.66)	
Enterprise investment, annualized	13.58	[0.39]	6.82	[0.56]
	(13.10)		(7.96)	
Panel C: Village-level				
Number of enterprises	0.02	[0.26]	0.01	[0.71]
	(0.01)		(0.01)	

Notes: This table presents randomization inference results for outcomes in Table 2.3. Column 1 reproduces the total effect for enterprises in treated villages (Column 2 in Table 2.3), and Column 3 reproduces the total effect for enterprises in control villages (Column 3 in Table 2.3). Columns 2 and 4 report randomization inference p-values from 500 repetitions (i.e. the share of replications where the estimated total effect is larger than the main estimate), where the cross-sectional distribution of cash distributed in radii bands across villages was randomly shuffled.

Table B.9.13: Randomization inference for price outcomes

		(1)	(2)	(3)	(4)
		Overall Effects		**ATE by market access)**	
		ATE	Average maximum effect (AME)	below median	above median
All goods		0.0010*	0.0042	0.0017*	0.0007
		(0.0006)	(0.0031)	(0.0009)	(0.0007)
		[0.066]	[0.130]	[0.120]	[0.480]
By tradability	More tradable	0.0014	0.0062	0.0023	0.0021
		(0.0015)	(0.0082)	(0.0023)	(0.0018)
		[0.278]	[0.382]	[0.276]	[0.356]
	Less tradable	0.0009	0.0034	0.0015	0.0001
		(0.0006)	(0.0032)	(0.0011)	(0.0008)
		[0.264]	[0.436]	[0.514]	[0.832]
By sector	Food items	0.0009	0.0036	0.0016	0.0002
		(0.0006)	(0.0033)	(0.0012)	(0.0008)
		[0.232]	[0.410]	[0.516]	[0.748]
	Non-durables	0.0014	0.0061	0.0026	0.0019
		(0.0017)	(0.0089)	(0.0026)	(0.0019)
		[0.334]	[0.434]	[0.234]	[0.414]
	Durables	0.0019*	0.0070	-0.0009	0.0034**
		(0.0011)	(0.0061)	(0.0011)	(0.0016)
		[0.034]	[0.158]	[0.640]	[0.128]
	Livestock	-0.0008	-0.0027	-0.0008*	-0.0017
		(0.0010)	(0.0052)	(0.0004)	(0.0020)
		[0.258]	[0.544]	[0.290]	[0.506]
	Temptation goods	-0.0011	-0.0112	-0.0008	-0.0003
		(0.0026)	(0.0143)	(0.0036)	(0.0035)
		[0.720]	[0.568]	[0.772]	[0.950]

Notes: This table presents randomization inference results for outcomes in Table 2.4. Randomization inference p-values from 500 repetitions (i.e. the share of replications where the estimated total effect is larger than the main estimate) are reported in brackets. In each iteration, we randomly re-generate cross-sectional and temporal roll out of transfers according to our actual treatment randomization as described in Section 2.2.

B.10 Study pre-analysis plans

We filed a series of pre-analysis plans as part of this study. These include: i) Haushofer et al. (2017a), and a follow-up amendment outlining spillover analyses, both of which focus on household outcomes; ii) Haushofer et al. (2016), which covered midline market price and enterprise data; and iii) Haushofer et al. (2018), which focused on macroeconomic quantities of interest. All pre-analysis plans can be accessed on the AEA trial registry: https://www.socialscienceregistry.org/trials/505. In this paper, we focus on primary outcomes for households, enterprises and prices, collected as part of our baseline and endline household and enterprise censuses and surveys, as well as our midline market price surveys.

Less relevant to this paper are: i) Walker (2017), which forms the basis of Walker (2018) on local taxes and public goods; and ii) Haushofer et al. (2017b), which conducts a separate exercise to study potential transfer targeting.

In the interest of space, we do not present an exhaustive list of every outcome component and analysis mentioned across these pre-analysis plans. A supplemental appendix containing the full set of pre-specified outcomes for these plans is available online at https://osf.io/r5q6v/.

Table B.10.1 presents the 10 primary household outcomes that we pre-specified as part of a single table, including FDR q-values accounting for multiple testing across these ten outcomes. In addition to the specifications reported in the main tables, we also report the pooled saturation effect, the average effect of being in a high saturation sublocation across all eligibility and village types. As outlined in Section 2.3, we prefer our spatial estimates as they take advantage of the full variation in treatment intensity in our data, but present these saturation results for completeness.

To calculate the pooled saturation effect, we use coefficient estimates from the following equation:

$$\begin{aligned} y_{hvs,t=1} &= \beta_0 + \beta_1 T_{vs} + \beta_2 E_{hvs} + \beta_3 H_s + \beta_4 T_{vs} \times E_{hvs} + \beta_5 T_{vs} \times H_s \\ &+ \beta_6 E_{hvs} \times H_s + \beta_7 T_{vs} \times E_{hvs} \times H_s + \delta_1 y_{hvs,t=0} + \delta_2 M_{hvs} + \varepsilon_{ihvs}. \end{aligned} \tag{B.5}$$

Here, h indexes the household, v indexes the village, s indexes the sublocation, and t indicates whether the variable was measured at baseline or endline. T_{vs} is an indicator for households residing in a treated village, E_{hvs} is an indicator for whether the household is eligible for transfers, and H_s is an indicator for living in a high-saturation sublocation; $\times$ denote interaction terms. Standard errors are clustered at the saturation group level. The pooled saturation effect is then a weighted average of $\beta_3, \beta_5, \beta_6$ and β_7 using population weights of all households across high-saturation sublocations.

We make two additional notes. First, in Haushofer et al. (2016), we were not clear whether we would focus on a balanced panel of market survey data or an unbalanced panel. For simplicity, we present results using a unbalanced panel, but results are robust to using a balanced panel. Second, our reduced form equations cluster standard errors at the village level, as pre-specified in Haushofer et al. (2017a), but results are also robust to clustering at the sublocation level. (Both sets of results are available upon request).

Table B.10.1: Pre-specified primary outcomes, household welfare plan

	(1)	(2)	(3)	(4)	(5)
	Recipient Households		**Non-Recipient Households**		
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Pooled saturation effect	Control, low saturation mean (SD)
Assets (non-land, non-house), net borrowing	178.78 ***	183.38 ***	133.06	42.06	1,131.66
	(24.66)	(45.59)	(81.79)	(42.33)	(1,419.70)
	[0.00]***	[0.00]***	[0.21]	[0.48]	
Household expenditure, annualized	293.59 ***	338.57 ***	334.77 ***	138.26 *	2,536.01
	(60.11)	(110.07)	(120.36)	(71.29)	(1,933.51)
	[0.00]***	[0.01]***	[0.04]**	[0.25]	
Household income, annualized	79.43*	135.70	224.96 ***	110.43 *	1,023.36
	(43.80)	(94.80)	(84.57)	(57.65)	(1,634.02)
	[0.06]*	[0.18]	[0.04]**	[0.25]	
Household revenue, annualized	77.44	183.61 **	53.50	116.22 **	933.19
	(52.63)	(89.34)	(103.81)	(56.10)	(1,697.97)
	[0.11]	[0.12]	[0.37]	[0.25]	
Psychological well-being index	0.09***	0.12*	0.08	0.04	0.01
	(0.03)	(0.07)	(0.05)	(0.03)	(1.01)
	[0.01]**	[0.14]	[0.21]	[0.26]	
Health index	0.03	0.06	0.01	-0.01	0.03
	(0.03)	(0.07)	(0.05)	(0.03)	(1.01)
	[0.14]	[0.29]	[0.46]	[1.00]	
Education index	0.09**	0.09*	0.10*	0.03	0.01
	(0.04)	(0.05)	(0.06)	(0.03)	(1.02)
	[0.02]**	[0.14]	[0.21]	[0.50]	
Female empowerment index	-0.01	0.08	0.09	0.02	0.05
	(0.07)	(0.14)	(0.14)	(0.08)	(0.94)
	[0.35]	[0.34]	[0.34]	[1.00]	
Food security index	0.10***	0.05	0.08	0.01	0.01
	(0.03)	(0.07)	(0.06)	(0.03)	(1.00)
	[0.01]***	[0.34]	[0.22]	[1.00]	
Hours worked last week (respondent)	1.28	-1.87	-1.80	1.05	34.06
	(1.02)	(1.86)	(1.47)	(0.95)	(27.11)
	[0.13]	[0.29]	[0.27]	[0.48]	

Notes: Each row represents regressions of a pre-specified primary outcome on different regressors. Column 1 reports the coefficient on an indicator for treatment village from a regression using data from eligible households (as classified by the GE census team), and includes an indicator for saturation status of the sublocation (Equation 2.1). Column 2 reports the total effect on treated households (eligible recipients) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a household's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the household (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer), as in Equation 2.2. For this analysis, the sample is restricted to eligible households. We have 5,168 to 5,423 observations (1,118 for the female empowerment index) for these columns. Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households, coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer), as in Equation 2.3. The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. We have 5,230 to 5,509 (978 for female empowerment) for column 3.. The number of radii bands included in columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the pooled saturation effect, the average saturation effect experienced by households in high-saturation sublocations, derived as a weighted average of $\beta_3, \beta_5, \beta_6$ and β_7 in Equation B.5. We have to 7,832 to 8,239 (1,535 for female empowerment). Column 5 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. Standard errors are clustered at the village in column 1, at the sublocation level in column 4, and calculated following Conley (2008) using a uniform kernel out to 10 km in columns 2 and 3. Minimum FDR q-values are reported in brackets. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

B.11 Additional welfare analysis

We provide here an illustrative mapping between household welfare and aggregate output, emphasizing that any increase in aggregate output must reflect some combination of (i) an increase in the *employment* of factors of production, which comes at an opportunity cost, and (ii) an increase in their aggregate *productivity*, which does not. We also illustrate how household welfare differs from household expenditure.

Consider a household i whose market interactions at time t involve the (net) purchase of a vector of commodities c_{it} at prices p_t, the supply of a vector l_{it} of labor services at wages w_t, and the supply of (net) savings to support capital investment equal to the difference between current-period income and expenditure. In addition, the household receives profit π_{it} from owned enterprises.[4] The household's problem is

$$\max_{\{c_{it}, l_{it}\}} u(\{c_{it}, l_{it}\}, \{c_{-it}, l_{-it}\}) + \lambda_i \left(T_i + \sum_{t=0}^{\infty} \delta^t (\pi_{it} + w_t \cdot l_{it} - p_t \cdot c_{it}) \right) \tag{B.6}$$

where λ_i is the Lagrange multiplier on the budget constraint and $\delta \equiv 1/r$ the discount rate on future funds.[5] The economy's capital stock at the beginning of period t is $k_t = \sum_i k_{it} = \sum_i \sum_{\tau=-\infty}^{t-1} \delta^\tau (\pi_{i\tau} + w_t \cdot l_{i\tau} - p_t \cdot c_{i\tau})$. The household's contribution to (real) output in period t measured using the income approach is equal its claims on firm profits plus the factor payments it receives, or $\pi_{it} + w_t \cdot l_{it} + r \cdot k_{it}$.[6]

Overall output is the sum of these contributions which is simply total enterprise value added, and the period-t contribution to the transfer multiplier is the effect of \$1 of transfers on this quantity. Whether distributed to households in the form of higher profits, wages, or interest payments, real output gains can be achieved only through increases in (i) the supply of labor or capital, or (ii) of productivity. In the case of labor supply this comes at a utility cost, since $\frac{\partial u}{\partial l_{it}} < 0$, so that a dollar increase in output must be worth less than a dollar in equivalent variation terms. Similarly in the case of capital, an increase in the period-t capital stock implies a decrease in consumption in some other period(s), so that again a dollar increase in output is worth less than a dollar in equivalent variation terms. In the case of a pure productivity gain, on the other hand, a dollar of output is worth a full dollar to the household(s) that receive it.

To contrast expenditure with welfare, assume for simplicity that first-order conditions are necessary and sufficient for a solution to the household's problem defined by (B.6),

[4] The term π_{it} could also capture other (net) transfers e.g. from peer households and from the government. We ignore these terms here as the estimated treatment effects on them in our data are negligible.

[5] One can generalize this formulation to allow for non-separable household production using non-marketed inputs such as family labor without changing the basic message. It is also straightforward to allow for discount rates to vary across agents, reflecting capital market imperfections.

[6] Its contribution measured using the consumption approach is its expenditure $p_t \cdot c_{it}$ plus its attributable share of firm investment which (assuming a closed economy) must equal household savings, i.e., $p_t \cdot c_{it} + (\pi_{it} + w_t \cdot l_{it} + r \cdot k_{it} - p_t \cdot c_{it})$, which is evidently equivalent.

and for simplicity we ignore externalities. The envelope theorem then implies

$$\frac{dv_i}{dT} = \lambda_i \sum_{t=0}^{\infty} \delta^t \left(\frac{t}{\delta} \frac{\partial \delta}{\partial T} s_{it} + \frac{\partial \pi_{it}}{\partial T} + \frac{\partial w_t}{\partial T} \cdot l_{it} - \frac{\partial p_t}{\partial T} \cdot c_{it} \right) \tag{B.7}$$

where $s_{it} = \pi_{it} + w_t \cdot l_{it} - p_t \cdot c_{it}$ is period t savings. In comparison, the welfare effect of a marginal change in T_i holding other transfers fixed is simply

$$\frac{\partial v_i}{\partial T_i} = \lambda_i \tag{B.8}$$

The (marginal) equivalent variation dEV_i/dT is the ratio of these expressions. To see how this relates to household expenditure, define (the present discounted value of) household expenditure as $e_i = \sum (1/(1+r))^t p_t \cdot c_{it}$. Differentiating the budget constraint, we have

$$\frac{de_i}{dT} = \sum_{t=0}^{\infty} \delta^t \left(\frac{t}{\delta} \frac{\partial \delta}{\partial T} (\pi_{it} + w_t \cdot l_{it}) + \frac{\partial \pi_{it}}{\partial T} + \frac{\partial w_t}{\partial T} \cdot l_{it} + w_t \cdot \frac{\partial l_{it}}{\partial T} \right) \tag{B.9}$$

Comparing the equations above, we see that

$$\frac{dEV_i}{dT} = \frac{de_i}{dT} - \sum_{t=0}^{\infty} \delta^t \left(\frac{t}{\delta} \frac{\partial \delta}{\partial T} (p_t \cdot c_{it}) + \frac{\partial p_t}{\partial T} \cdot c_{it} + w_t \cdot \frac{\partial l_{it}}{\partial T} \right) \tag{B.10}$$

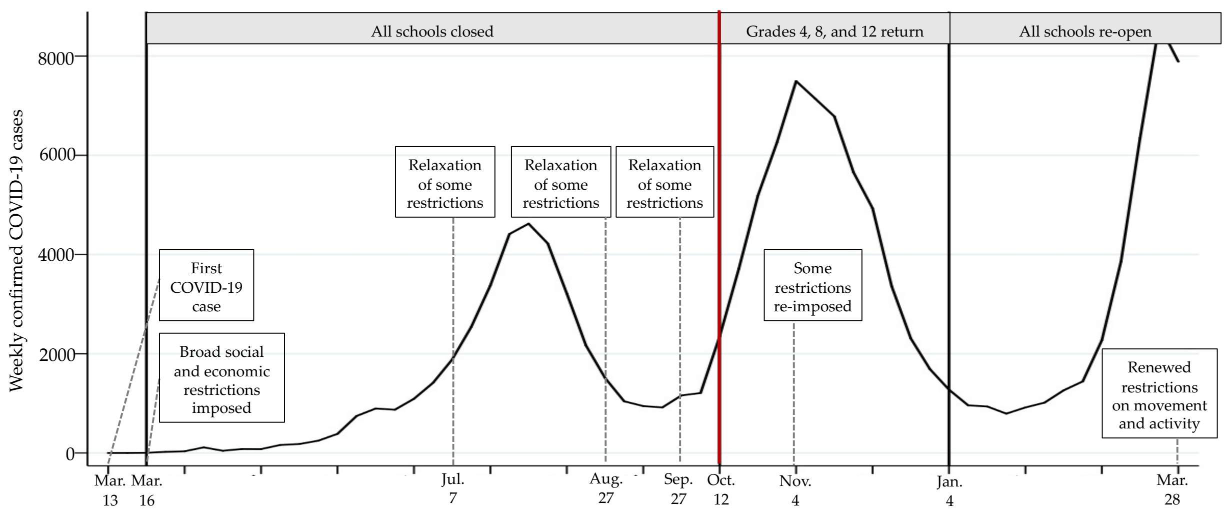
This expression shows that changes in expenditure are closely related to changes in equivalent variation, but with several intuitive (and correctable) sources of bias. First, (nominal) expenditure incorrectly counts appreciation of the price of the household's planned time path of consumption, whether due to appreciation of intra-period prices ($\frac{\partial p_t}{dT} \cdot c_{it}$) or of the inter-period interest rate $\frac{t}{\delta} \frac{\partial \delta}{\partial T} (p_t \cdot c_{it})$, as a welfare gain. This is why constant-dollar expenditure measures are preferable. Second, it incorrectly counts income gains due to behavioral responses such as increased labor supply ($w_t \cdot \frac{\partial l_{it}}{\partial T}$) as a welfare gain. Finally, if (more realistically) we were to examine expenditure over any finite period of time this would introduce a third bias, as this metric would count as a welfare gain any increases in current expenditure that were driven by decreases in future expenditure (i.e., by dis-saving).

Appendix C

Supplementary Appendix – Chapter 3

C.1 Additional Figures and Tables

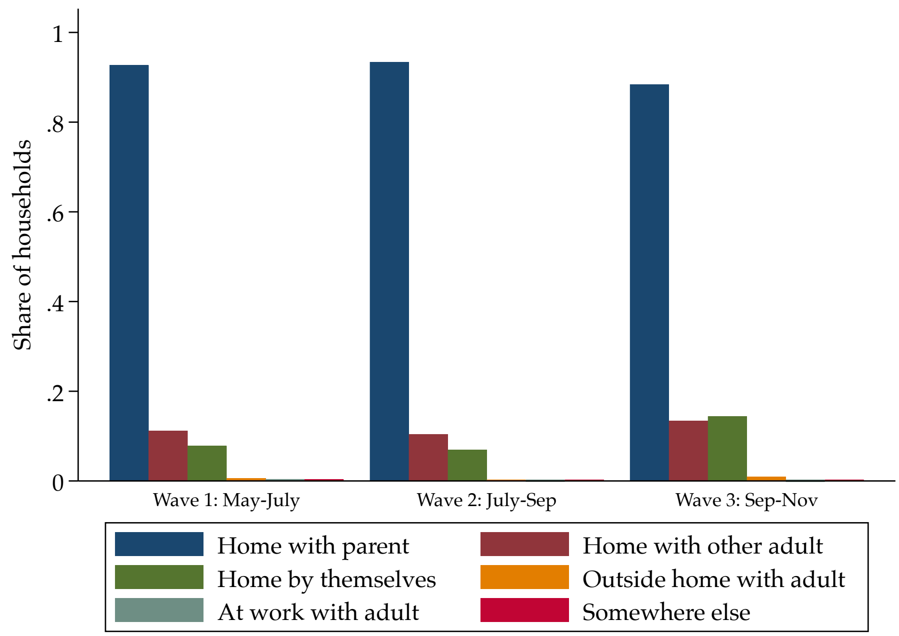
Figure C.1.1: Kenya COVID-19 cases, pandemic policy, and data collection timeline



Timing of Data Collection	Round 1 May – Jul.	Round 2 Jul. – Sep.	Round 3 Sep. – Nov.	Round 4 Jan. – Mar.
Respondents	4,062	4,504	4,993	4,906

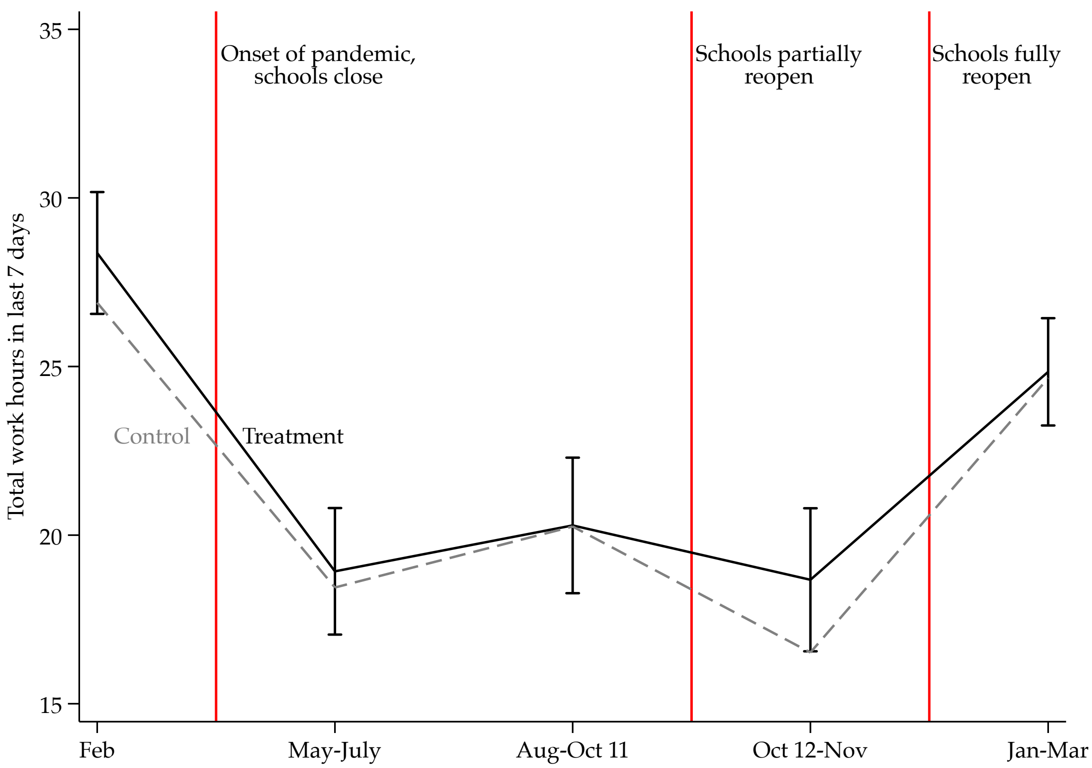
The figure shows the evolution of weekly confirmed COVID-19 cases in Kenya over time, along with the timing of key pandemic policy changes. The red bar indicates the partial school reopening on 12 October, the focus of the analysis. ‘Relaxation of some restrictions’ indicates that one or more of the initial pandemic constraints were at least partially reduced. Specific policy changes are outlined in Appendix C.
Sources: COVID-19 government response timeline for Kenya; Kenya COVID Tracker; Presidency of Kenya; Kenya Ministry of Education Twitter feed; COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University

Figure C.1.2: Childcare arrangements when children are out of school



Respondents are asked to specify all of the situations where a randomly selected child spent at least some time when out of school in the past week. 'Somewhere else' combines 'daycare/other childcare' and 'at home with a maid/domestic helper.' The figure uses information on childcare arrangements for all children, but the distribution is nearly identical when considering only children in grades 3-9.

Figure C.1.3: Respondent work hours in the last 7 days by survey round and treatment status



The figure shows raw means for household respondents' total work hours in the last 7 days by treatment status in each time period and 95% confidence intervals for the difference between treatment and control households. Means are shown for the respondent only due to missing data on pre-pandemic working hours for other household adults. Treatment households have a child enrolled in grades 4 or 8, and control households have a child enrolled in grades 3, 5, 6, 7, or 9. We do not show means for mixed households with children in both grade groups.

Data for February are based on recall from the first time a respondent is surveyed. We combine observations from the first two weeks of survey round 3, before the partial school reopening, with data from round 2. The red bars indicate changes in Kenya's school closures policy. The fall in hours after the partial reopening for control households reflects the end of main harvest period in Kenya, as 64% of households are engaged in agriculture.

Table C.1.1: Baseline balance by treatment status

	Control HH Mean	N	Mixed HH Mean	N	Treatment HH Mean	N	C-T p-value	C-M p-value
Respondent characteristics								
Age	40.02	948	41.26	361	41.28	335	0.102	0.079
Female	0.59	948	0.58	361	0.56	335	0.306	0.778
Completed primary school	0.88	945	0.84	361	0.87	335	0.524	0.070
Completed secondary school	0.48	945	0.42	361	0.47	335	0.753	0.048
Completed school beyond secondary	0.15	945	0.13	361	0.17	335	0.392	0.350
Married	0.74	937	0.8	356	0.72	328	0.396	0.020
Is the household head	0.63	948	0.63	361	0.65	335	0.516	0.791
Household characteristics								
Female household head	0.29	948	0.26	361	0.3	335	0.800	0.309
Age of household head	44.43	948	45.29	361	46.03	335	0.038	0.238
Count adults	2.55	948	2.75	361	2.64	335	0.289	0.015
More than 2 household adults	0.4	948	0.49	361	0.41	335	0.843	0.003
Only 1 household child	6.78	948	6.15	361	6.99	335	0.418	0.007
Age of youngest household child	0.42	948	0.43	361	0.36	335	0.033	0.648
Any young (0-4) children	0.66	948	0.71	361	0.66	335	0.800	0.058
Count young (0-4) children	0.56	948	0.63	361	0.46	335	0.040	0.199
Count school (5-17) children	2.47	948	3.26	361	2.35	335	0.143	0.000
Count adolescent (10-17) children	1.64	948	2.7	361	1.59	335	0.465	0.000
Household wealth index	-0.06	948	-0.15	361	0.03	335	0.169	0.113
Connected to electricity grid	0.46	948	0.41	361	0.51	335	0.111	0.105
Urban household	0.46	948	0.47	361	0.47	335	0.620	0.764
Household engaged in agriculture	0.61	948	0.65	361	0.59	335	0.573	0.217
Any child engaged in household farm labor	0.26	948	0.33	361	0.24	335	0.593	0.019
Household engaged in enterprise	0.15	948	0.16	361	0.19	335	0.134	0.683
Respondent labor participation								
Engaged in any work in last 7 days	0.68	948	0.67	361	0.7	335	0.534	0.888
Engaged in wage employment in last 7 days	0.1	948	0.08	361	0.13	335	0.211	0.150
Engaged in HH agriculture in last 7 days	0.55	948	0.6	361	0.53	335	0.479	0.171
Engaged in HH non-ag enterprise in last 7 days	0.09	948	0.1	361	0.13	335	0.100	0.563
Engaged in any work in February 2020	0.82	948	0.84	361	0.86	335	0.088	0.344

The table presents means for treatment households (T) with a child in grade 4 or 8, control households (C) with a child in grade 3, 5, 6, 7, or 9, and mixed households (M) with a child in both grade groups. Data are from the first time a household is observed, typically in survey round 1 (May-early July) while schools were fully closed. Individual-level data are for the survey respondent.

Columns on the right present differences and means and p-values for tests of equality for control households compared to treatment and mixed households, separately. The joint F-stat for differences across control and treatment households is 1.12, with p-value 0.305. It is 4.37 ($p<0.001$) for differences across control and mixed households.

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

Table C.1.2: Heterogeneity in impacts of partial reopening by grade of child eligible to return to school

Panel A: All analysis households

	(1) Any work	(2) Wage work	(3) HH ag. work	(4) HH ent. work	(5) Total hrs	(6) Wage hrs	(7) HH ag. hrs	(8) HH ent. hrs	(9) Childcare hrs
Post × Treat	0.041	0.011	0.037	0.015	3.630***	0.395	3.090***	0.281	0.460
	(0.026)	(0.013)	(0.027)	(0.016)	(1.365)	(0.565)	(1.091)	(0.625)	(5.031)
Observations	8538	8538	8538	8538	8538	8538	8538	8538	3,073
Mean, pre-reopen control	0.592	0.063	0.517	0.064	16.483	2.089	12.124	2.205	52.743

Panel B: Analysis households with a child in grades 2-6

	(1) Any work	(2) Wage work	(3) HH ag. work	(4) HH ent. work	(5) Total hrs	(6) Wage hrs	(7) HH ag. hrs	(8) HH ent. hrs	(9) Childcare hrs
Post × Treat	0.017	0.002	0.012	0.027	1.586	0.450	1.558	-0.278	-1.322
	(0.033)	(0.018)	(0.033)	(0.019)	(1.668)	(0.833)	(1.295)	(0.864)	(6.414)
Observations	6724	6724	6724	6724	6724	6724	6724	6724	2453
Mean, pre-reopen control	0.592	0.066	0.511	0.070	16.679	2.148	12.023	2.414	52.950

Panel C: Analysis households with a child in grades 6-10

	(1) Any work	(2) Wage work	(3) HH ag. work	(4) HH ent. work	(5) Total hrs	(6) Wage hrs	(7) HH ag. hrs	(8) HH ent. hrs	(9) Childcare hrs
Post × Treat	0.024	0.014	0.049	0.005	4.926***	0.152	4.724***	0.256	1.781
	(0.033)	(0.017)	(0.035)	(0.021)	(1.809)	(0.737)	(1.490)	(0.695)	(6.168)
Observations	6282	6282	6282	6282	6282	6282	6282	6282	2235
Mean, pre-reopen control	0.597	0.059	0.528	0.061	16.676	1.929	12.590	2.121	53.812

Panel D: Include grade 12 in treatment definition

	(1) Any work	(2) Wage work	(3) HH ag. work	(4) HH ent. work	(5) Total hrs	(6) Wage hrs	(7) HH ag. hrs	(8) HH ent. hrs	(9) Childcare hrs
Post × Treat	0.043	0.014	0.049*	0.015	3.057**	0.554	2.449*	0.131	0.705
	(0.028)	(0.014)	(0.029)	(0.018)	(1.556)	(0.592)	(1.285)	(0.664)	(4.873)
Observations	9407	9407	9407	9407	9407	9407	9407	9407	3387
Mean, pre-reopen control	0.586	0.063	0.509	0.063	16.033	2.076	11.804	2.081	52.151

Panel E: Analysis households with a child in grades 10-12

	(1) Any work	(2) Wage work	(3) HH ag. work	(4) HH ent. work	(5) Total hrs	(6) Wage hrs	(7) HH ag. hrs	(8) HH ent. hrs	(9) Childcare hrs
Post × Treat	0.018	-0.019	0.081	0.018	0.260	0.767	1.143	-1.284	4.021
	(0.066)	(0.052)	(0.071)	(0.053)	(5.984)	(1.917)	(5.433)	(1.770)	(11.853)
Observations	2547	2547	2547	2547	2547	2547	2547	2547	841
Mean, pre-reopen control	0.648	0.054	0.565	0.083	18.414	1.978	13.403	2.860	46.725

This table presents estimates of Equation 3.1 for different sub-samples. Panel A includes all households in the main analysis sample (with children in grades 3-9). Panel B focuses on the subset of analysis households with a child in grades 2-6, for which treatment means having a child in grade 4 eligible to return to school. Panel C is analogous for but children in grade 8. Panel D expands the sample to include households with a child in grade 12 in the treatment group and households with a child in grades 10 or 11 in the control group. Panel E focuses on households with a child in grades 10-12, for which treatment means having a child in grade 12. In panels B, C, and E, households with a child in another treated grade outside the focus range are categorized as 'mixed.' Dependent variables are defined over the last 7 days, and take a value of 0 for individuals not working in a particular activity. Childcare hours are observed for the household respondent only. All regressions include household and county by month fixed effects, and additional household and individual controls. SEs clustered at household level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.1.3: Robustness of results

Panel A: Individual fixed effects

	(1) Any work	(2) Wage work	(3) HH ag.	(4) HH ent.	(5) Total hrs	(6) Wage hrs	(7) HH ag hrs	(8) HH ent hrs
Post × Treat	0.041	0.009	0.037	0.016	3.735***	0.314	3.203***	0.368
	(0.026)	(0.013)	(0.027)	(0.016)	(1.367)	(0.565)	(1.089)	(0.628)
Observations	7765	7765	7765	7765	7765	7765	7765	7765
Mean, pre-reopen control	0.593	0.062	0.518	0.065	16.372	2.051	12.068	2.209

Panel B: Adults age 25-50

	(1) Any work	(2) Wage work	(3) HH ag.	(4) HH ent.	(5) Total hrs	(6) Wage hrs	(7) HH ag hrs	(8) HH ent hrs
Post × Treat	0.009	0.005	0.033	0.031*	4.017***	0.029	3.511***	0.807
	(0.028)	(0.018)	(0.026)	(0.016)	(1.509)	(0.757)	(1.142)	(0.761)
Observations	5362	5362	5362	5362	5362	5362	5362	5362
Mean, pre-reopen control	0.600	0.083	0.499	0.082	17.480	2.685	11.807	2.916

Panel C: Potential parents and sole caregivers

	(1) Any work	(2) Wage work	(3) HH ag.	(4) HH ent.	(5) Total hrs	(6) Wage hrs	(7) HH ag hrs	(8) HH ent hrs
Post × Treat	0.024	0.011	0.042*	0.016	3.448**	0.419	3.007***	0.197
	(0.026)	(0.017)	(0.024)	(0.015)	(1.436)	(0.705)	(1.123)	(0.671)
Observations	6118	6118	6118	6118	6118	6118	6118	6118
Mean, pre-reopen control	0.606	0.079	0.515	0.075	17.568	2.589	12.277	2.616

Panel D: Post defined by timing of reopening announcement, 21 Sept 2020

	(1) Any work	(2) Wage work	(3) HH ag.	(4) HH ent.	(5) Total hrs	(6) Wage hrs	(7) HH ag hrs	(8) HH ent hrs
Post × Treat	0.015	0.007	0.017	0.014	2.725**	0.292	2.113**	0.387
	(0.023)	(0.012)	(0.023)	(0.015)	(1.213)	(0.504)	(0.953)	(0.562)
Observations	8538	8538	8538	8538	8538	8538	8538	8538
Mean, pre-reopen control	0.604	0.062	0.531	0.065	16.730	2.024	12.439	2.211

This table presents estimates of variations of Equation 3.1. Panel A replaces household with individual fixed effects. Panel B focuses on adults age 25-50—the most likely to be parent caregivers and engaged in work. Panel C includes only adults identified as potential parents—between 14 and 55 years older than the oldest household child—or sole caregivers (the only household adult). Panel D defines *Post* not by the date schools reopened on 12 October 2020 but by the timing it was announced, 27 September.

Dependent variables are defined over the last 7 days, and take a value of 0 for individuals not working in a particular activity. Observations include data from May to November 2020, and include treatment households with children in grades 4 or 8 (indicated by 'Treat'), control households with children in an adjacent grade, and 'mixed' households with both (results not shown). 'Post' is a dummy for being observed on or after the partial school reopening on October 12. Regressions include household (individual in Panel A) and county by month fixed effects, and additional household and individual controls. SEs clustered at household level.

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

Table C.1.4: Heterogeneity in impacts of partial school reopening on working hours by individual/household characteristics

Interaction term Z	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Female	Ag HH	Urban	Large Urban	Above Mean Wealth	Any Kids Aged 0-4	>2 HH Adults
Post	-1.516	-1.047	0.994	0.994	-1.319	-1.322	-4.799*
	(2.330)	(2.222)	(3.214)	(3.227)	(2.644)	(2.737)	(2.710)
Post × Treat	3.793**	1.756	3.337	3.337	6.095**	5.035***	2.598
	(1.785)	(1.820)	(2.127)	(2.136)	(2.590)	(1.945)	(2.086)
Post × Z=1	1.570	-0.591	-1.100	0.698	1.056	-1.526	6.723*
	(1.739)	(3.738)	(4.117)	(5.162)	(3.705)	(3.696)	(3.732)
Post × Treat × Z=1	-0.687	5.485*	-0.280	-1.238	-4.209	-3.376	1.065
	(1.629)	(3.204)	(2.940)	(4.108)	(3.108)	(3.080)	(2.858)
Observations	8538	8538	8538	5172	8538	8538	8538
Mean, pre-reopen control	16.475	16.475	16.475	16.475	16.475	16.475	16.475

This table presents estimates of Equation 3.1 but interacting a characteristic Z with all right-hand side variables except the household fixed effects. The column label indicates which characteristic Z is being used. Values for household characteristics are from the first time they are observed in the data. 'Large Urban' is a dummy for location in one of Kenya's largest urban areas (Nairobi, Mombasa, Nakuru, Kisumu, Kiambu) relative to any rural area, while 'Urban' is a dummy for location in any urban area. 'Above Mean Wealth' is a dummy for whether and index of household wealth, based on housing and asset ownership, is above the sample mean.
The dependent variable is total working hours over the last 7 days, with individuals not working coded as working 0 hours. Observations include data from May to November 2020, and include treatment households with children in grades 4 or 8 (indicated by 'Treat'), control households with children in an adjacent grade, and 'mixed' households with both (results not shown). 'Post' is a dummy for being observed on or after the partial school reopening on October 12. Regressions include household and county by month fixed effects, and additional household and individual controls. SEs clustered at household level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.1.5: Heterogeneity in impacts of partial school reopening on adult agriculture hours by individual/household characteristics

Interaction term Z	(1) Female	(2) Ag HH	(3) Urban	(4) Large Urban	(5) Above Mean Wealth	(6) Any Kids Aged 0-4	(7) >2 HH Adults
Post	0.014	0.546	1.075	1.075	0.240	-1.232	-1.002
	(1.495)	(1.440)	(2.298)	(2.302)	(2.115)	(1.890)	(1.804)
Post × Treat	4.021***	0.984	3.030*	3.030*	5.386**	3.314**	3.030*
	(1.282)	(1.074)	(1.799)	(1.802)	(2.224)	(1.664)	(1.635)
Post × Z=1	-0.577	-2.876	-1.140	2.115	-0.933	0.133	1.556
	(0.890)	(2.804)	(2.917)	(3.629)	(2.773)	(2.690)	(2.769)
Post × Treat × Z=1	-0.989	6.278**	0.107	-3.624	-2.917	-0.414	0.312
	(0.914)	(2.558)	(2.337)	(2.570)	(2.550)	(2.485)	(2.301)
Observations	8538	8538	8538	5177	8538	8538	8538
Mean, pre-reopen control	16.475	16.475	16.475	16.475	16.475	16.475	16.475

This table presents estimates of Equation 3.1 but interacting a characteristic Z with all right-hand side variables except the household fixed effects. The column label indicates which characteristic Z is being used. Values for household characteristics are from the first time they are observed in the data. 'Large Urban' is a dummy for location in one of Kenya's largest urban areas (Nairobi, Mombasa, Nakuru, Kisumu, Kiambu) relative to any rural area, while 'Urban' is a dummy for location in any urban area. 'Above Mean Wealth' is a dummy for whether and index of household wealth, based on housing and asset ownership, is above the sample mean.

The dependent variable is household agriculture hours over the last 7 days, with individuals not working in household agriculture are coded as working 0 hours. Observations include data from May to November 2020, and include treatment households with children in grades 4 or 8 (indicated by 'Treat'), control households with children in an adjacent grade, and 'mixed' households with both (results not shown). 'Post' is a dummy for being observed on or after the partial school reopening on October 12. Regressions include household and county by month fixed effects, and additional household and individual controls. SEs clustered at household level.

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

Table C.1.6: Heterogeneity in impacts of partial school reopening on working hours by prior work

Interaction term Z	(1) Baseline Any	(2) Baseline Wage	(3) Baseline HH Ag	(4) Baseline HH Ent	(5) Closures Any	(6) Closures Wage	(7) Closures HH Ag	(8) Closures HH Ent
Post	-3.525	-0.345	-5.835	-2.436	1.248	-0.797	0.215	-0.806
	(2.767)	(1.833)	(4.320)	(2.586)	(2.301)	(1.772)	(2.481)	(1.843)
Post × Treat	3.474	2.918**	-0.953	4.500**	-1.281	3.290**	1.174	3.835***
	(2.429)	(1.420)	(3.188)	(2.154)	(1.873)	(1.394)	(1.937)	(1.421)
Post × Z=1	3.923	-10.741*	4.100	-14.419*	-2.055	6.325	-0.243	-0.336
	(2.704)	(6.105)	(5.422)	(7.998)	(3.016)	(6.224)	(3.339)	(6.113)
Post × Treat × Z=1	0.356	4.508	8.180**	0.782	5.290**	-1.348	4.150	-3.827
	(2.650)	(4.925)	(4.067)	(6.985)	(2.459)	(5.231)	(2.671)	(4.708)
Observations	8538	8538	2912	2912	8146	8146	8146	8146
Mean, pre-reopen control	16.475	16.475	16.475	16.475	16.475	16.475	16.475	16.475

This table presents estimates of Equation 3.1 but interacting a characteristic Z with all right-hand side variables except the household fixed effects. The column label indicates which characteristic Z is being used. 'Baseline' work participation is based on recall for February 2020, and is limited to the respondent for household agriculture and enterprise. 'Closures' work participation is based on any participation in a given sector from May-October 2020.
The dependent variable is total working hours over the last 7 days, with individuals not working are coded as working 0 hours. Observations include data from May to November 2020, and include treatment households with children in grades 4 or 8 (indicated by 'Treat'), control households with children in an adjacent grade, and 'mixed' households with both (results not shown). 'Post' is a dummy for being observed on or after the partial school reopening on October 12. Regressions include household and county by month fixed effects, and additional household and individual controls. SEs clustered at household level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

C.2 Major Pandemic Policy Changes in Kenya

The following outline summarizes when major nation-wide pandemic-related policies were implemented and relaxed over the course of 2020 after the first COVID-19 cases in Kenya on March 13.

The dates for the announcements of new restrictive policies are in *italics* and the dates when these policies were relaxed or ended are in **bold**. We also include announcements related to school closures, even though policies did not necessarily change with these announcements. Most policies were extended multiple times after first being imposed; we do not list the dates of policy extensions, except for school closures.

- *March 13-20*
  - Suspend all public gatherings, meetings, games, events
  - Ban on gatherings of more than 10 people
  - All schools closed
  - Recommend working from home where possible
  - Ban on foreigner entry; quarantine requirements for entry of nationals and visa holders
  - Public transport asked to reduce to 60% of capacity
- *March 24-27*
  - Ban on national and international flights
  - Closure of bars and restaurants for in-person service
  - Direct cash payments implemented for vulnerable citizens
  - Stay at home requirements imposed, except for 'essential' trips
  - Curfew imposed from 1700 to 0500 hours
  - Public transit closed between 'infected' and 'not infected' areas
- April 26: School closures extended to June 4
- **April 27**: Partial reopening of restaurants for take-out service
- June 6: School closures extended until further guidance from the Ministry of Health
- **June 7**: Nightly curfew revised to between 2100 and 0400 hours
- June 24: Announcement that school might reopen on September 1
- **July 7**
  - Phased reopening of religious gatherings
  - Up to 100 people permitted to attend weddings and funerals
  - Local air travel within Kenya to resume July 15
  - International air travel to resume August 1
- July 7: Announcement that schools will remain closed until January 2021, final exams are cancelled, and students would repeat the year; colleges and universities following strict guidelines might reopen in September
- **July 27**

  - Restaurants reopened, must close by 1900 hours
  - Ban on sale of alcoholic drinks and beverages in eateries and restaurants
- **August 27**
  - Restaurants may remain open until 2000 hours
  - Ban on sale of secondhand clothing lifted
  - Licensed hotels may sell alcohol
- September 15: Ministry of Education releases guidelines for safe reopening of schools
- September 21: Ministry of Education calls all teachers to report back to schools by September 28
- **September 27**
  - Nightly curfew revised to between 2300 and 0400 hours
  - Bars may reopen; restaurants and eateries may sell alcohol; bars, restaurants, and eateries may remain open until 2200 hours
  - Religious gatherings may open for up to 1/3 of capacity
  - Up to 200 people may attend funerals and weddings
- October 6: Ministry of Education announces that students in examination grades (4, 8, and 12) shall return to classes on October 12
- **October 12**: Students in examination grades (4, 8, and 12) to return to classes
- *November 4*
  - Requests for government work to be done remotely when possible
  - Political gatherings suspended
  - Nightly curfew revised to between 2200 and 0400 hours
  - Bars, restaurants, and eateries must close by 2100 hours
- November 4: Announcement that schools to fully reopen in January 2021
- **January 4**: Schools fully reopen

Other policies were implemented that specifically affected certain parts of the country. For example, on April 6 the government instituted a 21 day movement ban/lockdown for Nairobi, Kilifi, Kwale, and Mombasa, and Mandera was added soon after. This lockdown was extended multiple times. These were the only counties affected. The lockdowns for Kilifi and Kwale ended on June 7 and those for Nairobi, Mombasa, and Mandera ended on July 8.

Sources: COVID-19 government response timeline for Kenya; Kenya COVID Tracker; Presidency of Kenya; Kenya Ministry of Education Twitter feed

C.3 Data Details

Data come from the Kenya COVID-19 Rapid Response Phone Surveys (RRPS), collected by the Kenya National Bureau of Statistics with support from the World Bank. Pape et al. (2021) describe the survey methodology and implementation in detail.

The main RRPS sample is drawn from the nationally representative Kenya Integrated Household Budget Survey (KIHBS) conducted in 2015-2016: 9,009 households that were interviewed and provided a phone number served as the primary sampling frame for the RRPS. All households in the sample were targeted in each round regardless of whether they were reached in a previous round. By the fourth round of the RRPS, 5,499 KIHBS households had been successfully surveyed at least once. The KIHBS sample is supplemented by random digit dialing (RDD). From a sampling frame of 5,000 randomly selected numbers, of which 4,075 were active, 1,554 households had completed at least one survey by round four.

The sample is intended to be representative of the population of Kenya using cell phones. In the 2019 Kenya Continuous Household Survey 80% of households nationally report owning a mobile phone, though certain counties—notably in the northeast—have much lower mobile phone penetration. Pape et al. (2021) report that KIHBS households that provided a phone number and those that were successfully surveyed in the RRPS have better socioeconomic conditions—measured by housing materials and asset ownership—than households that did not provide a phone number or that did but were not reached for the RRPS.

The RRPS data include household survey weights adjusting for selection and differential response rates across counties and rural/urban strata, attempting to recover national representativeness. We do not apply these household weights for our individual-level regression analyses, but do apply them for population-level inference based on our results in the discussion.

The surveys include information on household composition, labor outcomes for household adults, and child schooling and care, as well as more general household information and COVID-specific modules. We use data from the first four rounds of the RRPS, covering May 2020-March 2021 and also construct measures for February 2020, before the first COVID-19 cases in Kenya, using recall questions from the first time a household was surveyed. Each round lasted approximately 2.5 months and covered a representative cross-section of households each week within each wave.

Data on childcare arrangements for a randomly selected child include questions on which household member has primary responsibility for the child's care, which household member was with the child in the last 15 minutes, and where and in whose company the child stayed during the day when out of school (from a set of general categories).[1] The surveys also ask respondents for their hours spent on childcare in the last 7 days.[2] Childcare

[1]Respondents are instructed to select all childcare arrangements used. Nevertheless, respondents might omit types of childcare that are used less frequently or that are seen as less socially acceptable (e.g., leaving a child at home by themselves).

[2]The survey asks "In the last 7 days, how many hours did you spend doing childcare?" and does not distinguish between time actively spent caring for a child and time spent on other activities while responsible

hours from other providers, including other household adults, all household children combined, and all non-household members combined are included in round 4 only.

---

for a child. We topcode reported childcare hours at 140, or 20 hours a day. Over 15% of respondents in our analysis sample indicate spending at least this many hours on childcare.

References

Auer, Daniel, and Johannes S Kunz. 2021. *Communication Barriers and Infant Health: Intergenerational Effects of Randomly Allocating Refugees Across Language Regions.* Technical report. Monash University, SoDa Laboratories.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. "Adaptive Linear Step-up Procedures That Control the False Discovery Rate." *Biometrika,* 491–507.

Conley, Timothy G. 2008. "Spatial Econometrics." In *The New Palgrave Dictionary of Economics,* Second Edition, edited by Steven N. Durlauf and Lawrence E. Blume, 7:741–47. Houndsmills: Palgrave Macmillan.

de Mel, S., D.J. McKenzie, and C. Woodruff. 2009. "Measuring microenterprise profits: Must we ask how the sausage is made?" *Journal of Development Econ.* 88, no. 1 (January): 19–31.

Deaton, Angus, and Salman Zaidi. 2002. *Guidelines for constructing consumption aggregates for welfare analysis.* Vol. 135. World Bank Publications.

Donaldson, Dave, and Richard Hornbeck. 2016. "Railroads and American Economic Growth: A 'Market Access' Approach." *QJE* 131 (2): 799–858.

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. 2020. "Pre-analysis Plan Report for General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya." Available at .

Haushofer, Johannes, Edward Miguel, Paul Niehaus, and Michael Walker. 2016. "Pre-analysis Plan for Midline Data: General Equilibrium Effects of Cash Transfers." May.

———. 2017a. "GE Effects of Cash Transfers: Pre-analysis plan for household welfare analysis." July.

———. 2017b. "GE Effects of Cash Transfers: Pre-analysis plan for targeting analysis." September.

———. 2018. "General Equilibrium Effects of Cash Transfers: Pre-analysis plan." June.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *QJE* 131 (4): 1973–2042.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-Out Estimation of Variance Components." *Econometrica* 88 (5): 1859–1898.

Pape, Utz Johann, Javier Baraibar Molina, Antonia Johanna Sophie Delius, Caleb Leseine Gitau, and Laura Abril Rios Rivera. 2021. *Socio-Economic Impacts of COVID-19 in Kenya on Households: Rapid Response Phone Survey Round 1.* The World Bank, January.

Walker, Michael. 2017. "Pre-Analysis Plan: Local Public Finance and Unconditional Cash Transfers in Kenya." February.

———. 2018. "Informal Taxation Responses to Cash Transfers: Experimental Evidence from Kenya." July.