General equilibrium welfare effects of cash transfers: Experimental evidence from Kenya*

Dennis Egger† Johannes Haushofer‡ Edward Miguel§ Paul Niehaus¶
Michael Walker†

September 13, 2019

*We thank the study participants for generously giving their time; Justin Abraham, Aakash Bhalothia, Christina Brown, Genevieve Deneoux, Tilman Graff, Max Lauletta, Michelle Layvant, Layna Lowe, Max Mueller, Priscilla de Oliviera, Robert On, Rachel Pizatella-Haswell, Zenan Wang and Francis Wong for excellent research assistance; Innovations for Poverty Action for fieldwork; the Busara Center for Behavioral Economics for piloting work; GiveDirectly for fruitful collaboration; and conference and seminar audiences at Harvard Business School, Harvard Kennedy School of Government, the Harvard/MIT Development seminar, the World Bank, University of British Columbia, UC Berkeley, UC San Diego, University of Chicago, University of Southern California, and the 2018 Y-Rise and 2019 AEA meetings for comments and discussion. This research was supported by grants from the International Growth Centre, CEPR/Private Enterprise Development in Low-Income Countries (PEDL), the Weiss Family Foundation, and an anonymous donor. Walker gratefully acknowledges financial support from the National Science Foundation Graduate Research Fellowship (Grant No. DGE 1106400). AEA Trial Registry RCT ID: AEARCTR-0000505

†University of California, Berkeley

‡Princeton University, NBER, Busara Center for Behavioral Economics, and Max Planck Institute for Collective Goods

§University of California, Berkeley, NBER, and CEGA

¶University of California, San Diego, NBER, and CEGA

Abstract

Tracing out 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 subjected to experimental examination. This study was prospectively 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 over 600 villages, with a large implied fiscal shock of roughly 15% of local GDP, and deliberately randomized the allocation 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. Enterprises in areas that receive larger 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 after roughly 18 months. Through monthly measurement of scores of commodities and consumer and durable goods, we document positive but minimal local price inflation (at 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 2.6. 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, consumption gains for untreated households 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.

1 Introduction

Tracing out the pattern of transactions in an integrated economy, and their contribution to aggregates of interest like overall output or to well-being, has long been one of the fundamental tasks of economic analysis. For instance, there has been interest in economics 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 are positive and large, based on non-experimental variation generated by policy changes (Chodorow-Reich 2019; Nakamura and Steinsson 2014; Auerbach, Gorodnichenko, and Murphy 2019). Until now, however, such 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 government cash transfer programs, which have now been implemented in scores of low and middle income countries (Haushofer and Shapiro 2016). 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, including consumption, earnings, assets, food security, child growth and schooling, self-reported health, female empowerment, and psychological well-being (Bastagli et al. 2016). Yet there is limited evidence on the aggregate economic impacts or welfare consequences of such policies, as the literature has typically focused on documenting impacts on treated households' behavior and has rarely assessed spillover or aggregate consequences (for exceptions, see Angelucci and De Giorgi (2009) and Bobonis and Finan (2009)).

The present study was prospectively designed to unite these two disparate literatures, applying experimental rigor to the study of aggregate economic issues in a novel way. Doing so is critical to understanding the overall consequences of cash transfer programs, which cannot be assessed without accounting for multipliers and externalities, and taking a stand on the determinants of household welfare. In particular, we designed and carried out a large-scale experiment in rural Kenya that provided one-time cash transfers worth roughly USD 1000 (distributed by the NGO GiveDirectly) to thousands of households across a sample of over 600 villages. The implied fiscal shock was large, as the cash transfers amounted to roughly 15% of local GDP during the peak 12 months of the program.

Beyond its fiscal scale, at least three aspects of the project represent advances on existing work. First, the project 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, increasing power to detect impacts on local economies. 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. In particular, we gathered detailed panel (longitudinal) data on household receipt of the transfer; household consumption expenditure patterns, among a representative sample of local households (both treated and untreated); local enterprise production, employment and revenue, based in part on local enterprise census data; labor market conditions; as well as especially high-

frequency (monthly) and spatially disaggregated market data on input and output prices. Third, we interpret the results through the lens of a theoretical framework which helps clarify which quantities and prices are likely to generate first-order impacts on welfare in this setting. We are then able to assess these empirically using the rich data collection described above.

Following earlier studies, we first document large direct impacts along a range of dimensions on households that received transfers, including increases in consumption expenditures and holdings of durable assets. While it is intuitive that recipients spend or save the money they receive, we emphasize that this mechanically implies the money must go somewhere, and thus that in a purely accounting sense we should expect to see correspondingly large changes elsewhere in the local economy. We do not document meaningful labor supply behavioral responses (decreases or increases) among treated households.

We then turn to data from detailed surveys of local firms. We find that enterprises in areas that receive larger cash transfers also experience meaningful gains in total revenues, in line with the increases in expenditure among households. Interestingly, sales increased without noticeable changes in firm investment behavior beyond a modest increase in inventories, and sales do not increase differentially for firms owned by cash recipient as opposed to non-recipient households. These patterns suggest a demand-led rather than an investment-led expansion in economic activity. Increased enterprise revenue in turn translates into a moderate increase in the wage bill paid to workers by local firms and a modest increases in overall profits.

We next examine how these changes affect untreated households. We find that they, too, show large consumption expenditure gains (despite not receiving a cash transfer): roughly one and half years after the transfers went out, annualized consumption expenditure is higher by an amount roughly comparable to the gains experienced by the treated households themselves. This result contrasts with earlier suggestive evidence of negative consumption and psychological spillovers from cash transfers within villages ((Haushofer, Reisinger, and Shapiro 2019)). The magnitude of our positive consumption spillovers is striking, and likely results in part from the income gains experienced by local firms' owners and workers. These income gains for non-recipients are in turn driven largely by increases in wage labor earnings, consistent with the fact that enterprise wage bills increased. In a reassuring check, the magnitude of per capita consumption gains (per dollar transferred) experienced by local households lines up roughly with the per capita revenue gains documented among local firms. This comparison is enabled by the project's unusually rich data collection on both the consumption and production side of local economies. On some level, this result should be unsurprising since the documented increase in local consumption expenditures needs to be spent somewhere, but ours is the first study (to our knowledge) to carefully document how such spending spreads through a local economy.

A central issue is, of course, 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. Through careful monthly measurement of scores of local commodities, consumer goods, and durable goods, we document that there was positive and

significant but minimal local price inflation in areas that received additional cash transfers. 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. Using the best of our measures, the overall effects on input prices – namely, the price of capital and the wage for labor – are often positive, and sometimes economically meaningful, but typically not statistically significant.

We next ask what these effects imply for the aggregate level of activity in the economy, and for the welfare of affected households. For the former, we compute a local fiscal multiplier, taking advantage of the fact that we observe the consumption expenditures of representative samples of both treated and untreated households, together with knowledge of the magnitude of the local cash transfer shock. Using this expenditure-based approach, we estimate a local fiscal multiplier that is positive and relatively large, at 2.6. Using a dual income-based approach that exploits different measures, we obtain a similar estimate of 2.7. These estimates are somewhat larger than most recent estimates derived from US policy variation, which often range from 1.5-2.0 (Chodorow-Reich 2019), although given standard errors we cannot rule out that magnitudes in Kenya and the US are the same.

Two methodological strengths of our approach to estimating the fiscal multiplier are (i) our ability to exploit true random assignment of cash transfers (rather than relying on quasi-experimental variation induced by policy), and (ii) the fact that the transfers came from donors outside the study area, rather than being internally debt-financed (which is typically the case in the US programs studied, and may complicate the interpretation of consumption responses due to Ricardian equivalence issues). To our knowledge, this is among the first studies to exploit a development economics RCT to estimate an important macroeconomic quantity, here the fiscal multiplier (see Muralidharan and Niehaus (2017) for a related discussion). One limitation, on the other hand, is that we observe limited data on consumption expenditures in the months immediately following the transfer in some communities, and this may lead to less precise estimates of the true multiplier as a result.

Finally, we develop a simple framework with household optimization in order to derive conditions under which a local cash transfer generates first-order gains in household welfare, beyond the direct benefits created by the transfer itself, which we treat as the numeraire (i.e., a USD 1 transfer is valued at one dollar). Intuitively, a household experiences indirect welfare changes if the influx of purchasing power into the economy induces changes in either their budget set (through changes in income or prices) or in behaviors of other households that generate externalities (such as crime, the provision of public goods, etc.). If the problem is sufficiently well-behaved, then the former changes can be summarized using logic similar to the sufficient statistics approach (Chetty 2009). We use this framework to highlight which of the treatment effects discussed above are central determinants of household well-being, rather than simply focusing on behavioral responses that likely entail offsetting costs – as would be the case, for instance, for an increase in individual income driven by higher labor supply. We also contrast the measure of welfare implied by the framework with household expenditure, which is often used in the literature as a proxy for welfare, identifying how

the two metrics diverge and how these wedges can be corrected.

As our framework (and much of classical price theory) emphasizes, the welfare consequences of goods or behaviors are hardest to quantify when they are unpriced – including many public goods, within household agency and female empowerment issues, and some forms of social comparison (e.g., inequality aversion). To begin to get a handle on these issues, the study also collected detailed information on a range of non-market outcomes that lie outside our main economic framework, including: measures of subjective well-being and happiness; domestic violence; child education, nutrition and health; and public goods. We find positive impacts of local cash transfers on several outcomes and null effects on others, but for none of the main measures in this domain did we document significant adverse spillover effects.

In our framework, the documented consumption gains for untreated households 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. This in turn suggests that concerns about cash transfer programs doing harm to non-recipients (Ozler 2018) are not borne out in our setting. Moreover, the large economic spillover gains we document among other households and firms in areas receiving large cash infusions implies that some existing evaluations of cash transfer programs that simply compare treated versus control households within a local area, but ignore aggregate effects, multipliers and externalities, may be under-estimating the overall gains generated by such programs. This insight may be of interest to public policymakers considering implementing or expanding such programs.

The constellation of empirical findings presents some potential puzzles regarding how local enterprise output and revenue increase substantially following the local cash transfer shock. The lack of meaningful local price inflation, even for local non-tradable goods, in areas that received the cash shock suggests that local firms were able to meet rising demand through a rapid supply response. Yet there is no evidence for a meaningful firm investment response to enable them to expand capacity, and evidence that firm hiring of labor increased only moderately. Where did this increased local output come from?

One intriguing – and somewhat speculative – possibility is that many local enterprises are characterized by substantial 'slack' in their utilization of factors of production, including both labor and machinery. This is plausible given the local environment: the typical firm has 1 to 2 employees (including the proprietor), which means that worker 'integer constraints' may often bind. In other words, it may be profit maximizing for firms to hire a worker to sell goods at a shop window, or be the sole worker at a grain mill, all day (relative to the alternative of shutting down the business). But this worker may relatively easily expand their labor effort and utilization of equipment (i.e., the milling machine) to meet higher demand with little in the way of extra labor or investment costs. The existence of such slack may help account for the existence of the large multiplier we document, as has also recently been argued in US data, especially in poorer US regions (Michaillat and Saez 2015; Murphy 2017).

The remainder of the paper is structured as follows. Section 2 describes the study design,

including the cash transfer intervention. Section 3 describes data collection and estimation methods for hosueholds, enterprises and prices. Section 4 presents results tracing out the path of the cash through the economy, starting with recipient households (Section 4.1), moving to enterprises (Section 4.2), and then turning to non-recipient households (Section 4.3) before investigating price effects (Section 4.4). Section 5 describes our method for using these data and results to calculate the fiscal multiplier. Section 6 presents results for key non-market outcomes (such as subjective well-being, health and food security) and externalities, including crime and public goods. Section 7 lays out a framework for mapping all of our results into changes in household welfare, summarizing effects on the household side. Section 8 discusses firm production and how we can make sense of the supply-side of our results. Section 9 concludes.

2 Study design

2.1 Setting: rural western Kenya

This study was registered and run under the name "GiveDirectly General Equilibrium (GE) Project" (Haushofer et al. 2014) in Siaya County, a rural area in western Kenya bordering Lake Victoria. Siaya County is predominantly Luo, the second largest ethnic group in Kenya. Siaya is relatively densely populated, with 395 people per km^2 compared to a Kenyan average of 91. GD selected both Siaya County itself, and within it the three subcounties Alego, Ugunja, and Ugenya, based on relatively high high poverty levels. Within this area, GD identified target villages for expansion; these were all rural villages (i.e., those not located in peri-urban areas) that were not part of a previous GD campaign.[1] This selection exercise resulted in a final sample of 653 villages, spread across 84 administrative sublocations (the geographic unit above a village), and three subcounties.[2] On average, villages consisted of about 100 households. The average household had 4.4 members, of which 2.3 were children. 97 percent of households were engaged in agriculture; 60 percent were also engaged in wage work, and 45 percent in self-employment. The average survey respondent was 48 years old and had about 6 years of schooling.

2.2 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, individuals had to live in homes with thatched

1. This selection was based on the 2009 national Population Census, which occurred prior to "devolution", a redrawing of administrative boundaries, creation of county governments, and transfer of significant power to these local governments, implemented across Kenya after elections in 2013, before the start of this study. Based on 2009 administrative boundaries, the study area covers 5 of the 7 divisions in Siaya District: Boro, Karemo, Ugunja, Ukwala and Uranga. These correspond to 3 current subcounties: Alego, Ugunja and Ugenya. The 2009 Census lists enumeration areas, which we refer to as villages.

2. Five villages were dropped after randomization: 4 villages, all of which contained the word "Town" in the name, were dropped for being too urban for GD to work in. One of these villages had previously been assigned to treatment, the three remaining had been assigned to control. In addition, field staff were unable to determine the boundaries of one control village, despite repeated efforts. This was an enumeration area created for the 2009 national Census that did not correspond to existing village boundaries. This village was also dropped.

roofs, a basic means-test for poverty. In treatment villages, GD enrolled all households that met this thatched-roof eligibility criterion ("eligible" households) as classified by their field staff through a village census, and confirmed via two additional visits to eligible households to complete the enrollment process (see Appendix B.1 for details).

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.[3] Households were able to select 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 made in this study are equivalent to approximately 15 percent of annual GDP (based on our data described below) in the treated areas during the peak 12 months of disbursements, and to 20 percent of annual GDP during full 24 month rollout period.[4]

Transfers were made in a series of three payments according to the following schedule: 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 token transfer), the second and final large installment of KES 40,000 was sent.[5] Transfers were typically sent at a single time per month (usually around the 15th) to all households in the study area scheduled to receive transfers. The transfers were non-recurring, i.e., no additional financial assistance was provided to recipient households after their final large transfer. Households in control villages did not receive transfers.

2.3 Experimental design

An overview of the experimental design and data collection is presented in Figure 1, and Figure 2 contains a map of the study area with treatment and control villages and sublocations identified, as well as markets.

We employed a two-level randomization design to generate variation that can be used to identify spillover effects at both the within-village and across-village levels. We randomly assigned the 68 sublocations in our study area (or in some cases, groups of sublocations) to high or low saturation status, resulting in 33 high saturation groups of villages, and 35 low saturation groups. Then, within high saturation groups, we randomly assigned two-thirds of villages to treatment status, while within low saturation groups, we assigned one-third of villages to treatment status. As noted above, within treatment villages, all eligible households received a cash transfer.

The randomization was conducted in two batches based on GD's expansion plans. The first

3. 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, the cost of which was deducted from the transfer amount. For more information on M-Pesa, see Mbiti and Weil (2015) and Jack and Suri (2011).

4. Note that this is a higher level as a share of local GDP than most government transfer programs, e.g., the ARRA programs studied in the recent US stimulus bill fiscal multiplier literature, see Chodorow-Reich (2019).

5. 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.

batch included villages in Alego subcounty, where GD had previously worked.[6] 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.

Our randomization design induces variation of treatment intensity across space. We have explicit variation at the sublocation level by virtue of being the highest level of randomization. We also have geographic variation induced by the two-level randomization: in both high and low saturation sublocations, some villages are (randomly) more likely to be surrounded by either more treatment villages or control villages.

This provides two distinct methods by which we could analyze our data. If most spillovers occur within the administrative boundaries of sublocations, then comparing effects in high versus

6. GD worked in 103 villages within Alego subcounty prior to 2014, and 90 villages from June to August 2014. Transfers to villages in our study began in September 2014. We began by compiling lists of rural villages in Alego in which GD had not previously worked. 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.

low saturation sublocations would be the most natural way to detect spillover effects. Sublocation-based spillovers could arise if there was re-allocation effort by public officials in the sublocation, or if households interact mainly with people throughout their sublocation but very rarely with those in other sublocations, even if located nearby. On the other hand, a geographic model of spillovers makes most sense if distance, regardless of underlying administrative boundaries, is the most salient feature for the likelihood of economic and social interactions. A geographic model is especially relevant for economic transactions that are not restricted by administrative boundaries. Given the fact that the sublocation boundaries are not "hard" in any sense, and that our data indicate that there is extensive economic interaction in nearby markets, regardless of whether or not it is located in the same sublocation as the household, we generally prefer estimating spatial spillovers based on the amount of cash distributed within various radii bands of one's village, as described in detail in Section 3.

Figure 2 displays the results of this randomization procedure spatially, and further emphasizes the importance of geography. There is visible clustering of treatment and control villages due both to the two-level design, as well as to chance. There is considerable variation across, but also within, sublocations in terms of the number of treated villages neighboring any one village. One can also see that villages bordering sublocations are still frequently contiguous. In addition, the nearest market for many villages will be across sublocation boundaries, making cross-sublocation spillovers likely. The proximity of treatment and control villages also emphasizes the point that there are no "pure control" areas.

Importantly (and as the figure shows), we do collect price data from market centers located in towns. So even if households travel to markets in towns, rather than ones in their villages, to make purchases, we will capture potential price changes in these markets. The variation in treatment intensity around markets is also visible, which we will use – along with variation in transfer timing – to estimate market price effects. There is also noticeable variation in the total number of villages (both treatment and control) located around markets. This implies variation in population density around markets, which is useful for studying heterogeneity in market price effects by market access, as discussed below.

One other geographic feature is important for putting these results in context. The main national road running from Mombasa (on the coast) to Nairobi and on to Kampala, Uganda passes through the study area, which is likely to help integrate the area into the national, regional and global economy.

3 Data and Empirical Specifications

We conducted four types of surveys: household surveys, enterprise surveys, market price surveys, and surveys of local public goods provision. Each source contributes to our ability to trace out the path of effects of the cash transfers through the local economy. Our household surveys include a random sample of eligible and ineligible households within each village; this allows us to measure

how the program affected all types of households in the study area. Enterprise data captures how any increases in spending affect the supply-side of the economy, and measures how enterprises respond. Market price data helps characterize whether areas with greater intensity of transfers experience inflation, a key input in any welfare analysis. Results from the local public goods surveys are presented in a separate paper (Walker 2018), and discussed briefly here. Below we describe the first three surveys in detail, which are the focus of the present paper, along with the main econometric specifications.

3.1 Household Data

A representative sample of households were surveyed in both treatment and control villages as follows. First, we conducted a baseline household census in all villages, which serves both as a sampling frame and allows us to classify households into eligible and ineligible groups. Second, we conducted baseline household surveys in advance of the distribution of any transfers to a village. Third, we collected endline household survey data from households, with a median survey date of 19 months after the transfer. We now go through each of these data sources in turn.

3.1.1 Baseline Household Census

Prior to the distribution of any transfers, we conducted a baseline household census and household survey with independent (non-GiveDirectly) enumerators. The census provided information on eligibility status of all households within a village, and served as a sampling frame for household surveys. The household census was designed to be comparable to GD's census. To ensure consistency across treatment and control villages, we conducted our own censuses in all villages, and base our analysis on village membership, household definitions and eligibility as classified by our project data collection field staff.[7] As described below, our main analysis is thus an intent-to-treat analysis (though we also sometimes make use of cash transfer distribution data from GD). Throughout this paper, when we refer to eligible households, we mean households classified as eligible in the census conducted by research team enumerators, unless otherwise noted.

3.1.2 Baseline Household Surveys

Inclusion in the household survey sample was done in the following manner. First, we determined household eligibility based on the research team's census data (i.e., whether households had a thatched roof), and used this information to randomly select 8 eligible households and 4 ineligible households per village to target for inclusion in the study. We refer to these households as "initially-sampled" households. When households contain a married or cohabiting couple, we randomly

7. GD field staff also visited recipient households three times in advance of distributing transfers. These additional visits can provide further insight into household eligibility, specifically related to attempts to game the eligibility criteria. GD estimates that, during the period of the GE study, a very small percentage of households classified as eligible at the time of the GD census were later deemed to be ineligible. These repeat visits could also lead to classification differences between GD and the GE study in rare cases; that said, incentives for deception are much smaller for the GE team, as this will not influence receipt of cash transfers.

selected either the male or female as the "target" survey respondent. This provided the list of individuals to survey.

Second, field teams visited the village in advance of transfer distributions to conduct baseline household surveys. Enumerators aimed to conduct all household surveys for a village in a single day. Due to this, if an initially-sampled household was not available to be surveyed on the day we visited the village for baseline surveys, we replaced the household with another randomly chosen eligible household until we had reached our target of 8 eligible households per village, or exhausted available eligible households. We refer to these as "replacement" households. In cases where we could not reach the target survey respondent, but the spouse/partner was available, we surveyed the spouse/partner rather than moving to a replacement household. We refer to households that were initially-sampled but not surveyed as "missed baseline" households.

At baseline, this procedure resulted in the surveying 6,510 initially sampled and 1,338 replacement households, with 1,302 initially-sampled households missed at baseline. Baseline was conducted between September 2014 and August 2015. The survey lasted about two hours per household, with detailed modules on economic activities, asset ownership, psychological well-being, health and nutrition.

A large number of baseline characteristics of eligible households are generally balanced across treatment and control villages, with the exception of an indicator for scoring above the median on a psychological well-being index (Table C.15, column (2), described in Appendix C). While surveys were completed before the distribution of transfers in all villages, in a few cases surveys were conducted after GD had held meetings in the village informing households that it would be a treatment village.

3.1.3 Endline Household Surveys

Endline household surveys targeted all "initially sampled" and "replacement" households, including those that were missed at baseline. For households that were surveyed at baseline, we attempted to survey the same respondent who had been surveyed at baseline. Endline surveys began in late May 2016 and concluded in June 2017.

At endline, we surveyed a total of 7,019 initially sampled (85 percent), 1,223 replacement (17 percent), and 1,015 missed baseline households (12 percent). Our main analysis focuses on "initially sampled" (which includes those missed at baseline) and "replacement" households; results are similar using only originally sampled households (see online appendix). We achieved high tracking rates at endline, reaching over 90 percent of eligible and ineligible households in both treatment and control villages, and these rates do not systematically vary by treatment status (Tables C.13 and C.14). The only subgroup difference of note is that we are slightly less likely to find ineligible households in high saturation sublocations that were initially surveyed at baseline.[8]

In addition to the modules collected at baseline, endline surveys also collected more detailed data on household expenditures, additional psychological scales (in particular related to stress and

8. More details on tracking and attrition are available in Appendix C.

hope), and more detailed information on crop production. Female respondents surveyed by a female enumerator were also administered a module on female empowerment and gender-based violence. (We describe all of these measures in more detail when discussing results.) Endline surveys lasted 2.75 hours on average.

Figure 3 illustrates this timeline graphically, both in calendar terms are relative to the "experimental start date" for each village. To ensure that a) we have an experimental start for both treatment and control villages and b) this is not driven by village-specific factors that may influence how long it took GD to enroll households within a village, we define the experimental start date as follows: we had generated a random order of villages in our study area by subcounty, and both GD and research field enumerators aimed to roll out project activities in this order. (In Ugunja and Ukwala subcounties, this was clustered by location). We calculate an "experimental start" for all villages (including treatment) based on this random ordering taking into account GD's mean monthly expansion pace across the subcounty.

Household endline surveys were conducted between 9 and 31 months after the experimental start date, with the 5th/95th percentiles of timing ranging from 12 to 27 months. 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 (see Figure 3 for more on survey timing).

3.2 Empirics: recipient households

Our first approach examines the reduced form effect of being an eligible household in a treatment versus a control village. We estimate:

$$y_{ivs} = \alpha_1 Treat_v + \alpha_2 HighSat_s + \delta_1 y_{ivs,t=0} + \delta_2 M_{ivs} + \varepsilon_{ivs}, \tag{1}$$

where y_{ivs} is an outcome of interest for household i in village v in sublocation s. $Treat_v$ is an indicator for residing in a treatment village at baseline, and thus α_1 is our main coefficient of interest.[9] Since all eligible households within treatment villages were targeted for treatment by GD, α_1 captures the total average treatment effect for eligible households in treatment versus control villages, made up of both the direct effect on recipient households and any within-village spillovers among eligible households; we cannot separately identify a direct effect on recipient households from within-village spillovers among eligible households.

We include several additional covariates in Equation (1). We include the sublocation saturation status $HighSat_s$, an indicator for being in high saturation sublocation at baseline. α_2 is a relatively coarse way to assess cross-village spillovers, as it assumes they are delimited by administrative boundaries and thus does not utilize all experimental variation. Hence we do not always report this coefficient in many specifications, and instead prefer our spatial analysis as described in the next section; the full results from this specification are presented in the appendix. We also include

9. In some cases, we look at individual-level results using Equation (1); we then define treatment status and eligibility on the basis of the household.

the baseline value of the outcome variable ($y_{ivs,t=0}$) when available in order to improve statistical precision (McKenzie 2012).[10] We cluster standard errors at the village level, and weight observations by inverse sampling probabilities to be representative of the population of eligible households.

Equation 1 does not take into consideration the spatial variation in treatment intensity visible in Figure 2. We would expect spillovers to be limited to, and uniform within, administrative boundaries if effects are mostly driven by the re-allocation of effort across areas, or if sublocations were largely self-contained units in terms of economic activity. As note above, in our context, this appears unlikely, since households across sublocations shop in many of the same market centers and there are no barriers to trade or movement across sublocations (nor are there meaningful ethnic differences that correlate with these borders). We therefore focus on estimating overall treatment effects based on the total amount of cash distributed in households' own villages, and other villages in radii bands around their villages.

In doing so, a challenge is to determine the "relevant" spatial distance at which cash transfer programs influence nearby economic outcomes. Given the lack of existing research knowledge on this question, we have no *a priori* reason to choose a particular set of rings as most empirically relevant. However, it is reasonable to expect weak monotonicity to hold (i.e., if we do not find meaningful effects based on treatment intensity at a distance of 2 to 4 kilometers from a market, we would not expect to find effects at 4 to 6 or 6 to 8 km from the market). In our pre-analysis plan, we therefore specified a procedure to select the "optimal" range of radii bands to include in our model.

We measure the spatial intensity of treatment by the total amount of cash distributed per capita as a share of per-capita GDP, within a series of donuts with inner radius r and outer radius $r+2$ around a village. We proxy for per-capita GDP with the annualized per-capita expenditure by control households in low saturation sublocations, where we weight eligible and ineligible households by their share of the population.[11] The population in each radius band 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 across each village in our study sample. To account for villages not included in our sample, but within radii bands of 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 the limited number of 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 (inflated by the overall average population growth rate in Kenya between 2009 and 2014 and net of the population in villages that our field team or GD had collected GPS data for) 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.

10. 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.

11. As we did not collect baseline expenditure data, the endline data from households in control villages and low saturation areas represent households least likely to be affected by any cross-village spillovers.

Next, we need to determine the range of radii bands to include in our specifications. We examine distance ranges from 0–2 km up to 18–20 km. To provide a sense of the relatively high population density in the study area, villages in the study area have on average 7.5 other villages located within a 2 km radius, and 27 within a 4 km radius. We then estimate the following model for eligible households:

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

Here, y_{iv} is the outcome of interest for household i in village v, and Amt_v is the total amount per capita transferred to village v over the entire study, measured as a share of per-capita GDP. $Amt_{\neg v,r}$ is the amount per capita transferred to villages other than v in the r to $r+2$ km buffer. Because the Amt variables depend not only on the random assignment of villages to treatment, but also on the share of households in a given buffer which are eligible, random assignment does not guarantee that the Amt variables are exogenous. We therefore instrument for the Amt variables by $Treat_v$ and the share $s^{elg,t}_{\neg v,r}$ of eligible households in villages other than v assigned to treatment (by buffer), respectively. $y_{iv,t=0}$ is the baseline value of the outcome variable, which we include when available.

This creates a series of nested models. To determine the "optimal" number of rings, we select the specification that minimizes the Schwartz Bayesian Information Criterion (BIC), while maintaining weak monotonicity, as described above. Let $\bar{R}$ denote the set of selected rings. We then use Equation (2), setting the maximum radii band to $\bar{R}$. Standard errors are calculated via Conley (1999, 2008) using a uniform kernel up to 10 km. Empirically, this algorithm selects only the closest radii (0–2 km) for almost all outcomes for both treated and untreated households.

We use the coefficient estimates from equations (1) and (2) to construct the following effects of interest. First, we present the *reduced form intent to treat (ITT) treatment effect* on treated households, α_1 from Equation (1). Second, we calculate the *average total effect on treated households* from Equation (2), where we multiply our coefficient estimates by the average per-capita transfer amount distributed in the relevant radius around eligible households. Concretely, we calculate the average total effect on the treated as:

$$\begin{aligned} &\hat{\beta} * \left(\overline{Amt}_v | i \text{ is a treated eligible}\right) \\ + &\sum_{r=0}^{R} \hat{\beta}_r * \left(\overline{Amt}_{\neg v,r} | i \text{ is a treated eligible}\right) \end{aligned} \tag{3}$$

for all radii bands up to $\bar{R}$ selected by our model. This second effect is not exactly like the first, since it allows for across-village spillovers in addition to direct effects and within-village spillovers.

3.3 Empirics: non-recipient households

Untreated households are comprised of eligible households in control villages as well as ineligible households in all villages. To test for spillover effects of cash transfers on this group, we estimate

the *total treatment effect on the untreated* using a spatial specification similar to Equation (2), but with several adjustments:

$$y_{iv} = \alpha + \sum_{r=0}^{R} \beta_r^1 Amt_r + \sum_{r=0}^{R} \beta_r^2 (Amt_r \cdot Elig_{iv}) + \gamma Elig_{iv} + y_{iv,t=0} + \varepsilon_{iv}. \tag{4}$$

First, because non-recipients do not experience direct effects, we no longer separate own-village effects and across-village spillovers: we drop Amt_v and replace $Amt_{\neg v,r}$ with Amt_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 that eligible households in control villages are untreated). Analogous to Equation (2), we instrument for Amt_r using the share of eligible households in buffer r assigned to treatment.

We calculate the average total effect on the untreated as:

$$\begin{aligned} & s^{elg,c} \left(\sum_{r=0}^{R} \left(\hat{\beta}_r^1 + \hat{\beta}_r^2 \right) * \left(\overline{Amt}_r | i \text{ is an untreated eligible} \right) \right) + \\ & s^{inelg} \left(\sum_{r=0}^{R} \hat{\beta}_r^2 * \left(\overline{Amt}_r | i \text{ is ineligible} \right) \right) \end{aligned} \tag{5}$$

where $s^{elg,c} = 1 - s^{inelg}$ is the population share of eligible control village households among untreated households, and the $\hat{\beta}_r^1$ and $\hat{\beta}_r^2$ terms come from Equation (4).

3.4 Enterprise Data

We have several sources of data on enterprises. First, the household surveys at baseline and endline include a detailed self-employment and an agricultural module. Second, we conducted censuses of all non-farm enterprises, and surveyed a representative subset of them at both baseline and endline. Each of these data sources offers their own advantages, which we discuss below.

3.4.1 Household survey data on enterprises

The household survey data provides insight into household agricultural enterprises, an important source of economic activity for almost all households: 97 percent are engaged in some form of subsistence agriculture. At endline, we measured crop-by-crop agricultural production, sales, employment as well as agricultural input costs. Given the high share of households in agriculture, we use our household census data when weighting agricultural enterprises.

The household self-employment module collects information on revenues, profits, hours worked, and some costs and investments for enterprises run by members of the household. Household survey data offers an advantage in that it provides a representative of universe of households (owners) living in study area, which makes it clear whom to attribute any profit gains to, which is especially important for determining welfare impacts. When decomposing household income by source (as follows in Table 3 below), we make use of these outcomes. However, household surveys miss local

businesses owned by people located outside of the study area, which our enterprise censuses and surveys do capture. For our main results on local production or GDP, we therefore also incorporate data collected directly from enterprises.

3.4.2 Enterprise Census and Survey

As part of our baseline household census, we included a question as to whether the household was running an enterprise from their homestead or from a fixed kiosk/shop. We also conducted an enterprise census designed to capture enterprises operating outside of homesteads. We then returned to survey enterprises operating outside of the homestead and open on the day of our visit (this coincided with baseline household surveys). In cases when villages had over 20 enterprises operating outside of homesteads (this often occurred in cases where villages overlapped with market centers), we randomly selected 20 enterprises to survey.[12] Enterprise surveys included questions on profits, revenues, a subset of costs, as well as inventories and investment.

The endline enterprise census was conducted in villages between November 2016 and April 2017 (see Figure 3), towards the end of household survey data collection. The approach to enterprise censuses and surveys was slightly different at endline. Enumerators visited villages with lists of all enterprises identified as part of baseline censusing activity, and sought to re-census all of these, as well as any new enterprises. We then randomly surveyed (from the endline census) up to 2 enterprises operating from within homesteads and up to 3 outside of homesteads, including those in market centers (if the village contains a market). This provides a representative snapshot of non-farm enterprises at endline. For some of these enterprises, we will also have baseline data, but being surveyed at baseline did not factor into the sampling at endline. We again collect information on revenue, profits, employees, wages, other costs and taxes paid.

A strength of our integrated approach, utilizing both household and enterprise surveys, is the ability to match most firms to their owners. In our context, enterprise activity is highly localized: 92% of total profits and 87% of revenues accrue to owners who live within a village. In addition, our matched dataset allows us to determine owners' eligibility status for transfers. This becomes useful when seeking to disentangle mechanisms, and when characterizing welfare effects. For instance, do enterprises in treatment villages have higher revenues and profits because enterprise owners received GD transfers, which relaxed credit constraints and accelerated investment? Or are enterprise impacts driven to a greater extent by changes in aggregate demand, which would affect all firms regardless of eligibility status? Likewise, the extent to which gains accrue to eligible versus ineligible households matter for understanding the broader welfare implications of the program.

Since agricultural enterprises are surveyed at the same time as households, we are able to match all of them to their owner. Among non-farm enterprises, we matched a total of 56% of those surveyed at endline to the baseline household census . This yields sensible patterns in the matched data. In particular, 28% of non-agricultural enterprises are owned by an eligible household, slightly

12. This number was selected as it was the number we estimated that an enumerator team could cover in a single day.

below their share in the population (33%), and enterprises owned by ineligible (and thus on average somewhat richer) households have 9% higher profits and 21% higher revenues on average than those owned by eligibles.

3.5 Empirics: enterprises

The empirical approach for enterprises mirrors our approach for households, with some adjustments reflecting differences in data sources. As outlined in the previous section, we combine data from three types of enterprises, which were selected via different sampling strategies: agricultural enterprises (with data from household surveys), non-farm enterprises operating within homesteads, and non-farm enterprises operating outside of homesteads (both from enterprise surveys). For outcomes that we have in common across all enterprise types, we pool data across all enterprise types; in some cases, we do not have analogous outcomes for agricultural enterprises.

As with households, we first estimate the reduced form effect of being an enterprise operating in a treatment village. We use a version of Equation (1), but add in interaction terms between enterprise type and the treatment and high saturation terms to reflect different sampling strategies, and allow for differential spillovers by enterprise type. Given that enterprise surveys are repeated cross-sections (with some enterprises in common), we include village-level means of the dependent variable when available in our ANCOVA specification rather than an enterprise-specific baseline value. Standard errors are clustered at the village level. Enterprises are weighted with inverse probability weights, taking into account their enterprise type. (When appropriate, for some outcomes we also weight by enterprise revenues, which we note below).

For enterprises, the spatial concentration of transfers may be especially important. When estimating spatial results at the enterprise level, we use a version of Equation (2), interacting both the cash transfer amount variables (and the underlying instrumental variables) with enterprise type. Radii bands are selected following the same algorithm as for households (described above), and enterprises are weighted in the same manner as for the reduced form equation. Standard errors are calculated following Conley (1999, 2008) using a uniform kernel out to 10 km.

As with households, we report both reduced form effects (α_1 from Equation (1)) and average total effects using results from Equation (2), including both direct and geographic spillover effects (where $Amt_v = 0$ for control villages). We estimate these effects separately for the three enterprise types g, and report results as effects per household for comparability to the household results:

$$\frac{1}{n_{hh}} \sum_{g} \widehat{\Delta y}_e^{\,g} * n_{ent}^{g}$$

where n_{hh} is the number of households (across treated and control villages), $\widehat{\Delta y}_e^{\,g}$ is the estimated effect for enterprise type g, and n_{ent}^{g} is the number of enterprises of type g (across all villages).[13] For the extensive margin, we use Equations (1) and (2) at the village level, with data on the number

13. As noted below, one exception to this is enterprise profit margin, where we compute a revenue-weighted average effect across all enterprises rather than summing up and generating a per-household measure.

of enterprises per household from our enterprises census.

Lastly, we estimate some results separately by recipient and non-recipient owners using the subsample of enterprises that we have been able to match to owners. For these regressions, since we know the treatment and eligibility status of the enterprise owners directly, we run the same specifications as for treated ((1) and (2)) and untreated (4) households.

3.6 Price Data

We make use of several sources of price data. First, we examine output prices facing consumers collected via monthly market price surveys. Second, we use of household and enterprise data on input factor prices for labor, capital and land.

The data on output prices comes from monthly market surveys in 61 markets across Alego, Ugunja, and Ugenya subcounties collected over the course of roughly two and a half years. These 61 markets constitute all markets in our study area (and neighboring towns) that have a weekly market day. There were, on average, 0.7 markets located within a 2 km radius of a given village, and 2.3 within 4 km, again indicating that this is a fairly densely populated area. In line with these figures, households report an average commuting time to their preferred market of 33 minutes, where more than 75 percent of respondents walk to the market.

The market surveys collected price information on a list of 72 products, including food (grains, vegetables, fruit, meat), livestock (goats, sheep), hardware (nails, paint), "duka" (kiosk store) products (non-food and packaged food), and other products (including fuel, health items, household items, and farming implements). Data were collected from three vendors of each product in each market.

The market surveys were conducted each month for 30 months, starting around 5 days after GD sent out transfers for the month to capture any very short-run price responses. Surveys began at least one month pre-treatment (before any local transfers had been made) for all markets, and continued through endline data collection to provide contemporaneous measures of prices at the time of the more detailed household and enterprise endline data collection.

We use these data to generate linear log-price indices by weighting the market price data by household expenditure shares. Due to time and budget constraints with baseline survey data collection, we did not conduct a full household consumption expenditure module at baseline. We instead make use of expenditure shares collected in the vicinity of our study that were collected as part of the Kenya Life Panel Survey.[14] We classify goods into tradables and non-tradables, where

14. The Kenya Life Panel Survey (KLPS) is a longitudinal dataset of nearly 10,000 Kenyan youth that participated in one of two previous randomized NGO programs — one which provided merit scholarships to upper primary school girls in 2001 and 2002 (Kremer, Miguel, and Thornton 2009), and one which provided deworming medication to primary school students during 1998–2002 (Miguel and Kremer 2004) in a neighboring county in western Kenya. The most recent complete survey round was KLPS Round 3, and respondents surveyed in rural areas of Siaya and Busia County (which neighbors Siaya) in 2013 or 2014 are used to compute the baseline expenditure shares. We prefer using existing expenditure data for our main specifications, rather than endline data from the present project, because changes in expenditure shares in our project could be endogenous; that said, results are nearly unchanged if we instead focus on consumption expenditure shares from untreated households in this project at endline.

tradables include easily transported, non-perishable items such as building materials (e.g. timber, cement, nails, iron sheets) or household goods (e.g. soap, firewood, charcoal, batteries, washing powder, mattress, soap), and non-tradables include more difficult to transport and perishable items, such as some food items (e.g. avocado, banana, cabbage, egg, pork, fish) and livestock. These classifications were undertaken based on feedback from local staff, but there may, of course, be some ambiguity about how to classify specific items.

Turning to the next source, the data on input prices was collected as part of household and enterprise surveys, as described in Sections 3.1 and 3.4 above. These include information on the monthly wages paid to workers, land prices, and interest rates on formal and informal borrowing and lending.

3.7 Empirics: prices

We make use of both spatial and temporal variation in the amount of cash distributed around each market to estimate output price effects. The richness of our panel data on prices allows us to control for both market and month fixed effects across all regression specifications. An added complication of our market price data is that we not only need to determine the relevant spatial distance, but also the relevant number of lagged time periods to include in the specification, as the temporal period over which effects may manifest is also an open question.

We follow an approach pre-specified in Haushofer et al. (2016): first, we select the optimal number of spatial buffer rings as we do for households, while fixing the time dimension. We estimate versions of the following model:

$$p_{mt} = \sum_{r=0-2km}^{18-20km} \beta_r \left(\tilde{Amt}_{mtr} + \tilde{Amt}_{m(t-1)r} + \tilde{Amt}_{m(t-2)r} \right) + \alpha_m + \lambda_t + \varepsilon_{mt} \tag{6}$$

Here, p_{mt} is the log of the aggregate price index of interest for market m in month t. $\tilde{Amt}_{mtr}$ is the per-household amount transferred within ring r around market m in month t, expressed as the fraction of GDP it represents (calculated in the same manner as for households). We aggregate it into a measure of the amount transferred in the last quarter (i.e., for months t, $t-1$, and $t-2$).

The regression specification also includes a market fixed effect α_m, and a month fixed effect λ_t. This approach thus accounts for any systematic price differences across markets, as well as differences in the share of eligible households located around markets. Monthly fixed effects account for seasonal differences and other time trends that affect all markets in the study area. Identification thus comes from differences over time in the amount transferred to households around a market. Standard errors are again calculated following Conley (1999, 2008) to account for spatial correlation. (In the market level specification, we do not instrument for treatment status and the share of eligible households since we include market fixed effects.) We select R to minimize the Schwartz BIC while imposing weak monotonicity, as with our household estimation.

We next choose the optimal number of monthly lags, while fixing the number of distance rings

at the optimum obtained in the first step. We estimate the following model:

$$p_{mt} = \sum_{r \in \bar{R}} \sum_{l=0}^{M} \beta_{rl} \tilde{Amt}_{m(t-l)r} + \alpha_m + \lambda_t + \varepsilon_{mt} \tag{7}$$

In this specification, we no longer sum the amount transferred to the area around a market across months. Instead, we separately consider the amounts transferred in each month (from 0 to 18 months) preceding the price measurement. We begin by including only contemporaneous transfers to the optimal radius ring around the market, and then gradually add temporal lags to this specification. The optimal number of lags is again determined by the specification that minimizes the BIC.

We perform this analysis for the overall, pre-specified basket of goods, and then separately for component categories of goods. The individual categories include: food items; non-food non-durables (such as soap, cooking fat, and firewood); livestock; durables (such as iron sheets, mosquito nets, and jerry cans used for transporting water or fuel); and temptation goods, i.e. alcohol and tobacco.

This approach yields optimal rings of 0–2 km, and for some outcomes 2–4 km, around a market, and an optimal temporal lag of 0 months in all cases. This implies that we only include contemporaneous transfers in estimating the effect of transfers on local prices. We thus estimate the effect of transfers on prices using the following model (dropping the 2–4 km term when not selected by the algorithm):

$$y_{mt} = \beta_0 + \beta_1 \tilde{Amt}_{m,t,0-2km} + \beta_2 \tilde{Amt}_{m,t,2-4km} + \alpha_m + \lambda_t + \varepsilon_{mt} \tag{8}$$

Again $\tilde{Amt}_{m,t,0-2km}$ is the amount transferred to households in the 0–*2* km ring around market m at time t, expressed as a share of GDP of low-saturation regions and weighted by the share of eligible and ineligible households. We control for both market and month fixed effects, α_m and λ_t.

To aid interpretation of the point estimates from Equation (8), we calculate two quantities of interest: i) the implied average treatment effect (ATE), and ii) the average maximum effect. The implied ATE is calculated by evaluating the regression specification at the average level of treatment intensity between the study period of September 2014 and March 2017, the time during which transfers went out to any market in the study area. In other words, it is the average price effect across all markets and all months in which transfers went out to any market in the study area. We calculate this effect by evaluating the optimal specification at the average treatment shares of GDP, in each buffer around each market, over the entire study period. On average, a 0–2 km buffer around a given market place experienced a cumulative cash inflow equivalent to 7.7 percent of local GDP, a large shock.

The average maximum effect corresponds to the average effect on the log price index in the month in which the maximum amount of cash (as a share of GDP) was distributed within a market's 0–2 km buffer. We first calculate this effect by determining the maximum GDP share

that went out within the 0–2 km buffer around each market (across all study months). We then average the maximum predicted effects across all markets.[15]

As our input price data comes from households, we use Equations (1), (2), and (4) to estimate these price effects, and report the corresponding average treatment effects.

3.7.1 Heterogeneous effects by market access

Standard theory in international trade predicts that more integrated markets are less likely to see meaningful price changes for commodities following a local aggregate demand shock. We thus examine output price heterogeneity with respect to a commonly used metric of market access MA_m:

$$MA_m = \sum_d \tau_{md}^{-\theta} N_d \approx \sum_{r=0}^{10} r^{-\theta} N_r \tag{9}$$

We use geographic distance r 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 (2016) in setting $\theta = 8$. We then classify markets into quartiles based on this measure, with lower quartiles representing more isolated markets. We then estimate effects separately for each quartile, and calculate average and maximum treatment effects in the same manner as above.

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 move onto enterprises, and untreated households. Throughout our results, units are USD PPP unless otherwise defined (where the transfer was worth USD 1,871 PPP). Also note that flow outcomes are annualized, and monetary outcomes are top-coded at the 99th percentile (as pre-specified) unless otherwise noted.

4.1 Recipient household effects

A natural first step in the causal chain involves understanding impacts on cash transfer recipients. What happens to recipient household spending, saving and economic activity in response to a large, unconditional cash transfer? These are key market outcomes through which the actions of recipient households may have effects on non-recipient households via price and other mechanisms.[16]

The main measure of total expenditure is calculated as the sum of total food consumption in last 7 days, frequent purchases in the last month and infrequent purchases over the last 12 months,

15. The transfers (as share of GDP) that generate these maximum effects may differ slightly in magnitude across baskets of goods, for instance, if the relevant radii bands differ.

16. This is not to minimize the importance of non-market outcomes, other behavioral responses, or externalities; many of these were included as primary outcomes as part of our pre-analysis plan, and we cover them in Section 6.

all converted to yearly values.[17] We also calculate durable expenditures as the sum of spending on home maintenance, home improvement, and household durables. The remainder is classified as non-durable.

Given time lags before endline data collection, these recall periods do not cover the immediate post-transfer period for many sample households. This is especially true for non-durable expenditures (given the shorter recall period), and for spending out of the first two cash transfers. Thus, while the consumption results accurately capture mean expenditure roughly 1.5 years after the start of transfers, they will likely understate overall consumption gains since the time of transfers if recipient households spent some of the transfers shortly after receipt. This is an important factor to keep in mind when interpreting these results, and a point we return to below in the discussion of the estimation of the multiplier.

Recipient households do indeed report significantly higher total expenditure: recipient households have USD PPP 290 more expenditure than eligible households in control villages (Table 1, column 1), an 11% increase over the control, low saturation mean of USD PPP 2,548. We find an even larger effect of USD PPP 348 (a 13.6% increase) when examining the total, spatial effect from Equation (2) (column 2). The fact that the total effect is larger than the reduced form effect is suggestive evidence for the existence of localized, positive cross-village spillovers for eligible households on expenditure.

The increases in total expenditure are driven by increases in both non-durable and durable consumption expenditure. Despite the fact that the potentially misses out on much of the immediate household spending on non-durables, we still observe increases of USD PPP 170 in the reduced form specification and USD PPP 216 from the spatial specification. Increased food expenditure accounts for a sizable portion of this increase in both specifications (45% and 65%, respectively). As with much of the existing literature (Evans and Popova 2017), we do not estimate a meaningful increase in spending on temptation goods (alcohol, tobacco and gambling) roughly 19 months after transfer disbursement. The fact that the control group mean is positive and non-trivial suggests that at least some households are comfortable reporting on such spending, at least partially mitigating concerns around reporting bias.[18]

Durable expenditure is also substantially higher for recipient households, with an increase of USD PPP 100 in annualized durable expenditure in the reduced-form specification; again this estimate is higher when using the spatial specification (in column 2). This is a notable increase relative to the control, low saturation mean (for eligible and ineligible households) of USD PPP 59.

Increases in durable expenditure translate into significantly higher stocks of household assets

17. 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; dowry or bride price.

18. That said, given that we do not have temptation good expenditure immediately after the transfers were distributed means we cannot fully rule out the possibility that larger effects occurred right after the distribution of transfers.

for recipient households (Table 1, Panel B).[19] 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. Importantly, this measure excludes the values of housing and land, given that these are harder to measure given thin local markets. We document an increase of USD PPP 162, or 26% relative to the mean of eligible households in control, low saturation villages.[20] The reduced form and total effects are similar in magnitude, which is consistent with minimal cross-village spillovers in terms of asset stocks.

Given existing research that documents that households often make home improvements in response to GD transfers (Haushofer and Shapiro 2016), these remain important outcomes to study, even if difficult to measure. We measure housing value as the respondent's self-reported cost to build a home like their own. We measure land value as the total land owned by the household times the household's report of the cost to buy one acre of land in their village of similar quality to the land that they own. We observe increases of 60% in the total effect for recipient households' housing values (USD PPP 377, p-value of 0.01), indicating that households in our sample may also be making home improvements.This is corroborated by the fact that twice the share of recipient households have metal roofs at endline compared to eligible households in control villages. We also find a positive, but not statistically significant, point estimate for the value of land owned for recipient households.

Taken together, the magnitude of the total effects on expenditures and assets (including housing value) totals up to roughly fifty percent of the transfer value. This does not yet account for other possible changes in recipient household balance sheets. Next, we examine these dimensions and assess if recipient households have offsetting responses that alter their overall budget constraint.

First, we examine labor supply. A classic concern about providing assistance to poor households is that the income effects from any assistance will result in reduced labor supply. However, consistent with Banerjee et al. (2017), we do not find evidence that recipient households have decreased their labor supply. If anything, the total hours worked by recipient households in agriculture, self-employment and employment have slightly increased. We find a moderate positive effect of 2.95 hours (p-value 0.10) on total household hours worked in the last seven days in the reduced form specification, and a positive but not significant estimate of 1.4 additional hours in the spatial specification among recipient households (Table 2, Panel A, columns 1 and 2). These positive point estimates for recipient households suggest that we may need to factor in some potential disutility due to increased labor supply among recipient households.

The overall budget envelop for recipient households also does not appear to have decreased: the point estimates indicate that total household income from all sources (excluding the GD transfer) increased (by USD PPP 81 and 141 in the two main specifications), and the reduced form effect is

19. Anecdotally, most respondents withdraw money from M-Pesa immediately and save via durable assets.

20. The mean for eligible household in control, low saturation sublocations is 679 (with SD 780), substantially less than the overall mean since ineligible households own more assets on average.

marginally significant.[21] The point estimate for net transfers received is negative, but not statistically significant, and, given the control mean, recipient households remain net receivers of transfers from others.

In sum, these results highlight that, as one might expect, cash transfer recipients have substantially increased their spending. Expenditure by recipient households remains elevated roughly 1.5 years after the transfers, and recipients have also accumulated more assets and improved the quality of their dwellings. There do not appear to be reductions in other sources of household income (if anything, recipient households have higher total incomes compared to non-recipient eligible households) or in labor supply. The data also suggest that a large share of household spending takes place locally: a majority of households have a market center within 2 km of their house, and enterprises report that 86 percent of their customers come from within the same village or sublocation. We now turn to the enterprise data to assess if this increased household spending affects local firm revenue and other outcomes.

4.2 Enterprise effects

We first examine effects on measures of enterprise profits, revenues and costs for both agricultural and non-agricultural enterprises. We measure (annualized) revenues and profits for non-agricultural enterprises by directly asking enterprise respondents about these quantities (Mel, McKenzie, and Woodruff 2009). Costs are calculated as the sum of the employee wage bill, rent and security costs. As this measure of enterprise costs is less comprehensive than our measure of enterprise revenues, we do not expect that profits will necessarily be equal to revenues minus costs.

For agricultural enterprises, we calculate total revenue 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.[22] Agricultural costs are the wage bill, agricultural inputs (e.g., seed and fertilizer), and land rental costs. Agricultural profits are then calculated as the total value of our measure of agricultural revenue minus agricultural costs. Recall that we include fixed effects for enterprise type in the regressions to help mitigate concerns about pooling data from somewhat different sources.

Table 4, Panel A documents that enterprises in both treatment and control villages in areas with a greater influx of transfers experience large revenue increases. Enterprises in treatment villages experience revenue gains of USD PPP 341, a 40% increase, while those in control villages experience gains of USD PPP 225 (26%). These revenue increases do not translate into a statistically significant increase in profits, as costs also increase, particularly the wage bill.

The revenue gains are concentrated in the retail and manufacturing sectors. Both treatment and control villages see statistically significant increases in manufacturing revenue of similar magnitudes

21. As is common in many low-income settings, we measure higher values of household consumption compared to household income.

22. In cases when crop output was reported in non-monetary units, we convert these to monetary values using the median crop output price measured in our market price surveys during the period of endline data collection in the subcounty where the household is located.

– USD PP 71 and 80, respectively – while treatment villages see larger gains in retail revenue (USD PPP 137 versus 77) (Appendix Table A.4).

In Panel B of Table 4, we examine three additional outcomes that were included in the enterprise survey: inventory, investment, and the profit margin. We find an increase in inventories of 11% in treatment villages (p-value < 0.05). While point estimates on enterprise investment are positive, none are large in magnitude or statistically significant, and in all cases are smaller that the effects on inventory. Given that we are finding the largest gains in the retail sector, a response in inventories, rather than investment, is sensible.

One concern when assessing the local investment response is that the measure of household assets may include some goods that are also productive for household enterprises. Our measure of household assets certainly includes some that are potentially productive for agriculture or non-agricultural enterprises, in addition to assets that largely provide consumption value. To better understand effects by category, we classify assets as "potentially productive", and in doing so we aimed to be inclusive. In particular, for potentially productive non-agricultural assets, we make no restrictions based on reported household self-employment activities. We consider agricultural tools and livestock as potentially productive agricultural assets.[23] Potentially productive non-agricultural assets are those that one could utilize for income-generating activities other than renting out the asset.[24] We consider assets that are most likely to provide an individual consumption benefit to the respondents' household to be non-productive (i.e., radio/cd player, kerosene lantern, bed, mattress, bednet, table, sofa, chair, cupboards, clock, television, and iron sheets).

Table A.1 presents results overall and by these categories.[25] About half of the increase in asset ownership comes through an increase in assets that appear to be purely non-productive. That said, we do document an increase in some potentially productive household assets (Table A.1) that is indicative of some increased investment in response to the transfers. We do not observe the timing of this spending, but the stock of potentially productive household assets has increased.[26]

Lastly, we study the extensive margin of enterprise creation at the village level, in Panel C; to normalize village size, we present results on the number of enterprises per household. Recall that essentially all households run agricultural enterprises, so a total control, low saturation mean of 1.12 indicates a mean of 0.12 non-agricultural enterprises per household per village. Perhaps surprisingly, we do not find significant effects on enterprise creation as a result of the cash transfers, with point estimates near zero for all specifications.

23. Agricultural tools include farm tools, such as hoes and machetes, hand carts, wheelbarrows and ox plows.

24. Namely, these include: bicycle, motorcycle, car, boat, kerosene stove, sewing machine electric iron, computer, mobile phone, car battery, solar (panels or system), and generators.

25. In this table, we do not include net borrowing as part of our measure of household assets, which are included in Table 1.

26. As these results did not restrict attention to only those involved in self-employment, they may be overestimating investment relevant to enterprises; on the other hand, they do not capture an exhaustive list of assets, so we may still be missing some.

4.3 Non-recipient household effects

How does the spending of recipient households, and the corresponding increases in enterprise revenue and wage bill, affect untreated households? We now turn to estimating effects for non-recipient households, including both eligible households living in control villages, and ineligible households living in both control and treatment villages.

We begin by examining expenditure and savings responses for untreated households. Column 4 of Table 1, Panel A presents results on expenditures for untreated households using Equation (4). We find positive and significant expenditure effects for untreated households, which recall were measured on average roughly a year and half after the start of cash transfers. Notably, the magnitude of these gains (USD PPP 337, p-value < 0.01) are quite similar to those of treated households (USD PPP 348). As previously noted, the recall period for expenditure does not cover the period immediately after the full transfer amount for all treated households, so the full integral of spending since the distribution of the transfer is likely still higher for treated households. Yet this finding does suggest that untreated households do eventually realize substantial gains from the program.

Where does the income to fund these consumption gains come from? One potential explanation is that untreated households are simply dis-saving, perhaps due to social pressure to keep up with the increased spending of recipient households in their communities. However, this does not appear to be the case: in Table 1, Panel B, we see that estimated treatment estimates for total assets, housing and land values are all positive (although not significant). In addition, we do not observe a borrowing response for untreated households from either formal and informal sources (Table 2, Panel C, column 3), which further supports the idea that these expenditure gains are not driven primarily by increased borrowing or dis-saving.

A second potential explanation is that these expenditure gains come from inter-household transfers to non-recipient households, as documented in Angelucci and De Giorgi (2009) for Mexico. This also does not seem to be the case, as we find no significant increase in net transfers received by untreated households (Table 3, row 1). In addition, the point estimate of USD PPP 8.75 is less than 3 percent of the expenditure gain for untreated households.

Where we do see a substantial increases is in wage earnings for untreated households (Table 3, row 4). The total average effect on wage earnings for untreated households is USD PPP 182 (p-value < 0.01), which serves as the key driver of an overall increase in household income for untreated households of USD PPP 216. Recall that enterprises operating in both treatment and control villages reported increases in wage bills of 60 and 50%, respectively (Table 4, row 4). These percentage changes are both greater than the 40% increase in wage earnings for untreated households that we document here. Hourly wages for untreated households also appear to experience a sizeable increase, at 0.19 log points, although the estimate is only marginally significant (Table 2, Panel A). The point estimate for total household labor supply for untreated households is actually somewhat negative, although there appears to be an increase in hours worked in wage employment for the respondent (Table A.2). The point estimate for the increase in total household income is

also in line with total consumption increase for untreated households; as both point estimates have standard errors, statistically we cannot reject that they are the same.

To sum up the results across this and the previous few subsections, we document that cash transfer recipient households receive and spend must of the transfer, leading to higher local enterprise revenues. This positive aggregate demand shock, in turn, appears to increases the income of even local untreated households, leading to higher spending on their part. This provides some initial suggestive evidence for a positive multiplier effect of the cash transfer program, an issue we return to in detail below.

4.4 Price effects

Whether the transfer program affects local prices is a first-order question for interpreting the welfare implications of our previous results, all of which have been reported in nominal terms. We begin by looking at output prices for consumer goods using our market price data. We then summarize results on input prices for labor, land and capital, using data collected as part of household and enterprise surveys.

4.4.1 Output prices

Our main measure of output prices is an expenditure-weighed log-index of market prices, constructed at the market level from data on a list of 72 products. We also create sub-indices for tradable and non-tradable goods, and by sector (see Section 3.6).

As can be seen in Table 5, both the ATE and average maximum effect are small and precisely estimated near zero across all product categories. Given our small standard errors, we can rule out even relatively small price effects: with 95 percent confidence, the ATE across the study period is below 0.0025 log points, or 0.25 percent. For the average maximum effect across markets, the upper bound of the 95 percent confidence interval is 0.013 log points, or 1.3 percent.

In addition to small overall price effects, we find small price effects across almost all product categories in our data, with some categories even experiencing price declines in response to greater transfer amounts as a fraction of GDP. In particular, food prices are in line with the overall price index, and durables do not increase dramatically in price. As predicted by theory, we observe a significant (but economically small) price increase for non-tradable goods, while the effect for tradable goods is smaller and not statistically significant.

In Columns 3 and 4, as described in Section 3.7.1, we classify markets by their level of market access, and report results separately for markets above and below the median of our market access measure. Estimated effects are frequently more positive in more remote markets. Given this, we look further into effects by market access in Figure 4, which presents the implied ATE by quartiles of market access, with lower values reflecting more isolated markets. Panels A-C show that there is a small amount of price inflation for non-tradables in the most isolated markets, and no detectable price changes elsewhere, or for other goods. Inflation for non-tradeables in isolated markets is nonetheless minimal, at 0.2-0.3% on average. These results are similar to findings from Cunha, De

Giorgi, and Jayachandran (2018), who study the price effects of an in-kind food and cash transfer program in Mexico. In-kind transfers lead to price decreases, while cash transfers lead to price increases, but the effects are small except in remote villages. In contrast, these price effects are far smaller than the Filmer et al. (2018) effects in the Philippines who find price increases of 5 - 7% for protein-rich foods such as eggs, fish and chicken.

To further emphasize the generally small nature of price effects, Figure A.1 presents implied ATEs for each product that we collect, with prices shaded by their classification into tradable and non-tradable sectors. Average price inflation is below 1% for every product in our data. This also helps mitigate concerns that results may be sensitive to our price index weights or product classification.

An important caveat to these results is that identification comes from variation in treatment intensity around different markets over time; we cannot fully rule out price increases throughout the entire study area compared to neighboring areas. However, we do not find much variation in prices across transfer amounts within our study area, despite sizable variation in treatment intensity: The 90-10 percentile range for total amount in annual per capita GDP going to the 0–2 km ring around each market over the entire period is [2%, 11%]. Moreover, the multi-year nature of our data for each market cover periods of intensive transfer distribution as well as periods where no transfers were going out. This suggests that we would not find price effects when comparing our study area to outside areas.

Overall, these results suggest a minimal output price response to a very large cash transfer program, and that one potential channel for negative spillovers to non-recipients (price inflation for consumer goods) is not a major concern in this study. This also suggests that much of the consumption expenditure and savings effects seen so far can be interpreted as effects in real terms. We next turn to price effects on input prices; in Section 8 we return to discuss the implications of these price effects for the production side of rural economies.

4.4.2 Input prices

While we have previously covered a number of aspects of input prices contained in Table 2, here we summarize the results in a single place. We focus on effects on prices and quantities for the inputs of labor, land and capital, measured through our household survey data.

We find some evidence of labor market effects, particularly for wages. In row 1 of Table 2, we look at wages for employees using household survey data. We include each household member that reports working for wages as an observation, and calculate their hourly wage based on their hours worked in the last 7 days and their monthly salary. In our reduced form specification, we find that eligible households in treatment villages earn USD PPP 0.09 more per hour, on a base of USD PPP 0.66. This effect becomes insignificant, however, when including across-village spillovers. For untreated households, the increase is even more marked at 0.19 USD PPP per hour, and this effect is significant at the 10% significance 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 insignificant. Together with the fact that enterprise wage bills increased significantly, this is another indication of positive wage effects (Table 4).

While land markets do exist, formal land sales are rare, making identifying land prices challenging. We therefore measure land prices from respondents' self-reports of how much an acre of land of similar quality to their own landholdings would sell for within their village. The advantage of this measure is that it provides land prices for all households in our sample. On the downside, these values are noisy. We find meaningful positive, but insignificant, effects on land prices (10-15%) for treated and untreated households(Table 2, Panel B.). As a robustness check, we look at land rental prices for households that rent in or rent out land; while this is only a subset of our respondents, this provides price data on the basis of actual land transactions. We do not find significant effects on land rental prices (Table A.3).

We likewise find little change in landholdings among treated households or those in more heavily treated areas. We also find no effects on total land rentals, nor the total amount of land used for agriculture (Table A.3).

We turn to capital markets next. We find no effects on household borrowing costs or total borrowing (Table 2, Panel C). We measure household borrowing from both formal (commercial banks, mobile credit services) and informal (family and friends, moneylender) sources. The total loan amount reports total borrowing across sources in the last 12 months, setting those that did not borrow equal to zero. 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.[27]

Taken together, there is some evidence for an increase in wages at unchanged levels of aggregate labor supply. Land and capital prices and usage, on the other hand, do not change significantly.

5 The fiscal 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 fiscal multiplier of cash transfers, where by 'local' we refer to the entirely of our study area. As is standard in the literature, we define this multiplier M as the cumulative effect of transfers on local GDP relative to the total amount T transferred:

$$\mathbb{M} = \frac{1}{T}\left(\int_{t=0}^{\infty} \Delta GDP_t\right) \tag{10}$$

The size of the fiscal 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 future taxation to change (e.g., whether Ricardian equivalence holds). Our setting is unusual in a useful way: because we observe a large one-time fiscal outlay that was made philanthropically and funded from outside of the economy we study, we can reasonably expect to measure a "pure" fiscal multiplier that should be independent of such

27. We include informal loans without interest, which brings down the average rate.

effects. This feature generates estimates that can be thought of as a model primitive against which estimates from other financing scenarios can be contrasted.[28]

An initial back-of-the-envelope calculation provides a first suggestion that the multiplier may be substantial: static estimates of annualized consumption, income, and revenue measured 18 months after a mean cash injection of 7.5% of GDP all increased by around 14 - 18%, suggesting a multiplier of roughly 2 to 2.5. 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 properly accounting for treatment effect dynamics. Given the precisely estimated and economically small price effects estimated above, we focus on impacts on nominal GDP.

Following national accounts definitions, our expenditure-based measure of local GDP is

$$GDP_t - NX_t = C_t + I_t + G_t$$

where C_t is consumption expenditure on non-durables and durables, measured as quarterly consumption plus accumulated assets at endline.[29] We exclude home and land values from our asset measure as our estimates of these may not be reliable, given thin local land and home rental markets, and because we think of land supply as being essentially fixed; doing so may lead us to err on the side of understating the multiplier. I_t is enterprise investment plus accumulated inventories at endline. We exclude local government expenditure, 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 also construct a dual income-based measure of local GDP as the sum of factor payments and profits:

$$GDP_t + NFI_t = W_t + R_t + \Pi_t + Tax_t$$

where W_t is the total household wage bill,[30] R_t are rental expenses of local enterprises,[31], Π_t are enterprise profits, and Tax_t is total enterprise taxes.[32]

The expenditure and income based measures of GDP we generate are based on unusually rich underlying data, but each has some potential limitations. In particular, a main concern is that they may misattribute transactions between agents located in our study area and counterparties located outside.

28. In fact, Chodorow-Reich (2019) shows that external financing and local debt-financing generate quantitatively similar multipliers in practice. Our estimate, therefore, may be closely related to what he calls the no-monetary-policy-response deficit-financed national multiplier.

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

30. We use the household rather than enterprise wage bill as our household sample is larger and includes individual-level wage earnings data.

31. We omit land rental income because we do not see any significant evidence of this above. This may lead us to understate the multiplier.

32. 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.

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. These would likely not be due to direct imports by households themselves: on average only 10% of households report shopping at a market outside our study area, and overall the impacts we see on household spending and 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. Since many local firms are retail establishments, however, imports of intermediate goods (including for example packaged consumer goods ready for sale) would likely increase. This suggests that the expenditure-based approach might be upwardly biased away from zero. (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 which our concept of the local multiplier does not capture but which are a part of broader impact of the intervention.

In the income case, on the other hand, potential bias could arise if there are changes in net wage income (NFI_t) earned outside the study area, since this not considered part of local GDP. This bias seems unlikely to be quantitatively important in our setting: 86% of all employees of study-area enterprises are family labor (and therefore presumaby 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.

Overall, we view the expenditure- and income-based multipliers as two distinct measures of the same underlying concept, each with its own potential data limitations, omissions and biases. Reflecting this, below we estimate them jointly and test joint as well as individual hypotheses across the two measures.

5.1 Empirical strategy

To capture the full dynamic response of GDP, we estimate and then integrate effects on each component discussed above over time. For each flow variable x (e.g. consumption, investment, wage income, etc.), we first estimate the following specification, which is a dynamic extension of our previous estimating equations:

$$x_{it,v} = \alpha_t + \sum_{s=0}^{10} \beta_s \tilde{Amt}_{v,t-s} + \sum_{s=0}^{10} \gamma_s \tilde{Amt}^{\text{0-2km}}_{\neg v,t-s} + \varepsilon_{it,v}$$

Here, $x_{it,v}$ is the value of x for household or enterprise i in village v in quarter t, $\tilde{Amt}_{v,t-s}$ is the per-capita amount of transfers going to village v in quarter $t-s$, $\tilde{Amt}^{\text{0-2km}}_{\neg v,t-s}$ is the per-capita amount going to villages other than v in the 0 to 2km buffer, and α_t is a quarter fixed effect.[33] We instrument $\tilde{Amt}_{v,t-s}$ with the treatment indicator multiplied by the transfer amount to village v in quarter $t-s$ as a share of total amount going to village v overall. As an instrument for $\tilde{Amt}^{\text{0-2km}}_{\neg v,t-s}$, we use the share of eligible households in the buffer assigned to treatment, multiplied by the transfer amount to villages other than v in the 2 km buffer in quarter $t-s$ as a share of the total amount going to other villages than v in the buffer overall. Where available, we control for the baseline value of x, or its mean and an indicator for being missing. Regressions employ inverse-sampling weights.

The coefficients in this model are well-identified since treatment (and survey times) were not only randomized across villages, but also rolled out to different villages in a randomized order. The main challenge is that the first household surveys started around 12 months after the experimental start date in each village, while enterprise surveys began about 18 months out (see Figure 3). 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 to a village, which is when we might expect to see some of the largest impacts on expenditure. We can still estimate the own-village response during these early quarters using variation in transfers coming into the village gradually over time, as the model treats every transfer dollar symmetrically. Similarly, we can estimate neighborhood effects using variation in the timing in which nearby villages were treated. We tend to obtain less precise estimates of responses in those early quarters immediately following transfers, however, as they are identified using less variation in treatment intensity compared to later quarters. Given this, we also examine (and present in the appendix) estimates of the multiplier that assume no impacts occurred in the first three quarters following a transfer; these estimates are almost surely biased downwards, as they miss any activity in the initial period following transfers, but are considerably more precisely estimated and thus may be attractive to consider if the goal is to minimize mean squared error.

We then integrate dynamic effects on flow variables over time up to 27 months (9 quarters) after treatment. We compute the dynamic profile of treatment effects (or impulse response function, IRF) using the coefficients estimated above and assuming that the treatment rolled out to treated households as planned (to avoid any variation in transfers induced endogenously by local factors): the timing is a token transfer at time 0, a first lump-sum 2 months later, and a second after 8 months. We compute this IRF separately for treated and untreated household,s and separately for three categories of enterprise: own-farm, and non-farm enterprises operating within, and those operating from outside the homestead in treatment and control villages. We then aggregate our 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.

For two components of GDP, we are instead able to measure impacts on the integral of flows over

33. We use only the 0 to 2 km buffer since our pre-specifed algorithm almost universally selected this specification.

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 household durable goods, and effects on inventory investment using effects on current inventory stocks. One drawback is that these figures may somewhat under-estimate cumulative spending to the extent that some assets depreciated between the time of purchase and measurement, although over the timeframe considered this may be a second-order concern. In the graphical presentation below, we assume that any effects on these stocks occurred equally across all post-treatment quarters we observe.

We conduct inference using the wild bootstrap clustered by sublocation, the highest unit of randomization (Cameron, Gelbach, and Miller (2008)). We test two one-sided hypotheses, namely, that each multiplier is less than zero and is less than one, as well as the joint hypothesis that both multiplier estimates satisfy these conditions. There has been extensive debate in the macroeconomic literature about whether or not the fiscal multiplier is greater than one (see Chodorow-Reich (2019) for a detailed discussion of existing estimates).

5.2 Multiplier estimates

Table 6 presents the two main multiplier estimates, as well as their components. In line with the rough back-of-envelope figure derived above, we estimate a sizeable multiplier using both approaches.

To start, the estimated expenditure multiplier is 2.62. Over 65% of this effect is driven by non-durable consumption expenditures. Durable asset purchases and enterprise investment make up another 15% each, while enterprise inventories are not quantitatively important. While part of the asset response could potentially reflect productive investments by household-operated enterprises, at least two thirds of the asset response appears to come from non-productive assets, across both treated and untreated households (see Table A.1). Taking this into account, consumption alone leads to an estimated multiplier above 2, 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 multiplier is quite similar in magnitude, at 2.71. This is notable since it is calculated using a completely distinct set of component measures. Of this total multiplier, we find that 59% reflects increased enterprise profits, 37% increased wages, and a much smaller contribution (4%) from capital income and taxes taken together. The increase in consumption, and the smaller increase in investment, we noted above 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.

One of the advantages of our "macro-experimental" approach to estimating the multiplier is the ability to conduct inference. To start, we reject the null of a negative multiplier (with a value less than zero) at the 10% level using either approach (Table 6), which is reassuring if not

surprising. Testing the null hypothesis of a multiplier less than or equal to 1 speaks more directly to recent debates. Using either the expenditure or income based approach alone, the p-value on this hypothesis is marginally significant at 0.18. Since the two measures exploit distinct data, we gain statistical power by testing the joint restriction that the null holds for both measures; when we do so, the null hypothesis that the multipliers are jointly less than or equal to one is rejected at the $p = 0.08$ level.

As noted above, the alternative approach of excluding effects on GDP during the first three quarters after transfers arrive (and thus assuming they are zero) may be attractive in a mean squared error sense since it yields more precise, if somewhat biased, estimates. Under this assumption, the estimates range from 1.7 to 2.0 (smaller than our preferred estimates in Table 6, as expected), and the null hypothesis that both multipliers estimates are jointly less than one is rejected at the $p = 0.06$ level (Table A.5).

Figure 5 breaks up the aggregate multipliers into quarters. The increase in GDP is fairly stable. If anything, it increases slightly up to a peak after 9 months (when the second lump-sum transfer has been received), and then slowly decreases. Interestingly, we reject a null effect even as far as two years after the transfer. In fact, we cannot reject that the expenditure response is constant across all quarters (p-value of 0.88). The income multiplier, on the other hand, visually appears to fluctuate more over time: 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 0.89). Future research could usefully investigate the dynamics of economic activity following a cash infusion in further detail.

Overall, we estimate large multipliers using two distinct sets of measures, in terms of both expenditures and income, in the range of 2.6 to 2.7. We are able to reject the null hypothesis that the multipliers are jointly less than one at over 90% confidence. These estimates are somewhat larger than the higher end of recent 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. The finding that multiplier estimates tend to be largest in the poorest US regions seems broadly consistent with our even larger estimates in rural Kenya, which is far poorer still. Beyond effects driven by the relative level of economic development, the differences between our estimates and existing US estimates may reflect other structural differences between the US and Kenyan economies, differences in the data used, as well as any effects on (or expectations of effects on) either monetary policy or future tax policy in the US, response effects that our experimental design allows us to avoid.

6 Non-market outcomes and externalities

The analysis so far has focused on outcomes such as income, expenditure and prices that affect the household budget constraint. We turn next to examining variables that arguably directly influence (or proxy for) well-being, and in particular to outcomes that may reflect externalities either between

or within households.

We focus in particular on six indices, constructed following Anderson (2008) and with components signed so that positive values indicate better outcomes. The *psychological well-being index* is calculated from depression, happiness, life satisfaction, and stress scales for the respondent. We interpret this as an overall measure of subjective well-being, capturing both monetary impacts and externalities. The *health status index* is made up of self-reported health, an index of common symptoms, and an indicator for whether the respondent has experienced a major health problem since the time of the baseline survey. The *food security index* is composed of measures for the number of times a) adults and b) children are skipping meals, going entire days without food, and going to bed hungry in the last week, as well as the number of meals with protein yesterday. We also utilize the survey questions about children to construct a subindex of child food security. This may capture intra-household externalities to the extent children do not have say in the household's allocation of resources. The *education index* is composed of total education expenditure by the household, and the proportion of school-aged children in school, and may similarly capture intra-household externalities. The *female empowerment index* is calculated by combining a gender violence index and a gender attitudes index. Due to the sensitive nature of some of these questions and cultural considerations, information on gender-based violence was only collected from female respondents being surveyed by female enumerators. Finally, the *security index* is higher the lower are measured levels of crime.[34]

For treated households, we find positive and significant effects for four of the six indices: psychological well-being, food security, education and security. We find statistically insignificant effects for our health index and female empowerment index. When looking at total effects, we find a similar pattern for all but the security index.[35] For untreated households, we find no significant changes except for the education index, which is higher by 0.1sd ($p < 0.10$). Importantly, we do not find evidence of adverse effects for untreated households on any of the indices, with point estimates positive for all but the security index, which is indistinguishable from zero (-0.02 SD, SE 0.07).

6.1 Inequality

One potential form of "psychic externality" has to do with the distribution of economic status within affected communities. If households care not just about their absolute levels of consumption or wealth, but also about comparisons with their neighbors, then these distributions may enter directly into utility functions. From the results above it is not entirely clear how consumption and wealth distributions changed; transfers were targeted to relatively poor households within each

34. The first five of these were pre-specified as primary outcomes in Haushofer et al. (2017); for the sixth, the security index, all components were pre-specified as part of a family of outcomes, though combining into a single summary index was not. Results for components of each index are presented in Appendix A.

35. It is interesting that we do not find that treatment leads to an increase in female empowerment for treated households. The point estimates on both the reduced form and total effect are negative but not statistically significant (-0.03 and -0.13, respectively). This contrasts with the results from Haushofer and Shapiro (2016), which found improvements in female empowerment and reductions in domestic violence among households receiving the larger transfer.

community, but we find large spillovers onto their neighbors.

To examine this more carefully, we first calculate Gini coefficients for consumption expenditure and wealth by village and estimate village-level treatment effects. We find no significant differences in either measure, and in fact estimate a positive and significant effect on the wealth Gini (Table A.10). These results reflect the magnitude of the spillovers documented above. As a benchmark, we also calculate the effects on inequality we might have expected to observe had there been *no* spillovers. This is especially meaningful for assets, as we can generate counterfactuals based on baseline asset ownership patterns, whereas all endline expenditure measures are potentially influenced by spillover effects. Specifically, we generate counterfactual expenditure and asset outcomes for recipient households assuming a marginal propensity to consume (save) out of transfer income of 0.4 (0.6). This equals the ratio of estimated impacts on consumption expenditure to estimated impacts on consumption expenditure, assets and housing values that we actually observe in our data (Table 1). We then multiply these marginal propensities by the transfer value and apply these to recipient households, while setting untreated households to their baseline values. In these counterfactual simulations, treatment does in fact reduce the Gini coefficient for assets, and we can reject the hypothesis that the observed and counterfactual treatment effects are the same for both treatment and control villages. We do not find significant reduction in the consumption expenditure counterfactual, suggesting endline measures, even from control, low saturation areas, have also been influenced by spillover effects.

Overall, these results highlight again the role of large spillovers onto non-recipient households. These spillovers mean that wealthier non-recipients benefit along with poorer recipients, but on some dimensions (e.g., wealth) they benefit so much that inequality increases, which may itself be a concern to the extent that wealth inequality is undesirable per se.

7 Conceptual framework for household welfare

The purchasing power shock we study evidently had wide-ranging effects on the local economy. To help interpret these effects from a welfare perspective, we next describe a framework that, while fairly generic, categorizes the ways in which these effects translate into changes in the welfare of affected households, including both households that did and did not receive transfers. We also use the framework to examine the relationship between well-being as typically conceptualized by economists, and short-hand metrics such as consumption expenditure, which are commonly used in applied work.

We are interested in understanding the indirect utility function $v_i(T_i, T)$ which defines the utility level achieved by household i when it receives a (possibly zero) transfer T_i while other eligible households in area receive transfers of T each. Define by T_i^* the transfer that would make household i indifferent between receiving T_i^* on the one hand, and receiving the full intervention we study on the other:

$$v_i(T_i^*, 0) = v_i(T_i, T) \tag{11}$$

We aim to characterize the ways in which T affects T_i^*. If there were no general equilibrium effects, then T would be irrelevant and we would simply have $T_i^* = T_i$, i.e. the tautology that a dollar is worth a dollar.

Generally speaking, we can write v_i as the value of some underlying optimization problem

$$v_i(T_i, T) = \max_{x_i} u_i(x_i, x_{-i}(T)) \text{ s.t. } x_i \in X(T_i, T) \tag{12}$$

Here u_i captures the household's preferences over variables x_i, which it chooses, as well as variables x_{-i}, that others choose and that directly effect i (e.g. through externalities, the provision of public goods, preferences over inequality, etc.). X is household i's constraint set. This formulation highlights the fact that a change in the amount T transferred into the local economy can alter the utility of household i (and hence its indifference point T_i^*) in two broad ways. First, it may change the constraint set X, for example by changing the prices facing i, or its income from various sources. Second, it may change behaviors of other households that directly affect i's well-being independent of its constraint set. If we interpret i as an individual rather than a household, then the same point applies to intra-household externalities such as parental investment in children's education. We organize our empirical analysis below around this broad categorization.

The specification of the constraint set determines which specific general equilibrium effects matter for household well-being. Henceforth let

$$v_i(T_i, T) = \max_{\{x_{it}\}} u_i(\{x_{it}\}, \{x_{-it}\}) + \lambda_i \left(T_i + \sum_{t=0}^{\infty} (\delta_i)^t (y_{it} - p_t \cdot f(x_{it})) \right) \tag{13}$$

Preferences here are over a sequence of x's; these need not be time separable, which allows for non-standard time discounting, the existing of durable goods that generate a flow of benefits, etc. The household optimizes subject to an intertemporal money budget constraint which depends on the transfer T_i it receives at time $t = 0$; the idiosyncratic discount rate δ_i it faces; any transfer income y_{it} it receives from other households, government, or claims on the profits of local enterprises; the prices p_t it faces, and the (net) quantities $f(x_{it})$ of each good or service it purchases (i.e. the sale of goods or services is captured by negative components of $f(x_{it})$). In the simplest case $f(x) = x$ and the household simply buys or sells all commodities in the market, but the more general structure here allows for non-separable household production using non-marketed inputs such as family labor (LaFave and Thomas 2016).[36]

The specification in (13) implies that the constraint set depends on two sets of factors: prices, including both the per-period prices p_t of goods and services and the intertemporal price δ_i, and transfer income y_{it}. In particular, it implies that it is important to distinguish between earned labor income (which appears in the $p_t \cdot f(x_{it})$ term and is a function of wages) and unearned transfer income such as economic profit from an owned enterprise (which appears as part of y_{it}).

Note also that (13) allows for some but not all forms of credit constraint. There may be arbitrary

36. For simplicity we exposit the framework under certainty about the future. The analysis that follows would be effectively the same, however, if we were to introduce uncertainty and take expectations.

credit constraints within the firms that produce profits which enter into y_{it}, and there may also be credit market imperfections such that different households face different saving or borrowing rates δ_i. If we further modelled hard constraints on household borrowing (e.g. a non-negative assets constraint) then the shadow value of money would potentially vary from period to period for some households. This would not change the qualitative result that welfare effects are a function of price and income changes, but would alter the relative quantitative value of price and income changes at different points in time.

7.1 Trading off price and income effects

In cases where the solution to the household's problem is characterized by first-order conditions (i.e. $\partial u_i/\partial x_{it} = \lambda_i \delta_i^t p_t \partial f/\partial x_{it}$) we can go further, applying the envelope theorem to calculate the relative value of the various changes to its budget set. The envelope theorem implies

$$\frac{dv_i}{dT} = \frac{\partial u_i}{\partial x_{-it}} \cdot \frac{\partial x_{-it}}{\partial T_i} + \lambda_i \left(\sum_{t=0}^{\infty} (\delta_i)^t \left(\frac{\partial y_{it}}{\partial T} - \frac{\partial p_t}{\partial T} \cdot f(x_{it}) + \frac{t}{\delta_i} \frac{\partial \delta_i}{\partial T} s_{it} \right) \right) \tag{14}$$

where $s_{it} = y_{it} - p_t \cdot f(x_{it})$ is period t savings. In comparison, the welfare effect of a marginal change in T_i holding other transfers fixed is simply

$$\frac{\partial v_i}{\partial T_i} = \lambda_i \tag{15}$$

The (marginal) equivalent variation is the ratio of these two, or

$$\frac{dEV_i}{dT} = \frac{1}{\lambda_i} \frac{\partial u_i}{\partial x_{-it}} \cdot \frac{\partial x_{-it}}{\partial T_i} + \sum_{t=0}^{\infty} (\delta_i)^t \left(\frac{\partial y_{it}}{\partial T} - \frac{\partial p_t}{\partial T} \cdot f(x_{it}) + \frac{t}{\delta_i} \frac{\partial \delta_i}{\partial T} s_{it} \right) \tag{16}$$

The second term in this expression captures changes in i's budget set. These can be priced in dollar terms: an increase in profits from owned enterprises of \$1, for example, is simply worth \$1, while an increase of \$1 in the unit price of a good of which the household purchases 10 units is worth $-$\$10. The second term captures externalities. If we know what behaviors have external effects then we can examine the relative comparative statics $\frac{\partial x_{-it}}{\partial T_i}$. These are difficult to price as we do not know the value $\frac{1}{\lambda_i} \frac{\partial u_i}{\partial x_{-it}}$, which the household places on these behaviors relative to money.

In principle, (16) could be used to calculate numeric estimates of the welfare consequences we see in our data for each household i. In practice this would be quite demanding of the data we have, as it requires observation or interpolation of the full time path of all outcomes and for all counterfactual transfer amounts $T \in [0, \overline{T}]$. Realistically this would require that we impose some degree of structure (e.g. homotheticity) on preferences, something we do not pursue here.[37] Note

37. The total equivalent variation induced by a transfer of $\overline{T}$ is given by

$$EV_i(\overline{T}) - EV_i(0) = \int_0^{\overline{T}} \frac{dEV_i}{dT} dT = \int_0^{\overline{T}} \sum_{t=0}^{\infty} (\delta_i)^t \left(\frac{\partial y_{it}}{\partial T} - \frac{\partial p_t}{\partial T} \cdot f(x_{it}) + \frac{t}{\delta_i} \frac{\partial \delta_i}{\partial T} s_{it} \right) dT \tag{17}$$

also that (16) does not correctly value the effects of price changes if first-order conditions do not apply, as would be the case if for example the household made important discrete choices such as occupation. In practice, we find small and precisely estimated effects on most prices with the possible exception of wages, which we discuss below.

7.2 A dual approach using expenditure

Changes in household consumption expenditure are often used as proxies for changes in well-being. To see how this metric relates to well-being in the sense of equivalent variation defined above, it is helpful to decompose the household's choices

$$f(x_{it}) = x_{it} = c_{it} + l_{it}$$

into the goods and services $c_{it} = \max\{0, x_{it}\}$ that it buys (e.g., clothes) and those $l_{it} = \min\{0, x_{it}\}$ that it sells (e.g. labor services). Expenditure is

$$e_i = \sum (\delta_i)^t p_t \cdot c_{it}$$

Differentiating the budget constraint,

$$\frac{de_i}{dT} = \sum (\delta_i)^t \left(\frac{\partial y_{it}}{\partial T} - \frac{\partial p_t}{\partial T} \cdot l_{it} - p_t \cdot \frac{\partial l_{it}}{\partial T} + \frac{t}{\delta_i} \frac{\partial \delta_i}{\partial T} (y_{it} - p_t \cdot l_{it}) \right) \tag{18}$$

Re-arranging terms (and for simplicity ignoring externalities for the moment) we obtain

$$\frac{dEV_i}{dT} = \frac{de_i}{dT} - \sum (\delta_i)^t \left(\frac{\partial p_t}{\partial T} \cdot c_{it} - p_t \cdot \frac{\partial l_{it}}{\partial T} + \frac{t}{\delta_i} \frac{\partial \delta_i}{\partial T} p_t \cdot c_{it} \right) \tag{19}$$

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, expenditure incorrectly counts appreciation of the price of consumption goods ($\frac{\partial p_t}{dT} \cdot c_{it}$) or the time path of consumption ($\frac{\partial \delta_i}{\partial T} p_t \cdot c_{it}$) as a welfare gain. Second, it incorrectly counts income gains due to behavioral responses such as increased labor supply ($p_t \cdot \frac{\partial l_{it}}{\partial T}$) as a welfare gain. One additional source of bias does not appear here but would if we measured expenditure over any finite period: it incorrectly counts an increase in spending financed by dis-saving as a welfare gain. Overall, this confirms that expenditure can be a useful metric to use for welfare analysis, with the caveats that it is important to measure increases in real rather than nominal spending, and to check that these are not financed by increased labor supply or decreased saving.

7.3 Implications & interpretation

Overall, the framework outlined here emphasizes the roles of the budget set and of externalities in driving welfare outcomes, and the conditions under which expenditure can serve as a good proxy for the former. Viewed through his lens, we can interpret our empirical results as follows.

First, all evidence points to a welfare-improving expansion in the budget sets of non-recipient households. Changes in consumer prices were small, while earnings increased substantially, driven primarily by an increase in labor earnings with a smaller and insignificant increase in enterprise profits. While the data are not dispositive, the increase in labor earnings seems likely to reflect at least in part an increase in wages. Working from expenditure yields a similar conclusion: both real flow expenditure and stocks of consumer durables (which reflect past expenditures) increase, results which are not accounted for purely by changes in consumer prices or in labor supply.

Second, what data we have on welfare changes via externalities suggests that these were either zero or in some cases slightly positive. Public goods provision at the village level was unchanged (Walker 2018) as were levels of crime (Table 7, row 7). Non-market outcomes also improved: non-recipient households scored insignificantly higher on an index of psychological well-being. Within households, children – who likely may have less say in decision-making than adults – appeared better off, with significantly higher food security and education index scores (Table 7, rows 4 and 5). An index of female empowerment, on the other hand, did not change significantly among either recipient or non-recipient households (Table 7, row 7).

8 Discussion: productive responses

One of the more striking features of the results above is that that enterprise sales increase substantially with only negligible increases in consumer prices, implying that real output increased. This raises the question what features of the productive economy enabled it to respond elastically to a very large shock to aggregate demand.

Purely as an accounting decomposition, increased output must be attributable to some mix of (a) increased use of intermediate goods (including goods ready for re-sale, as in retail enterprises) and (b) increased value added. Increases in value added must in turn reflect some mix of (i) increased employment of factors of production (in particular labor and capital, assuming land is in fixed supply) and (ii) increased *utilization* of factors already employed.

Increased use of intermediates seems likely to be a part of the story, though because we did not measure enterprise spending on intermediates we cannot examine this directly. Retail makes up a sizeable share of the economy (42% by revenue, Table A.11), and the enterprise revenue gains we document are largely concentrated in retail along with manufacturing (Table A.4). A large share of intermediate goods in the value of sales is consistent with the small effects we find on consumer goods prices.

We find relatively limited evidence of increased employment of factors, on the other hand. We find no significant change in enterprise investment in fixed capital (Table 4, row 6), and while

inventories increased somewhat they actually decrease in proportion to sales (26 to 21%, Table 4, row 5). Moreover, if output increased due to investment activity 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 A.12).[38] Turning to labor supply, we find no significant change in total household labor supply (Table 2). Wage bills do increase, accounting for 27% of increased revenue (Table 4), but given the absence of a labor supply response this seems likely to primarily reflect higher wages, and we find marginally significant but large increases for untreated households (Table 2, row 1). Lastly, land is in fixed supply and (unsurprisingly) agricultural households do not report owning or renting more of it (Table A.3).

Overall, these results suggest that there may have been a degree of "slack" in the local economy, in the sense that it was possible to substantially increase output without substantially increasing the employment of factors of production. This is broadly consistent with what we know about the structure of the economy, and in particular the fact that output gains were concentrated in retail and manufacturing. Both industries in this area typically involve "on-demand" production: a shopkeeper waits to work until customers arrives, and similarly many small-scale manufacturing enterprises such as posho (grain) mills require someone to "man the shop" even when there are no customers to serve.[39] And they appear to spent substantial amounts of time waiting: the mean non-agricultural enterprise in our sample reports an average of just 1.7 customers per hour they are open. Moreover, it seem likely that at the small scales they operate integer constraints in hiring bind for many; a majority of non-agricultural enterprises (72%) have only a single employee.

This interpretation is consistent with other recent findings in East African settings. In ongoing work in Uganda, for example, Bassi & Porzio find that employees in on-demand manufacturing (welding, furniture-making) spend about 25% of time "waiting for customers" or "eating and resting." More broadly, it harkens back to classic arguments in development economics that it might be possible to expand production without notable price inflation due to the availability of "surplus labor" (Lewis 1954). A growing literature has also found evidence of excess capacity in rich countries, especially in periods of recession (e.g., Michaillat and Saez (2015); Chodorow-Reich (2019)).

9 Conclusion

We find that a large-scale cash transfer program in rural Kenya leads to sharp increases in the consumption expenditures of treated 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 untreated and treated households approximately a year and a half after the initial transfers. Local firms do not show meaningful increases in investment, and there is minimal local price inflation, with quite precisely

38. Recall that we can match all agricultural all enterprises to eligible or ineligible owners, and have matched 56% of non-agricultural enterprises.

39. Posho mills and welding, both of which share this need to wait for customers, account for over 60 percent of manufacturing revenue.

estimated effects of far less than 1% on average across a wide range of goods. Taken together, two different empirical approaches – one using consumption data and the other income data – both yield an estimated fiscal multiplier of 2.6 to 2.7, and the hypothesis that the multiplier is less than equal to 1 is rejected at 90% confidence. This multiplier is somewhat larger than recent estimates from the United States, which fall in the range of 1.5-2.0. Several suggestive patterns in the data in our context are consistent with the existence of considerable "slack" on the supply side, which can also help account for the large positive estimated multiplier.

Some authors working in the recent cash transfer literature in development economics have been concerned about the possibility that large-scale programs like the one we study could lead to a range of potentially adverse consequences for non-beneficiaries, however, these concerns are not born out in our data. The consumption expenditures of untreated households and firms rise substantially in areas receiving large cash transfers, there is little to no price inflation, overall income, consumption and asset inequality does not increase meaningfully in treated areas, nor are there negative effects along a range of other outcomes that we measure, including in terms of gender norms and domestic violence, health, education and local public goods. An important further implication of our finding of, if anything, positive spillover effects along a number of important dimensions, including consumption, is that some other recent RCT's of cash transfer programs that simply compare outcomes across treatment and control villages may be understating true overall impacts by ignoring the general equilibrium effects that we capture (Miguel and Kremer 2004)

This study is among the first to our knowledge to exploit experimental randomized control trial methods to directly estimate macroeconomic parameters and more broadly capture large-scale aggregate effects of a development program. The parameter we focus on here – the fiscal multiplier – has been of intense interest in the field since at least the seminal work of Keynes (1936). In doing so, we provide a novel counter-example to some well-known critiques of RCT methods in development economics, including by Deaton (2010), Bardhan (2005), and Easterly (2006), who have claimed that experimental methods are not well-suited to studying the big questions in economics. In doing so, we hope to demonstrate that there need not be a trade-off between a study's rigor and its relevance: economics research can increasingly hope to achieve both (Muralidharan and Niehaus 2017; Burke, Bergquist, and Miguel 2019).

A traditional perspective would be that, once a local aggregate demand shock (like the cash transfers in our study) ends, the economy would eventually return to the previous equilibrium. However, other theoretical perspectives from international trade and economic geography (Marshall (1890), Krugman (1980), among others) would suggest there could be persistent local effects of a temporary cash infusion, due to agglomeration effects, changes in market structure and firm specialization, and even shifts in the social networks of traders and suppliers during the period of the shock. Temporary cash transfers and other forms of assistance have also been shown to have effects on long-run human capital accumulation and earnings in low income settings (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 and novel experimental empirical test of these theories. Additional measures, alternative research designs and structural econometric methods may be needed to evaluate long-run effects in the presence of large positive spillovers, as these may tend to equalize impacts across space over time.

References

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association* 103 (484): 1481–1495.

Angelucci, Manuela, and Giacomo De Giorgi. 2009. "Indirect effects of an aid program: how do cash transfers affect ineligibles' consumption?" *American Economic Review* 99 (1): 486–508.

Auerbach, Alan J, Yuriy Gorodnichenko, and Daniel Murphy. 2019. *Local Fiscal Multipliers and Fiscal Spillovers in the United States.* Working Paper, Working Paper Series 25457. National Bureau of Economic Research, January. doi:10.3386/w25457.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. 2016. "Worms at Work: Long-run Impacts of a Child Health Investment*." *The Quarterly Journal of Economics* 131, no. 4 (July): 1637–1680. ISSN: 0033-5533.

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.

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

Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, and Tanja Schmidt with Luca Pellerano. 2016. *Cash transfers: what does the evidence say?: A rigorous review of programme impact and of the role of design and implementation features.* Overseas Development Institute, July.

Bobonis, Gustavo, and Frederico Finan. 2009. "Neighborhood Peer Effects in Secondary School Enrollment Decisions." *The Review of Economics and Statistics* 91 (4): 695–716.

Bouguen, Adrien, Yue Huang, Michael Kremer, and Edward Miguel. 2019. "Using RCTs to Estimate Long-Run Impacts in Development Economics." *forthcoming in Annual Review of Economics.*

Burke, Marshall, Lauren Falcao Bergquist, and Edward Miguel. 2019. "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets." *The Quarterly Journal of Economics* 134, no. 2 (December): 785–842. ISSN: 0033-5533.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *The Review of Economics and Statistics* 90 (3): 414–427.

Chetty, Raj. 2009. "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods." *Annual Review of Economics* 1 (1): 451–488.

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

Conley, Timothy G. 1999. "GMM estimation with cross sectional dependence." *Journal of Econometrics* 92, no. 1 (September): 1–45.

———. 2008. "Spatial Econometrics." In *The New Palgrave Dictionary of Economics,* Second Edition, edited by Steven N. Durlauf and Lawrence E. Blume, 7:741–747. Houndsmills: Palgrave Macmillan.

Cunha, Jesse M, Giacomo De Giorgi, and Seema Jayachandran. 2018. "The Price Effects of Cash Versus In-Kind Transfers." *The Review of Economic Studies* 86, no. 1 (April): 240–281. ISSN: 0034-6527.

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

Donaldson, Dave, and Rik Hornbeck. 2016. "Railroads and American Economic Growth: A 'Market Access' Approach." *Quarterly Journal of Economics* 131 (2): 799–858.

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.

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

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

Haushofer, J., J. Reisinger, and J. Shapiro. 2019. "Is Your Gain My Pain? Effects of Relative Income and Inequality on Psychological Well-being." Working Paper.

Haushofer, Johannes, Edward Miguel, Paul Niehaus, and Michael Walker. 2014. "General Equilibrium Effects of Cash Transfers in Kenya." AEA Trial Registry. November. `https://www.socialscienceregistry.org/trials/505/history/3031`.

———. 2016. "Pre-analysis Plan for Midline Data: General Equilibrium Effects of Cash Transfers." May.

———. 2017. "GE Effects of Cash Transfers: Pre-analysis plan for household welfare analysis." July.

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *The Quarterly Journal of Economics* 131 (4): 1973–2042. ISSN: 0033-5533, 1531-4650.

Jack, William, and Tavneet Suri. 2011. "Mobile Money: The Economics of M-PESA." NBER Working Paper No. 16721, January.

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

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *Review of Economics and Statistics* 91 (3).

Krugman, Paul. 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade." *American Economic Review* 70, no. 5 (December): 950–959.

LaFave, Daniel, and Duncan Thomas. 2016. "Farms, Families, and Markets: New Evidence on Completeness of Markets in Agricultural Settings." *Econometrica* 84 (5): 1917–1960.

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

Marshall, Alfred. 1890. *The Principles of Economics.* McMaster University Archive for the History of Economic Thought.

Mbiti, Isaac, and David N. Weil. 2015. "Mobile Banking: The Impact of M-Pesa in Kenya." In *African Successes, Volume III: Modernization and Development,* 247–293. NBER Chapters. National Bureau of Economic Research, Inc, March.

McKenzie, David. 2012. "Beyond baseline and follow-up: The case for more T in experiments." *Journal of Development Economics* 99 (2): 210–221.

Mel, Suresh de, David J. McKenzie, and Christopher Woodruff. 2009. "Measuring microenterprise profits: Must we ask how the sausage is made?" *Journal of Development Economics* 88, no. 1 (January): 19–31.

Michaillat, Pascal, and Emmanuel Saez. 2015. "Aggregate Demand, Idle Time, and Unemployment." *Quarterly Journal of Economics* 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.

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

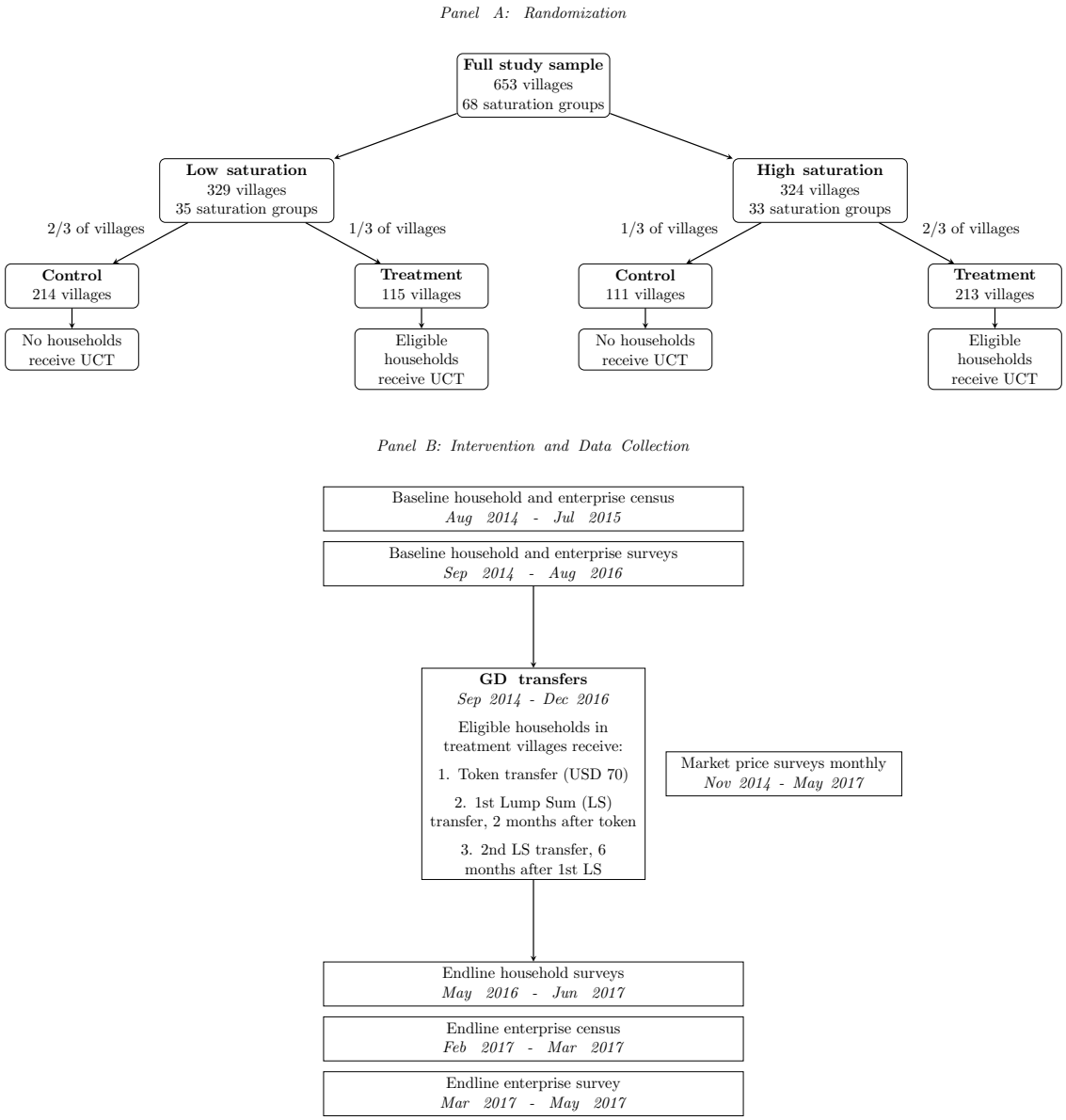
Murphy, Daniel. 2017. "Excess capacity in a fixed-cost economy." *European Economic Review* 91 (C): 245–260.

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.

Ozler, Berk. 2018. "Most good you can do. But for whom?" October 22. https://blogs.worldbank.org/impactevaluations/most-good-you-can-do-whom.

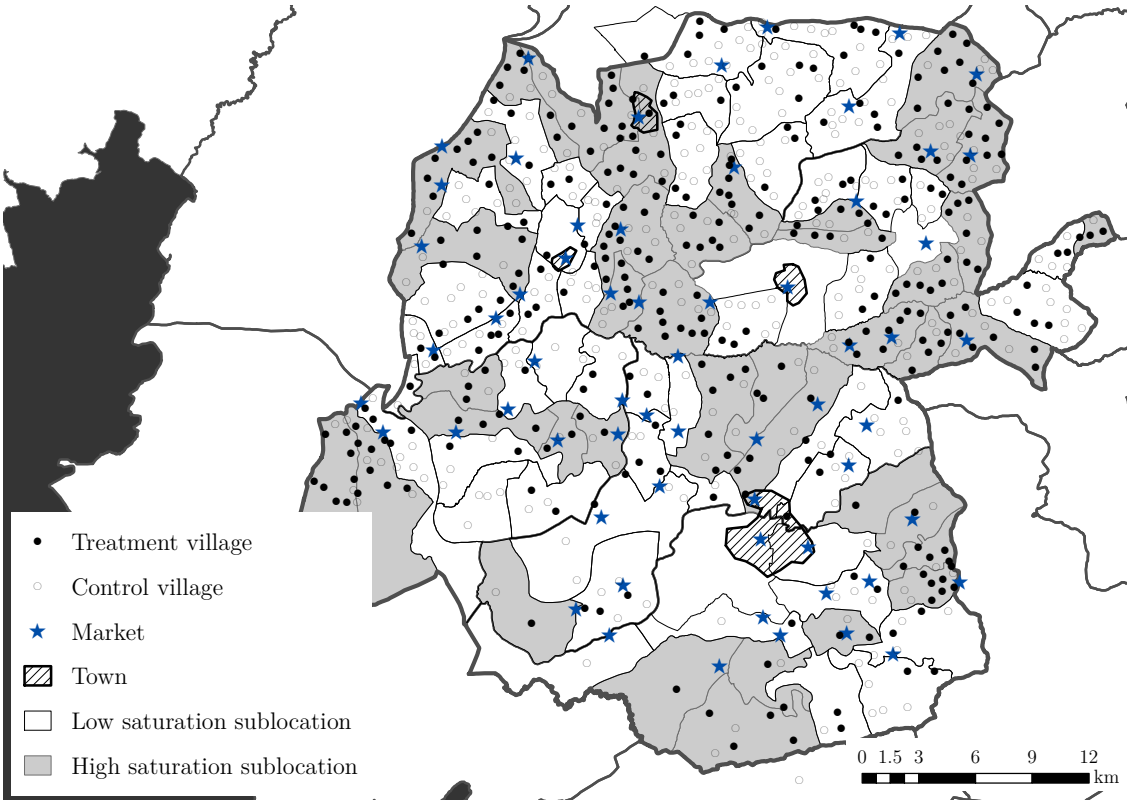
Walker, Michael. 2018. "Informal Taxation Responses to Cash Transfers: Experimental Evidence from Kenya." July.

Figure 1: Study design



Notes: This figure 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. 99 percent of transfer value was sent by December 2016.

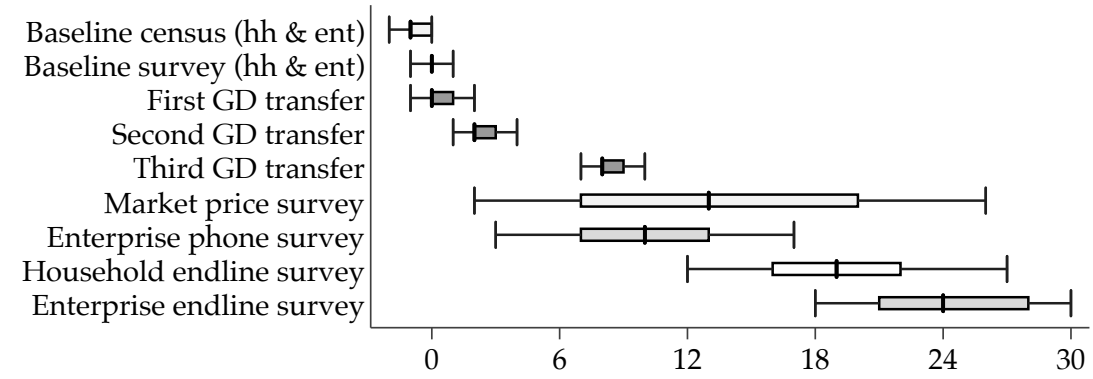
Figure 2: Study area



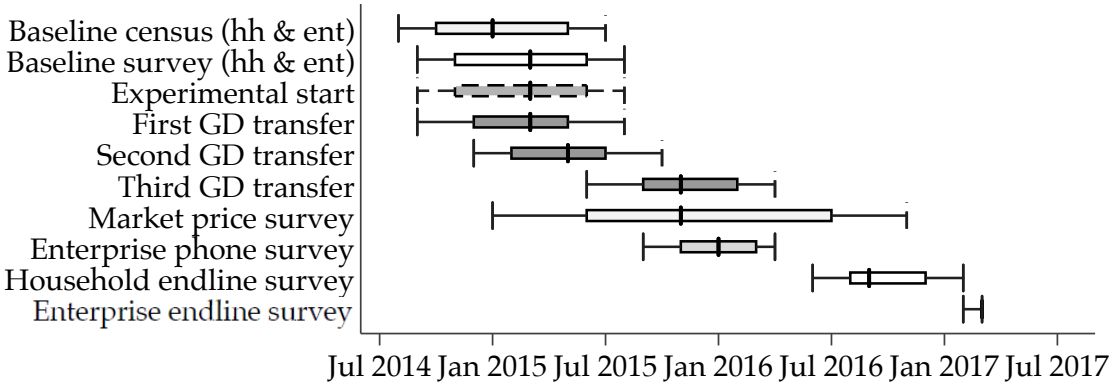
Notes: This figure plots study villages, sublocation boundaries, and weekly markets in the study area in Siaya County, Kenya. Control villages are denoted by hollow circles, treatment villages are denoted by solid circles, and blue stars indicate the locations of markets. High saturation sublocations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

Figure 3: Study timeline

(a) Timing relative to experimental start

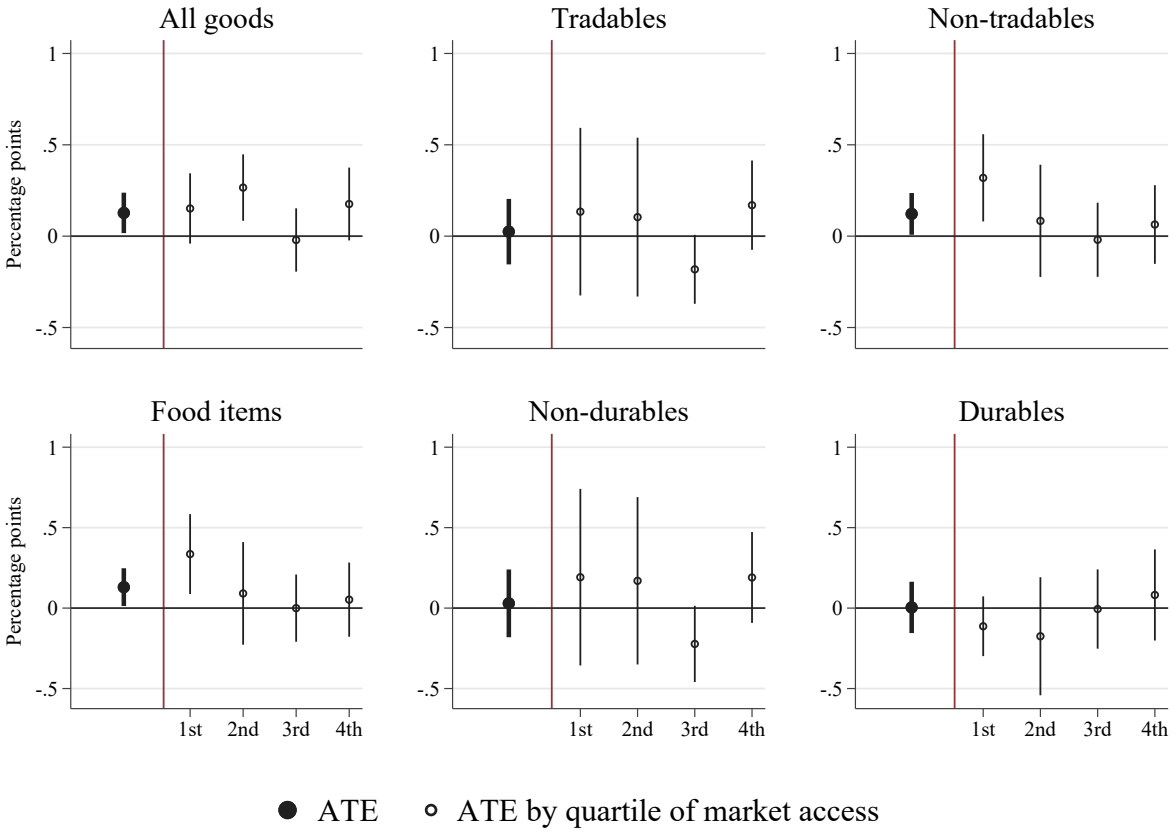


(b) Calendar timeline



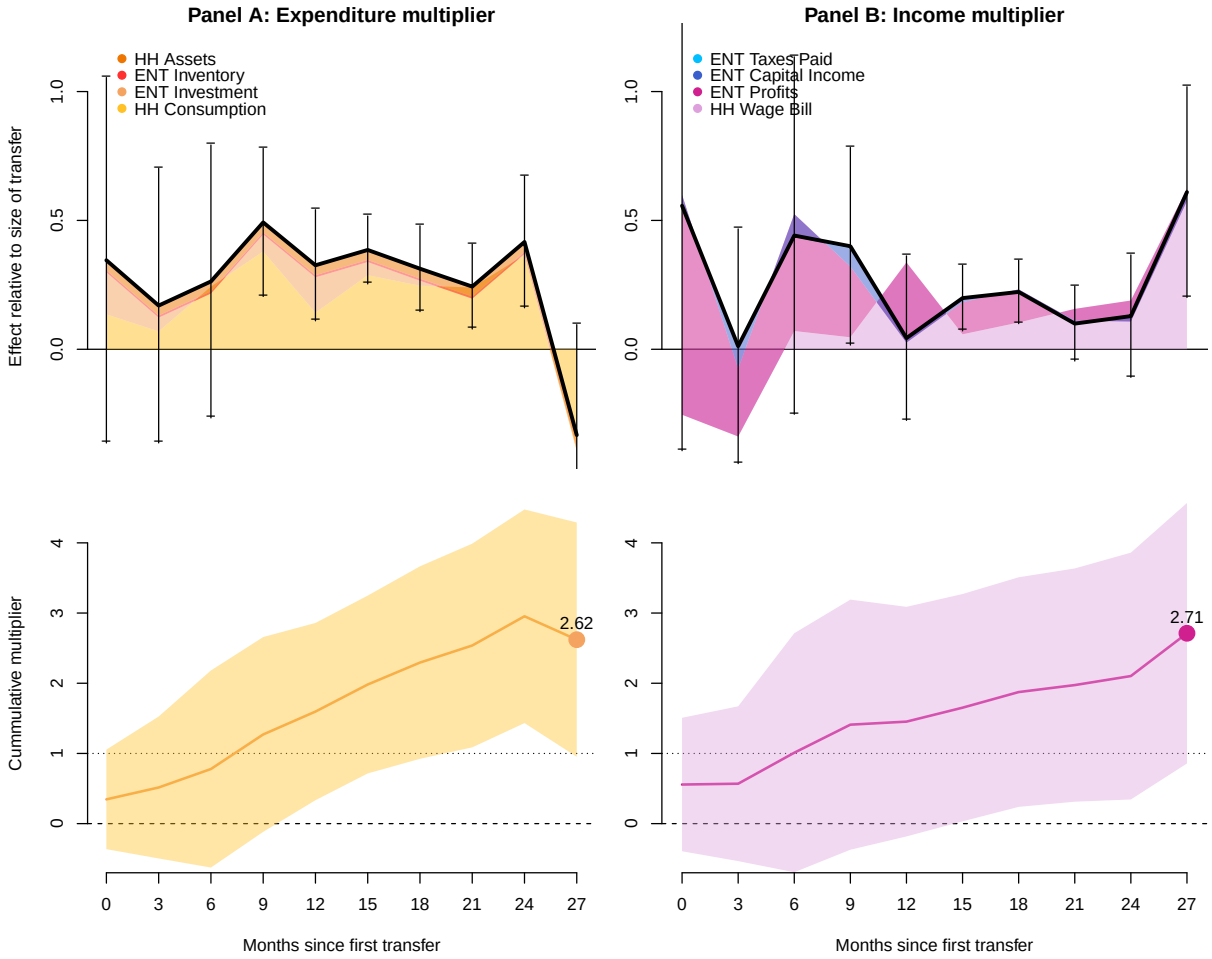
Notes: This figure plots the 5th, 25th, 50th, 75th and 95th percentiles of study activities. In Panel A, timing is reported relative to the anticipated start of activities in each village (the "experimental start"). The experimental start for a village is calculated based on the random ordering of treatment 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. In Panel B, the calendar timing of activities is repored.

Figure 4: Output prices by market access



Notes: Each panel represents a regression of the logarithm of a price index (multiplied by 100) on the "optimal" number of lags and distance buffers of per capita Give Directly transfers in each buffer. 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. 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 (in increasing order), 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. Bars represent 95% confidence intervals based on standard errors as in Conley (1999, 2008), where we allow for spatial correlation up to 10km and autocorrelation up to 12 months.

Figure 5: Fiscal multiplier over time



Notes: Both expenditure and fiscal multipliers broken down by components and over time. Panel A prints the expenditure multiplier throughout the first 27 months after start of the transfers. The upper graph reports the total multiplier impulse response function each quarter as a thick black line. It also prints components of this figure each quarter as colored areas below the curve. Note that some components turn negative from time to time, leading to some areas to overlap, which is indicated by darker shading of the components. Standard errors around the joint multiplier estimates are obtained by 2000 replications of the wild bootstrap and reported as brackets of $\pm$1SE. The integral under the multiplier curve adds up to our main multiplier estimate of 2.62. The bottom graph of Panel A traces out this integral over time. It also prints $\pm$1SE confidence intervals around the cumulative estimate at each quarter, obtained by 2000 bootstrap replications. Panel B repeats the same exercise for the income multiplier.

Table 1: Expenditures and Savings

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village)	Total Effect	Total Effect	Control, low saturation
	Reduced form	IV	IV	mean (SD)
Panel A: Expenditure				
Total expenditure, annualized	290.12***	347.56***	337.06***	2,548.48
	(57.36)	(111.64)	(122.21)	(1,940.21)
Total non-durable expenditure, annualized	170.10***	215.82**	291.80***	2,414.02
	(52.68)	(96.34)	(110.45)	(1,803.34)
Food expenditure	77.62**	139.54**	132.85**	1,594.24
	(35.02)	(66.75)	(58.58)	(1,079.05)
Temptation goods expenditure	7.53	4.37	−0.71	39.30
	(5.23)	(9.18)	(6.50)	(128.08)
Total durable expenditure, annualized	6.06***	4.92***	−0.45	6.33
	(0.97)	(1.59)	(1.03)	(20.01)
Panel B: Assets				
Total assets (non-land, non-house), net borrow	169.50***	161.69***	113.04	1,142.53
	(28.27)	(54.06)	(90.52)	(1,406.93)
Housing value	378.10***	377.52***	67.75	2,020.61
	(25.70)	(40.46)	(311.56)	(4,808.15)
Land value	−9.82	330.28	2,408.48	6,759.16
	(232.86)	(349.71)	(1,510.54)	(18,651.34)

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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2: Input Prices and Quantities

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Panel A: Labor				
Hourly wage earned by employees	0.09***	0.03	0.19*	0.66
	(0.03)	(0.05)	(0.10)	(0.82)
Household total hours worked, last 7 days	2.95*	1.40	−4.70	62.51
	(1.64)	(3.89)	(3.17)	(53.46)
Panel B: Land				
Land price per acre	202.10	437.25	611.35	3,905.67
	(204.41)	(296.04)	(416.00)	(3,132.63)
Acres of land owned	−0.14	−0.07	−0.11	1.51
	(0.09)	(0.11)	(0.12)	(3.35)
Panel C: Capital				
Loan-weighted interest rate, monthly	−0.01	0.01	0.00	0.06
	(0.01)	(0.01)	(0.01)	(0.07)
Total loan amount	3.83	4.68	2.39	57.39
	(4.10)	(7.52)	(12.84)	(156.79)

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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3: Household Balance Sheet

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Net value of household transfers received, annualized	−1.57	−11.46	8.75	136.59
	(6.56)	(13.78)	(19.11)	(281.16)
Total tax paid, annualized	1.94	0.55	2.18	16.45
	(1.22)	(2.13)	(2.01)	(35.33)
Total profits (ag & non-ag), annualized	31.77	38.83	33.14	492.58
	(24.23)	(54.21)	(49.41)	(813.06)
Total wage earnings, annualized	43.48	72.83	182.49***	456.87
	(31.55)	(64.40)	(68.44)	(1,147.02)
Total household income, annualized	81.18*	140.53	216.00**	990.03
	(42.75)	(100.85)	(94.65)	(1,586.70)

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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 4: 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	14.55	67.22	31.71	361.14
	(19.96)	(44.72)	(41.22)	(772.91)
Enterprise revenue, annualized	6.84	341.21**	224.87**	847.67
	(81.90)	(143.98)	(106.68)	(2,797.15)
Enterprise costs, annualized	−11.73	99.47***	83.66*	134.22
	(28.49)	(38.24)	(43.47)	(611.07)
Enterprise wagebill, annualized	−12.82	81.93**	69.69*	134.18
	(23.27)	(34.59)	(37.40)	(550.89)
Panel B: Non-agricultural enterprises				
Enterprise inventory	10.85	34.69**	17.06	217.01
	(8.10)	(14.46)	(10.48)	(566.01)
Enterprise investment, annualized	4.05	13.38	6.69	200.91
	(6.54)	(15.54)	(8.70)	(721.03)
Enterprise profit margin	0.01	−0.06*	−0.04*	0.44
	(0.01)	(0.03)	(0.02)	(0.61)
Panel C: Village-level				
Number of enterprises	0.01	0.00	0.00	1.12
	(0.01)	(0.02)	(0.02)	(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. 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 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 5: Output Prices

		(1)	(2)	(3)	(4)
		Overall Effects (in %)		**ATE by market access (in %)**	
		ATE	Average maximum effect (AME)	below median	above median
All goods		0.13**	0.68**	0.19**	0.10
		(0.06)	(0.28)	(0.09)	(0.07)
By tradability	Tradables	0.02	0.18	0.10	0.02
		(0.09)	(0.67)	(0.16)	(0.08)
	Non-tradables	0.12**	0.65**	0.20*	0.04
		(0.06)	(0.29)	(0.11)	(0.07)
By sector	Food items	0.13**	0.70**	0.21*	0.05
		(0.06)	(0.30)	(0.11)	(0.08)
	Non-durables	0.03	0.22	0.16	0.01
		(0.11)	(0.79)	(0.19)	(0.10)
	Durables	0.00	0.03	−0.12	0.06
		(0.08)	(0.59)	(0.11)	(0.09)
	Livestock	−0.09	−0.46	−0.08*	−0.17
		(0.10)	(0.47)	(0.04)	(0.19)
	Temptation goods	−0.22	−1.60	−0.14	−0.27
		(0.19)	(1.37)	(0.28)	(0.20)

Notes: Each row represents a regression of the logarithm of a price index (multiplied by 100) on the "optimal" number of lags and distance buffers of per capita Give Directly transfers in each buffer. 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 as the average across all markets of the maximum predicted effect (in absolute value) across all months for each market. 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 Conley (1999, 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 6: Fiscal Multiplier Estimates

	(1) CTM Estimate	(2) H_0: CTM < 0 p-value	(3) H_0: CTM < 1 p-value
Expenditure multiplier	**2.62**	**0.07***	**0.18**
	(1.67)		
Household consumption	1.71	0.14	0.33
	(1.55)		
Household assets	0.39	0.00***	1.00
	(0.04)		
Enterprise investment	0.44	0.16	0.89
	(0.43)		
Enterprise inventory	0.08	0.02**	1.00
	(0.04)		
Income multiplier	**2.71**	**0.07***	**0.18**
	(1.84)		
Household wage bill	0.99	0.22	0.49
	(1.24)		
Enterprise profits	1.61	0.10*	0.32
	(1.28)		
Enterprise capital income	0.07	0.35	1.00
	(0.18)		
Enterprise taxes paid	0.04	0.10	1.00
	(0.03)		
Joint test of both multipliers		**0.02****	**0.08***

Notes: Results from 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 (consumption, investment, wages, profits, capital income, and taxes) are obtained by dynamically estimating effect sizes over the first 27 months after the first transfer and computing the integral under this curve. Effects on remaining stock variables are approximated by the total endline treatment effect. Standard errors are computed by 2,000 replications of a wild clustered bootstrap technique, which re-allocates within-sublocation 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 estimate by counting the number of times the wild bootstrap produced an estimate complying with the null hypotheses and compute p-values as the fraction the null was satisfied. Finally, the last row reports p-values from joint tests against the same nulls, following an identical procedure. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

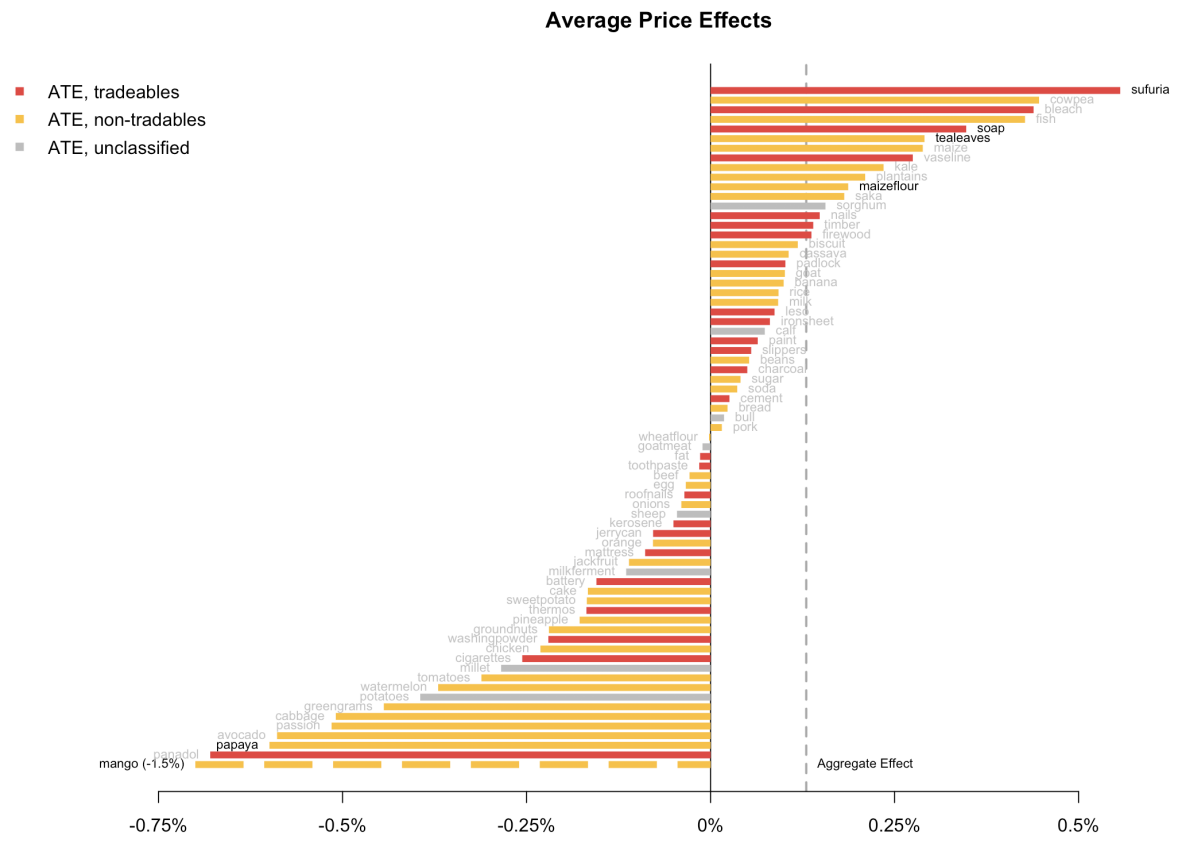
Table 7: Non-market Outcomes and Externalities

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Psychological well-being index	0.09***	0.13*	0.09	0.01
	(0.03)	(0.07)	(0.06)	(1.00)
Health index	0.03	0.06	0.01	0.01
	(0.03)	(0.06)	(0.05)	(1.02)
Food security index	0.09***	0.11	0.09	0.00
	(0.03)	(0.08)	(0.05)	(1.00)
Children food security	0.11***	0.16**	0.08	0.01
	(0.03)	(0.07)	(0.08)	(1.01)
Education index	0.08**	0.11*	0.10*	0.00
	(0.03)	(0.05)	(0.06)	(0.99)
Female empowerment index	−0.03	−0.13	0.09	0.01
	(0.07)	(0.15)	(0.15)	(0.98)
Security index	0.09***	0.02	−0.02	0.00
	(0.03)	(0.07)	(0.07)	(1.02)

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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

A Supporting Figures & Tables

Figure A.1: Price Effects by Product



Notes: Each row represents a regression of the logarithm of the median price of a good on the "optimal" number of lags and distance buffers of per capita Give Directly transfers in each buffer. The number of radii bands and lags is the same across goods, and chosen sequentially for the aggregate price index by minimizing the BIC, as pre-specified. Regressions include a full set of market and month fixed effects. Bars denote 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 as the average across all markets of the maximum predicted effect (in absolute value) across all months for each market. 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 parantheses) are as in Conley (1999, 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 A.1: Household assets by productivity status

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Total household assets (non-land, non-house)	163.73***	145.17**	94.42	1,189.33
	(32.10)	(61.91)	(82.81)	(1,337.25)
Potentially Productive Ag Assets	40.93**	9.20	−8.55	501.54
	(20.32)	(35.26)	(45.57)	(752.20)
Livestock Assets	37.01*	5.26	−8.83	466.48
	(19.79)	(33.82)	(43.69)	(728.46)
Ag Tools Assets	3.64***	3.29	0.39	32.88
	(0.99)	(2.26)	(2.59)	(39.53)
Potentially Productive Non-Ag Assets	43.42***	41.16*	31.52	217.20
	(10.98)	(24.66)	(26.01)	(420.99)
Non-Productive Assets	76.63***	94.60***	50.31	450.37
	(9.21)	(16.60)	(31.72)	(466.30)

Notes: This table presents results on household asset ownership based on classifications of assets by productivity status. Potentially productive agricultural assets include agricultural tools and livestock. Potentially productive non-agricultural assets include: 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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.2: Input prices and quantities: additional labor supply outcomes

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village)	Total Effect	Total Effect	Control, low saturation
	Reduced form	IV	IV	mean (SD)
Panel A: Quantities				
Household hours worked on own farm	1.61	0.05	−6.75***	35.35
	(1.15)	(2.69)	(2.59)	(38.57)
Individual hours worked in self-employment	1.96*	5.19**	−1.37	26.95
	(1.10)	(2.15)	(1.76)	(23.88)
Individual hours employed last week	0.84	−2.06	2.49	23.33
	(0.99)	(2.47)	(2.67)	(25.86)
Individual hours employed last week in agriculture	−1.36**	−2.02***	0.33	6.61
	(0.53)	(0.78)	(1.11)	(13.31)
Individual hours employed last week not in agriculture	1.81*	−0.24	1.91	16.24
	(1.02)	(2.46)	(2.65)	(26.25)
Panel B: Prices				
Hourly wage earned by employees	0.09***	0.03	0.19*	0.66
	(0.03)	(0.05)	(0.10)	(0.82)
Hourly wage earned by employees in agriculture	0.13**	0.23***	−0.06	0.67
	(0.06)	(0.09)	(0.13)	(0.68)
Hourly wage earned by employees not in agriculture	0.03	0.05	0.20	1.05
	(0.07)	(0.11)	(0.23)	(1.37)

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 and baseline values of the outcome when available. Column 3 presents the pooled average spillover effect on eligible households in control villages as well as ineligible households, coming from the "optimal" spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household, weighted by sampling weights of eligible/ineligible households. The reported average effect comes from evaluating the regression at the average amounts per capita GDP in each buffer among non-recipient households. The number of radii bands included in column 3 is chosen by minimizing the BIC, and we control for the share of eligible GE households in each radii band / a house hold's own village in both specifications. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct.

Table A.3: Input prices and quantities: additional land outcomes

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Panel A: Quantities				
Acres of land owned	−0.14	−0.07	−0.11	1.51
	(0.09)	(0.11)	(0.12)	(3.35)
Acres of land rented out	−0.04	−0.18	0.28	1.05
	(0.11)	(0.22)	(0.22)	(1.28)
Acres of land rented in	0.03	0.06	0.07	0.70
	(0.03)	(0.06)	(0.07)	(0.64)
Acres of land used for crops	0.02	−0.02	0.00	0.97
	(0.03)	(0.05)	(0.06)	(1.24)
Panel B: Prices				
Land price per acre	202.10	437.25	611.35	3,905.67
	(204.41)	(296.04)	(416.00)	(3,132.63)
Monthly land rental price per acre	0.01	−0.28	1.80	10.01
	(0.55)	(0.98)	(1.41)	(9.60)
Total ag land rental costs	5.46**	7.48	11.86	54.19
	(2.66)	(5.32)	(10.36)	(51.18)

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 and baseline values of the outcome when available. Column 3 presents the pooled average spillover effect on eligible households in control villages as well as ineligible households, coming from the "optimal" spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household, weighted by sampling weights of eligible/ineligible households. The reported average effect comes from evaluating the regression at the average amounts per capita GDP in each buffer among non-recipient households. The number of radii bands included in column 3 is chosen by minimizing the BIC, and we control for the share of eligible GE households in each radii band / a house hold's own village in both specifications. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct.

Table A.4: Enterprise revenue effects by sector

	(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)
Retail revenue	27.76	136.81*	77.29	313.72
	(48.67)	(83.06)	(55.17)	(520.73)
Manufacturing revenue	−15.09	70.99**	79.69**	68.73
	(18.81)	(30.54)	(34.16)	(127.68)
Services revenue	−2.89	20.56	10.80	90.66
	(21.45)	(37.74)	(29.10)	(160.11)
Agriculture revenue	21.13	45.75	19.16	346.49
	(14.38)	(30.42)	(23.26)	(416.41)

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 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. 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.5: Fiscal Multiplier Estimates: Quarter 4 and onwards

	(1) CTM Estimate	(2) H_0: CTM < 0 p-value	(3) H_0: CTM < 1 p-value
Expenditure multiplier	**1.98**	**0.01*****	**0.11**
	(0.81)		
Household consumption	1.27	0.05**	0.36
	(0.77)		
Household assets	0.39	0.00***	1.00
	(0.04)		
Enterprise investment	0.25	0.06*	1.00
	(0.16)		
Enterprise inventory	0.08	0.02**	1.00
	(0.04)		
Income multiplier	**1.70**	**0.01*****	**0.19**
	(0.77)		
Household wage bill	1.51	0.01***	0.21
	(0.61)		
Enterprise profits	0.04	0.46	0.99
	(0.40)		
Enterprise capital income	0.13	0.07*	1.00
	(0.09)		
Enterprise taxes paid	0.01	0.02**	1.00
	(0.01)		
Joint test of both multipliers		**0.00*****	**0.06***

Notes: Results from joint estimation of expenditure and income multipliers using data starting four quarters after the distribution of transfers. 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 (consumption, investment, wages, profits, capital income, and taxes) are obtained by dynamically estimating effect sizes over the first 27 months after the first transfer and computing the integral under this curve. Effects on remaining stock variables are approximated by the total endline treatment effect. Standard errors are computed by 2,000 replications of a wild clustered bootstrap technique, which re-allocates within-sublocation 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 estimate by counting the number of times the wild bootstrap produced an estimate complying with the null hypotheses and compute p-values as the fraction the null was satisfied. Finally, the last row reports p-values from joint tests against the same nulls, following an identical procedure. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.6: Psychological Well-Being and Health components

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Psychological well-being index	0.09***	0.13*	0.09	0.01
	(0.03)	(0.07)	(0.06)	(1.00)
Depression	−0.03	0.04	−0.04	−0.01
	(0.03)	(0.07)	(0.06)	(1.00)
Happiness	0.01	0.14**	0.05	0.00
	(0.03)	(0.07)	(0.06)	(1.00)
Life satisfaction	0.14***	0.07	0.04	0.02
	(0.03)	(0.08)	(0.06)	(1.00)
Perceived stress	−0.02	−0.04	−0.09	0.00
	(0.03)	(0.06)	(0.05)	(1.00)
Health Index	0.03	0.06	0.01	0.01
	(0.03)	(0.06)	(0.05)	(1.02)
Self-reported health	0.05	0.02	0.02	3.32
	(0.03)	(0.08)	(0.06)	(1.04)
Index of recent health symptoms	0.00	0.01	−0.04	−0.02
	(0.03)	(0.06)	(0.06)	(1.02)
Has experienced a major health problem since baseline	0.00	−0.03**	0.02	0.18
	(0.01)	(0.02)	(0.02)	(0.38)

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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.7: Education and Food Security components

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Education Index	0.08**	0.11*	0.10*	0.00
	(0.03)	(0.05)	(0.06)	(0.99)
Total education expenditure (annualized)	25.19**	47.09***	55.22	272.80
	(10.43)	(17.01)	(35.01)	(568.03)
Proportion of school aged children in school	0.01	0.00	0.02*	0.96
	(0.01)	(0.01)	(0.01)	(0.18)
Food security index	0.09***	0.11	0.09	0.00
	(0.03)	(0.08)	(0.05)	(1.00)
Children food security	0.11***	0.16**	0.08	0.01
	(0.03)	(0.07)	(0.08)	(1.01)
Number of days adults skipped meals in the last week	−0.11**	−0.06	−0.08	0.98
	(0.05)	(0.11)	(0.08)	(1.56)
Number of days adults went without food in the last week	−0.06**	−0.04	−0.04	0.20
	(0.02)	(0.05)	(0.05)	(0.72)
Number of days adults went to bed hungry in the last week	−0.09***	−0.14*	−0.08	0.37
	(0.03)	(0.08)	(0.05)	(0.86)
Number of days children skipped meals in the last week	−0.12**	−0.12	−0.12	0.62
	(0.05)	(0.11)	(0.08)	(1.30)
Number of days children went without food in the last week	−0.03**	−0.05*	−0.04	0.08
	(0.01)	(0.03)	(0.03)	(0.48)
Number of days children went to bed hungry in the last week	−0.07***	−0.13***	0.00	0.20
	(0.02)	(0.05)	(0.05)	(0.62)
Number of meals eaten yesterday with meat, fish, or eggs	0.01	0.01	0.03	0.44
	(0.02)	(0.04)	(0.03)	(0.60)

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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.8: Female empowerment index components

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	$\mathbb{1}$(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Female Empowerment Index	−0.03	−0.13	0.09	0.01
	(0.07)	(0.15)	(0.15)	(0.98)
Violence Index	0.00	0.08	−0.04	−0.01
	(0.08)	(0.15)	(0.17)	(0.99)
Frequency of physical violence (last 6 months)	0.01	1.02	0.21	0.96
	(0.68)	(1.52)	(0.30)	(6.21)
Frequency of emotional violence (last 6 months)	0.13	1.27	−0.01	1.29
	(0.58)	(1.24)	(1.10)	(5.74)
Frequency of sexual violence (last 6 months)	−0.10	−0.43	−0.49	0.74
	(0.19)	(0.40)	(0.35)	(3.83)
Attitudes Index	0.04	0.02	−0.09	0.00
	(0.07)	(0.15)	(0.12)	(0.99)
Male-oriented attitudes	−0.09	−0.24	0.12	1.64
	(0.09)	(0.18)	(0.35)	(1.27)
Justifiability of domestic violence	0.25*	0.44	0.18	2.29
	(0.13)	(0.28)	(0.26)	(1.95)

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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.9: Security index components

	(1)	(2)	(3)	(4)
	Treated Households		**Untreated Households**	
	𝟙(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation mean (SD)
Security index	0.09***	0.02	−0.02	0.00
	(0.03)	(0.07)	(0.07)	(1.02)
No. times victimized by theft, last 12 months	0.00	0.09**	0.02	0.29
	(0.02)	(0.04)	(0.05)	(0.84)
No. times victimized by assault/arson/witchcraft, last 12 months	−0.13**	−0.13	−0.05	0.34
	(0.06)	(0.11)	(0.06)	(1.41)
Has experienced a crime but did not report it, last 12 months	−0.03**	−0.02	−0.02	0.16
	(0.01)	(0.02)	(0.02)	(0.37)
Is worried about crime or safety in neighborhood	−0.04**	−0.01	0.05	0.40
	(0.02)	(0.03)	(0.03)	(0.49)

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. 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). For this analysis, the sample is restricted to eligible households. 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). 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 calculated following Conley (1999, 2008) using a uniform kernel out to 10 km. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.10: 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	0.6	0.5	0.1	32.5
	(0.7)	(1.2)	(0.6)	(8.0)
Counterfactual Gini coefficient (MPC = .4)	−0.7	0.4	0	32.3
	(0.7)	(1.3)		(7.8)
P-value: effect = counterfactual effect	p=0.16	p=0.95	p=0.89	
Panel B: Assets				
Gini coefficient	−1.1	2.0*	2.8**	45.4
	(0.9)	(1.2)	(1.4)	(10.1)
Counterfactual Gini coefficient (MPS = .6)	−7.8***	−5.6***	0	45.8
	(0.9)	(2.0)		(10.8)
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 and assets distribution by transferring its baseline distribution (for assets) or by imputing based on endline non-missing values in control and low-saturation villages (for expenditure). We add in the associated gain from the assumed MPC (0.4) and MPS (0.6), calculated 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. Standard errors are clustered at sublocation level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.11: Composition of enterprises by sector

	Overall		**Non-Ag**	
Sector	Count Share	Revenue Share	Count Share	Revenue Share
Retail	0.10	0.42	0.64	0.64
Manufacturing	0.03	0.12	0.17	0.19
Services	0.03	0.11	0.19	0.17
Agriculture	0.84	0.35		

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 (agricultural enterprises). Data on revenue shares for non-agricultural sector comes from endline enterprise surveys, while data on agricultural revenue shares comes from endline household surveys.

Table A.12: Enterprise outcomes by owner eligibility

	(1)	(2)	(3)	(4)
	Treated Owners		**Untreated 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	10.93	31.13	51.58**	361.14
	(13.44)	(21.30)	(23.96)	(772.91)
Enterprise revenue, annualized	72.09**	172.77*	210.87***	847.67
	(29.23)	(101.66)	(47.04)	(2,797.15)
Enterprise costs, annualized	24.65**	36.32**	39.12**	134.22
	(9.92)	(14.82)	(17.45)	(611.07)
Enterprise wagebill, annualized	25.29***	46.54***	45.07***	134.18
	(9.37)	(14.51)	(17.42)	(550.89)
Panel B: Non-agricultural enterprises				
Enterprise inventory	2.84	8.04	5.40	217.01
	(2.79)	(7.46)	(5.53)	(566.01)
Enterprise investment, annualized	−5.09	−17.40	5.41	200.91
	(5.59)	(17.44)	(8.27)	(721.03)
Enterprise profit margin	−0.04**	−0.09***	−0.02	0.44
	(0.02)	(0.03)	(0.03)	(0.61)
Panel C: Village-level				
Number of enterprises	0.00	0.03**	0.00	1.12
	(0.01)	(0.01)	(0.01)	(0.14)

B Details on study design, attrition and tracking

B.1 Cash transfer intervention

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; in practice, if the husband was at home, he was typically registered, while if the husband was out, the wife was registered. 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.[40] Households were instructed 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. Backcheck: All registered households were backchecked to confirm eligibility in advance of transfers being sent. Importantly, the census, registration, and backcheck 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.

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

Transfers were typically sent at a single time per month (usually around the 15th) to all households scheduled to receive transfers.

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.

Due to the large number of villages and households involved in the study, GD worked on a rolling basis across villages in the study area following the random order described above. GD generally began sending transfers to eligible households within a village once 50 percent of the eligible households (as identified in the census) had completed the enrollment process. A majority of households within a village all received transfers at the same time. Nevertheless, there are several reasons why transfer timing may vary somewhat within villages relative to this predetermined timing. First, all households may not have completed enrollment at the same time as the majority of households. Second, if households delayed in signing up for M-Pesa, this would also introduce delays in their transfers and differences across villages. Third, once transfers began, if households reported marital problems or other conflicts arising due to the transfers, transfers would be delayed while seeking a resolution.

C Tracking, Attrition and Balance

C.1 Household Data

C.1.1 Analysis of potential attrition bias

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 conduct the following analyses. Let r_{hvs} be an indicator for whether household h in village v in sublocation s is observed at endline. We investigate whether this indicator of non-attrition varies with treatment status by estimating:

$$r_{hvs} = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \varepsilon_{hvs} \tag{20}$$

Columns (1), Panel A of Tables C.13 and C.14 present results for eligible and ineligible households, respectively. 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 baseline (Column 2), or being reached twice (Column 3). It is also robust to restricting the sample to only households reached at endline (Panel B) or only households surveyed at baseline (Panel C). Looking at ineligible households in Table C.14, we find

that ineligible households in high-saturation sublocations are significantly less likely to be reached twice (Table C.14, Panel A, Column 3). This is mostly due to households reached at baseline not being surveyed again at endline (Panel C). This result is significant only at the 10 percent level and only applies to high-saturation sublocations.

C.1.2 Baseline balance

To test whether there are differences in baseline characteristics across treatment arms for respondents observed at endline, we estimate the following model:

$$(X_{hvs,t=0}|r_{hvs} = 1) = \beta_0 + \beta_1 T_{vs} + \beta_2 H_s + \varepsilon_{hvs} \tag{21}$$

T_{vs} is an indicator for households residing in a treated village. H_s is an indicator for living in a high-saturation sublocation, which we control for as it was part of the research design; however, our focus is on β_1. Table C.15 presents results. In Columns (2) and (4) we report baseline differences between eligible and ineligible households in treatment and control villages for a number of variables. We observe insignificant differences for most variables, with the exception of an indicator for scoring above the median scores on a psychological well-being index, which on average is higher in eventually treated villages at baseline, both among eligible and ineligible households. Ineligible households in treated villages also have a significantly higher food security index at baseline, are more likely to have children in the household, and are less likely to be employed (the latter two differences being significant at the 10 percent level).

Table C.13: Tracking and attrition: eligible households

	(1) Surveyed at endline	(2) Surveyed at baseline	(3) Surveyed at baseline and endline	(4) Initially sampled household	(5) Replacement household
Panel A: All households targeted at endline					
Treatment Village	0.006	0.010	0.006	0.005	−0.008
	(0.009)	(0.011)	(0.013)	(0.010)	(0.013)
High Saturation Sublocation	−0.003	−0.020	−0.019	0.006	−0.003
	(0.009)	(0.014)	(0.015)	(0.011)	(0.013)
Control, Low Sat Mean (SD)	0.903	0.879	0.814	0.853	0.168
	(0.297)	(0.326)	(0.389)	(0.355)	(0.374)
Observations	6,039	6,039	6,039	6,039	5,196
Panel B: Among households surveyed at endline					
Treatment Village		0.001	0.001	0.013	−0.015
		(0.011)	(0.011)	(0.010)	(0.013)
High Saturation Sublocation		−0.018	−0.018	0.005	−0.002
		(0.013)	(0.013)	(0.011)	(0.013)
Control, Low Sat Mean (SD)		0.902	0.902	0.845	0.172
		(0.298)	(0.298)	(0.362)	(0.377)
Observations		5,425	5,425	5,425	4,768
Panel C: Among households surveyed at baseline					
Treatment Village	−0.004		−0.004	0.008	−0.008
	(0.009)		(0.009)	(0.013)	(0.013)
High Saturation Sublocation	0.000		0.000	0.003	−0.003
	(0.009)		(0.009)	(0.013)	(0.013)
Control, Low Sat Mean (SD)	0.926		0.926	0.832	0.168
	(0.262)		(0.262)	(0.374)	(0.374)
Observations	5,197		5,197	5,197	5,196

Notes: This table reports tracking and attrition rates for households classified as eligible to receive GD transfers by GE project field staff. Panel A includes all eligible households that were targeted for endline sureys. Panel B looks at eligible households that completed endline surveys, and serves as our main analysis sample. Panel C looks at eligible households that completed baseline surveys, and provides information on households that attrited from baseline to endline. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table C.14: Tracking and attrition: ineligible households

	(1) Surveyed at endline	(2) Surveyed at baseline	(3) Surveyed at baseline and endline	(4) Initially sampled household	(5) Replacement household
Panel A: All households targeted at endline					
Treatment Village	0.011	0.008	0.017	0.006	−0.009
	(0.011)	(0.014)	(0.016)	(0.013)	(0.018)
High Saturation Sublocation	−0.016	−0.020	−0.036**	0.000	0.005
	(0.011)	(0.016)	(0.018)	(0.015)	(0.018)
Control, Low Sat Mean (SD)	0.908	0.869	0.814	0.835	0.190
	(0.289)	(0.337)	(0.390)	(0.372)	(0.393)
Observations	3,111	3,111	3,111	3,111	2,649
Panel B: Among households surveyed at endline					
Treatment Village		0.008	0.008	0.012	−0.016
		(0.015)	(0.015)	(0.014)	(0.018)
High Saturation Sublocation		−0.025	−0.025	−0.001	0.007
		(0.017)	(0.017)	(0.015)	(0.018)
Control, Low Sat Mean (SD)		0.896	0.896	0.833	0.187
		(0.306)	(0.306)	(0.374)	(0.390)
Observations		2,817	2,817	2,817	2,458
Panel C: Among households surveyed at baseline					
Treatment Village	0.012		0.012	0.009	−0.009
	(0.011)		(0.011)	(0.018)	(0.018)
High Saturation Sublocation	−0.021*		−0.021*	−0.004	0.005
	(0.011)		(0.011)	(0.018)	(0.018)
Control, Low Sat Mean (SD)	0.936		0.936	0.810	0.190
	(0.244)		(0.244)	(0.393)	(0.393)
Observations	2,651		2,651	2,651	2,649

Notes: This table reports tracking and attrition rates for households classified as ineligible to receive GD transfers by GE project field staff. Panel A includes all ineligible households that were targeted for endline sureys. Panel B looks at ineligible households that completed endline surveys, and serves as our main analysis sample. Panel C looks at ineligible households that completed baseline surveys, and provides information on households that attrited from baseline to endline. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table C.15: Household balance

	Eligibles			Ineligibles		
	(1) Control, Low Sat Mean (SD)	(2) Treatment Effect	(3) N	(4) Control, Low Sat Mean (SD)	(5) Treatment Effect	(6) N
Female	0.67	0.01	4,768	0.79	−0.02	2,458
	(0.47)	(0.02)		(0.41)	(0.02)	
Respondent aged 25 or older	0.83	0.01	4,755	0.97	−0.01	2,448
	(0.38)	(0.01)		(0.18)	(0.01)	
Is married	0.64	0.02	4,768	0.42	0.03	2,458
	(0.48)	(0.01)		(0.49)	(0.02)	
Completed primary school	0.41	0.01	4,768	0.29	0.02	2,458
	(0.49)	(0.02)		(0.45)	(0.02)	
Has child	0.80	0.01	4,768	0.68	0.04*	2,458
	(0.40)	(0.01)		(0.47)	(0.02)	
Self-employed	0.27	0.00	4,768	0.28	0.01	2,458
	(0.45)	(0.01)		(0.45)	(0.02)	
Employed in wage work	0.36	−0.03	4,768	0.21	−0.03*	2,458
	(0.48)	(0.02)		(0.41)	(0.02)	
Total non-land, non-home assets, net loans (z-scored) (USD)	−0.18	0.01	4,768	0.41	0.08	2,458
	(0.65)	(0.02)		(1.36)	(0.06)	
Total household income in the last 12 months (z-scored) (USD)	0.01	0.01	4,768	0.07	0.02	2,458
	(0.89)	(0.03)		(1.02)	(0.05)	
Total business revenue in the last 12 months (z-scored) (USD)	−0.08	0.03	4,768	0.13	0.05	2,458
	(0.75)	(0.03)		(1.20)	(0.06)	
Psychological wellbeing index	0.00	0.06*	4,765	0.00	0.08*	2,458
	(0.99)	(0.04)		(0.99)	(0.05)	
Food security index	−0.01	0.00	4,768	−0.01	0.09**	2,458
	(1.00)	(0.03)		(1.01)	(0.04)	

Notes: Differences in baseline outcomes between households in treatment and control villages. Column (2) shows the baseline difference between eligible households in treated and untreated villages, and Column (5) shows the same difference for ineligible households. Columns (1) and (4) show the respective control means and standard deviations. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

C.2 Enterprise Data

Our enterprise surveys are analyzed as repeated cross-sections, so we do not report attrition rates between baseline and endline for enterprises. We do check baseline balances for enterprises, using an enterprise-level version of Equation (21), and report results in Table C.16. We do not find statistically significant results for most outcomes, though enterprise costs are marginally significantly higher for enterprises in high saturation sublocations at baseline.

Table C.16: Enterprise Balance

	(1) 𝟙(Treat village) Reduced form	(2) 𝟙(High Saturation) Reduced form	(3) Control, low saturation weighted mean (SD)
Number of enterprises	0.00	0.01	1.10
	(0.01)	(0.01)	(0.14)
Number of enterprises, non-ag operated outside hh	0.00	0.01	0.08
	(0.01)	(0.01)	(0.14)
Number of enterprises, non-ag operated from hh	0.00	0.00	0.05
	(0.00)	(0.00)	(0.03)
Number of enterprises, own-farm agriculture	0.00	0.00	0.96
	(0.00)	(0.00)	(0.01)
Enterprise profits, annualized	−16.01	13.48	431.11
	(18.12)	(22.94)	(731.59)
Enterprise profit margin	0.02	−0.03	0.50
	(0.02)	(0.02)	(0.38)
Enterprise revenue, annualized	−89.78	120.05	868.44
	(90.68)	(96.14)	(4,171.18)
Enterprise costs, annualized	0.11	0.62*	4.61
	(0.35)	(0.35)	(13.16)
Enterprise wagebill, annualized	4.50	3.65	61.16
	(7.31)	(6.16)	(267.72)

Notes: Columns 1 and 2 report the coefficients on an indicator for treatment village and an indicator for saturation status of the sublocation from a regression of each outcome at baseline on these indicators. 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. Each regression is weighted by inverse sampling weights. For monetary values, 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. 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 level. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.