General equilibrium effects of cash transfers: experimental evidence from Kenya

Dennis Egger Johannes Haushofer Edward Miguel
Paul Niehaus Michael Walker*

December 18, 2019

Abstract

How large economic stimuli generate individual and aggregate responses is a central question in economics, but has not been studied experimentally. We provided one-time cash transfers of about USD 1000 to over 10,500 poor households across 653 randomized villages in rural Kenya. The implied fiscal shock was over 15 percent of local GDP. We find large impacts on consumption and assets for recipients. Importantly, we document large positive spillovers on non-recipient households and firms, and minimal price inflation. We estimate a local fiscal multiplier of 2.7. We interpret welfare implications through the lens of a simple household optimization framework.

*Egger: University of California, Berkeley; Haushofer: Princeton University, NBER, Busara Center for Behavioral Economics, and Max Planck Institute for Collective Goods; Miguel: University of California, Berkeley, NBER, and CEGA; Niehaus: University of California, San Diego, NBER, and CEGA; Walker: University of California, Berkeley, and CEGA. Niehaus is a co-founder, former president (2012 - 2017) and chairman of the board of *GiveDirectly*. Acknowledgements: 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, Anya Marchenko, Max Mueller, Priscilla de Oliviera, Robert On, Rachel Pizatella-Haswell, Emaan Siddique, Zenan Wang, Francis Wong and Kejian Zhao for excellent research assistance; Innovations for Poverty Action for fieldwork; the Busara Center for Behavioral Economics for piloting work; GiveDirectly for fruitful collaboration; and Vittorio Bassi, Abhijit Banerjee, Chris Blattman, Giacomo DiGiorgi, Esther Duflo, Yuriy Gorodnichenko, Seema Jayachandran, Adriana Lleras-Muney, Atif Mian, Emi Nakamura, Ben Olken, Jon Steinsson, and numerous conference and seminar participants at Barcelona GSB, Geneva, Harvard Business School, Harvard Kennedy School of Government, the Harvard/MIT Development seminar, Maastricht, Stanford, the World Bank, University of British Columbia, UC Berkeley, UC Santa Cruz, UC San Diego, University of Chicago, University of Southern California, University of Washington, Zurich, the 2019 BREAD/NBER meeting, the 2019 Korean Economic Review international conference, the 2019 American Economic Association annual meetings, and the 2018 Y-Rise 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). The study received IRB approval from Maseno University and the University of California, Berkeley. AEA Trial Registry RCT ID: AEARCTR-0000505.

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 a central task of economic analysis. For instance, there has been interest in understanding the aggregate impacts of fiscal stimulus and cash infusions for decades (Keynes 1936), and a growing body of empirical evidence from rich countries shows that fiscal multipliers 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, these issues have not been subjected to experimental examination.

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

The present study was prospectively designed to unite these two disparate literatures by studying the aggregate impacts of large cash stimulus programs experimentally. 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 over 10,500 poor households in a sample of 653 villages with a population of roughly 300,000. The implied fiscal shock was large, as the cash transfers amounted to over 15% of 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, we generated substantial spatial variation in the intensity of transfers by deliberately randomizing the allocation of cash transfers not just across households or villages (as is typical), but also across geographic sublocations (groups of 10–15 villages), thereby increasing power to detect aggregate impacts. Second, we carried out unusually extensive original data collection, giving us greater visibility into the chain of causal effects

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

linking cash transfers to aggregate outcomes in a complex and interconnected economy. Our household and enterprise censuses of the study area count 65,385 households and 12,095 non-farm enterprises. Within this unusually large sampling frame, we gathered detailed panel (longitudinal) data on household receipt of the transfer; household consumption expenditure patterns (representative for both treated and untreated households); local enterprise production, employment and revenue; labor market conditions; as well as especially high-frequency (monthly) and spatially disaggregated market data on prices. Third, we interpret the results through the lens of a theoretical framework 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 effects empirically using the rich dataset.

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

We then turn to data from detailed surveys of local enterprises. Enterprises in areas that receive more cash transfers also experience meaningful gains in total revenues, in line with the increases in household expenditure. Interestingly, sales increased without noticeable changes in firm investment behavior (beyond a modest increase in inventories), and sales do not increase differentially for firms owned by cash recipient households relative to non-recipients. Both 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 increase in overall profits. Methodologically, one important feature of the enterprise (and to a lesser extent household) results is that they are largely driven by the overall intensity of treatment in nearby communities, as opposed to the treatment status of the village in which the enterprise is located alone. This suggests that common study designs which aim to identify spillover effects by clustering treatment at the village level and assuming no spillovers across villages may be mis-specified, at least in densely populated areas such as the one we study.[2]

We next examine how these changes affect untreated households. Despite not receiving transfers, they too exhibit large consumption expenditure gains: their annualized consumption expenditure is higher by 13% eighteen months after transfers began, an increase roughly comparable to the gains contemporaneously experienced by the treated households. This result contrasts with earlier suggestive evidence of negative consumption and psychological spillovers from cash transfers within villages (Haushofer, Reisinger, and Shapiro 2019). Increased spending is not financed by dissaving, but more likely results in part from the income

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

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

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

We next ask what these effects imply for the aggregate level of economic activity. 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, as well as investment by local firms. Using this expenditure-based approach, we estimate a local fiscal multiplier of 2.5. A dual income-based approach, relying on distinct and complementary measures of labor and capital income, enterprise profits, and taxes, yields a similar estimate of 2.8. These estimates are somewhat larger than (though not statistically distinguishable from) recent multiplier estimates derived from US policy variation, which often range from 1.5-2.0 (Chodorow-Reich 2019). They are also somewhat larger than those from a structural simulation exercise, which predicted that local income multipliers from cash transfers in rural Kenya could range from 1.6 to 1.9 (Thome et al. 2013), and are similar in magnitude to non-experimental estimates from a cash transfer program in Mexico, which range from 1.5 to 2.6 (Sadoulet, Janvry, and Davis 2001).

Thus, a further contribution of this study is to exploit a randomized experiment to estimate an important macroeconomic quantity (see Muralidharan and Niehaus (2017) for a related discussion). A methodological strength of our approach is the fact that transfers came from donors outside the study area, rather than being internally debt-financed; the latter is typically the case in the US programs studied, and may complicate the interpretation of consumption responses due to Ricardian equivalence issues. One weakness, on the other hand, is that we observe limited data in the months immediately following the transfers in some communities, which limits the precision of our estimates.

In section 6, we interpret the welfare implications of our results using a simple theoretical framework. It describes an optimizing household that experiences direct benefits from transfers it receives, and also indirect welfare changes to the extent that the influx of purchasing power into the economy induces changes either in its 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.). We also characterize, under some additional assumptions, the wedge between welfare as defined by equivalent variation and as measured by expenditure, a commonly-used proxy for well-being in the development literature.[3]

Interpreted through this lens, the results generally suggest that non-recipients were made better off by an expansion in their budget sets: their incomes and consumption increase without a corresponding increase in total labor supply or local consumer goods prices, and without evidence of dissaving. Turning to externalities, we find either positive or null effects, but no negative effects, on domestic violence; child education, nutrition and health; crime; and public goods, as well as measures of subjective well-being. This result suggests that concerns about cash transfer programs doing harm to non-recipients are not borne out in our setting. One possible exception, to the extent households have preferences over their *relative* socioeconomic standing, is inequality: we estimate that positive spillovers were so large that village-level inequality might have slightly increased on some measures, despite the fact that transfers were initially targeted to the relatively poor.

The constellation of empirical findings raises an intriguing question about how the economy absorbed such a large shock to aggregate demand. Real output increased, and yet there is at most limited evidence of increases in the employment of land (which is in fixed supply), labor, or capital. One plausible, albeit speculative, possibility is that the *utilization* of these factors was "slack" in at least some enterprises (Lewis 1954). This seems plausible because in the retail and manufacturing sectors, where output responses were concentrated, the typical firm has a single employee (i.e., the proprietor), suggesting that integer constraints may often bind. In addition, many enterprises operate "on demand" in the sense that they produce only when they have customers, and the average non-agricultural enterprise sees just 1.7 customers per hour. In addition to retail, much manufacturing in this setting is on demand; for example, a mill owner waits for customers to bring grain and then grinds it for them. The existence of slack may help account for 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).

3. Specifically, we follow the sufficient statistics approach of Chetty (2009) and others.

The remainder of the paper is structured as follows. Section 1 describes the study design, including the cash transfer intervention. Section 2 describes data collection and estimation methods for households, enterprises and prices. Section 3 presents results tracing out the path of the cash through the economy. Section 4 describes the fiscal multiplier methodology and results. Section 5 presents results for key non-market outcomes (such as subjective well-being and health) and externalities, including crime and public goods. We lay out a household optimization framework for understanding welfare effects in Section 6. Section 7 discusses firm production and how we can make sense of the supply-side of our results. The final section concludes.

1 Study design

1.1 Setting: rural western Kenya

The study took place within three continguous subcounties of Siaya County, a rural area in western Kenya bordering Lake Victoria. The population in Siaya is predominantly Luo, the second largest ethnic group in Kenya, and while rural is also relatively densely populated, with 395 people per km^2 compared to a Kenyan average of 91. The main national road running from the port of Mombasa to Nairobi and on to Kampala, Uganda passes through the study area, likely helping to integrate it into the national, regional and global economy.

The NGO GiveDirectly (GD) selected the study area based on its high poverty levels. Within this area, GD selected for expansion villages that were rural (i.e., not located in peri-urban areas) that it had not previously worked in.[4] This yielded a final sample of 653 villages spread across 84 sublocations (the administrative unit above a village). The mean village consists of 100 households, and the average household had 4.3 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.

Transfers and data collection took place from mid-2014 to early 2017, a period of relative economic prosperity and political stability in Kenya. The World Bank reports annual per capita GDP growth rates ranged between 2.4 to 3.4 percent. All data collection concluded months prior to the August 2017 national election.

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

1.2 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, households had to live in homes with thatched roofs, a simple means-test for poverty. In treatment villages, GD enrolled all households that met this 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. (Registering for M-Pesa was a prerequisite for receiving transfers; households without a mobile phone were given the option to purchase one from GD staff with the cost deducted from their transfer.) Households 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 17 percent of annual GDP (based on our data described below) in the treated areas during the peak 12 months of disbursements, and to 24 percent of annual GDP during the full 24 month rollout period. Although small in relation to overall Kenyan GDP in 2015 (<0.1%), locally this is a larger relative shock than most government transfer programs, e.g., the ARRA programs studied in the recent US stimulus bill fiscal multiplier literature, see Chodorow-Reich (2019).

Transfers were made in a series of three payments as follows: a token transfer of KES 7,000 (USD 151 PPP) was sent once a majority of eligible households within the village had completed the enrollment process, followed two months later by the first large installment of KES 40,000 (USD 860 PPP). Six months later (and eight months after the token transfer), the second and final large installment of KES 40,000 was sent.The transfers were non-recurring, i.e., no additional financial assistance was provided to recipient households after their final large transfer, and they were informed of this up front. Households in control villages did not receive transfers.

1.3 Experimental design

To identify spillovers both within and across villages, we employed a two-level randomization design; Figure 1, Panel A illustrates the design. First, we randomly assigned 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. Within high (low) saturation groups, we then randomly assigned two-thirds (one-thirds) of villages to treatment status. Within treatment villages, all eligible households received a cash transfer.[5] This design induces variation in treatment intensity across space due both to (i) the variation in sublocation treatment intensity, and (ii) random variation in the location of treated villages within sublocations. Figure 2 illustrates: there is considerable variation across, but also within, sublocations in terms of the number of treated villages neighboring any given village.[6]

The most appropriate method for estimating effects depends on the nature of spillovers (Baird et al. 2018). Comparing high versus low saturation sublocations is most appropriate if most spillovers occur within the administrative boundaries of sublocations, for instance, if there was re-allocation of effort across villages by public officials within the sublocation, or if households rarely interact with individuals outside their sublocation. On the other hand, a geographic model of spillovers makes most sense if distance is the most salient feature for economic and social interactions, regardless of underlying administrative boundaries. Given the fact that the sublocation boundaries are not "hard" in any sense in Kenya, do not reflect salient ethnic or social divides here, 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, we generally prefer estimating spatial spillovers based on the amount of cash distributed within various radii bands around one's village. The proximity of treatment and control villages also highlights the plausibility of cross-village spillovers, even across sublocation boundaries.

2 Data and empirical specifications

We conducted four types of surveys: household surveys, enterprise surveys, market price surveys, and surveys of local public goods provision. Results from the public goods surveys are presented primarily in a separate paper (Walker 2018), and discussed briefly here. We filed several pre-analysis plans covering the data and analyses collected as part of this project. This paper focuses on primary outcomes for households, enterprises and prices, and we note where we go beyond these plans; for more details, see Appendix G.

2.1 Household data

We first conducted a baseline household census in all villages, which serves as a sampling frame and classifies household eligibility status. The census was designed to mimic GD's censusing procedure but was conducted by independent (non-GD) enumerators across both treatment and control villages for consistency. Throughout this paper, we base our analysis

5. The randomization was conducted in two batches based on GD's expansion plans (Appendix B.2).

6. Figure B.1 provides a higher-resolution example for two villages.

on village membership, household definitions and eligibility as classified by our project data collection field staff. In all, the census identified 65,385 households with a total baseline population of 280,000 people in study villages.

We conducted baseline household surveys within one to two months after the census and before the distribution of any transfers to a village (Figure 1, Panel B).[7] We used census information to sample at random eight eligible and four ineligible households per village to target for surveys. When households contain a married or cohabiting couple, we randomly selected one of the partners as the target survey respondent. Due to time and budget constraints, we sought to complete all baseline household surveys in a single day. If a household on our sampling list was not available on that day, we instead surveyed a randomly-selected replacement household with the same eligibility status.

We conducted a total of 7,848 baseline household surveys between September 2014 and August 2015.[8] The survey contained detailed modules on economic activities, asset ownership, psychological well-being, health and nutrition. A large array of baseline characteristics appear 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.2, column 2, in Appendix C.2).

Endline household survey data was collected between 9 and 31 months after each household's "experimental start date," meaning the month in which transfers were expected to start in a village assigned to treatment, regardless of their actual treatment timing.[9] The 5th/95th percentiles of timing ranged from 12 to 27 months, and the median survey was conducted 19 months after the experimental start month, or about 11 months after the distribution of the last lump sum transfer (Figure 1, Panel B).

Endline household surveys targeted all households on our initial sampling lists (including those that were missed at baseline), along with replacement households that were surveyed at baseline. For households that had been surveyed at baseline, we attempted to survey the individual who was the baseline respondent. We conducted a total of 8,242 endline household surveys between May 2016 and June 2017.[10] We achieved high tracking rates at endline,

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

8. Of this total, 6,510 households were on the initial sampling list, and 1,338 were randomly-selected replacement households.

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

10. This includes 7,019 initially sampled and 1,223 replacement households. Of the initially sampled households, 1,015 had been missed at baseline. Our main analysis focuses on the "initially sampled" (which includes those missed at baseline) and "replacement" households; results are similar using only originally sampled

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 (Table C.1). 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; more information on tracking and attrition is in Appendix C.2.

In addition to the modules collected at baseline, endline surveys also collected more detailed data on household expenditures and crop production, additional psychological scales (in particular, related to stress and hope), and female respondents surveyed by a female enumerator were also administered a module on female empowerment and gender-based violence; we describe these measures in more detail when discussing results.

2.2 Empirics: recipient households

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

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

where y_{ivs} is an outcome of interest for household i in village v in sublocation s.[11] $Treat_v$ is an indicator for residing in a treatment village at baseline, and $HighSat_s$ an indicator for being in a high-saturation sublocation. Here α_1 captures the total average treatment effect for households in treatment versus control villages, including both the direct effect of treatment (for eligible households) and any within-village spillovers; note that our design does not allow us to identify these separately. α_2 is a relatively coarse way to assess cross-village spillovers, as it does not utilize all experimental variation. We include the baseline value of the outcome variable ($y_{ivs,t=0}$), when available, to improve statistical precision (McKenzie 2012).[12] We cluster standard errors at the village level, and weight observations by inverse sampling probabilities to be representative of the population of eligible households.

Overall, we view Equation 1 as a useful benchmark but unlikely to capture well the spatial variation in treatment intensity evident in Figure 2. To better capture spillovers, we estimate models in which an (eligible) household's outcomes depend on the amount of

households (available upon request).

11. When we examine individual-level outcomes using Equation (1), we define treatment status and eligibility on the basis of the household in which the individual lives.

12. 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.

money distributed in its own village and in other geographically proximate villages:

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

The novel terms here are the amount Amt_v of cash transferred to one's own village v over the entire study, and the amount $Amt_{\neg v,r}$ of cash transferred to villages other than v in a series of bands with inner radius r km and outer radius $r+2$ km. We normalize both to be measured as a share of per-capita GDP.[13] The Amt variables depend on both the random assignment of villages to treatment and also on the endogenous share of households in those villages eligible for transfers, so we instrument for them using the own-village treatment indicator $Treat_v$ and the share $s^{e,t}_{\neg v,r}$ of eligible households in each band assigned to treatment. To account for spatial correlation, we calculate standard errors using a uniform kernel up to 10 km (Conley 1999; 2008).[14]

Because we had no a priori knowledge of the relevant distances over which general equilibrium effects might operate, we pre-specified an approach in which we first estimate a series of nested models varying the outer limit R of the spatial bands from 2 km to 20 km in steps of 2 km, and then select the one which minimized the Schwartz Bayesian Information Criterion (BIC). We report estimates of Equation (2) using the selected outer limit $\bar{R}$. As it turns out, this algorithm selects only the innermost 0–2 km band for almost all outcomes.[15]

Equation (2) correctly identifies the overall effects of the intervention if treatment intensity within bands less than $\bar{R}$ distance is a sufficient statistic for household i's outcomes. In particular, this implies that transfers delivered outside the radius $\bar{R}$ have no effect on i. If this assumption is violated – if, for example, all households were affected to some extent by all transfers in the study area – then the effects we estimate are *relative* to these "ambient" effects. The fact that our design generates substantial variation in treatment intensity, including some areas with minimal local transfers, helps to partially address this concern: transfers in the 2 km buffer around each village range from 0 to 26% of GDP, with a 10-90 percentile range of [4%, 15%]. Even in the 4 to 6 km buffer, the 10-90 percentile range remains wide, at [3%, 10%]. Together with the fact that our algorithm suggests spillovers are highly localized (we select only the 2 km buffer in the vast majority of cases), this sug-

13. We use an expenditure-based measure of GDP that is described in Section 4, which we convert to a per-capita measure based on household census data from our study area, and augmented with 2009 Kenya National Population Census data when necessary. Per capita GDP in low saturation control villages is 641 USD PPP (or 2744 USD PPP per household); see Appendix B.3 for details.

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

15. Results are robust to adjusting inference to account for this model selection step, as described below.

gests that a subset of our villages can reasonably serve as "pure controls." We also typically estimate spillovers of the same 'sign' as the direct effects, which suggests that any remaining bias in our estimates likely leads us to understate, rather than overstate, overall effects. All told, we view the problem of estimating spatial effects as unlikely to admit a perfect solution, but believe that our study design and econometric specification allow us to advance meaningfully relative to existing work.

An overall goal is to estimate the total effect of treatment, that is, how each unit's outcomes differ from what they would have been had there been no intervention. As a benchmark, we first report estimates of α_1 from Equation (1), which is the total treatment effect if all spillovers are contained within villages (a common identifying assumption). Next, we calculate the average total effects on eligible households implied by Equation (2), multiplying the estimated coefficients by the average values of the regressors. Concretely, we calculate the average total effect on the treated as $\hat{\beta} \cdot \big(\overline{Amt}_v | i \text{ is a treated eligible}\big) + \sum_{r=0}^{R} \hat{\beta}_r \cdot \big(\overline{Amt}_{\neg v,r} | i \text{ is a treated eligible}\big)$ for all radii bands up to the selected $\bar{R}$. This second effect allows for across-village spillovers in addition to direct effects and within-village spillovers.[16]

2.3 Empirics: non-recipient households

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

$$y_{iv} = \alpha + \sum_{r=0}^{R} \beta_r^1 Amt_r + \sum_{r=0}^{R} \beta_r^2 (Amt_r \cdot Elig_{iv}) + \gamma Elig_{iv} + y_{iv,t=0} \cdot \delta + \varepsilon_{iv}. \tag{3}$$

This specification modifies Equation (2) as follows. 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). As above, we instrument for Amt_r using the share of eligible households in the corresponding band assigned to treatment. When available, we include the baseline value of the outcome variable interacted with eligibility status. We then report the

16. We also consider the possibility that effects are non-linear in the per-capita amounts transferred. Specifically, Figure A.1 presents non-linear estimates of equation 2 for total consumption and firm revenue, key summary measures of local economic activity. We cannot formally reject linearity of these outcomes with respect to treatment intensity, and the relationships are visually roughly linear. We conduct the same test for our 10 pre-specified primary outcomes and eligible / ineligible households separately, and cannot reject linearity at the 10% significance level for any of them.

average total effect on the untreated as a population-weighted average of effects for eligible and ineligible households.[17]

2.4 Enterprise data

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

Specifically, we conducted censuses and surveyed a representative subset of all non-farm enterprises at both baseline and endline. We conducted the baseline enterprise census on the same day as the household census. The baseline household census included a question on whether the household was running an enterprise from their homestead or from a fixed kiosk/shop. The enterprise census targeted enterprises operating outside of homesteads. We then returned to survey enterprises operating outside of the homestead and open on the day of our visit, coincident with baseline household surveys. In villages with over 20 enterprises operating outside of homesteads, e.g., those that overlapped a market center, we randomly selected 20 enterprises to survey. Enterprise surveys cover profits, revenues, and a subset of costs (including the wage bill), as well as inventories and investment.

We conducted the endline enterprise census between November 2016 and April 2017. Enumerators sought to re-census all enterprises identified in the baseline census, as well as any newly established enterprises. This updated census served as the endline sampling frame. In each village we then randomly sampled to survey up to 2 enterprises operating from within homesteads and up to 3 outside of homesteads, including those in market centers in villages containing a market. Surveys again covered revenue, profits, employees, wages, some other costs, and taxes paid. Our main endline sample includes 1,699 enterprises operated from within and 1,442 from outside the homestead (both from enterprise surveys), as well as 7,899 agricultural enterprises from the household survey,

This integrated approach to household and enterprise surveying helps us match firms to

17. We calculate this as $s^{e,c}\left(\sum_{r=0}^{R}\left(\hat{\beta}_r^1+\hat{\beta}_r^2\right)*\left(\overline{Amt}_r | i \text{ is an untreated eligible}\right)\right)$ + $s^{i}\left(\sum_{r=0}^{R}\hat{\beta}_r^1*\left(\overline{Amt}_r | i \text{ is ineligible}\right)\right)$, where $s^{e,c}=1-s^i$ 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 (3).

their owners; see Appendix D for more details on the enterprise matching approach. We match all agricultural enterprises we located (as we find these via household surveys), and 56% of non-agricultural enterprises, for a total of 93% of all enterprises. Based on this match we estimate that enterprise activity is quite localized, with 92% of total profits and 87% of revenues accruing to owners who live within the village in which the enterprise operates. Patterns with respect to the eligibility status of the owner are generally sensible: 28% of non-agricultural enterprises are owned by an eligible household, slightly below their share in the population (33%), and enterprises owned by ineligible (and thus on average somewhat richer) households have 9% higher profits and 21% higher revenues on average than those owned by eligibles.

2.5 Empirics: enterprises

To estimate impacts on enterprises, we combine data from 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). We typically pool data across all enterprise types, except when we do not observe some outcomes for agricultural enterprises.

We estimate enterprise-level effects using versions of Equations (1) and (2), with radii bands selected as above, interacting right hand side variables with enterprise type (Appendix D.1 lists the full specifications). We include village-level means rather than enterprise-level values of the lagged dependent variable given that the enterprise surveys were repeated cross-sections. We carry out estimation using inverse probability weighting, accounting for enterprise type, except in some cases where we also revenue-weight outcomes. As above, we calculate and report average total effects, weighting effects for the three enterprise types. To facilitate comparisons between the household and enterprise results, we also normalize effects as per-household rather than per-enterprise.[18] To examine extensive margin effects, we estimate village-level analogues to this approach with the total number of enterprises censused (per household) as the dependent variable.

2.6 Price data

We measure consumer goods prices using monthly surveys of commodity prices in local markets. These surveys were conducted over the course of 2-2.5 years in all 61 markets in the

18. Specifically, we calculate $\frac{1}{n_{hh}} \sum_g \widehat{\Delta y}_e^g * n_{ent}^g$, where n_{hh} is the total number of households in the study area (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 in the area.

study area (and neighboring towns) with at least a weekly market day. These include market centers located in towns, and so will appropriately reflect the impacts of households (potentially) traveling to towns to spend their transfers. Figure 2 shows the substantial variation in treatment intensity around markets, as well as the heterogeneity in village proximity to markets. The average village had 0.7 markets located within 2 km and 2.3 markets within 4 km, again indicating the rather high density of residential settlement. Households report an average commuting time to their preferred market of 33 minutes, where more than 75 percent of respondents walk to the market.

Market surveys collected prices for 72 relatively homogeneous products, including food (grains, vegetables, fruit, meat), livestock (goats, sheep), hardware (nails, paint), "duka" (kiosk store) products (non-food and packaged food), and others (e.g., fuel, health items, household items, and farming implements). We collected quotes from three vendors of each product in each market in each month, and use the median for each product-market-month. We then calculate linear log-price indices by weighting prices by household expenditure shares.[19] We also examine effects on component subcategories of goods, which include: food items; non-food non-durables (such as soap, cooking fat, and firewood); durables (such as iron sheets, mosquito nets, and jerry cans used for transporting water or fuel); livestock; and temptation goods, i.e., alcohol and tobacco.

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

2.7 Empirics: prices

We estimate effects on consumer goods prices using both spatial and temporal variation in the amount of cash distributed around each market. Specifically, we estimate

$$p_{mt} = \sum_{r \in \bar{R}} \sum_{l=0}^{M} \beta_{rl} \tilde{Amt}_{m(t-l)r} + \alpha_m + \lambda_t + \varepsilon_{mt} \tag{4}$$

where p_{mt} is a price outcome for market m in month t. $\tilde{Amt}_{m(t-l)r}$ is the per-household amount transferred within band r around market m in month $t-l$, expressed as a fraction

19. We use expenditure data from the the Kenya Life Panel Survey (Baird et al. 2016) conducted in 2013-2014 in rural areas of Siaya and neighboring Busia county. We use the KLPS data because we did not collect a full expenditure module at baseline due to time and budget constraints, and prefer not to use endline expenditure data which are potentially endogenous. That said, results are nearly unchanged if we use endline consumption expenditure shares from untreated households.

of GDP. We exploit our panel setup by conditioning on fixed effects for both markets (α_m) and months (λ_t). The former absorb any systematic price differences across markets (as well as differences in the share of eligible households located around those markets), while the latter account for seasonal differences and other time trends common to all markets. We again account for spatial correlation in calculating standard errors (Conley 1999; 2008).

We determine both the relevant spatial distance R and the relevant temporal lag M over which price effects persist by minimizing an information criterion, similar to above. Specifically, and following a pre-specified approach (Haushofer et al. (2016)), we first select the optimal number of spatial buffer rings by estimating models of the form

$$p_{mt} = \sum_{r=0-2km}^{18-20km} \beta_r \left(\tilde{Amt}_{mtr} + \tilde{Amt}_{m(t-1)r} + \tilde{Amt}_{m(t-2)r} \right) + \alpha_m + \lambda_t + \varepsilon_{mt} \tag{5}$$

Note that in this specification we impose the assumption that transfers over the preceding quarter are what determine prices in month t. We select the value $R = \bar{R}$ that minimizes the Schwartz BIC while imposing weak monotonicity. We then select the number of monthly lags by estimating Equation (16) with $R = \bar{R}$ and selecting the model that minimizes the Schwartz BIC. With few exceptions, this procedure selects a single 0-2 km band around each market and a single temporal effect, implying that we only include contemporaneous transfers in estimating price effects.

Identification in Equation (16) comes from the roll-out of treatment across space *and* time, and the project's research design leads to substantial variation in both dimensions. The 10-90 percentile range for total amounts in annual per capita GDP going to the 0–2 km ring around each market over the entire period is [2%, 13%]. Moreover, the multi-year nature of the data for each market cover periods both of intensive transfer distribution as well as times when no transfers were going out. As above, we are unable to capture price increases that radiate throughout the entire study area (compared to neighboring counties) over the entire period, but the highly localized nature of the price effects that we do detect suggests that any such effects are unlikely to be large.

We use estimates of Equation (16) to calculate two price effects. The implied *average treatment effect* (ATE) is the average price effect across all markets and all months in which any transfers went out to any market in the study area, i.e., during the study period of September 2014 to March 2017. This is simply equal to the estimated coefficients multiplied by the mean of the corresponding regressors of interest. The *average maximum transfer effect* is the average across markets of the estimated effect in the month in which the maximum amount of cash (as a share of GDP) was distributed into the selected radii bands (in other words, out to $\bar{R}$) from the market.

We estimate effects on input prices using Equations (1), (2), and (3), as our input price data come from household surveys, and report the corresponding average treatment effects.

We focus on two sources of heterogeneous price effects. First, we classify goods into tradables and non-tradables, where tradables include relatively easily transported, non-perishable items, and non-tradables include more difficult to transport or perishable items.[20] Second, we classify markets into those with better or worse *market access*. Standard theory in international trade predicts that more integrated markets should be less likely to experience meaningful price changes 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):

$$MA_m = \sum_d \tau_{md}^{-\theta} N_d \approx \sum_{r=0}^{10} r^{-\theta} N_r \tag{6}$$

Geographic distance r is used to proxy for trade costs between origin market m and destination d, i.e., $\tau_{md} = r$. Destinations are 1 km radii bands around each market, with total population N_r in each buffer, and we follow Donaldson and Hornbeck (2016) in setting $\theta = 8$. Within quantiles of this metric, we calculate average and average maximum transfer treatment effects in the manner described above.

3 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. Monetary units are USD PPP unless otherwise defined (where the transfer was worth USD 1,871 PPP), flow outcomes are annualized, and monetary outcomes are top-coded at the 99th percentile (as pre-specified), unless otherwise noted.[21]

3.1 Recipient household effects

We first present impacts on the spending, saving and economic activity of recipients themselves. The main expenditure measure is the (annualized) sum of total food consumption in last 7 days, frequent purchases in the last month, and infrequent purchases over the last

20. For instance, tradables include building materials (e.g., timber, cement, nails, iron sheets) and some household goods (soap, firewood, charcoal, batteries, washing powder, mattress, soap), while non-tradables include 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 best classify specific items. The full classification (as pre-specified) is included in Appendix E.

21. Our main measures were pre-specified, though the groupings in some cases vary from the pre-analysis plan in order to ease exposition.

12 months.[22] Durables expenditures are measured as the sum of spending on home maintenance, home improvement, and other household durables, and the remainder is classified as non-durable spending.

As expected, recipient households report significantly higher total expenditure: USD PPP 293 more expenditure than eligible households in control villages (Table 1, column 1), an 11.5% increase over the control village in low saturation area mean of USD PPP 2,537. The estimated total treatment effect, including spatial effects, is larger at USD PPP 343, a 13.5% increase (column 2). This pattern between columns 1 and 2 is a first piece of evidence for the existence of localized, positive cross-village spillovers which we will see repeated across other outcomes.

The pattern of expenditure increases is broadly consistent with earlier work (Haushofer and Shapiro 2016). Both non-durable and durable spending increase substantially. Food expenditure accounts for a sizable portion of the overall increase in non-durable expenditure in both columns (41% and 65%, respectively). We can reject meaningful increases in reported spending on temptation goods (alcohol, tobacco and gambling), consistent with earlier work (Evans and Popova 2017).[23]

Consistent with increased expenditure on durables, asset stocks also increase (Table 1, Panel B). Anecdotally, many recipients withdrew money from M-Pesa immediately and saved via durable assets. The main pre-specified measure of assets includes livestock; transportation (bicycles, motorcycles, and cars); electronics; farm tools; furniture; and other home goods. We add in net household lending to, and borrowing from, both formal and informal sources. This measure of assets increases by USD PPP 174, or 24% of the mean for eligible households in control villages in low saturation sublocations.[24] This measure excludes the values of housing and land, which are harder to measure given thin local markets, but also likely important given existing work shows that households often use GD transfers to invest in housing (Haushofer and Shapiro 2016). We separately measure housing value as the

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

23. While there is surely some under-reporting of temptation goods, the fact that the control group mean is non-trivial demonstrates that at least some households do feel comfortable report such spending. Given that we do not generally observe expenditure immediately after transfers were distributed, we cannot fully rule out the possibility that temptation good spending increased temporarily at that time.

24. The mean for eligible household in control villages and low saturation sublocations is USD PPP 724 (with SD 863), substantially less than the overall mean, which is not surprising since ineligible households own more assets on average.

respondent's self-reported cost to build a home like theirs, and land value as landholdings multiplied by the household's report of the per-acre cost of land of similar quality (in their village). Estimated housing value increases by USD PPP 391, or 65% of the control mean, and estimated land value increases, but this effect is not statistically significant.

A back-of-the-envelope calculation suggests that increases in expenditure on non-durables, durable assets and housing improvements a year and a half after the transfer together account for roughly sixty percent of the transfer value, assuming consumption effects on the treated were constant over time between the first transfer and the endline household survey. In the most basic Keynesian model, this marginal propensity to consume of 0.6 would imply a substantial multiplier of $\frac{1}{1-MPC} \approx 2.5$; we return to empirically estimating this multiplier in more detail in section 4.

Theoretically, the effect of a wealth transfer on earnings is ambiguous: it may reduce labor supply through an income effect, but may also enable productive investment. In the data, recipient households' income from all sources (excluding the GD transfers) does not appear to have decreased: point estimates are positive (USD PPP 78 and 131 in the two main specifications) and the reduced form effect is marginally significant.[25] Examining labor supply specifically, we do not find evidence that recipient households worked less; if anything, total hours worked by recipient households in agriculture, self-employment and employment increased slightly (Table 2, Panel A, columns 1 and 2). This is consistent with the studies reviewed by Banerjee et al. (2017), which generally have found that cash transfers in low and middle income countries do not reduce labor supply.

Interestingly, we also see little heterogeneity in estimated treatment effects on measures of assets, expenditure, income, and hours worked for eligible households across eight pre-specified dimensions of baseline characteristics (Figure A.2): respondent gender, an indicator for being over age 25, marital status, primary school completion, an indicator for having a child in the household, an indicator for being above the median on our measure of psychological well-being, and work status (in self-employment or wage employment).

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

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

Overall, these results highlight that cash transfer recipients substantially increased their expenditure on a broad range of goods. This spending was likely financed primarily by the initial transfers themselves, with possibly some contribution from higher earnings. A large share of this spending likely takes place locally: enterprises report that 86 percent of their customers come from within the same village or sublocation. We therefore turn next to examining impacts on local enterprises.

3.2 Enterprise effects

We measure (annualized) revenues and profits for non-agricultural enterprises directly by asking respondents about these quantities (Mel, McKenzie, and Woodruff 2009). We calculate costs as the sum of the employee wage bill, rent and security costs; this is not a comprehensive measure of all costs, and hence we do not expect the revenue measure to equal our measure of profits plus measured costs. For agricultural enterprises, total revenue is calculated as the sum of crop output (measured at the crop level) plus the value of pastoral and poultry output sold, and the value of the household's own consumption of pastoral and poultry output.[26] Agricultural costs are the wage bill, all agricultural inputs (e.g., seed and fertilizer), and land rental costs. We then calculate agricultural profits as total agricultural revenue minus agricultural costs. Recall that we include firm sector fixed effects to absorb any systematic differences in measurement by sector.

There are large increases in revenue for enterprises in both treatment and control villages (Table 3, Panel A). Revenues in treated villages increased by USD PPP 348 per household, a 46% increase, while those in control villages increased by USD PPP 231 (30%). Revenue gains are concentrated in the retail and manufacturing sectors. Both treatment and control villages experience statistically significant increases in manufacturing revenue of similar magnitudes – USD PPP 107 and 105, respectively – while treatment villages see larger gains in retail revenue (USD PPP 132 versus USD PPP 54) (Appendix Table A.3).

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

26. 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 the market price surveys during endline data collection in the household's subcounty of residence.

(control) villages increase spending on labor by USD PPP 85 (73), a sizable change relative to the control mean.

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

One caveat to this point is that many household assets are difficult to categorize into "productive" assets as opposed to consumer durables. For example, bicycles may be used for personal transportation (i.e., to visit friends), but could also be used for operating a bicycle taxi to generate income. We therefore inclusively categorize as "potentially productive" both livestock as well as a number of non-agricultural assets that could potentially be used for income-generating activities (beyond simply renting out the asset).[27] When we do so, overall roughly half of the increase in household asset ownership documented above is in what we believe to be purely non-productive assets, with small gains in productive agricultural assets (e.g., farm tools) and a modest gain for potentially productive assets (Table A.2).[28] We also fail to detect any investment response for non-agricultural enterprises owned by treated households: neither investment nor inventories increase relative to eligible owners in control villages (Table A.4, Panel B). Taken together, these patterns are also consistent with the cash transfer program generating only a limited local investment response.

3.3 Non-recipient household effects

We next examine how the spending behavior of recipient households, and its impacts on local enterprises, affected non-recipient households. Recall that non-recipients include both eligible households living in control villages, and ineligible households living in both control and treatment villages.

We find positive and significant expenditure effects for untreated households, measured on average roughly a year and half after the start of cash transfers. Column 3 of Table 1, Panel

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

28. For comparability across categories, results in this table do not include net borrowing, which is included in Table 1.

A presents results based on Equation (3). Notably, the magnitude of these gains (USD PPP 334, p-value < 0.01) are quite similar to those of treated households (USD PPP 343). The pattern of expenditure increases is also broadly similar to that for treated households, except that spending on durables does not increase among untreated households. One possible reason for the similarity in overall spending impacts is that the timing of effects on recipient and non-recipient households may be different, with recipient households showing impacts earlier than non-recipient households, but effects converging by roughly one year after the final transfer was received.

How did non-recipients fund these consumption gains? One possibility is that they are dis-saving, perhaps due to social pressure to "keep up with the Joneses", their neighbors who received the transfer. However, this does not appear to be the case: estimated treatment estimates for total assets, housing and land values are all positive, although not significant (Table 1, Panel B). Nor do we observe a borrowing response for untreated households from either formal and informal sources (Table 2, Panel C, column 3). A second potential explanation is that expenditure gains reflect inter-household transfers to non-recipient households, as documented in Angelucci and De Giorgi (2009) for Mexico. This also does not seem to be the case, as we find no significant increase in net transfers received by untreated households, and the point estimate of USD PPP 8.75 is less than 3 percent of the expenditure gain for untreated households; this mirrors the lack of an effect on net transfers among recipient households noted earlier.

Rather, the data suggest that consumption gains are driven by higher earned income: total annualized income increases by USD PPP 229. It is often argued in development economics that survey estimates of consumption are better measured and often substantially larger than estimates of income, particularly for poor households (Deaton and Zaidi 2002; Deaton 2018). While this is true in our case, we cannot reject that the total effect on income is the same as the effect on consumption expenditure for untreated households ($p = 0.21$). Income gains come largely from wage earnings, which increase by USD PPP 182, with a smaller and not significant contribution from profits from owned enterprises. These results are broadly in line with the enterprise results, in which profit increases were modest and marginally significant while the wage bill expanded significantly, by 71 and 61% in treatment and control villages, respectively (Table 3, row 4). Higher wage earnings appear more likely to reflect higher wages than increased labor supply, as the point estimate for overall household labor supply is actually somewhat negative (although there does appear to be an increase in hours worked in wage employment for the respondent, Table A.5). Hourly wages earned by untreated household increase meaningfully, although the estimate is only marginally significant (Table 2, Panel A).

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

3.4 Price effects

We turn next to effects on consumer goods prices – which are important for understanding the extent to which other monetary impacts are real as opposed to nominal, among other things – followed by the prices of major factors of production: labor, land and capital.

3.4.1 Output prices

Overall, we find small, positive and precisely estimated effects on consumer goods prices. As described above, the main measure is an expenditure-weighed log-index of market prices, constructed at the market level from data on a list of 72 products, with sub-indices for tradable and non-tradable goods. Both the ATE and average maximum transfer effect are small and precisely estimated near zero across all product categories (Table 4). The tight standard errors allow us to rule out even relatively small price effects: with 95 percent confidence, the ATE across the study period is below 0.0025 log points, or 0.25 percent. For the average maximum transfer effect across markets, the upper bound of the 95 percent confidence interval is 0.011 log points, or 1.1 percent. In addition to these overall price effects, price effects are small across almost all product categories. In particular, food prices are in line with the overall price index, and durables do not increase meaningfully in price. To help mitigate concerns that results may be sensitive to the price index weights or product classification, we find that average price inflation is below 1% for every product in the data.

Observed variation in price responses is generally in line with theoretical predictions. We find a significant (but economically small) price increase for non-tradable goods, while the effect for tradable goods is smaller and not statistically significant. We also observe somewhat larger increases in markets less integrated into the local economy, as defined by our measure of market access (Section 2.7). Columns 3 and 4 split markets into those above and below median market access, with estimated effects frequently more positive in more remote markets. Figure A.3 further breaks this pattern down by quartile of market access, with lower values reflecting more isolated markets. Panels A-C show a small amount of inflation for non-tradables in the most isolated markets, and no detectable price changes elsewhere or for other goods. Inflation for non-tradables in isolated markets nonetheless remains limited, at 0.2-0.3% on average. We also carried out enterprise phone surveys of a subset of enterprise types during the period in which transfers were going out, which collected price data on a limited number of products; price effects on these local manufacturing and services prices are also limited (see Appendix E.2).

These patterns are qualitatively 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 there lead to price decreases, while cash transfers lead to price increases, but their estimated effects are small except in remote villages. In contrast, these price effects are far smaller than those Filmer et al. (2018) estimate in the Philippines, which are 5 to 7% for protein-rich foods, although lower for other product categories. Burke, Bergquist, and Miguel (2019) show that a credit intervention impacting the supply of staples also affects local grain market prices in a different region of Kenya. Reconciling these divergent results with ours is a task for future research.

3.4.2 Input prices

Table 2 presents estimated effects on input prices measured in the household survey data (some of which we have previously mentioned). We find some evidence of higher wages. In row 1 of Table 2, we examine 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 hours worked in the last 7 days and their monthly salary (adjusted to weekly scale). In the reduced form specification, eligible households in treatment villages earn USD PPP 0.11 more per hour, on a base of USD PPP 0.70. This effect is no longer significant, however, when we also estimate across-village spillovers. For untreated households, the increase is even more marked at 0.19 USD PPP per hour, and this effect is significant at the 10% level. These potentially large wage effects do not seem to be driven by large labor supply responses. In row 2, we calculate the total hours worked by adult household members in agriculture, self-employment and employment, and estimate effects at the household level. Effects are relatively small and not significant. Together with the fact that enterprise wage bills increased significantly, these patterns are strongly suggestive of positive local wage effects (Table 3).

Effects on estimated land prices are positive and economically meaningful (at 9-14%), but not statistically significant (Table 2, Panel B). Our measure of land prices is a noisy one; as noted above, land prices are difficult to measure as formal sales are rare, and so we use respondents' self-reports of the amount per acre land like theirs in the same village would sell for. We also examine land rental prices as a robustness check, which yield data on actual land transactions for a subset of respondents; we do not find significant effects on land rental prices (Table A.6). Unsurprisingly (given land should be in relatively fixed supply), we find little change in total landholdings among treated households or those in more heavily treated areas. We also find no effects on total land rentals, nor on the total amount of land used for agriculture (Table A.6).

Finally, we observe fairly precisely estimated null effects on interest rates and total borrowing (Table 2, Panel C). We measure household borrowing from both formal (e.g., commercial banks, mobile credit services) and informal (family and friends, moneylenders) sources. The total loan amount reports total borrowing across sources in the last 12 months, setting those that did not borrow equal to zero. Note that the loan-weighted interest rate is the monthly interest rate on the most recent loan by source, weighted by the total amount of borrowing by source; we include informal loans without interest, which brings down the average rate.

4 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 'local' refers to the entire study area. As is standard in the literature, we define this multiplier $\mathbb{M}$ as the cumulative effect of transfers on local GDP relative to the total amount T transferred:

$$\mathbb{M} = \frac{1}{T}\left(\int_{t=0}^{\infty} \Delta GDP_t\right) \tag{7}$$

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 effects. This feature generates estimates that can be thought of as a model primitive, and with which estimates from other financing scenarios can be contrasted.[29]

An initial back-of-the-envelope calculation suggests that the multiplier may be substantial: static estimates of annualized consumption, income, and revenue measured 18 months

29. 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. Ramey (2011) reviews the literature and concludes the multiplier for temporary, debt-financed government spending is between 0.8 and 1.5, and Ramey and Zubiary (2018) find smaller estimates of the fiscal multiplier using different methods and data.

after a mean cash injection of 8.5% of GDP all increased by around 14-18%, suggesting a multiplier of roughly 2. In this section, we refine this estimate by both accounting as fully as possible for effects on all components of GDP, including spillover effects, and 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, the expenditure-based measure of local GDP is $GDP_t = C_t + I_t + G_t + NX_t$, where C_t is consumption expenditure on non-durables and durables, measured as quarterly consumption plus accumulated assets at endline.[30] To avoid potential double-counting, we exclude home durables from the consumption measure as part of this expenditure may be reflected in an accumulation of assets. In addition, we exclude net lending as well as home and land values from the asset measure because changes in home and land values may not be driven purely by investment, and because we think of land supply as being essentially fixed. These exclusions may lead us to err on the side of understating the multiplier, for instance, dropping home durables lowers the expenditure multiplier by 0.2, and home and land values have an even bigger impact. I_t is enterprise investment plus accumulated inventories at endline. We exclude local government expenditure, G_t, as Walker (2018) shows that the intervention had a precisely estimated null effect on it.

Since we also measure household and enterprise income, we can construct a dual income-based measure of local GDP as the sum of factor payments and profits: $GDP_t = W_t + R_t + \Pi_t + Tax_t - NFI_t$, where W_t is the total household wage bill, R_t are rental expenses of local enterprises, Π_t are enterprise profits, and Tax_t is total enterprise taxes.[31]

The expenditure- and income-based measures of GDP we generate are based on unusually rich underlying data, but each has potential limitations. In particular, each may misattribute transactions between agents located in the study area and counterparties located outside.

In the expenditure case, the main concern is that we do not directly observe net exports (NX_t). Imports show up as expenditure but are not local GDP, while exports do not show up in expenditure but are part of local GDP. To the extent that cash transfers decrease (increase) net exports from the study area, our expenditure multiplier would overstate (understate) the multiplier. Intuitively, we might expect net exports to fall following a large external income transfer: since many local firms are retail establishments, imports of intermediate goods

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

31. We employ the household rather than enterprise wage bill, as the household survey sample is larger and includes individual-level wage earnings data. We omit land rental income because we do not see any significant evidence of effects on this above; this may lead us to understate the multiplier. 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.

(including packaged consumer goods ready for sale) would likely increase. This suggests that the expenditure-based approach might be upwardly biased away from zero.[32] 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. We present an extension below assessing how sensitive our multiplier estimate is to different assumptions about net exports. Of course, increases in net imports could in part reflect increases in economic activity outside of the study area due to the GD transfers, which our concept of the local multiplier does not capture but which are a part of the broader impact of the intervention.

In the income case, potential bias could arise if there are changes in net wage income (NFI_t) earned outside the study area, since this is not considered part of local GDP. This bias seems unlikely to be quantitatively important in our setting: 86% of all employees of study-area enterprises are family labor (and therefore presumably overwhelmingly local), and among individuals employed for a wage, only 6% report an employment contact address outside the study area. To the extent some bias remains, we would expect it to be negative (towards zero), if net labor income earned outside the study area decreases in response to higher local business revenue, employment and wages. This suggests that the income-based approach may yield a lower bound on the multiplier.

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

4.1 Empirical strategy

To capture the full dynamic response of GDP, we estimate and then integrate effects on each component discussed above over time.[33] For each flow variable x (e.g., consumption, investment, wage income, etc.), we first estimate the following specification, which is a dynamic extension of previous estimating equations:

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

32. Note that direct imports by households themselves are unlikely to increase because 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 local enterprise revenue are fairly similar, suggesting that consumer spending was quite localized. Similarly, non-farm businesses report only 5% of customers coming from outside the study area, and that share does not change significantly in response to treatment.

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

The coefficients in this model are well-identified: both treatment status and survey times were randomized across villages, and 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 after about 18 months (see Figure 1). With a few exceptions, recall periods are less than or equal to one month, so we often do not directly observe the initial response in flow variables in the months immediately after the first transfers went to a village, which is when we might expect to see some of the largest impacts on expenditure. However, given that our specification treats each dollar transferred symmetrically, we can still estimate the local response during these early quarters because transfers to recipients rolled out over 8 months.[34] Similarly, we can estimate neighborhood effects using the substantial variation in the timing with which nearby villages were treated. We tend to obtain less precise estimates of responses in early quarters immediately following transfers, however, as they are estimated 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 during the first three quarters following a transfer; these estimates are almost surely biased downwards, as they miss all economic activity in the initial period following transfers, but are considerably more precisely estimated and thus may be attractive to consider, for instance, if a goal is to minimize mean squared error.

We then integrate dynamic effects on flow variables over time up to 29 months (10 quarters) after treatment. We compute the dynamic profile of treatment effects (or the impulse response function, IRF) using the coefficients estimated above and assuming that the treatment rolled out to treated households as planned: 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 households, and separately for three categories of enterprise in both treatment and control villages: own-farm, non-farm enterprises operating within households, and those operating from outside the homestead. We then aggregate the quarterly estimates across all villages using inverse population weights from our household and enterprise census to compute the study area-wide IRF for each flow component.

For two components of GDP, we are instead able to measure impacts on the integral of flows over time by simply measuring impacts on accumulated stocks, simplifying the problem. Specifically, we measure effects on durable consumption expenditure using effects on the stock of endline household durable goods, and effects on inventory investment using

34. While the first transfer is largely unanticipated, the second and third are not and responses may differ. However, we do not know of any borrowing against future GD payments, and credit markets are imperfect in our context. We therefore pool all transfers in our dynamic regressions, and leave the analysis of potentially differing effects of anticipated transfers for future research.

effects on current inventory stocks at endline. One drawback is that these figures are likely to under-estimate cumulative spending to the extent that some assets depreciated (or were resold) between the time of purchase and measurement, although over the limited timeframe considered this may be a second-order concern; any such bias would tend to reduce the estimated multiplier. In the graphical presentation, we assume that any effects on these stocks occurred equally across all post-treatment quarters.

Inference is conducted 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 (i) each multiplier is less than zero, and (ii) each is less than one, as well as the joint hypothesis that both multiplier estimates satisfy these conditions. There has been extensive debate in macroeconomics about whether or not the fiscal multiplier is greater than one (Chodorow-Reich 2019), making the second hypothesis of particular interest.

4.2 Multiplier estimates

Table 5 presents the two main multiplier estimates. We estimate a sizeable multiplier using both approaches, in line with the back-of-envelope figure derived above.

The estimated expenditure multiplier is 2.53 (Panel A). 61% of this effect is driven by consumption expenditures. Household asset purchases and enterprise investment make up another 17% and 19% respectively, and enterprise inventories are not quantitatively important. While part of the asset response could potentially reflect productive investments by household-operated enterprises, at least 42% of the asset response comes from non-productive assets, across both treated and untreated households (see Table A.2). Taking this into account, consumption alone leads to an estimated multiplier of 1.7, underscoring the overall point that cash transfers appear to have led to a predominantly demand-side driven increase in local economic activity.

As argued in section 4, the expenditure-based measure may partly reflect imports of intermediate inputs from outside our study area, which do not constitute local GDP. As a robustness check to gauge the magnitude of this potential bias, we first assign each component in the consumption and asset measures to enterprise types at which the good is most likely to be purchased (using revenue-shares of different enterprise types, where appropriate). Because of the incomplete nature of our firm cost measure, we can determine that at most a fraction $1 - \frac{cost_i + profit_i}{revenue_i}$ of the revenue of firm i is spent on intermediate goods. For each firm type, we then get a revenue-weighted average upper bound for the share of intermediates in its production function. Next, we make assumptions about what share of intermediate goods is likely imported from outside the study area, conservatively erring on the side of assuming a high share. We assume inventories are proportional to consumption expenditure and come entirely in terms of intermediate goods; for enterprise investment, we assume its import share is equivalent to that of household assets, since enterprise and household assets are fairly similar in this context. This conservative methodology yields an upper bound of 21% for the expenditure-weighted share of local consumption that may reflect expenditure on imported intermediate goods. If imports scale linearly with expenditure, this suggests a multiplier of at least 2.0 on local expenditure alone (see Appendix F).

The estimated income-based multiplier is quite similar in magnitude to the expenditure-based multiplier, at 2.81 (Panel B), and we cannot reject that they are the same ($p = 0.86$). This is notable since it is calculated using a completely distinct set of component measures. Of this total effect, we find that 60% reflects increased enterprise profits, 36% increased wages, and a much smaller contribution (5%) comes from 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.

We can also look at the relative contribution of treated and untreated households to both multipliers, as a share of the total household contribution to the multiplier. Untreated households account for 80% of the household contribution to the expenditure multiplier, and 85% of the contribution to the income multiplier, both of which are roughly their share in the local population. This suggests that studies focusing only on recipient households may be missing sizable shares of program benefits.

An advantage of this "macro-experimental" approach to estimating the multiplier is the ability to conduct statistical inference. To start, we reject the null of a negative multiplier (with a value less than zero) at the 10% level using either approach (Table 5), and reject the null at $p = 0.02$ when testing the joint restriction, and at $p = 0.04$ when testing the average of both multipliers. Since the two measures exploit distinct data, we gain statistical power by examining both measures together. Testing the null hypothesis of a multiplier less than one speaks more directly to recent debates. Using either the expenditure or income-based approach alone, the p-value on this test is 0.17 and 0.18. The average of the two multipliers is 2.67 ($SE = 1.52$, $p = 0.14$), and the hypothesis that the multipliers are jointly less than one is rejected at the marginally significant $p = 0.07$ 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 statistically attractive in a mean squared error sense since it yields more precise, if surely somewhat downward biased, estimates. Under this assumption, the estimates range from 1.8 to 2.0 (smaller than the preferred estimates in Table 5, as expected), and the null hypothesis that both multipliers estimates are jointly less than one is rejected at $p = 0.04$ (Table F.2).

Figure 3 presents these results graphically, breaking up the aggregate multipliers into quarters after transfers went out. Panel A presents the expenditure-based multiplier. The increase in GDP is fairly stable over time; in fact, we cannot reject that the expenditure response is constant across all quarters (p-value of 0.85). 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 as late as two years after the transfer. The less precisely estimated effects, with larger confidence intervals, during the first three quarters afters transfers go out are visually apparent. The income multiplier, on the other hand, visually appears to fluctuate more over time (Panel B): it is marked by a strong early response in profits, while wages appear to take longer to rise. Yet as with the expenditure measure, we cannot reject equality of all quarterly coefficients (p-value of 0.88). Future research could investigate the dynamics following a cash infusion in further detail.

Overall, we estimate large multipliers in the range of 2.5 to 2.8, using two distinct sets of measures, in terms of both expenditures and income. We reject the null hypothesis that the multiplier estimates are jointly less than one at over 90% confidence, and the results are robust across several alternative approaches. 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 from that literature 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 poorer still. Beyond effects driven by the relative level of economic development, the differences between our results and existing estimates may reflect other structural differences between the Kenyan and US economies, differences in the data used, as well as any effects on (or expectations of effects on) either monetary policy or future taxes in the US, response effects that this study's experimental design usefully allows us to avoid.

5 Non-market outcomes and inequality

We next examine other variables that arguably directly influence (or proxy for) well-being, and may thus capture externalities (either between or within households). We focus 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, possibly 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 skipped 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 sub-index of child food security; this may capture intra-household externalities to the extent children do not have meaningful 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 currently 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 culturally sensitive nature of some of these questions, information on gender-based violence was only collected from female respondents being surveyed by female enumerators, resulting in a somewhat smaller sample size. Finally, the *security index* captures measured levels of crime (and positive values denote less crime).[35]

For treated households, we find positive and significant effects for four of the six indices: psychological well-being, food security, education and security. Estimated effects are close to zero and not significant for the health index and female empowerment index. When looking at total effects including spillovers for the treated, we find a similar pattern for all but the security index.[36] For untreated households, we find no significant effects of local cash transfers except for the education index, which is higher by 0.1 SD ($p < 0.10$). Importantly, we do not find evidence of adverse spillover 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).

5.1 Inequality

One potential form of "psychic externality" has to do with the distribution of economic status within 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 immediately clear how

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

36. It is notable that we do not find that treatment increases female empowerment in treated households. The point estimates on both the reduced form and total effect are negative but not statistically significant (-0.01 and -0.13, respectively). This contrasts with Haushofer and Shapiro (2016) who found increases in female empowerment and reductions in domestic violence among households receiving a similar transfer.

consumption and wealth distributions changed; transfers were targeted to relatively poor households within each community, but we find large spillovers onto their neighbors, who are generally somewhat better off.

To assess this, we first calculate Gini coefficients for consumption expenditure and wealth by village and estimate village-level treatment effects. We find no significant reductions in either measure; in fact, we estimate a small positive and marginally significant effect of transfers on the wealth Gini in control villages ($p < 0.1$, Table A.7). This is driven by a) ineligibles (who were richer at baseline) experiencing larger spillovers than control eligibles (Table A.8), and b) within each group, effects on initially wealthier households being slightly larger. As a benchmark, we also calculate the effects on inequality we might have expected to observe had there been *no* spillovers. Specifically, we generate counterfactual expenditure and asset outcomes for recipient households assuming they spend 66% of their transfer on consumption, and 34% on durable assets (this equals the ratio of effects we find in our data, Table 1), while setting untreated households to their baseline values.[37] In these counterfactual simulations, treatment reduces the Gini coefficients for assets and consumption, but less for the latter as eligible and ineligible households differ more strongly in terms of assets at baseline than consumption. We can reject the hypothesis that the observed and counterfactual treatment effects on consumption inequality are the same in treatment villages; for assets, we reject equality for both treatment and control villages.

These patterns again highlight the large spillover gains for non-recipient households: these mean that wealthier non-recipients benefit along with recipients, and on some dimensions (e.g., assets) they benefit so much that inequality may even slightly increase. Other programs (e.g. Banerjee et al. (2015)) have found larger direct effects for less-poor households; here we show that the same pattern can hold for spillover effects.

6 Conceptual framework for household welfare

To help interpret the welfare implications of the effects documented above, we next examine a simple theoretical framework. While fairly generic, the framework categorizes the ways in which these effects translate into changes in household welfare, including both households that did and did not receive transfers. We also use this framework to examine the relationship between (i) the classic "equivalent variation" concept of welfare and (ii) household expenditure, a widely used proxy measure in development economics.

Consider an indirect utility function $v_i(T_i, T)$ which defines the utility level achieved by

37. Since we did not collect consumption expenditure at baseline, we use random draws of low-saturation control villages for the consumption counterfactuals. We also construct counterfactuals based on the total effects in Table 1 (not scaled to sum up to the full transfer); our conclusions do not change fundamentally.

household i when it receives a (possibly zero) transfer T_i while other eligible households in the 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:

$$v_i(T_i^*, 0) = v_i(T_i, T) \tag{8}$$

We aim to characterize the ways in which T affects T_i^*. Notice that if there were no general equilibrium effects, in the sense that v_i did not depend on T, then we would simply have $T_i^* = T_i$, i.e., the tautology that the value of receiving a dollar is simply a dollar.

Write v_i as the value of some generic underlying optimization problem

$$v_i(T_i, T) = \max_{x_i} u_i(x_i, x_{-i}(T)) \text{ s.t. } x_i \in X(T_i, T) \tag{9}$$

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 affect i (e.g., through externalities, public goods provision, preferences over inequality, etc.). X is household i's constraint set. Preferences here are over a sequence of x's; these need not be time separable, which allows for non-standard time discounting, the existence of durable goods that generate a flow of benefits, etc. This formulation highlights the fact that a change in T 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.

In the data, we observe impacts on prices and income from various sources that together determine the household's money budget constraint. To model this, let

$$v_i(T_i, T) = \max_{\{x_{it}\}} u_i(\{x_{it}\}, \{x_{-it}\}) + \lambda_i \left(T_i + \sum_{t=0}^{\infty} (\delta_i)^t (y_{it} - p_t \cdot f(x_{it})) \right) \tag{10}$$

Here the budget constraint depends on the transfer T_i the household 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).[38] Overall, this specification implies that we need to understand effects on the exogenous (from the household's point of view) variables prices (p_t and δ_i) and transfer income y_{it} to assess welfare impacts.[39]

6.1 A dual approach using expenditure

Household consumption expenditure is often used as a proxy for well-being. To see how consumption relates to well-being in the equivalent variation sense defined above, it will be helpful to assume that the solution to the household's problem is well-approximated by first-order conditions (i.e., $\partial u_i/\partial x_{it} = \lambda_i \delta_i^t p_t \cdot \partial f/\partial x_{it}$). This allows us to apply the envelope theorem to calculate the marginal effects of T_i and T on realized welfare; doing so and taking their ratio yields the marginal equivalent variation of an increase in transfers T (where derivations are in Appendix H):

$$\frac{dEV_i}{dT} = \frac{1}{\lambda_i}\frac{\partial u_i}{\partial x_{-it}} \cdot \frac{\partial x_{-it}}{\partial T} + \sum_{t=0}^{\infty}(\delta_i)^t \left(\frac{\partial y_{it}}{\partial T} + \frac{t}{\delta_i}\frac{\partial \delta_i}{\partial T} s_{it} - \frac{\partial p_t}{\partial T} \cdot f(x_{it})\right) \tag{11}$$

The first term here captures welfare effects through externalities and public goods, which are of course not captured by household expenditure. The second term captures effects on the net present value of the household's budget constraint.[40] To see how this quantity is related to (the net present value of) expenditure, consider the simple case where households buy and sell goods and services $f(x_{it}) = x_{it} = c_{it} + l_{it}$, where we have decomposed x_{it} into the goods and services $c_{it} = \max\{0, x_{it}\}$ the household buys (e.g., clothes) and those $l_{it} = \min\{0, x_{it}\}$ that it sells (e.g., labor services). The NPV of expenditure is $e_i = \sum(\delta_i)^t p_t \cdot c_{it}$. Then

$$\frac{dEV_i}{dT} = \frac{de_i}{dT} - \sum(\delta_i)^t \left(\frac{\partial p_t}{\partial T} \cdot c_{it} + \frac{t}{\delta_i}\frac{\partial \delta_i}{\partial T} p_t \cdot c_{it} - p_t \cdot \frac{\partial l_{it}}{\partial T}\right) \tag{12}$$

38. For simplicity, we consider the case where the future is certain; the analysis that follows would be effectively the same if we introduced uncertainty and took expectations.

39. Equation (10) allows for some but not all forms of credit constraint. There may be arbitrary 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 modeled hard constraints on household borrowing (e.g., a non-negative assets constraint), then the shadow value of money could 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.

40. While we do not pursue it here, one could in principle use Equation (11) to calculate numeric estimates of the equivalent variation for each household i. The exercise would be quite demanding of the data, as it requires observation or interpolation of the full time path of all outcomes and for all counterfactual transfer amounts $T \in [0, \overline{T}]$, and would require imposing some structure (e.g., homotheticity) on preferences.

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. Finally, if (more realistically) we were to examine expenditure over any finite period of time this would introduce a third bias, as this metric would count as a welfare gain any increases in current expenditure that were driven by decreases in future expenditure (i.e., by dis-saving).

6.2 Implications & interpretation

The empirical results can be interpreted through the lens of this framework. 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 not significant 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, rather than being driven by increased labor supply. Working from the expenditure side 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 labor supply.

Second, what data we have on welfare changes via externalities suggests that these were generally zero or in some cases slightly positive. Public goods provision at the village level was unchanged (Walker 2018) as were levels of crime (Table 6, row 7). Non-market outcomes also improved: non-recipient households scored (not significantly) higher on an index of psychological well-being. Within households, children – who have less say in decision-making than adults – appeared better off, with significantly higher food security and education scores (Table 6, 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 6, row 7). One potential exception concerns inequality (Table A.7); if households have preferences over their relative socioeconomic standing within the community, then the fact that positive spillovers were so large as to generate a small net increase in asset inequality (despite the means-testing of the initial transfers) may have been a source of disutility.

7 Discussion: productive responses

The results above, and in particular the modest price responses and large multiplier estimates, raise the question what features of the local economy enabled it to respond as elastically as it did to a large shock to aggregate demand. While fully addressing this question is beyond the scope of the present project, we outline here what we can say given the available data.

We note first that any explanation of these patterns must apply to the retail and manufacturing sectors specifically, as this is where output gains are concentrated. Table A.3 reports this decomposition. Estimated increases in services and agriculture are more modest and not significantly different from zero.

Next, we find limited evidence of increased *employment* of the main factors of production: land, labor, or capital. Land is of course in relatively fixed supply; agricultural households do not report owning or renting more of it (Table A.6) and we would not expect it to be a limiting factor in retail or manufacturing production. We find no significant change in total household labor supply (Table 2), though we do see a net shift out of self-employment and into wage employment (Table A.5, Panel A), with the latter increasing by 1.7 hours per person per week on average. These estimates are not statistically different from zero, however, and even under generous assumptions can explain only around a 5% increase in real output, well below the total response we observe.[41] Finally, for the non-agricultural enterprises that increased their output, we observe no significant changes in investment in fixed capital (Table 3, row 7), and while inventories increased somewhat they actually decrease in proportion to sales (from 26% to 21%, Table 3, row 6). Moreover, if output increased due to investment activity, then we would expect to see these increases concentrated in enterprises owned by transfer recipients (who gained access to a new source of capital), but if anything we find the opposite (Table A.4).

In a purely accounting sense, this means that a large share of the increase in output must be due to an increase in the throughput of intermediate goods and in the utilization of factors of production used to process them. This is consistent with observation of the retail and manufacturing enterprises in the area, which typically involve some degree of "on-demand" production. A retail establishment, for example, requires premises and an employee to "mind the shop," but once these are in place the volume of goods it retails depends largely on consumer demand. Similarly, many small-scale manufacturing enterprises require equipment and staff to be in place but then produce only when customers arrive. In fact, about 60% of manufacturing revenue accrues to just two enterprise types, grain (posho) mills and welding establishments, both of which often operate "on-demand."

These examples suggest retail and manufacturing sectors in which there are meaningful

41. Specifically, an increase of 1.7 hours per person is a 7.4% increase in wage labor hours. Assuming a Cobb-Douglas production function with a labor share of 2/3, and no productive value of time given up from self-employment, this implies a 4.9% increase in real output.

costs that are fixed over the relevant ranges – e.g. the costs of a building, of a mill, or of hiring a single employee – and as a result meaningful slack capacity that can be better utilized as demand increases. While we did not collect direct measures of capacity utilization, some indirect evidence suggests the existence of meaningful slack. The average non-agricultural enterprise saw just 1.7 customers per hour, in between which all other inputs (i.e. employee time, fixed capital, inventories, etc.) may sit idle. For labor inputs in particular, 72% of non-agricultural enterprises have just a single employee, which suggests that (due to integer constraints) the labor input into these business is essentially fixed over the relevant range.

Given this structure of production, we would expect the revenue from additional sales to be paid out to (i) the suppliers of intermediate goods, (ii) the suppliers of elastic factors of production, whose marginal product increased as they became better utilized, and finally to (iii) enterprise owners to the extent they are able to extract economic profits. We do not directly measure purchases of intermediates, but upper bounds for the expenditure-weighted share of intermediate inputs in total sales are sizeable, at 55% in the retail and 24% in the manufacturing sector (see Appendix F). Consistent with (ii), we do see an increase in wage bills which accounts for 27% of increased revenue (Table 3). Estimated effects on profits, meanwhile, are modest and not statistically significant (and may in any case be better interpreted as returns to the owners' capital which, as usual, are difficult to distinguish from true economic profits).

While only suggestive, this interpretation of the supply side response to a demand shock is consistent with other recent findings in East African settings. In ongoing work in Uganda, for example, Bassi et al. (2019) 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 an old idea in development economics that it might be possible to expand production without notable price inflation due to the availability of slack capacity. Classic arguments focused on "surplus labor" (Lewis 1954), while here it was apparently possible to increase the utilization of both labor and capital. The issue here may have been not that wages are kept artificially high as in Lewis (1954), but rather that there are real constraints that limit the ability to flexibly scale the employment of capital and labor to match demand. Mechanisms to address this through better coordination, such as periodic markets, do so imperfectly, leaving some degree of residual excess capacity.[42]

42. A growing literature has also found evidence of excess capacity in rich countries, especially in periods of recession (e.g., Murphy (2017), Michaillat and Saez (2015), and Chodorow-Reich (2019)).

8 Conclusion

A large-scale cash transfer program in rural Kenya led 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 estimated effects of far less than 1% on average across a wide range of goods. Taken together, two different empirical approaches – using consumption data and income data – both yield an estimated local fiscal multiplier of 2.5 to 2.8, and the hypothesis that the multipliers are less than or equal to 1 is rejected with 90% confidence. This multiplier is somewhat larger than recent estimates of 1.5-2.0 from the United States. Several suggestive patterns in the data are consistent with the existence of "slack" on the supply side in our context, which can also help account for the large 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 adverse consequences for non-beneficiaries. However, these concerns are not supported by our results. The consumption expenditures of untreated households and revenues of firms rise substantially in areas receiving large cash transfers; there is little to no price inflation; overall income, consumption and asset inequality do not increase meaningfully in treated areas; nor are there negative effects in terms of 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 RCTs of cash transfer programs that simply compare outcomes in treatment versus control villages may be understating true overall impacts by ignoring the general equilibrium effects that we capture (along the lines that Miguel and Kremer (2004) argue in the context of a health program).

This study is among the first, to our knowledge, to exploit randomized controlled 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 the subject of intense interest since at least the seminal work of Keynes (1936). In studying it using an RCT, we provide a novel counter-example to some well-known critiques of RCT methods in development economics, including by Bardhan (2005), Easterly (2006), and Deaton (2010), 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 achieve both (Muralidharan and Niehaus 2017; Burke, Bergquist, and Miguel 2019).

The extent to which the multiplier results apply to other settings merits further discussion. It seems natural that the findings would be particularly relevant for other low or middle income economies that share structural and institutional features with Kenya, including many other African settings. One open question is the extent to which the relatively large estimated multiplier is driven by the targeting of transfers to the poorest rural households. It is plausible that spillover effects would be more muted if the program were targeted to better-off households, who would likely have a lower marginal propensity to consume out of a one-time transfer. Thus, the distribution of spending capacity could have implications for aggregate output. An especially noteworthy feature of our study setting is the fact that we estimate a large multiplier even during a period when the Kenyan economy was experiencing steady economic growth, rather than a recession. Hence it appears that any under-utilization of supply side capacity is not simply temporary or cyclical in rural Kenya, but rather may be more persistent. A recent body of research argues that there may be a related phenomenon of steady-state "liquidity traps" or "secular stagnation" in advanced economies as well (e.g., Rachel and Summers 2019, Mian, Straub, and Sufi 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 steady-state. However, other theoretical perspectives from international trade, economic geography, and development (e.g., Marshall 1890, Rosenstein-Rodan 1943, Krugman 1980, Murphy, Shleifer, and Vishny 1989), as well as the liquidity traps literature mentioned above, would suggest there could be persistent local effects of a temporary cash infusion, due to agglomeration effects, increasing returns, changes in income inequality, market structure and firm specialization, and even shifts in the social networks of traders and suppliers. Temporary cash transfers and other forms of assistance have also been shown to have effects on long-run human capital accumulation and earnings 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.

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, J. Aislinn Bohren, Craig McIntosh, and Berk Özler. 2018. "Optimal Design of Experiments in the Presence of Interference." *The Review of Economics and Statistics* 100 (5): 844–860.

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.

Baird, Sarah, Craig McIntosh, and Berk Ozler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *Quarterly Journal of Economics* 126 (4): 1709–1753.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. "A multifaceted program causes lasting progress for the very poor: Evidence from six countries." *Science* 348 (6236).

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).

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

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.

Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. "Adaptive Linear Step-up Procedures That Control the False Discovery Rate." *Biometrika:* 491–507.

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.

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.

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

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

Deaton, Angus, and Salman Zaidi. 2002. *Guidelines for constructing consumption aggregates for welfare analysis.* Vol. 135. World Bank Publications.

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

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

———. 2017b. "GE Effects of Cash Transfers: Pre-analysis plan for targeting analysis." September.

———. 2018. "General Equilibrium Effects of Cash Transfers: Pre-analysis plan." June.

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

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

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.

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.

Mian, Atif, Ludwig Straub, and Amir Sufi. 2019. *Indebted Demand.* Unpublished.

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.

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

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.

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

Ramey, Valerie A. 2011. "Can Government Purchases Stimulate the Economy?" *Journal of Economic Literature* 49 (3): 673–685.

Ramey, Valerie A., and Sarah Zubiary. 2018. "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data." *Journal of Political Economy* 126 (2): 850–901.

Rosenstein-Rodan, Paul N. 1943. "Problems of Industrialisation of Eastern and Southeastern Europe." *Economic Journal* 53:202–211.

Sadoulet, Elisabeth, Alain de Janvry, and Benjamin Davis. 2001. "Cash Transfer Programs with Income Multipliers: PROCAMPO in Mexico." *World Development* (6): 1043–1056.

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

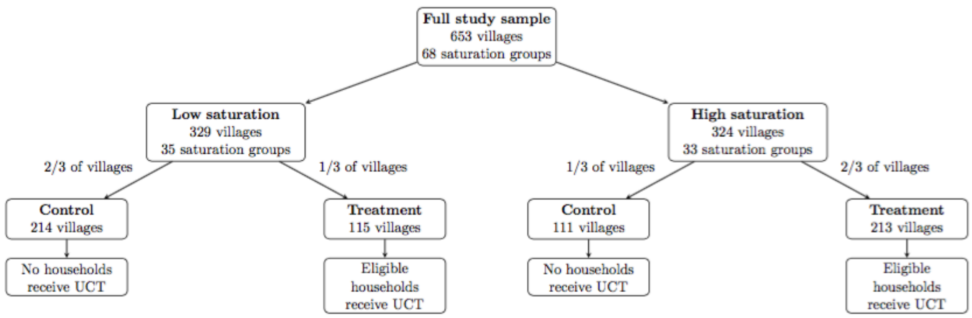
Walker, Michael. 2017. "Pre-Analysis Plan: Local Public Finance and Unconditional Cash Transfers in Kenya." February.

———. 2018. "Informal Taxation Responses to Cash Transfers: Experimental Evidence from Kenya." July.

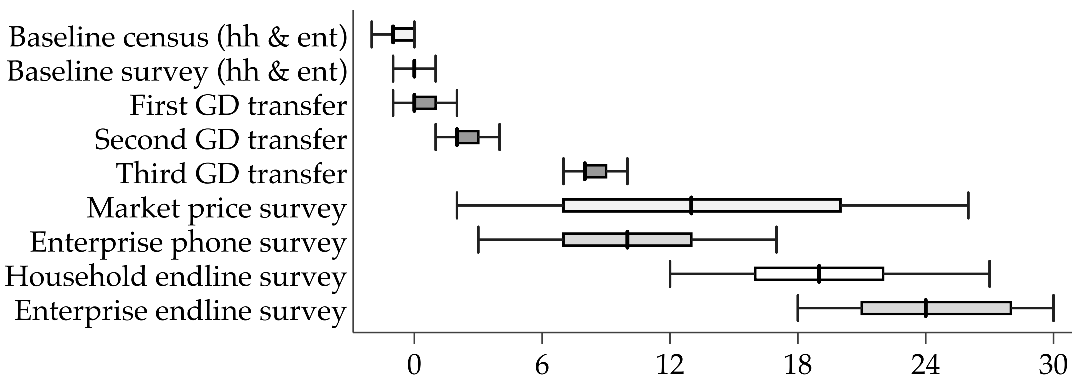
World Bank. 2017. *Closing the Gap: The State of Social Safety Nets 2017.* Technical report. Washington, D.C.: World Bank Group, April.

Figure 1: Study design and timeline

(a) Randomization

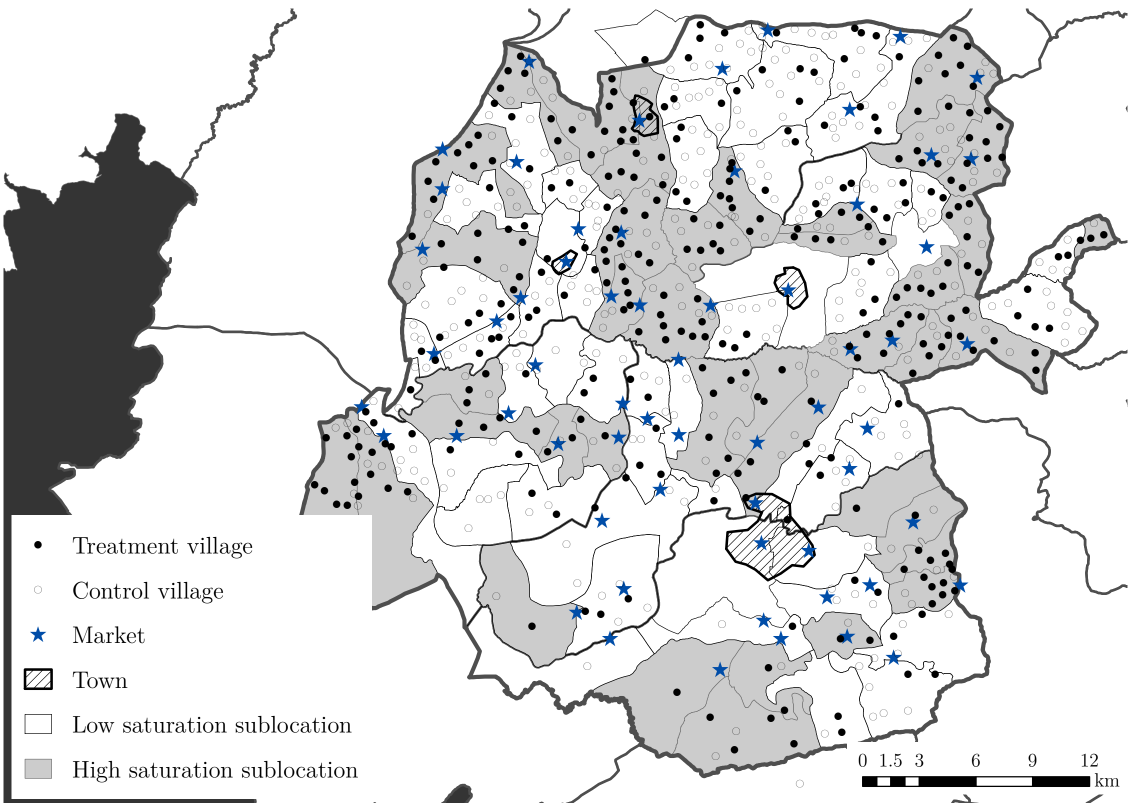


(b) Timing relative to experimental start



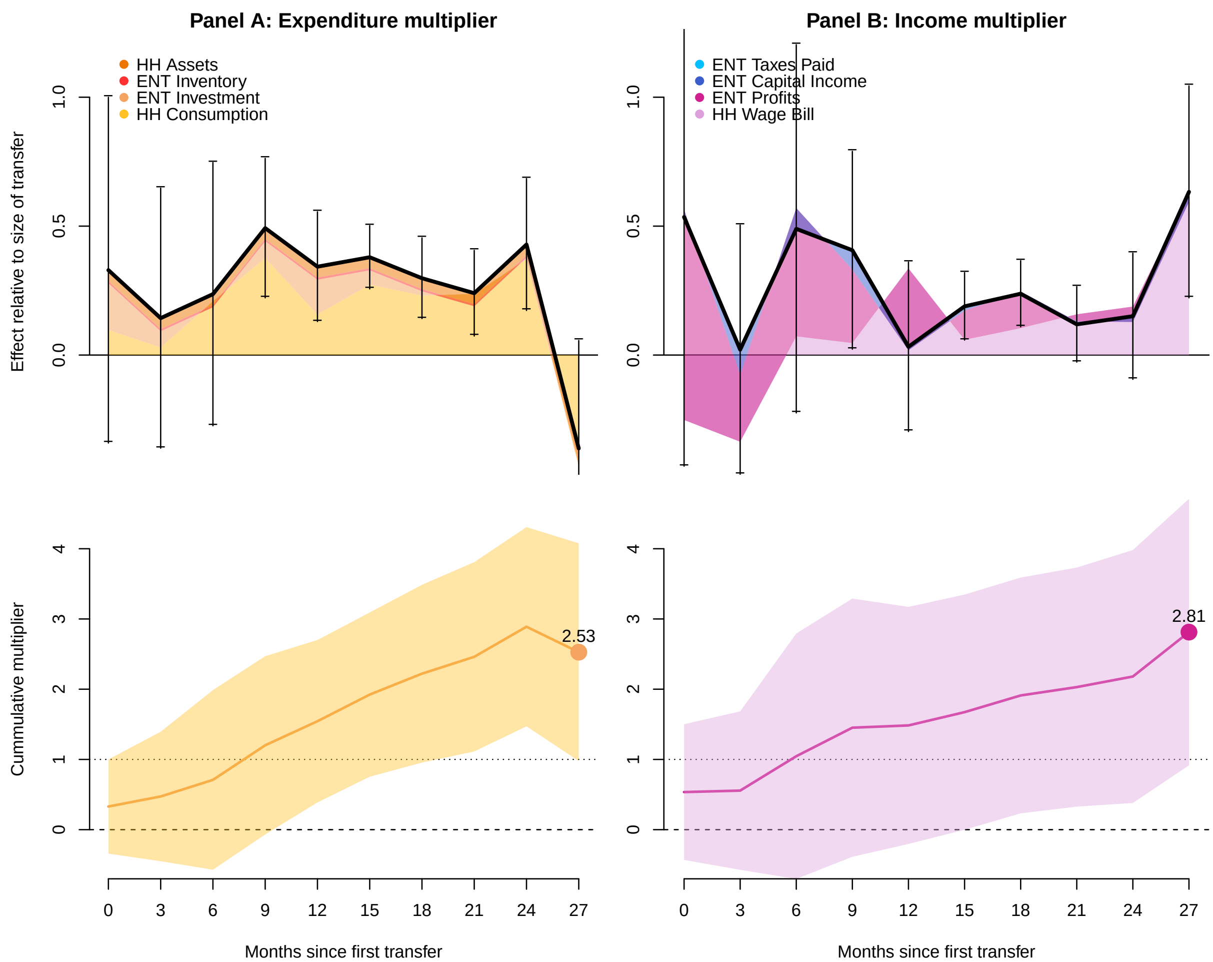
Notes: Panel A illustrates the two-level randomized controlled trial experimental design. 653 villages were grouped into 68 saturation groups based on the sublocation (the administrative unit directly above the village level) in which they are located. Saturation groups were then randomly assigned to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

Figure 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: Fiscal multiplier over time



Notes: Both expenditure and fiscal multipliers broken down by components and over time. Panel A shows the expenditure multiplier over 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. Colored areas below the curve represent the different components of expenditure. Note that some components turn negative from time to time, leading to some areas to overlap, which is indicated by darker shading. Standard errors around the joint multiplier estimates are obtained by 2000 replications of the wild bootstrap and reported as brackets of ±1SE. The integral under the multiplier curve adds up to our main multiplier estimate of 2.46. The bottom graph of Panel A traces out this integral over time. It also prints ±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, Savings and Income

	(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)
Panel A: Expenditure				
Household expenditure, annualized	292.98***	343.34***	333.66***	2,536.86
	(60.09)	(112.02)	(123.22)	(1,934.09)
Non-durable expenditure, annualized	174.99***	211.90**	288.46***	2,402.43
	(55.41)	(96.75)	(111.44)	(1,801.59)
Food expenditure, annualized	71.61*	138.57**	132.81**	1,578.43
	(36.93)	(66.75)	(58.57)	(1,072.31)
Temptation goods expenditure, annualized	6.51	4.48	−0.71	37.10
	(5.79)	(9.17)	(6.50)	(123.59)
Durable expenditure, annualized	95.18***	106.29***	8.40	59.44
	(12.64)	(21.44)	(12.50)	(230.90)
Panel B: Assets				
Assets (non-land, non-house), net borrowing	178.47***	174.10***	132.63*	1,132.15
	(24.63)	(47.09)	(78.31)	(1,420.22)
Housing value	377.14***	390.59***	78.93	2,033.72
	(26.37)	(40.29)	(215.76)	(5,030.37)
Land value	49.50	112.92	543.71	5,030.72
	(186.30)	(277.47)	(459.46)	(6,607.61)
Panel C: Household balance sheet				
Household income, annualized	77.62*	131.48	229.42***	1,023.45
	(43.66)	(100.78)	(88.58)	(1,634.70)
Net value of household transfers received, annualized	−1.68	−11.49	8.75	130.18
	(6.81)	(13.78)	(19.10)	(263.75)
Tax paid, annualized	1.81	−0.18	1.90	16.93
	(1.28)	(2.13)	(1.93)	(36.51)
Profits (ag & non-ag), annualized	28.61	32.15	44.09	485.20
	(23.75)	(54.31)	(45.05)	(787.10)
Wage earnings, annualized	42.52	73.15	182.24***	495.37
	(32.24)	(64.23)	(65.54)	(1,231.56)

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, including between 5,372 and 5,424 observations. Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households (5,448 to 5,509 observations), coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer). The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. Standard errors are clustered at the sublocation in column 1, and calculated following Conley (1999), 2008 using a uniform kernel out to 10 km in columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 2: Input Prices and Quantities

	(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)
Panel A: Labor				
Hourly wage earned by employees	0.11***	0.03	0.19*	0.70
	(0.03)	(0.05)	(0.10)	(0.89)
Household total hours worked, last 7 days	2.44	1.43	−4.70	63.20
	(1.71)	(3.89)	(3.17)	(54.14)
Panel B: Land				
Land price per acre	166.84	346.75	556.73	3,952.86
	(201.20)	(304.27)	(412.26)	(3,148.52)
Acres of land owned	−0.19	−0.07	0.08	1.42
	(0.14)	(0.11)	(0.10)	(2.37)
Panel C: Capital				
Loan-weighted interest rate, monthly	−0.01	0.01	−0.01	0.06
	(0.01)	(0.01)	(0.01)	(0.07)
Total loan amount	5.55	2.22	6.09	80.61
	(4.95)	(9.18)	(13.23)	(204.36)

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 transfered 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, including between 2,828 and 5,423 observations for variables at the household level, and 2,832 observations at the individual level for wages. Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households, coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer). We have between 2,781 to 5,509 observations at the household level and 2,391 wage observations at the individual level. The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. In addition, prices are quantity-weighted. That is, wages are weighted by the number of hours worked, land prices by the number of acres purchased, and interest rates by size of each loan. Standard errors are clustered at the sublocation in column 1, and calculated following Conley (1999), 2008 using a uniform kernel out to 10 km in columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 3: Enterprise Outcomes

	(1)	(2)	(3)	(4)
	Treatment Villages		**Control Villages**	
	$\mathbb{1}$(Treat village) Reduced form	Total Effect IV	Total Effect IV	Control, low saturation weighted mean (SD)
Panel A: All enterprises				
Enterprise profits, annualized	15.76	68.13	31.21	322.91
	(23.51)	(44.48)	(41.18)	(691.28)
Enterprise revenue, annualized	5.93	348.10**	230.98**	757.56
	(94.49)	(143.24)	(106.88)	(2,499.92)
Enterprise costs, annualized	−9.47	98.42**	79.96*	147.39
	(29.19)	(42.21)	(44.68)	(550.78)
Enterprise wagebill, annualized	−12.69	85.16**	73.08*	120.24
	(26.19)	(37.51)	(41.29)	(492.65)
Enterprise profit margin	0.01	−0.05**	−0.06***	0.44
	(0.02)	(0.03)	(0.02)	(0.61)
Panel B: Non-agricultural enterprises				
Enterprise inventory	10.85	34.69**	17.06	193.87
	(9.12)	(14.46)	(10.48)	(505.65)
Enterprise investment, annualized	4.05	13.38	6.69	179.49
	(7.08)	(15.54)	(8.70)	(644.15)
Panel C: Village-level				
Number of enterprises	0.01	0.01	0.00	1.12
	(0.01)	(0.02)	(0.01)	(0.14)

Notes: Column 1 reports the coefficient on an indicator for treatment village, and includes an indicator for saturation status of the sublocation. Column 2 reports the total effect on enterprises in treatment villages (own-village effect plus across-village spillover) from the "optimal" IV spatial regression of each outcome on the amount transferred per capita to a enterprise's own village v (instrumented by village treatment status), and to villages other than v in each 2km radii band around the enterprise (instrumented by the share of eligible households assigned to treatment in villages other than v inside the buffer). Column 3 reports the total effect on enterprises in control villages (across-village spillover only). For each column, we stack 3 separate regressions for own-farm enterprises, non-agricultural enterprises operated within the household, and non-agricultural enterprises operated outside the household, due to our independent sampling across these enterprise categories. We have between 10,015 and 10,284 observations for all enterprises, and 2,414 to 2,423 for variables we collect for non-ag enterprises only. 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 clustered at the sublocation in column 1, and calculated following Conley (1999), 2008 using a uniform kernel out to 10 km in columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 4: 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**	0.54*	0.19**	0.10
		(0.06)	(0.30)	(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**	0.52	0.20*	0.04
		(0.06)	(0.32)	(0.11)	(0.07)
By sector	Food items	0.13**	0.55*	0.21*	0.05
		(0.06)	(0.33)	(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	−0.29	−0.08*	−0.17
		(0.10)	(0.50)	(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. For each price index, we include a balanced panel of 1,586 market-by-month observations. The number of radii bands and lags is chosen sequentially by minimizing the BIC, as pre-specified. Regressions include a full set of market and month fixed effects. Column 1 reports the implied ATE, calculated by evaluating the "optimal" regression specification at the average level of treatment intensity between September 2014 and March 2017, the time during which transfers went out. Column 2 reports the average maximum effect, calculated at the average across all markets of the month in which the largest per capita transfers went into a market's neighborhood (up to the largest buffer selected by the algorithm). Columns 3 and 4 break down the ATE by market access, defined as $MA_m = \sum_{r=0}^{10} r^{-\theta} N_r$, 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 5: Fiscal Multiplier Estimates

	(1) $\mathbb{M}$ Estimate	(2) H_0: $\mathbb{M} < 0$ *p*-value	(3) H_0: $\mathbb{M} < 1$ *p*-value
Panel A: Expenditure multiplier	**2.53** (1.55)	**0.05***	**0.17**
Household consumption	1.55		
Household assets	0.42		
Enterprise investment	0.47		
Enterprise inventory	0.08		
Panel B: Income multiplier	**2.81** (1.90)	**0.07***	**0.18**
Household wage bill	1.00		
Enterprise profits	1.68		
Enterprise capital income	0.09		
Enterprise taxes paid	0.04		
Panel C: Both multipliers			
Average of both estimates	2.67 (1.52)	0.04**	0.14
Joint test of both multipliers		0.02**	0.07*

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 29 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, which re-allocates within-sublocation Rademacher-perturbed residuals from the main population regressions to fitted outcome values to create perturbed samples. Columns (2) and (3) conduct one-sided tests of each 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. Panel C conducts two tests regarding both multipliers. The first row computes the average of both estimates and conducts tests on this average in much the same bootstrapped fashion as above. Finally, the last row reports p-values from joint tests against the same nulls, following an identical procedure. Appendix table F.2 additionally reports standard errors and bootstrapped p-values for all subcomponents of each multiplier. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table 6: Non-market Outcomes and Externalities

	(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)
Psychological well-being index	0.09***	0.12*	0.08	0.01
	(0.03)	(0.07)	(0.06)	(1.01)
Health index	0.04	0.06	0.01	0.03
	(0.03)	(0.06)	(0.05)	(1.01)
Food security index	0.10***	0.08	0.08	0.01
	(0.03)	(0.08)	(0.06)	(1.00)
Children food security	0.13***	0.17**	0.09	–0.04
	(0.04)	(0.08)	(0.09)	(1.12)
Education index	0.09**	0.10*	0.10*	0.01
	(0.04)	(0.05)	(0.06)	(1.02)
Female empowerment index	–0.01	–0.13	0.09	0.05
	(0.07)	(0.15)	(0.15)	(0.94)
Security index	0.11***	–0.02	–0.02	0.03
	(0.04)	(0.07)	(0.07)	(0.96)

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, including between 4,121 and and 5,423 observations (and a subset of 1,118 for female empowerment). Column 3 presents the average spillover effect on eligible households in control villages as well as ineligible households, coming from a stacked spatial regression of each outcome on the amount transferred per capita GDP to each 2km radii band around each household (instrumented by the share of eligibles assigned to treatment in each buffer). We have between 4,048 and 5,309 observations (and a subset of 978 for female empowerment). The reported average effect comes from a population-share-weighted average effect experienced by those two groups, and is representative of the average untreated household. The number of radii bands included in columns 2 and 3 is chosen, as pre-specified, by minimizing the BIC. Column 4 reports the weighted mean and standard deviations of the outcome variables in low-saturation control villages (across eligible and ineligible households). Each regression is weighted by inverse sampling weights and contains baseline values of the outcome when available. Standard errors are clustered at the sublocation in column 1, and calculated following Conley (1999), 2008 using a uniform kernel out to 10 km in columns 2 and 3. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.