Slack and Economic Development: Structural Pre-analysis Plan*

Dennis Egger Tilman Graff Edward Miguel Felix Samy Soliman
Nachiket Shah Michael Walker

March 22, 2024

Summary

This document pre-specifies a general structural model to be used for understanding slack in economic development in the context of the General Equilibrium Effects (GE) project in western Kenya. The project is a two-level randomized controlled trial of the NGO GiveDirectly's unconditional cash transfer program. Transfers were distributed from 2014-16, with a first set of findings described in Egger et al. (2022). We outline key moments and parameters that will be used for calibrating the model, as well as our current thinking on criteria for evaluating model fit and objectives for counterfactual analyses. We note that pre-specifying such analyses in macroeconomics is currently not standard practice, and we may need to adjust the approach described here in response to initial findings. We also anticipate conducting analyses beyond those pre-specified here. This document is not meant to preclude model adjustments, extensions, or additional analyses, but will clarify this aspects of the final analysis were pre-specified and which were conceived of at later points.

*Acknowledgements: This pre-analysis plan is registered as part of AEA Trial Registry 13210. Egger: University of Oxford, dennis.egger@economics.ox.ac.uk; Graff: Harvard University, tgraff@g.harvard.edu; Miguel: University of California, Berkeley, emiguel@berkeley.edu; Samy Soliman: felix.samysoliman@econ.uzh.ch; Shah: University of California, Berkeley, nachi365@berkeley.edu; Walker: University of California, Berkeley, mwwalker@berkeley.edu.

1 Intro and context

The General Equilibrium Effects (GE) project is a randomized controlled trial of an unconditional cash transfer program by the NGO GiveDirectly (GD). In villages selected for treatment, GD transferred around USD 1,000 (nominal) to all eligible households in the village, about 75% of annual expenditure for recipient households. The transfers constituted a shock of about 15% of local annual GDP at the time that they were distributed. Only households with grass-thatched roofs were eligible to receive transfers, a basic means-test for poverty; we find about one-third of households eligible in the study area. The intervention involved over USD 11 million in transfers and 653 villages in Siaya County, Western Kenya.

Treatment assignment was randomized at the village level, and within treatment villages, all households meeting GD's eligibility requirement received the unconditional cash transfer. A second level of randomization provided variation in treatment intensity: sublocations, an administrative unit directly above the village including about ten villages on average, were randomly assigned to high or low saturation status. In high saturation sublocations, two-thirds of villages were treated, while in low saturation sublocations, only one-third of villages were treated. Figure 1 gives an overview of the study area and experimental design.[1]

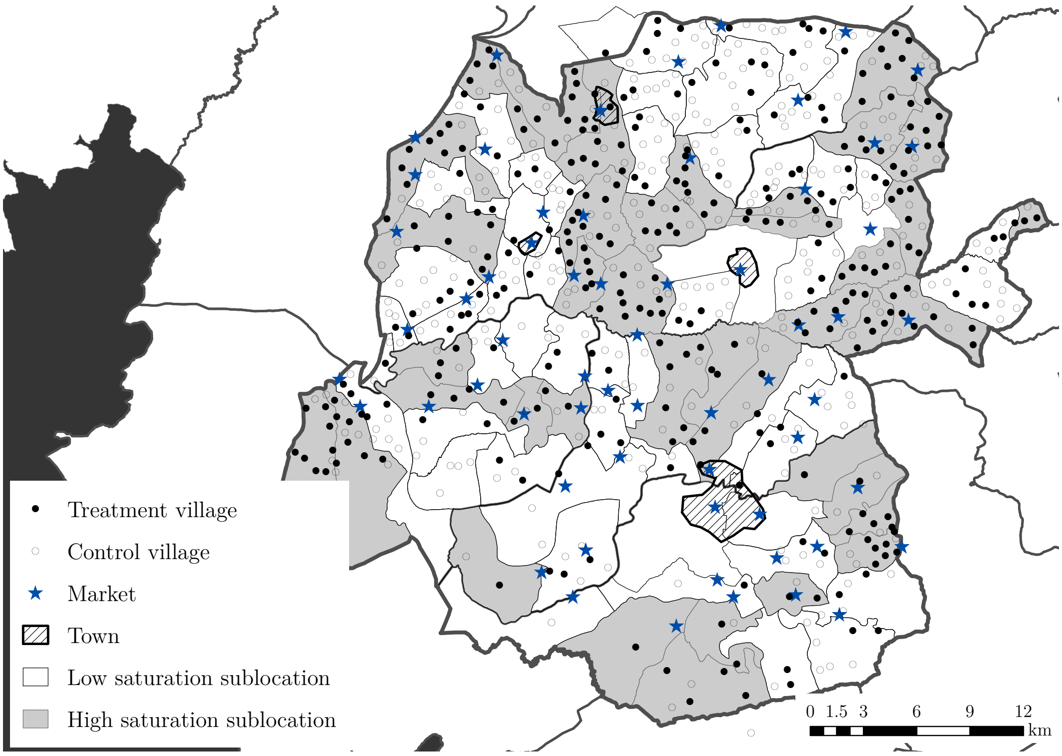

Figure 1: Study area and experimental design (from Egger et al. (2022)

Egger et al. (2022) documents the shorter-run findings of the program through data collected about 1.5 years after the start of transfers. A key finding of the study was a positive local (real) transfer multiplier of 2.5, with minimal overall price inflation (0.2% on average). An open question was how to rationalize the relatively large multiplier. Egger et al. (2022) hypothesized that this may be as a result of slack capacity in the local economy. This argument was based on the fact that there were no documented increases in inputs of production (hours worked or investment), and reasonable assumptions on imports of intermediate goods were also unable to account for the large increase in output.

These findings motivated the authors of the current pre-analysis plan to further investigate "slack" and under-utilization in poor economies, including through detailed measurement of utilization along several dimensions, both in the study area, and urban markets in central Nairobi. Preliminary data analyses show that slack was significantly higher in poor economies, remote and rural economies, small firms, and sectors with indivisible

[1]More details are available in Egger et al. (2022) and a series of previously registered pre-analysis plans for this project on its own AEA Trial Registry AEARCTR-0000505. See `https://www.socialscienceregistry.org/trials/505`.

inputs. Motivated by these new empirical facts and our earlier findings, **this pre-analysis plan seeks to pre-specify a spatial general equilibrium model with input indivisibilities as a key new friction**. Together with the new descriptive facts on the determinants of underutilization, this model will be presented in more detail in the paper “Slack and Economic Development”. the This pre-analysis plan specifies the model developed to date, methods for structurally estimating the model based on data from Egger et al. (2022), and key moments and facts that we intend to target to assess model fit.

We note that pre-specifying structural, spatial and/or macroeconomic models, the method of estimation and calibration of such models, or the measures against which the goodness-of-fit is assessed, is not standard in the current macroeconomic and related literature. Typically, these models have many degrees of freedom, several alternative modeling or functional form assumptions that may be similarly justifiable, and often highly non-linear behavior. Extensive and careful back-and-forth between model and data is therefore required. However, while we fully recognize that we may need to make further adjustments to the model described below, we feel that pre-specifying our current thinking, initially limiting our degrees of freedom at least partially, and clearly defining the targets of this overall exercise may provide greater transparency to the process of model calibration, estimation, and adaptation. It may also contextualize and strengthen the interpretability of goodness-of-fit tests and model validation ex-post. Pre-analysis plans have become widespread in empirical research in applied economics in the last decade (see Miguel (2021) for some discussion of the patterns), and we hope to begin to bring some of the potential benefits of this methodological approach to macroeconomic research.

Throughout this PAP, we list avenues for plausible extensions of the model. These can be understood as promising directions we might take the study if the analysis of the baseline model deems them relevant. If we determine that a particular set of extensions is not relevant, we will not carry them out nor necessarily report them.

1.1 Data analyzed to date

This project builds on existing data collected as part of the GE project. In particular, it utilizes data from Egger et al. (2022), which included baseline (pre-transfer) and follow-up data collected an average of 1.5 years after the start of transfers, as well as from a second round of follow-up data collected in 2019-2022. It also includes a market survey in Nairobi in late 2023 that was modelled after the initial GE data collection. A number of analyses have already been conducted using these data, including from previously-filed pre-analysis plans for the second endline survey Egger et al. (2021a,b).[2]

Importantly, the research team has generated a number of summary statistics to help inform modelling choices and approaches, including for instance a) the documentation of substantial slack (or under-utilized capacity) in enterprises, of more slack in poor, rural, and remote economies, small firms and sectors with indivisible inputs and b) the empirical verification of a relatively linear log on log gravity relationship between commuting time and household shopping patterns.

However, we have not yet calibrated the model using these data. All quantification of the model to date has been on simple “toy” geographies and using rather arbitrary parameter choices. The appendix includes slides from a recent conference presentation that summarizes the current status of the project’s analysis. The entire research team commits to not estimating the model in advance of filing this pre-analysis plan.

[2]These pre-analysis plans are available via the GE study registry AEARCTR-0000505, `https://www.socialscienceregistry.org/trials/505`.

2 Model

2.1 The baseline model

2.1.1 Demand

There are two sectors, traded agriculture and non-traded manufacturing / services.[3]. Consumers have Cobb-Douglas preferences over agriculture X and services Y:

$$U = X^{1-\alpha} Y^{\alpha}$$

The term Y itself is an aggregate of differentiated goods and services $y(\omega)$:

$$Y = \left(\int_{\omega} y(\omega)^{\frac{\theta-1}{\theta}} d\omega \right)^{\frac{\theta}{\theta-1}}$$

Demand for each variety ω is

$$D(\omega) = I_Y P^{\theta-1} p(\omega)^{-\theta} \equiv \zeta p(\omega)^{-\theta}$$

where I_Y is the total budget spent on manufacturing and services and $P \equiv \left(\int_{\omega} p(\omega)^{1-\theta} d\omega \right)^{\frac{1}{1-\theta}}$ is the ideal price index of the non-agricultural sector. The cost of living index for the aggregate economy is

$$P_U = \alpha^{-\alpha} (1-\alpha)^{-(1-\alpha)} P^{\alpha} P_x^{1-\alpha}.$$

Unknown parameters to be estimated from data: θ and α.

2.1.2 Supply

Agriculture. Agricultural products X are produced using a decreasing returns to scale production function using labor N:

$$X = \tilde{A} N^{\beta}$$

with $\beta < 1$. $\tilde{A}$ represents composite technology and will be parameterised using a "true" technology parameter A and land Λ, so that $\tilde{A} \equiv A \Lambda^{1-\beta}$. Agricultural produce is freely traded with the rest of the world at the international numeraire price of $p_X = 1$.

Unknown parameters to be estimated from data: Λ, A, and β.

Non-agricultural sector. Each non-agricultural product variety $y(\omega)$ is produced with labor input n and effort e, as well as firm-specific productivity $\varphi(\omega)$:[4]

$$y(\omega) = \varphi(\omega) \min\{e, n\}, \qquad n \in \mathbb{N}$$

This value-added production function is subject to *indivisibility constraints*, since n can only be hired in full integer units. Effort e is allowed to float. Labor input n costs a wage w (which is endogeneous) and effort is renumerated in fixed units ν (which we estimate).[5]

Non-agricultural firms compete in monopolistic competition and maximise profits:

[3]We note that much of manufacturing in this economy is on-demand, looking relatively similar to services in terms of cost structure. For instance, the largest manufacturing business in terms of revenue is grain milling, typically provided by a mill located within the village, and attended by one employee who operates the mill on behalf of local customers.

[4]In services, the predominant input is labor, and e represents labor effort. In manufacturing, the indivisible input may also be thought of as capital/equipment, and e may then reflect variable costs, such as fuel or electricity.

[5]This wage structure matches labor contract in our setting, which predominantly take the form of a fixed wage plus a piece rate for each unit of output sold.

1. **Unconstrained firms.** If feasible for a given level of n, an *unconstrained* firm with productivity φ will charge the monopolistic price p_u, sell quantity y_u make profits π_u:

$$\begin{aligned} p_u(\varphi, n) &= \frac{\theta}{\theta-1}\frac{\nu}{\varphi} \\ y_u(\varphi, n) &= \zeta\left(\frac{\theta}{\theta-1}\frac{\nu}{\varphi}\right)^{-\theta} \\ \pi_u(\varphi, n) &= \zeta\frac{\theta^{-\theta}}{(\theta-1)^{1-\theta}}\left(\frac{\nu}{\varphi}\right)^{1-\theta} - wn \end{aligned}$$

but this is only feasible if $y_u(\varphi) \leq \varphi n$ or

$$\zeta\left(\frac{\theta}{\theta-1}\nu\right)^{-\theta}\varphi^{\theta-1} < n$$

2. **Constrained firms.** if $y_u(\varphi) > \varphi n$, the firm becomes *constrained* and has the following profit maximising prices, quantities, and profits:

$$\begin{aligned} p_c(\varphi, n) &= \zeta^{1/\theta}(\varphi n)^{-1/\theta} \\ y_c(\varphi, n) &= \varphi n \\ \pi_c(\varphi, n) &= \zeta^{1/\theta}(\varphi n)^{\frac{\theta-1}{\theta}} - (\nu + w)n \end{aligned}$$

Define the quantities p, y, π as the unconstrained (constrained) expressions above, depending on whether the firm is constrained or not.

3. **Cutoffs.** we can compute the *full capacity cutoff* φ_n^f as the point where $y_u(\varphi_n^f) = \varphi_n^f n$ and the *upgrade cutoff* φ_{n+1}^u as the point where $\pi(\varphi_{n+1}^u, n+1) \geq \pi(\varphi_{n+1}^u, n)$. The cutoff for entry is the lower bound of the productivity distribution. This follows because

$$\pi(\varphi) = \max_n \pi(\varphi, n) > 0 \;\forall\varphi.$$

All firms have the option not to to hire additional workers. These firms are owner operated and face no further fixed cost of hiring, guaranteeing positive profits.

4. **Aggregation.** We assume there is a mass M of firms who differ in their productivity φ. In particular, φ is drawn from a Pareto distribution with scale q and shape η:

$$\begin{aligned} G(\varphi) &= \begin{cases} 1 - \left(\frac{q}{\varphi}\right)^{\eta} & \varphi \geq q, \\ 0 & \text{else.} \end{cases} \\ g(\varphi) &= q^{\eta}\varphi^{-\eta-1} \qquad \text{if } \varphi \geq q \text{ and 0 otherwise.} \end{aligned}$$

Then, we can express aggregate labor demand

$$N_Y = M\sum_{n\in\mathbb{N}} n\left[G(\varphi_{n+1}^u) - G(\varphi_n^u)\right]$$

and the aggregate price index

$$P^{1-\theta} = M\sum_{n\in\mathbb{N}}\left[\int_{\varphi_n^u}^{\varphi_n^f} p_u(\varphi, n)^{1-\theta}g(\varphi)d\varphi + \int_{\varphi_n^f}^{\varphi_{n+1}^u} p_c(\varphi, n)^{1-\theta}g(\varphi)d\varphi\right]$$

and aggregate revenue

$$R = P^{1-\theta}$$

and expected profits

$$\mathbb{E}\,\pi = \sum_{n\in\mathbb{N}} \left[\int_{\varphi_n^u}^{\varphi_n^f} \pi_u(\varphi, n) g(\varphi) d\varphi + \int_{\varphi_n^f}^{\varphi_{n+1}^u} \pi_c(\varphi, n) g(\varphi) d\varphi \right]$$

Entry. Households can freely choose between wage employment and entrepreneurship. We propose two different ways to model entry into entrepreneurship:

1. In the baseline, non-frictional case, the payoff to wage employment needs to exactly equal the expected profit from agriculture, i.e.

$$\mathbb{E}\,\pi = w.$$

2. In the second formulation, which we call frictional entry, the payoff to wage employment includes an idiosyncratic preference shock $w + \epsilon$ where $\epsilon \sim N(\mu_\epsilon, \sigma_\epsilon)$. Agents enter as entrepreneurs whenever

$$\mathbb{E}\pi > w + \epsilon.$$

Given a continuum of households L, the mass of entrepreneurs is

$$M = \mathbb{P}(\epsilon < \mathbb{E}\pi - w)L.$$

Upon entry, these entrepreneurs get a productivity draw $\varphi \sim G$. Regardless of their productivity draw, all entrants become active in the market and earn $\pi(\varphi)$. This follows because firms with $n = 1$ are owner operated, guaranteeing $\pi(\varphi) > 0\ \forall \varphi$.

Equilibrium. In equilibrium

- wages in the agricultural and services sector must be equalised
- Consumers maximize utility
- Households enter as entrepreneurs if it is beneficial to them
- the labor market clears

We start off with four (4) endogeneous variables, ζ, w, P, and M. We note, that we can express ζ immediately as a function of the other variables:

$$\zeta = \frac{\alpha I}{P^{1-\theta}} = \frac{\frac{\alpha}{1-\alpha}\left[\tilde{A}N_X^\beta + \Delta\right]}{P^{1-\theta}}$$

where Δ is a transfer from outside the economy, and N_X is the labor demand of the agricultural sector. This we get from the wage equalisation condition:

$$\beta\tilde{A}N_X^{\beta-1} = w \implies N_X = \left(\frac{\beta\tilde{A}}{w}\right)^{\frac{1}{1-\beta}}$$

so that we are left with three equilibrium conditions:

$$N = N_X + N_Y$$
$$P = \left[M \sum_{n \in \mathbb{N}} \left[\int_{\varphi_n^u}^{\varphi_n^f} p_u(\varphi, n)^{1-\theta} g(\varphi) d\varphi + \int_{\varphi_n^f}^{\varphi_{n+1}^u} p_c(\varphi, n)^{1-\theta} g(\varphi) d\varphi \right] \right]^{1-\theta}$$
$$w = \mathbb{E}\, \pi$$

In the case with frictional entry, we replace the last equilibrium condition with

$$M = \Phi_{\mu_\epsilon, \sigma_\epsilon}(\mathbb{E}\, \pi - w) N$$

i.e., the mass of entrants is the entry probability times total population. Unknown parameters to be estimated from data: ν, q, η, Δ, and N.

2.2 Spatial Economy

We can embed the above outlined production structure into a multi-location spatial economy. Suppose there are L locations indexed by o, d. In each location, N_o homogeneous people live and work, but they must travel to different locations to consume (which we call "shopping"). In the data, we can leverage the fact that we observe shopping patterns (i.e. expenditures by product by destination location for each origin location) at different points in time, based on the random roll-out of household surveys. We can match these observations with our contemporaneous price surveys to recreate price menus every household faces at the time of the household survey. In every period t, every consumer in a location has the same expenditure I_{ot}. Since non-agricultural goods and services are not traded, consumers need to travel to shop to different markets around them. Agriculture X is freely traded and does not require travelling. In particular, we follow Ahlfeldt et al. (2015) and assume that a consumer c in location o who shops in location d in period t receives indirect utility

$$u_{codt} = \frac{I_{ot}}{P_{dt}} \frac{z_{codt}}{D_{od}}$$

where I_{ot}/P_{dt} is real consumption, $D_{od} = \exp(\kappa \tau_{od})$ is disutility from travel (which does not change between periods), τ_{od} is travel time in minutes between o and d, and $z_{codt} \sim \text{Frechet}(\sigma)$ is an idiosyncratic shock.

Through the usual Frechet algebra, we can express *expenditure shares*, i.e. the share of total spending from location o that occurs at location d in period t as

$$\pi_{odt} = \frac{(P_{dt} D_{od})^{-\sigma}}{\sum_{d'} (P_{d't} D_{od'})^{-\sigma}} \tag{1}$$

In *spatial equilibrium*, all the above equilibrium conditions need to be satisfied for all locations, plus consumers' shopping patterns comply with equation (1). Importantly, there is no labor mobility or migration, as the paper by Egger et al. (2022) does not report major migration or labor market movements, at least in the short to medium run (which is our focus).

Unknown parameters to be estimated from data: κ, σ.

2.3 Possible extensions

- Assuming a Pareto distribution of productivity is standard and leads to nice (semi) closed-form expressions for wages, etc. However, it also implies a fat right tail, ie. there are lots of *very* productive firms. This could become a problem in estimation, as e.g. a case with $\eta << \theta$ leads to infinite labor demand. For this reason, and to improve model fit, we may resort to a different productivity distribution, such as the log-normal distribution. The estimation strategy would be the same as below, adapted to the necessary parameters of the distribution.

- Multiple non-agricultural sectors: Our rich data environment allows us to differentiate further within the non-agricultural sector. Our enterprise surveys are representative of all sectors in the economy, our household surveys collect shopping patterns between all origins and destinations at the product level. Our price surveys focus primarily on commodities but allow some further differentiation within non-agricultural products. We may therefore extend the model to include different sectors within the non-agricultural composite Y, assuming a Cobb-Douglas functional form over consumption of different sectors.
- Intermediate inputs: The model above is specified in value-added fashion. Since, in practice, a large share of intermediate inputs are imported and represent a form of leakage of spending out of the study economy, we may decide to explicitly model imported intermediate inputs. Currently, such flows would be captured in terms of the agricultural good X. On the data side, we collect intermediate goods expenditure by origin (i.e. imported vs. local) in the follow-up surveys in 2019, which we can use to inform such modelling choices and calibration/estimation.
- Factor mobility: While we shut down factor mobility at this point (based on the fact that no mobility was detected in Egger et al. (2022)), we may in the future introduce such mobility. We think of this as providing insight into short-run (no mobility) vs. long-run dynamics in the spatial economy, and may come back to test some of these differences using longer-term follow-up data.
- Lastly, we might investigate longer-run dynamics of entry, capital, and investment in an economy characterised by slack, though this may be beyond the scope of the current paper.

3 Quantification of the environment

We quantify the study environment in Western Kenya in two ways:

1. for estimation, we calibrate the *aggregate economy* (i.e., treat our study region, comprised of 653 villages and 61 markets, as one single location with aggregate characteristics)
2. to evaluate impacts of cash transfers in our counterfactuals, we calibrate the *spatial economy* (i.e., treat 61 markets and their catchment areas as individual locations in spatial equilibrium)

3.1 Aggregate economy

We calibrate the aggregate economy with the following characteristics:

- **Total population** N: we normalise this to 1
- **Total population working in agriculture** N_X: the share of total hours worked in agriculture from representative household surveys at baseline.
- **Total baseline inflows from abroad** Δ_0: average household level net remittance inflows from outside the study area collected in our baseline household surveys
- **Total baseline ag production** X: agricultural yields from representative baseline household surveys
- **Total agricultural land** Λ **per capita**: data on agricultural land holdings from representative household surveys at baseline

3.2 Spatial economy

We further divide the study region into 61 market locations, each indexed by i and with the following characteristics

Parameter	What is it	How to get it	Moment if SMM / value if calibrated
α	Cobb Douglas share of ag consumption	read from the data	0.606
β	DRS parameter for ag production	take from literature	0.49
η	Shape of productivity Pareto	estimate via SMM	revenue share of top quintile
ν	Marginal cost of effort	estimate via SMM	share of firms reporting $mc = 0$
θ	CES elasticity within locations	estimate via SMM	variable profit share
σ	Spatial gravity parameter	estimate directly	6.25
κ	Spatial sensitivity to travel times	estimate directly	0.05
Λ	Land in the economy	normalise	1
A	Agricultural technology	match data	total ag. production
q	Scale of the Pareto distribution	normalise	1

Table 1: **Parameter overview and plan**

- **Location working population** N_i: the shares of total study area population working in each location.[6]
- **Location baseline inflows from outside the study area** $\Delta_{0,i}$: Average net remittances received by each location (from all other locations j in the study area, and from outside the study area) from our baseline household surveys.
- **Location agricultural land** Λ_i: shares of total agricultural land for each location from baseline household surveys
- **Profit matrix** Π: this apportions what percentage of profits made in location i are made by owners in location j. To calculate this, we use our detailed owner-household matched data structure – each enterprise censused at our follow-up surveys is matched in real time to the contemporaneous household census.
- **Cash transfer to each region** Δ_i: this is the experimental cash transfer sent to each location by GiveDirectly.
- There are several market centers around or within the subcounties (often in towns) in which we work, but that are not a part of the original study. For these, we need at least the following information to estimate the model: (i) the percentage of firms owned by people that are within that town (i.e. outside of the study); (ii) the percentage of workers that work in the town that are not a part of the study area; (iii) disposable income of people in those towns. In the absence of additional data, we plan to proxy for (i) and (ii) with analogous shares from villages/market centers within our study area. For (iii), we intend to explore options of a) multiplying mean estimates of disposable income from households in our study (potentially conditioned to look more like those in towns) by estimated population within each town, and b) estimating the relationship between population density and disposable income within our study area.

4 Estimation / Calibration routine

We follow a variety of strategies to estimate and calibrate the parameters in our model. Table 1 presents an overview. We provide more detail below:

4.1 Reduced form parameters

Some parameters we can read directly off the data or related work:

1. Cobb-Douglas Share of agricultural consumption α is the budget share on tradeable agriculture from the consumption surveys. Importantly, this includes expenditures on the "service content" of food expenditures. We get this by combining direct survey evidence on expenditures on non-tradeable, non-food consumption with expenditure on food multiplied by the cost share of food sellers on everything that isn't

[6]Note: the vast majority of people work at their residence location. We will, however, be able to account for any cross-location flows using location data from the employment module of household surveys

intermediate inputs. In particular, the total budget share on non-food is 0.32 of which 81% is locally-produced, and the share of intermediate inputs in local food stands at 0.49. This leaves us with $\alpha = 0.32 * 0.81 + 0.68 * (1 - 0.49) = 0.606$

2. DRS parameter for ag production β. We plan to take a reasonable value from the literature. Porting the production function estimates from Adamopoulos and Restuccia (2020) to our setting, we get $\beta = 0.49$.

4.2 Structural parameters

We conduct a series of estimation steps to arrive at our structural parameters.

Spatial shopping elasticities σ and κ. Taking logs of equation (1) and including origin, destination and time fixed effects yields,

$$\log \pi_{odt} = -\sigma \log P_{dt} - \sigma\kappa\tau_{od} + f_o + f_d + f_t$$

where f_o, f_d, f_t are origin, destination, and time fixed effects. In this regression, we can separately identify the sensitivity to price changes σ, as well as the sensitivity to travel time κ. We can use our data on shopping patterns across space, matched to panel data on prices to create price menus across time and space. Since the travel cost τ_{od} is not assumed to vary across time, we can use our panel data to separately identify κ and σ. This approach is related to the estimation strategy used in Ahlfeldt et al. (2015).[7]

When we run a preliminary version of this estimation using PPML in order to determine the feasibility of this approach, we get $\sigma = 6.25$ and $\sigma\kappa = 0.314$, implying $\kappa = 0.05$. We get similar values when we estimate it using OLS and excluding zero-flows. For robustness, we may vary the set of fixed effects we include to estimate these parameters.

Implied parameters. Agricultural productivity can be backed out through the production function

$$Y_X = A\Lambda^{1-\beta} N_X^{\beta}$$
$$Y_X/N = A(\Lambda/N)^{1-\beta}(N_X/N)^{\beta}$$

this pins down A as prices of agricultural goods are fixed in our setting.

Structural parameters ν, θ, η. We plan to conduct simulated method of moments estimation to obtain structural estimates for the elasticity of substitution θ, the variable cost of effort ν, and the shape of the productivity Pareto distribution η. To do so, we plan to match four main data moments to inform the three parameters. At the time of writing this PAP, we have experimented with the following moment-parameter pairings:

- The "variable profit-share", (i.e., revenues minus variable payments excluding wages, divided by revenues). Intuition: Standard monopolistic competition arguments imply that the ratio of variable profits to revenues is proportional to the elasticity of substitution $\frac{\pi^V}{R} = \frac{1}{\theta}$.
- The share of firms below full capacity (ie. reporting $mc = 0$). Intuition: The level of the utilisation rate informs us about the cost of utilisation ν.
- Revenue shares of each quintile of the revenue size distribution within each sector are informative about the variance of firm productivity and therefore map to η.

The intuition for the mapping from data to parameter is not a formal identification argument. All relationships are jointly estimated. We also reserve the possibility to switch targeted moments in case the above do not imply identification, or our estimates turn out to be excessively sensitive to the chosen moments (Andrews et al., 2017).

[7] They do not use panel data, but instead use the variation of equilibrium wages to identify σ.

4.3 Possible extensions

1. We specify 61 marketplaces as locations. In reality, there are 653 rural villages (part of our study area) and 4 peri-urban towns (not part of our study area, but hosting a market). We can (with considerable computational effort) also run our spatial model on the basis of these locations.

2. Similarly, we might at some point estimate the "step size" of the indivisibility constraint. Currently, this is integers, so the step size is 1.0. But we might return to it and estimate this in a later iteration.

3. We also want to check robustness of our chosen estimation moments. In particular, we want to try how robust our estimation is to using

   - the overall utilisation rate (instead of the share of firms with $mc = 0$)
   - different percentiles of the revenue distribution (other than just the top 20%)

5 Counterfactual

To assess the fit of the model, we plan to compare the model to the following untargeted moments in the data. We view these as the key moments / targets of interest and the model's success in matching them is a test of the hypotheses we set out when we began working on this model. **We view this as a core part of this pre-specification exercise, as it allows us to transparently assess the "success" of the pre-specified model, without selectively choosing what such success looks like ex-post.**

- The relationship between firm size and utilization rates: The empirical counterparts we will compare the model to are 1) the average utilisation in owner-operated enterprises versus those with hired workers, 2) the correlation between the utilisation rate and firm revenues.

- The correlation between market access and slack.

- We will replicate the Egger et al. (2022) as closely as possible within the model structure and generate, within the model, the following untargeted moments, which we then compare to their causally estimated empirical counterparts from the large-scale RCT:

  1. the nominal and real multiplier reported in Egger et al. (2022). For this, we will replicate the treatment effect regressions from Egger et al. within the model. This will allow us to compare model implied treatment effects to those in the data.
  2. the inflation rate reported in Egger et al., where we again replicate the empirical specifications on model-generated data. If possible, we may also split effects between tradeables and non-tradeables, though we view this as more exploratory.
  3. the wage inflation rate reported in Egger et al.
  4. the structural transformation – the share of hours worked in agriculture vs. non-agricultural sectors – as reported in Egger et al.
  5. treatment effects reported in Egger et al. on household expenditures, enterprise revenue, enterprise profits.
  6. We may additionally conduct further empirical analyses to validate the model, including investigating empirically and within the model: a) the differential impact of cash transfers on sectors/economies based on their average level of utilization, and b) the differential impact of cash transfers on inflation by market-access.

- Lastly, we will run additional counterfactuals without empirical analogues. We view these as interesting for economic policy:

1. the same cash transfer program (same size and same size as % of GDP) to households in urban Nairobi (where market access is far higher), contrasting the size of the multiplier (both real and nominal) there versus the original rural study region
2. double the size of the cash transfer in Western Kenya
   - we can also trace out the size of the multiplier (both real and nominal) as the size of the transfer increases
3. a more targeted cash transfer, which goes to the highest-slack regions
   - we view this as contributing to the literature and debate on place-based policies and the desirability of targeting of cash transfers at the regional level.

References

Adamopoulos, Tasso and Diego Restuccia (**2020**). Land Reform and Productivity: A Quantitative Analysis with Micro Data. *American Economic Journal: Macroeconomics* 12(3), pp. 1–39

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf (**2015**). The Economics of Density: Evidence From the Berlin Wall. *Econometrica* 83(6), pp. 2127–2189

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro (**2017**). Measuring the Sensitivity of Parameter Estimates to Estimation Moments. *The Quarterly Journal of Economics* 132(4), pp. 1553–1592

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker (**2022**). General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya. *Econometrica* 90(6), pp. 2603–2643

Egger, Dennis, Johannes Haushofer, Edward Miguel, and Michael Walker (**2021**a). GE Effects of Cash Transfers: Pre-analysis plan for Endline 2 Household. `https://www.socialscienceregistry.org/trials/505`

——— (**2021**b). GE Effects of Cash Transfers: Pre-analysis plan for Endline 2 Local Economy Analyses

Miguel, Edward (**2021**). Evidence on research transparency in economics. *Journal of Economic Perspectives* 35(3), pp. 193–214