Job Search in Developing Countries:
Crowd-In and Crowd-Out Externalities

Mitchell VanVuren (Yale)

June 2024

Please click here for the latest version

Abstract

Productive wage work is often difficult to find in developing countries. Many policies aim at assisting searchers and expanding the wage sector, but the rationale for intervening is unclear. This paper develops a search-and-matching model that incorporates key features of developing economies including a large self-employment sector, savings-constrained households, and capital-constrained firms. Four search externalities — two positive and two negative — emerge, leading to inefficiency. After estimating the model using an experiment that provided search subsidies to job seekers in Ethiopia, I find that the optimal policy is a *tax* that roughly doubles the cost of search, rather than a subsidy.

For their helpful comments, I thank David Argente, Kevin Donovan, Joe Kaboski, Guiseppe Moscarini, Joseph Mullins, David Lagakos, Remy Levin, Alessandra Peter, Tommaso Porzio, Valerie Ramey, Martin Rotemburg, Stephen Ross, Todd Schoellman, Jedediah Silver, Daniela Vidart, Johannes Weiland, and various seminar participants at the NEUDC, Midwest Macro, Y-RISE Annual Conference, UCSD, University of Connecticut, and Yale. Any errors remain my own. Mitchell VanVuren: Yale Research Initiative on Innovation and Scale (Y-RISE) email: mitchell.vanvuren@yale.edu

1. Introduction

In developing countries, reliable wage-sector jobs are often difficult to find, and many individuals spend months or even years alternating between self (or marginal) employment and job search before finally finding long-term wage work (Poschke 2019, Donovan, Lu & Schoellman 2023). This is commonly understood to give rise to "subsistence self-employment" (Schoar 2010, Duflo & Banerjee 2011, Herreño & Ocampo 2021) in which individuals who would prefer to work a wage job, could they find one, instead engage in less productive self-employment. Consequently, there has been substantial interest in the impacts and effectiveness of policies aimed towards helping these individuals move into wage work, including subsidies to labor search, transport, and (temporarily) wages (e.g. Levinsohn et al. 2014, Franklin 2018, De Mel et al. 2019, Abebe et al. 2021, and many others).

Frictional search models are the standard tools for understanding such movements across labor market states. Workhorse search models generate inefficiencies but lack many features central to developing economies, including large shares of self-employment and substantial credit market frictions (which give rise to subsistence self-employment). Building and evaluating labor market policies for developing countries requires understanding how the inefficiencies arising from labor market frictions manifest in such an environment.

This paper develops a model of frictional labor markets, incorporating these key characteristics of developing countries, and studies the externalities that arise in this setting. Individuals have access to a self-employment option for subsistence and are savings-constrained, limiting their ability to fund job search. Entrepreneurs run firms and face financial constraints that restrict their growth and distort the allocation of resources, and the two interact through a labor market exhibiting canonical Diamond-Mortensen-Pissarides search-and-matching frictions, which generates externalities from search.

Individuals in the model desire higher-paying wage jobs but must pay a search cost (e.g. commuting) and give up a period of income in order to search. Because they face idiosyncratic job-finding risk, only sufficiently self-insured individuals will choose to search while others will opt for the guaranteed but lower income of self-employment (as in Feng, Lagakos & Rauch 2018, Herreño & Ocampo 2021). The model thus reproduces the empirical fact that individuals frequently and stochastically shift between self employment and job search before finding wage work.

Entrepreneurs operate a constant returns to scale production technology and consequently desire to be large but are restricted in size by a collateral constraint that prevents them from financing capital beyond some multiple of their wealth (as in Itskhoki & Moll 2019, Buera, Kaboski & Shin 2021). They hire workers by posting vacancies, but any funds spent paying vacancy posting costs are funds that can no longer be used as collateral in the future. Thus labor market frictions act as a constraint to firm growth.

Despite the complexities of borrowing-constrained individuals, credit-constrained firms, and frictional labor markets, substantial insight into the externalities stemming from labor market frictions can be gained analytically. Motivated by the emphasis within the development literature on worker-side interventions (often called "Active Labor Market Policies"), I consider the problem of a dynamic Ramsey planner who maximizes average worker utility subject to the constraint that it must respect individual and entrepreneur budget and credit constraints, must respect the matching technology, and cannot dictate the behavior of entrepreneurs (as in e.g. Itskhoki & Moll 2019). Examining the planner's problem reveals four search externalities that cause the individually optimal search decision rule to differ from the planner's.

The first externality, Congestion, corresponds to the well-known externality of the same name in typical models. An individual who searches exerts downward pressure on labor market tightness, lowers the job finding probability, and crowds out other searchers (as those on the margin now opt for self-employment instead). As a result, the planner desires a lower level of search than arises in the competitive equilibrium.

Unlike in typical models, however, this negative externality is offset by a positive Firm Size externality in which an individual's search also crowds in additional workers. This externality arises due to the fact that entrepreneur size is limited by a collateral constraint. When an individual searches and finds a job, the resulting output is split between worker and entrepreneur via bargaining. The entrepreneur then uses a portion of this output as collateral to finance further expansion of the firm, including posting additional vacancies and hiring workers. Thus, in the longer run, the individual's search decision directly increases future employment, the value of which is not reflected in the individual's wage, and the planner desires a higher level of search.

The second positive externality, the Allocative Efficiency externality, is somewhat more involved. As search pushes labor market tightness down, hiring costs fall and entrepreneurs grow their firms faster, as some of these cost savings are used to finance growth. Further, this increase in growth turns out to be larger for more productive entrepreneurs. Because they wish to grow faster (than less productive entrepreneurs), hiring costs make up a larger share of their total costs. Thus, a decline in hiring costs frees up (proportionally) more resources for growth. As a result of their faster relative growth, the share of capital and labor allocated to the productive entrepreneurs increases over time, raising allocative efficiency, Total Factor Productivity, and average wages, resulting in a positive search externality.

The final externality, Capital Shallowing, is a variant of the so-called monopolist externality that commonly arises in models that treat workers and entrepreneurs as distinct agents (e.g. Itskhoki & Moll 2019). Here, it arises due to the fact that downward pressure on labor market tightness (from search) lowers entrepreneurs' cost of hiring and, decreasing the cost of labor relative to capital and lowering the capital-labor ratio. Although such a change is not typically considered an externality, the planner (who values only worker welfare) wishes to act as a monopolist on the behalf of workers and thus views this change as a negative externality as it leads to lower wages.

After identifying the four externalities, I turn to the question of optimal policy. Interestingly, despite having access to a wide variety of complex tax instruments that condition on household heterogeneity, the planner's optimal solution balancing the externalities can be implemented using only a single tax (or subsidy) on search (along with lump-sum transfers to restore budgets). In particular, the optimal policy does not involve a subsidy to savings, implying that individuals' inability to borrow does not interact with labor market frictions to exert additional externalities from savings. Intuition might suggest that a planner who wants to induce individuals to search more will also want to make individuals save more in order to fund this search; however, this turns out not to be the case. Conditional on a subsidy that fully internalizes the externalities, individuals' consumption-savings decisions align with those of the planner, highlighting the fact that it is not households' inability to borrow *per se* that justifies policy intervention.

In order to quantify the four externalities and compute the optimal policy, I estimate the model using simulated method of moments to match search behavior from weekly data collected as part of an experimental evaluation of a labor search subsidy in Ethiopia (Abebe et al. 2021). The model is estimated exclusively using data on control individuals (those not receiving a subsidy) while the outcomes of treated individuals receiving the subsidy are reserved for model validation. The model passes this validation check — while the subsidy is offered, treated individuals are about 5 percent more likely to search for work in both the data and model.

Surprisingly, the optimal subsidy in the estimated model is equal to -101 of total search costs — that is, the optimal policy is actually a tax on search that roughly doubles the search cost (an increase equal to about 20 percent of average self-employment earnings), indicating that the negative externalities from search (Congestion and Capital Shallowing) outweigh the positive externalities (Size and Allocative Efficiency). The gains from implementing the optimal policy are substantial — on average, welfare increases by 1.5 percent of consumption. In order to decompose the contribution of each externality to this overall gain, I start from optimal equilibrium and consider counterfactual equilibriums in which the size of each externality is moved marginally closer to its size in the competitive equilibrium,[1] in essence, removing the impact of the externality. The difference in welfare between the optimal equilibrium and one of these counterfactuals then quantifies the welfare impact of the corresponding externality.

Decomposing the overall welfare gains into the individual contribution of each externality (via a simple decomposition exercise) reveals two main takeaways. First, the externalities that influence employment are quantitatively much larger than those that influence wages. The Congestion and Firm Size externalities account for 1.6 and -0.8 percent of the welfare gains respectively, while Allocative Efficiency and Capital Shallowing account for only -0.2 and 0.2 percent.[2] The second, evident from the same results, is that the negative (crowd-out) externalities dominate the positive (crowd-in), and, in particular, the large size of the Congestion externality relative to the Firm Size externality is the primary reason that the optimal policy takes the form of a tax rather than a subsidy.

The final section of this paper pivots from normative to positive analysis and uses the estimated model to examine the aggregate impact of a subsidy to search. The purpose of this section is to highlight the importance of accounting for these

[1] Using the marginal, rather than the total, change is necessary as equilibria may become degenerate when the impact of a single externality is removed completely.

[2] Due to the nature of the decomposition exercise, the sum of these individual impacts is not necessarily equal to the overall welfare impact of 1.5 percent.

externalities, and their resulting general equilibrium impact, in evaluating labor market policies. In partial equilibrium without the externalities, the subsidy increases the size of the wage sector dramatically (from 30 to 47 percent) and leads to higher welfare (1.3 percent of consumption); however, once the externalities are account for in general equilibrium, the policy is half as effective, growing the wage sector to only 40 percent, and carries a substantial welfare cost of -2.5 percent of consumption.

Overall, the surprising conclusion of this paper is that labor markets in developing countries (at least those similar to the market in Ethiopia used to estimate the model) are characterized by workers who search too much rather than too little. Consequently, policies aimed at helping and encouraging workers to search, such as search subsidies, are counterproductive. While they do manage to increase the size of the wage sector and may even yield promising experimental results, they do so fairly ineffectively once externalities are taken into account and carry substantial welfare costs.

Related Literature: This paper is closely related to the macroeconomic development literature studying the impact of entrepreneur-level credit constraints on growth and development such as Buera, Kaboski & Shin (2011), Moll (2014), Itskhoki & Moll (2019), and Buera, Kaboski & Shin (2021). This paper also builds on recent work drawing distinctions between subsistence self-employment and entrepreneurship (such as Feng & Ren 2021) or otherwise studying unemployment in developing countries (such as Feng, Lagakos & Rauch 2018, Poschke 2019). Closely related is Herreño & Ocampo (2021) who use a model in which households use self-employment to cope with the risks of wage employment (the same mechanism as this paper) to study the macroeconomic effects of microloans and cash transfers.

The model dynamics in which workers flow freely between self/marginal employment and labor search before finding a long-term wage job are very similar to those documented in Donovan, Lu & Schoellman (2023). In a similar vein, Banerjee et al. (2021) find that skilled workers in developing countries exhibit higher unemployment rates relative to unskilled workers and show that this difference leads to differences in occupational choice. Porzio, Rossi & Santangelo (2021) use a model with frictional reallocation of labor from (self-employment dominated) agriculture to (wage work dominated) non-agriculture to quantify the importance of human capital in explaining the process of structural change.

This paper is also closely related to the microeconomic literature on Active Labor Market Policies, which are intended to help grow the wage sector. Abebe et al. (2021) and Franklin (2018) both study the effects of cash transfers (the same policy studied in the quantitative portion of this paper). De Mel et al. (2019), Algan et al. (2020), and Alfonsi et al. (2020) all study firm-side interventions (although the last includes an additional worker-side treatment arm) also intended to help workers find jobs. McKenzie (2017) provides an excellent review of this literature, which is too exhaustive to list here.

2. Model

Time is discrete. There is measure one of individuals (workers) and an endogenous measure of entrepreneurs. Households consume, save, and choose between working in self-employment or participating in the labor market. Entrepreneurs operate firms, consume profits, and accumulate capital and labor for future periods.

2.1. Search and Matching Technology

The labor market for wage work exhibits typical search-and-matching frictions. Workers must search for jobs and entrepreneurs must hire by posting vacancies. The cost of searching for a job and the cost of posting a vacancy are denoted by b and c respectively. Each period, the number of worker-firm matches is given by a homogeneous of degree 1 matching function $m(u, v)$ where u is the measure of individuals searching for a job and v is the number of vacancies posted by firms. As is typical, $\theta = \frac{v}{u}$ is defined to be labor market tightness so that $p(\theta) \equiv m(\frac{1}{\theta}, 1) = \frac{m(u,v)}{v}$ is the probability that any vacancy is filled and $\theta p(\theta) = \frac{m(u,v)}{u}$ is the probability that any searcher finds a job. Finally, matches between workers and firms are separated with exogenous probability λ at the end of every period.

2.2. Workers

A unit measure of infinitely-lived workers are indexed by their wealth a, their employment status e, and their self-employment productivity y. Their lifetime utility is given by

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\sigma}}{1-\sigma} \tag{1}$$

Individuals are endowed with one unit of time each period which they supply inelastically and indivisibly to either work or search.[3]

Labor Decisions: Any individual can engage in self-employment and operate the self-employment technology

$$y_t = A_s l_t \tag{2}$$

An individual's self-employment productivity l_t follows an exogenous Markov process described by transition matrix M. For simplicity, I normalize A_s to unity so that self-employment earnings are simply given by $y_t = l_t$. By assumption, self-employment uses only an individual's own labor and does not involve hiring workers from outside the household. Thus, this option most closely corresponds to the concept of "subsistence self-employment".

Instead of engaging in self-employment, an individual can choose to pay a search cost b and search for a wage job. A searcher earns nothing in the current period and finds a permanent job with probability $\theta p(\theta)$. After finding a job and becoming employed, the individual can either work in their wage job or return to self-employment (in equilibrium, all employed workers will choose to engage in wage work). Wages are determined through bargaining (discussed later) and depend on the productivity of the entrepreneur with whom the individual is matched, given by z_t.[4]

Budgets: Workers face incomplete markets a la Aiyagari (1994), Bewley (1977), and Huggett (1993) and accumulate assets for self-insurance. Each period, assets pay an exogenous rate of return r (i.e. this is a small open economy). Individuals cannot borrow (i.e. $a_t \geq 0$). Their budget constraint is then

$$a_{t+1} + c_t = (1+r)a_t + (1-e_t)\big((1-s_t)y_t - s_t b\big) + e_t w_t(z_t) \tag{3}$$

where $s_t \in \{0,1\}$ is a choice variable with $s_t = 1$ corresponding to the decision to search in period t and $e_t \in \{0,1\}$ is an indicator variable with $e_t = 1$ indicating that the individual is employed in period t.

[3]This assumption can be justified by the fact that a model period is one week. Additionally, quantitative experiments with allowing interior choices of time allocation suggest that the optimal policy is fairly close to "bang-bang", with individuals largely choosing to allocate their entire time budget to either work or search rather than a mix of the two, for reasonable parameters.

[4]Section 2.5 shows that the bargained wage depends only on the productivity of the entrepreneur and not on other entrepreneur or individual state variables.

Search: Search is undirected, and every vacancy has an equal probability of being filled. An individual's probability of matching with a job that will pay $w(z)$ (conditional on matching with any job), denoted $H(z;X)$ where X is a vector of aggregate state variables, is given by the share of vacancies posted by z-type entrepreneurs. Although, in principle, w and H depend on all the state variables of both the individual and the entrepreneur to which they are matched, they are written here to depend only on the matched entrepreneur's productivity z. A later section will show this to be the case, justifying this notation.

Employed workers are separated from their jobs with probability λ. Additionally, an individual can lose their job if the entrepreneur employing them dies (probability $1-\Delta$, discussed below) or chooses to downsize its labor force. Under generous parameter conditions (satisfied in the quantified model), it can be shown that downsizing never occurs in equilibrium, which I assume throughout the rest of the paper. Thus the probability that an employed worker retains their job at the end of the period is given by $(1-\tilde{\lambda}) = \Delta(1-\lambda)$.

Bellman Equation: Taking all of the above, the individual's optimization problem can be written recursively as

$$V_u(a,y;X) = \max_{c,a',s\in\{0,1\}} \frac{c^{1-\sigma}}{1-\sigma} + \beta\Bigg(\big(1-s\theta p(\theta)\big)E_{y'}[V_u(a',y';X')|y]+ \\ s\theta p(\theta)\big(E_z[V_e(a',z;X')]\Bigg)$$

$$V_e(a,z;X) = \max_{c,a'} \frac{c^{1-\sigma}}{1-\sigma} + \beta\Bigg((1-\tilde{\lambda})V_e(a',z;X')+ \\ \tilde{\lambda}E_{y'}[V_u(a',y';X')]\Bigg) \tag{4}$$

$$\begin{aligned} s.t.\ \ a'+c =&(1+r)a+(1-s)y-sb \qquad \text{for } V_u \\ a'+c =&(1+r)a+w(z) \qquad \text{for } V_e \\ X' =&G(X) \\ y' \sim& M(y) \\ z \sim& H(z;X) \end{aligned}$$

where X is a vector of aggregate state variables and G is the perception function for the evolution of the aggregate state. V_u and V_e denote the value function of the individual while unemployed and employed respectively. For simplicity, an individual who moves from employment to unemployment draws their self-employment productivity y from the stationary distribution of M.

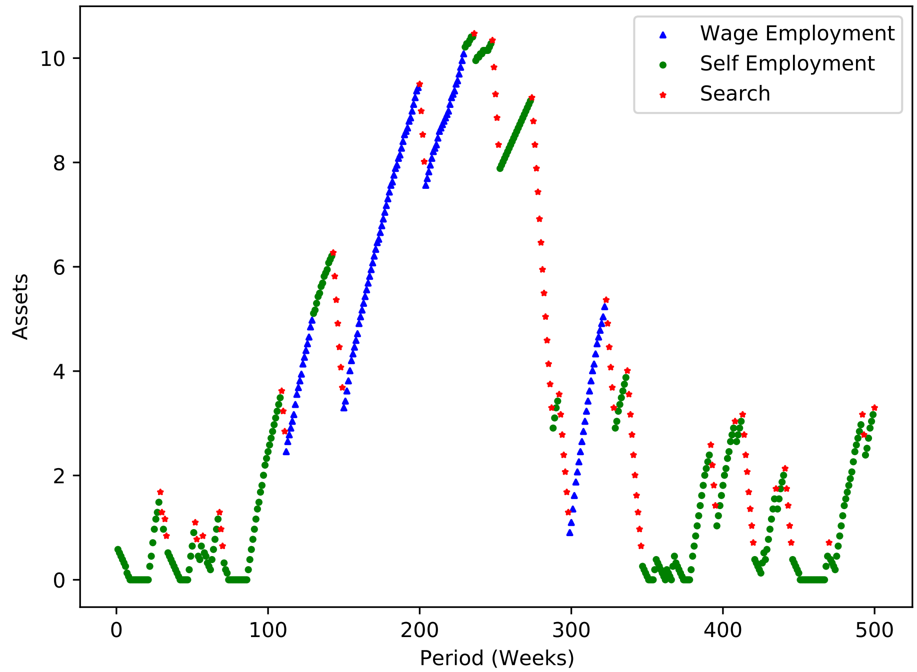
2.3. Worker Behavior

Workers decide whether to engage in self-employment or search for a wage job by weighing the benefits of search against the costs. In addition to the explicit search cost b and the opportunity cost of forgone self-employment earnings, the presence of borrowing constraints means that the higher risk of job search also serves as a cost, particularly if the probability of finding a wage job is small as it is in many developing countries. Thus only individuals who are sufficiently self-insured will opt to pay the search cost and search. Those without much self-insurance will enjoy the safety of lower but guaranteed income in self-employment. For those that search, the search cost quickly diminishes their savings and reduces their self-insurance, eventually driving them to self-employment until they can re-accumulate sufficient self-insurance.

The result is that individuals near the threshold of self-insurance spend a few periods working in self-employment and accumulating assets, then switch to searching for a wage job for a few periods, and return to self-employment once their savings have been depleted. Of course the exact cutoff in savings above which households decide to search depends on their self-employment productivity y_t (which is stochastic), leading to some unpredictability in the exact timing of these switches.

Figure 1 displays an example of this behavior for a single individual simulated for 500 periods (each period corresponds to a week). The x-axis displays time while the y-axis displays the individual's stock of assets. Green points correspond to weeks where the individual is engaging in self-employment, red points represent search, and blue points represent wage work. At the start, the individual is near the threshold of self-insurance and alternates between working in self-employment and searching for wage work depending on their particular level of assets and self-employment productivity. Shortly after period 100, their search is successful, and they acquire a high-earning wage job and quickly accumulate assets. They eventually separate from their employer but use their stock of assets to fund extensive search and remain in the wage sector. This behavior continues for quite some time until approximately week 350 when the individual exhausts their

Figure 1: Worker Self-Employment and Wage Sector Behavior over Time



Note: This figure plots a simulated individual's search, wage work, and self-employment behavior as well as assets over 500 periods. The model used to perform the simulation is based the estimated model described in Section 4 with some features exaggerated to make the underlying behavior clearer.

assets without finding a job and returns to self-employment punctuated by brief periods of search.

2.4. Entrepreneurs

While individuals work in either self-employment or the wage sector, entrepreneurs operate firms and employ workers. Including entrepreneurs as distinct agents (as opposed to an occupational choice for individuals, as in Buera, Kaboski & Shin 2021), reflects the qualitative difference between "subsistence self-employment" (which individuals can flow in and out of fairly freely as in Donovan, Lu & Schoellman 2023) and productive entrepreneurship with the potential to grow and employ many workers, in addition to providing a dramatic increase in tractability.

There are N entrepreneurs each of size $\frac{M}{N}$ born every period, and the model considers the limit $N \to \infty$.[5] At the end of a period, entrepreneurs die with probability

[5]The assumption that there are an infinite number of atomic entrepreneurs rather than a mea-

Δ. Entrepreneurs are born with idiosyncratic ability z drawn from some distribution with bounded support $h(z)$ and an initial level of financial wealth $\underline{f}$ (taken to be exogenous). They discount the future at rate β (the same rate as workers) and receive lifetime utility from consumption (labeled d_t for "dividends") given by

$$\sum_{t=0}^{\infty}(\beta\Delta)^t \frac{d_t^{1-\sigma}}{1-\sigma} \tag{5}$$

Each entrepreneur operates a Cobb-Douglas production technology that depends on their ability:

$$y_t = z k_t^{\alpha} n_t^{1-\alpha} \tag{6}$$

Entrepreneurs rent capital from the international capital market at an exogenous rental cost $(r+\delta)$ and pay workers wage w_t, determined by bargaining, but must use their own assets f_t as collateral to finance capital. Their collateral constraint is given by

$$k_t \leq \gamma f_t \tag{7}$$

where $\gamma \geq 1$ is a parameter summarizing the degree of financial market frictions, with $\gamma = 1$ representing the case of full self-financing and $\gamma \to \infty$ representing no financial frictions.[6]

To hire labor and adjust n_t, entrepreneurs post vacancies v_t. Each vacancy costs c units of output to post and is filled at the end of the period with probability $p(\theta)$. The evolution of n_t is dictated by the equation

$$n_{t+1} = (1-\lambda)n_t + p(\theta)v_t \tag{8}$$

where λ is the exogenous separation rate. Here, it is worth clarifying that while individuals face idiosyncratic risk in job finding and separation, entrepreneurs do not — an entrepreneur with n_t workers can ensure a labor force of precisely n_{t+1} next period by posting $\frac{n_{t+1}-(1-\lambda)n_t}{p(\theta)}$ vacancies.

sure of non-atomic entrepreneurs is not typical but eliminates many technical difficulties in the discussion of wage bargaining. Other than this, there are no substantive differences between the two assumptions.

[6] While this constraint is exogenous, it can be thought of as arising from the unenforceability of contracts or other institutional features that make uncollateralized lending risky and microfounded as such (see e.g. Buera, Kaboski & Shin 2021).

An entrepreneur's period profits are given by

$$\pi_t(z, k_t, n_t) = z k_t^{\alpha} n_t^{1-\alpha} - (r+\delta)k_t - w_t n_t \tag{9}$$

Due to the collateral constraint, an entrepreneur will earn positive profits each period. They split these profits between consumption, posting vacancies, and accumulating additional collateral f_{t+1} and face a budget constraint given by

$$d_t + f_{t+1} = \pi_t(z, k_t, n_t) + f_t - c v_t \tag{10}$$

2.5. Wage Bargaining

Each period, entrepreneurs and their hired workers bargain over wages. Because capital acts as a fixed factor of production (the collateral constraint always binds in equilibrium), firm output exhibits decreasing returns to scale in labor. To accommodate this, I follow Stole & Zwiebel (1996) and model production as a cooperative game between workers and entrepreneurs in which each agent is paid their Shapley value (see Smith 1999, Acemoglu & Hawkins 2014, for other papers using this approach).[7]

The entrepreneur enters the game with capital k and workforce n. Any worker that chooses not to cooperate will engage in self-employment for a period and then return to the bargaining table in the next period (i.e. the outside option is a shirking of duties for a period, rather than termination of the match). Defectors draw their self-employment productivity from the stationary distribution of M, but negotiation occurs before these productivity draws are realized so that workers are treated symmetrically.

If the entrepreneur and x of their n workers choose to cooperate, they form a coalition, operate the entrepreneur's production technology, and produce $zk^{\alpha}x^{1-\alpha}$. The remaining $n-x$ workers form their own coalition and produce $(n-x)\bar{y}$ (where $\bar{y}$ is average self-employment productivity). Each agent is paid their Shapley value arising from this game, so that the wage per worker is given by

$$w = \chi z k^{\alpha} n^{-\alpha} + (1-\chi)\bar{y} \tag{11}$$

[7] It is worth noting that while I model the production game directly, the microfoundations in Stole & Zwiebel (1996) contain an error and do not actually justify the use of Shapley values in the presence of decreasing returns. Brügemann, Gautier & Menzio (2019) note this error and provide an alternative bargaining protocol that correctly microfounds the Shapley values.

where χ is a parameter governing the bargaining power of the entrepreneur relative to workers. [8]

The resulting wage determination equation is intuitive; workers are simply paid some linear combination of their average product of labor and their outside option $\bar{y}$, with the weight determined by bargaining power.

2.6. The Entrepreneur's Problem and Behavior

Combining equations (5) - (10) and the wage bargaining equation (11), the entrepreneur's problem can be written recursively as

$$
\begin{aligned}
V(z,f,n;X) = \max_{f',n',k,v,d} \quad & \frac{c^{1-\sigma}}{1-\sigma} + \beta\Delta V(z,f',n';X) \\
s.t.\ \ d + f' =& (1-\chi)zk^{\alpha}n^{1-\alpha} - (r+\delta)k - (1-\chi)\bar{y}n + f - cv \\
n' =& (1-\lambda)n + p(\theta)v \\
k \leq& \gamma f \\
v \geq& 0 \\
X' =& J(X)
\end{aligned}
$$

where X is a vector of aggregate state variables and J is the entrepreneur's perceptions function for the evolution of the aggregate state. It is important to note that the wage bargaining equation has been substituted into the entrepreneur's budget constraint and does not depend on the composition of their workforce, eliminating the need to track this as a state variable.

Entrepreneur Behavior: One important result (stemming from the fact that the user costs of both capital and labor are linear) is that an entrepreneur's capital-labor ratio depends only on their productivity z and aggregate state variables X

[8] At a technical level, the game is between an atomistic entrepreneur and a continuum of workers; the parameter χ is the relative size of the atomistic entrepreneur. It is also worth noting that because the Shapley value results in workers being paid a linear combination of their average product (rather than marginal product), the model does not nest perfectly competitive wages as a special case.

(see Appendix B for the derivation).[9] Denote this value as η so that

$$\eta(z;X) = \frac{\gamma f'^*}{n'^*} \tag{12}$$

where f'^* and n'^* are the entrepreneur's optimal policy functions.

This result lends the model substantial tractability. In general, the bargained wage w depends on all entrepreneur state variables, which can change over time due to accumulation of collateral. However, a constant capital-labor ratio (combined with the wage bargaining equation 11) implies that wages depend only on entrepreneurs' productivity (which is fixed over their life), justifying the use of $w(z)$ and $H(z)$ in the household problem above.

A second useful result is that entrepreneurs will pursue a constant productivity-dependent growth rate. That is, f'^* will satisfy

$$\begin{aligned} f'^* =& g(z;X)f \\ \frac{\partial g}{\partial z} >& 0 \end{aligned} \tag{13}$$

for some function g. Intuitively, g is increasing in z; more productive entrepreneurs will grow quicker. Together, the two functions η and g are sufficient to fully characterize entrepreneur behavior as a function of their productivity z and the aggregate state X.

2.7. Some Initial Intuition on Externalities

The two functions η and g can be used to gain some initial insight into the crowd-in and crowd-out externalities of search. Because these occur largely through changes in labor market tightness, it is useful to abuse notation and write $\hat{\eta}(z;\theta)$ and $\hat{g}(z;\theta)$ to represent "the steady-state values of η and g for a z productivity entrepreneur facing steady-state labor market tightness θ." This notation is possible because entrepreneur policy functions depend on the aggregate state only through current and future values of θ. We can then make the following comparative-static-like statements:

[9]This statement holds in universally in steady-state and holds for any transition path under the parameter restriction that $\lambda > 1 - \beta\Delta$ which is satisfied in the quantitative model. It is also worth noting that entrepreneurs with sufficiently low z will choose $k = 0, n = 0$ (i.e. will disengage from the economy and eat their cake rather than operate at a loss), leading to an undefined capital-labor ratio.

Proposition 1 *Let $\hat{g}$ and $\hat{\eta}$ be defined as above. Then*

$$\frac{d\hat{\eta}}{d\theta} > 0$$
$$\frac{d\hat{g}}{d\theta} < 0 \text{ and } \frac{\partial}{\partial z}\left(\frac{\partial \hat{g}}{\partial \theta}\right) < 0$$

where partial derivatives denoted by ∂ are taken while holding other endogenous outcomes (i.e. $\hat{\eta}$) constant.

The first claim of Proposition 1 ($\frac{d\hat{\eta}}{d\theta} > 0$) is that an entrepreneur's capital-labor ratio is increasing in labor market tightness (as a tighter labor market increases the cost of labor relative to capital). As a result of this, an individual choosing to search loosens the labor market and puts downwards pressure on the capital-labor ratio, leading to a reduction in wages for everyone — the Capital Shallowing (or monopolist) externality.

The second claim is that an entrepreneur's growth rate is decreasing in market tightness ($\frac{d\hat{g}}{d\theta} < 0$) and, more interestingly, that productive entrepreneurs are more sensitive to changes in θ ($\frac{\partial}{\partial z}\big(\frac{\partial \hat{g}}{\partial \theta}\big) < 0$). This effect arises from the fact that hiring costs make up a larger share of total costs for faster growing (i.e. more productive) firms.

Total costs for a firm wishing to grow at rate g^* are given by

$$r\eta^* + \chi z\eta^* + (1-\chi)\bar{y} + \big(g^* - (1-\lambda)\big)\frac{c}{p(\theta)} \tag{14}$$

and, consequently, a reduction in hiring costs (due to a looser labor market) represents a larger proportional reduction in total costs for higher g firms. Because these cost savings are used (in part) to fund growth, this leads to larger increases in growth for firms with large baseline growth rates (i.e. the most productive firms).

As a result, reductions in labor market tightness improve allocative efficiency in the economy. As the growth rate of productive firms increases more than unproductive ones, the share of resources in the economy allocated to the productive firms increases, and misallocation is reduced. From workers' perspective, this increases expected wages (as productive firms pay more). This link between an individual's search decision and average wages is ultimately the source of the Allocative Efficiency externality.

The two remaining externalities (Congestion and Firm Size) are not directly apparent from the entrepreneur policy functions. The next task is then to formalize the problem of a social planner in order to examine an exhaustive list of externalities.

3. Efficiency and the Social Planner

It is not immediately clear what the appropriate social planner's problem is. As in much of the labor search literature, the problem of an all-powerful planner free from any financial constraints or labor market frictions is uninteresting (except perhaps as a benchmark); this planner would simply allocate all labor and capital to the most productive entrepreneur and divide output in a way that equalizes marginal utility across all households and entrepreneurs. This teaches us nothing about the externalities generated by households' labor search or how these externalities interact with borrowing constraints.

Instead, I follow the traditional approach and consider the problem of a constrained social planner who must respect the search-and-matching technology, as well as individuals' borrowing constraints (as in Davila et al. 2012). Further, because the stated goal of most labor market policies (e.g. so-called "Active Labor Market Policies") is to improve outcomes for workers, I focus on a social planner who values only worker welfare and places no weight on the welfare of entrepreneurs. This approach has the additional benefit of being somewhat typical in macro-development models with multiple types of agents (e.g. Itskhoki & Moll 2019).

To prevent the planner from simply forcing entrepreneurs to hand over consumption to households, I impose that the social planner can only dictate the decisions of households and cannot control the behavior of entrepreneurs (i.e. the planner's choice of g and η must be consistent with entrepreneur optimization), who continue to solve their optimization problem each period. This has the added benefit of preventing the planner from improving outcomes by "picking winners" and forcing unproductive entrepreneurs to shut down in order to relieve collateral constraints for productive entrepreneurs. Allocations satisfying these three constraints make up the set of feasible allocations for the planner.

Definition: A path of household policy functions $\{c_t(a,y,z), a'_t(a,y,z), s_t(a,y,z)\}_{t=0}^{\infty}$, entrepreneur policy functions $\{g_t(z), \eta_t(z)\}_{t=0}^{\infty}$, distributions of households across savings and matched-employer productivities $\{m_t(a,z)\}_{t=1}^{\infty}$, and labor market tightness $\{\theta_t\}_{t=0}^{\infty}$ is **feasible** given an initial distribution $m_0(a,z)$ and market tightness

θ_{-1} if

1. It respects the household budget constraint for all a, y, z

$$\begin{aligned} a'_t + c_t &= Ra + w_t(z_t, \theta_t) && \forall a, y, t \text{ when } z \geq 0 \\ a'_t + c_t &= Ra + (1 - s_t)y + s_t b && \forall a, y, t \text{ when } z = 0 \\ a'_t &\geq 0 \end{aligned} \tag{15}$$

2. It respects the labor market matching technology

$$\begin{aligned} \frac{v(m_t, \eta_t, \eta_{t+1}, g_t)}{\theta_t} &= \int \int s_t(a, 0) m_t(a, 0) j(y) dy da \\ m_{t+1}(a', z) &= (1 - \tilde{\lambda}) m_t(a, z) + H(z, m_t, \eta_t, \eta_{t+1}, g_t) p(\theta_t) v(m_t, \eta_t, \eta_{t+1}, g_t) \end{aligned} \tag{16}$$

   where v is the total number of posted vacancies as a function of entrepreneur policy functions, and H is the probability that an individual who finds a job is matched with a firm of productivity level z.[10]

3. The entrepreneur policy functions $\{g_t(z), \eta_t(z)\}_{t=0}^{\infty}$ solve the entrepreneurs' problem (Appendix equation 25), conditional on θ_{-1} and $\{\theta_t\}_{t=0}^{\infty}$.

The task of the social planner is to maximize average household welfare subject to these feasibility conditions. Formalizing the statement of this problem is straightforward but cumbersome and is relegated to Appendix C.

There are two details worth noting. The first is that this definition of the planner's problem implicitly imposes the assumption that there is no autocorrelation in individuals' self-employment productivity y (i.e. the distributions m_t are only defined over (a, z)). This assumption, maintained throughout the rest of the paper, substantially reduces notation and improves readability and does not at all change any of the core mechanisms at play. The second is that the planner's problem features full commitment (they choose the entire sequence$\{\theta_t\}_{t=0}^{\infty}$ simultaneously) abstracting from any potential complications of dynamic games between the planner and model agents.

[10]Both v and H are formally defined in Appendix C.

3.1. Privately- vs Socially- Optimal Search Decision Rules

With the planner's problem specified, Proposition 2 finally formalizes the externalities that have been discussed only intuitively up until now. The proposition below makes the simplifying assumption that $\sigma \to 0$ (i.e. linear utility). This assumption is not necessary, and Appendix C provides the statement of the proposition valid for any (time-separable) utility function.[11]

Proposition 2 *Under the assumption that $\sigma \to 0$, the optimal steady-state search policies $s(y)$ of an individual and the constrained social planner depend only on an individual's self-employment productivity and are to search if and only if self-employment productivity y falls below the thresholds defined by s_c and s_p (respectively) below.*

$$\textit{Individual: } s_c + b \leq \beta\bar{\theta}p(\bar{\theta}) \int_z \frac{w(z,\bar{\theta}) - \big(\int_{y>s_c} yj(y)dy + J(s_c)(s_c - b)\big)}{1-\beta(1-\tilde{\lambda})} \bar{H}(z)dz \quad (17)$$

$$\textit{Planner: } s_p + b \leq \beta\bar{\theta}p(\bar{\theta}) \int_z \frac{w(z,\bar{\theta}) - \big(\int_{y>s_p} yj(y)dy + J(s_p)(s_p - b)\big)}{1-\beta\Delta g(z,\bar{\theta})} \bar{H}(z)dz \;+\; \mu \quad (18)$$

$$\mu = \frac{\bar{\theta}/\bar{S}}{\frac{\partial \log v}{\partial \log g}\frac{\partial \log g}{\partial \log \theta} - 1}\Bigg(\underbrace{\int_z \frac{\partial w}{\partial \eta}\frac{\partial \eta}{\partial \theta}\bar{m}(z)dz}_{\textit{Capital Shallowing}} + \underbrace{\int_z \bar{\lambda}(z)\bar{\theta}p(\bar{\theta})\bar{S}\frac{\partial H}{\partial g}\frac{\partial g}{\partial \theta}dz}_{\textit{Efficiency}} \quad (19)$$

$$+ \underbrace{\int_z \bar{\lambda}(z)\bar{H}(z)\bar{S}p(\bar{\theta})\big(1 + \frac{\partial \log p}{\partial \log \theta}\big)dz}_{\textit{Congestion}} + \textit{Anticipation Term}\Bigg)$$

where bars denote steady-state values of the competitive equilibrium and planner's problem respectively, J is the CDF of y, $\bar{S}$ is the steady-state number of searchers, and $\lambda(z)$ is the planner's shadow price denoting the marginal value of an additional worker being

[11] The fully general statement is not substantively different than the statement in Proposition 2 in the sense that the planner's optimal search decision rule differs from the individual's only by precisely the same four externalities. However, the assumption of linear utility substantially improves the readability of equations (17) and (18), providing clearer insight into the core intuition of the result.

matched with a productivity z entrepreneur. The anticipation term is described further in the appendix.

The privately optimal search policy simply weighs the total (opportunity-cost-inclusive) cost of search $y + b$ against the benefits, which are given by the expected excess earnings while employed, discounted over the expected duration of the employment spell.

Relative to this rule, the planner's decision rule differs in two ways, both of which are discussed further below. The first is that the planner discounts the excess earnings from employment using the expected growth rate of the firm (with probability Δ the entrepreneur will survive and grow their workforce by g) rather than the separation rate. The second is that the planner carries an additional term μ which internalizes changes in labor market tightness.

Search Externalities: As μ contains all but one of the externalities, it seems intuitive to start the discussion of externalities there. All three of the externalities contained in μ manifest through changes in labor market tightness. Consequently, μ is weighted by the net change in labor market tightness due to a change in the number of searchers after accounting for the response of vacancies.[12] This weight is negative, reflecting the fact that an increase in the number of searchers leads to a decrease in labor market tightness.

The term labeled "Congestion" is typical in labor search models and reflects the fact that an additional searcher pushes down labor market tightness, reducing the job-finding probability for all searchers. As in textbook search models, the size of this externality depends on the elasticity of the matching function ($\frac{\partial \log p}{\partial \log \theta}$) as a high elasticity implies that an additional searcher leads to a large reduction in the job-finding probability. Intuitively, this externality is also increasing in the steady-state number of searchers $\bar{S}$.

The term labeled "Efficiency" contains the impact of the improvement in allocative efficiency that occurs when the labor market loosens as a result of search (discussed above). A looser labor market increases the growth rate of firms ($\frac{\partial g}{\partial \theta}$) but does so by disproportionately more for high-productivity firms. As a result, the probability of a worker matching with a high-productivity entrepreneur increases ($\frac{\partial H}{\partial g}$ is increasing in g) which the planner values according to the shadow price λ. At an aggregate level, this manifests as higher average wages (productive

[12]To see that this expression indeed gives the net change, note that $\theta = \frac{v(\theta)}{s} \Rightarrow \frac{d\theta}{ds} = \frac{\theta/s}{\frac{d \log v}{d \log \theta} - 1}$.

entrepreneurs pay more) and higher aggregate TFP.

The term labeled "Capital Shallowing" reflects the decline in the capital-labor ratio, and consequently the average wage, arises from a decline in labor market tightness. Interpreting this effect as an externality may seem to hinge on the assumption that the planner does not value the consumption of entrepreneurs (while the decline capital-labor ratio reduces wages, it also increases profits), but the presence of such a term is general. It arises even if the planner puts positive weight on entrepreneurs' consumption (an extension done in Appendix C.2) or workers and entrepreneurs are assumed to be the same agents (as in Buera, Kaboski & Shin 2021). In all cases, the fundamental source of this externality is the fact that the planner's marginal value of consumption varies across agents and/or states so that the impact of the redistribution arising from a change in the capital-labor does not necessarily sum to zero. Thus this term is akin to the monopolist externality in Itskhoki & Moll (2019) or the pecuniary externality in Davila et al. (2012).

The final externality, which I call the "Firm Size" externality, is not present in μ, as it does not operate through labor market tightness, and is instead present in the different discount rates used to discount the excess earning while employed in the individually- and socially- optimal search decision rules. It is fairly easy to see that $\Delta g(z, \theta) > \Delta(1 - \lambda) = (1 - \tilde{\lambda})$ in any steady-state. Thus the planner's valuation of a job is higher than an individual's, even fixing labor market tightness.

This externality is a result of the fact that individuals do not capture all the output created by their job match. Some of it is captured by the entrepreneur and used to finance future growth and, as a result, the hiring of additional workers tomorrow. This effect is easiest to see by considering an entrepreneur growing at rate g who is exogenously matched with a unit measure of workers from outside the economy (say, immigrants). Without these extra workers, the entrepreneur would hire $(g - 1)n$ workers (on net) for the next period, but with these workers, the entrepreneur hires $(g-1)(n+1)$. The extra workers crowd-in $(g-1)$ additional workers in the next period (and $(g-1)^2$ the next, etc.), conditional on entrepreneur survival.

3.2. The Dual Problem and Optimal Search Subsidies

The problem of selecting the welfare maximizing path subject to a set of dynamic constraints in a heterogeneous agent economy is similar to other Ramsey-type problems often found in the literature dealing with welfare and efficiency in heterogeneous agent models (e.g. Itskhoki & Moll 2019, Dávila & Schaab 2023). Like all Ramsey problems, the primal problem of choosing paths of consumption (or consumption functions) subject to feasibility constraints can be equivalently formulated as a dual problem in which the planner selects optimal tax rates from a sufficiently rich set of instruments to decentralize the optimal allocation in competitive equilibrium.

Implementing the planner's solution in this way is deceptively complex. At first glance, Proposition 2 seems to suggest a simple implementation — the competitive and planner search rules both weigh the value of self-employment (on which the two agree) against the expected value of search (on which they disagree). A single subsidy (or tax) should be enough to resolve this disagreement and align the competitive rule with the planner's. However, such a subsidy also alters budget constraints, and the resulting competitive equilibrium does not solve the planner's problem by virtue of being infeasible. A complex set of state-contingent lump-sum transfers are needed to restore feasibility.

As an alternative, one may also be interested in the optimal policy under a restricted set of simpler tax instruments. A natural choice of instruments is a subsidy (or tax) to search funded (τ_b) by a single proportional tax rate on self-employment earnings (τ_y) in a balanced-budget manner. Taxing self-employment earnings rather than something like wages has the advantage of aligning more closely with the planner's (unrestricted) solution by minimizing the impact of changes in the budget constraints. Under this set of tax instruments, the optimal search decision rule s_r (for "restricted") takes a form very similar to the planner's full solution.

$$
\begin{aligned}
s_r + b \leq \beta\bar{\theta}p(\bar{\theta}) \int_z \frac{w(z,\bar{\theta}) - \left(\int_{y>s_r} yj(y)dy + J(s_r)(s_r - b)\right)}{1 - \beta\Delta g(z,\bar{\theta})} \bar{H}(z)dz \ + \ \mu \\
+ \ \underbrace{\frac{\int \lambda(y)\big((1-s_r)y\tau_y + s_r b\tau_b\big)j(y)dy}{\tau_y \frac{ds_r}{d\tau_y} + \tau_b \frac{ds_r}{d\tau_b}}}_{\text{Impact on Budgets}}
\end{aligned}
\tag{20}
$$

The only difference between the optimal policy under the restricted set of instruments s_r and the planner's optimal policy s_p is an extra term accounting for the impact of τ_y and τ_b on budget constraints, discounted by the responsiveness of search behavior (highly responsive search behavior means only a small subsidy is necessary, so the effect on budget constraints is small).[13]

The quantitative analysis below largely focuses on this optimal set of restricted policies, partially because they are easier to interpret than policies involving complex lump-sum transfers and partially because they are substantially easier to compute. Fortunately, difference between the optimal restricted policy and the planner's full solution turns out to be quantitatively negligible (the "Impact on Budgets" will be an order of magnitude smaller than any of the four externalities).

4. Model Estimation

Quantifying these externalities requires bringing the model to data. This, in turn, requires focusing on a particular labor market (as many model parameters are likely to vary from country to country and even from city to city). To this end, I opt to estimate the model to match the labor market of Addis Ababa, Ethiopia, largely because some experiments useful for model estimation happened to be conducted there.

Model parameters fall into two categories. The first are parameters that are directly estimated from data (such as collateral requirement γ) or set to standard values (such as the discount rate β). The second are parameters that are more difficult to measure directly (such as the search cost b). These parameters are estimated using the simulated method of moments (SMM) to match data moments from the aforementioned experiments, as well as some aggregate moments. Before going into the details of the SMM estimation, it is worth briefly discussing a handful of the key non-SMM parameters whose values are important.[14]

4.1. Key Directly Estimated Parameters

Importantly, the rate of return on individuals' savings R is taken to be less than unity with an annual value of 0.9 (chosen to roughly match the Ethiopia inflation rate, suggesting that individuals' savings a are best thought of as cash). Because the model is estimated to a weekly frequency, this corresponds to a value of $0.9^{\frac{1}{52}}$. The assumption that the return to savings is less than one, and thus that saving is costly, is typical in models of developing countries (see e.g. Donovan 2021, Fu-

[13] Note that the assumption of linear utility (i.e. $\lambda(y) = 1$) and the balanced-budget constraint imply that this additional term is equal to zero. In other words, the competitive equilibrium under the optimal feasible policy has the same decision rule as the planner's problem. This is not true in general though, and fully expressing this term provides intuition that extends to the general case.

[14] Although not all parameters are discussed here, Appendix D and, in particular, Appendix Table D.1 provide an exhaustive list of these parameters, their values, and some further discussion.

jimoto, Lagakos & VanVuren 2023). Here, encoding this assumption is important as the difficulty of maintaining a cushion of savings is an oft-cited justification of the need for search subsidies (although the analysis in Section 3 revealed that this does not directly justify intervention).

For similar reasons, the income process of the self-employed is also important. This is measured directly using weekly data on workers and job seekers collected by Abebe et al. (2021) as part of an experiment in Addis Ababa (details below). In the context of Addis, the majority of the variation in earnings (among those without a permanent job) comes from whether an individual is currently working a temporary, gig-style job or not. Consequently, self-employment productivity is modeled as a binary Markov process (with the high state corresponding to "working" and the low state to "not working") whose transition matrix can be estimated directly (the probability of remaining in one's current state is roughly 89 percent per week for both high and low states). The ratio of earnings between the high and low state can also be measured directly and is set to 2.63. Thus the model closely matches observed volatility in self-employment earnings.

The matching function is chosen to be a simple urn-ball matching function, $p(\theta) = \frac{1-e^{-\zeta\theta}}{\theta}$, with an efficiency parameter of ζ that can be adjusted to target any elasticity $\frac{d\log p}{d\log\theta}$ in the model steady-state.[15] Absent detailed estimates of this elasticity in the context of Addis Ababa and lacking the necessary data to estimate it, I choose ζ to generate an elasticity of -0.3. This is a fairly typical value and roughly in line with the estimates of Hall & Schulhofer-Wohl (2018) for the United States.

Finally, many of the entrepreneurs' parameters can be estimated using data from the World Bank. MIX Market data contains financial information on microcredit providers in Ethiopia and suggests a rough average yield of 25 percent which, combined with an 8 percent depreciation rate, suggests a user cost of capital equal to 33 percent annually.[16] The average collateral requirement in Addis Ababa (computed using the World Bank Enterprise Survey for Ethiopia in 2015) is 350 percent — a firm that owned 350,000 Birr of capital could finance a 100,000 Birr loan — suggesting a value of 1.29 ($1 + \frac{1}{3.5}$) for the collateral parameter γ. Finally, fitting a geometric distribution to the firm age distribution via maximum likelihood yields

[15] The choice of functional form is unimportant beyond the fact that it includes a free parameter that can be used to target the desires elasticity for p. A functional form exhibiting a constant elasticity, such as Cobb-Douglas, would be ideal, but the use of discrete time limits sensible choices for p to those that lead both $p(\theta)$ and $\theta p(\theta)$ to be bounded between zero and one.

[16] Loan loss rates in Ethiopia are negligible for the purposes of this calculation.

an annual entrepreneur death probability $(1-\Delta)$ of 0.08.

4.2. Parameters Estimated using the Simulated Method of Moments

Table 1: Parameter Estimates from Simulated Method of Moments

Parameter	Estimate	Corresponding Moment
σ	5.2	% wage work
λ	0.01	Unemployment rate
A_s	0.34	Wage sector premium
b	0.05	% of expenditure on search
M$\underline{f}$	.001	Control wage employment after 16 weeks
c	0.37	Cost to hire as % of wage
χ	0.62	Elas. of avg. wage to output per worker
$\bar{z}$	0.35	Avg. growth rate

Note: This table displays the parameters estimated using simulated method of moments, their estimates, and the moment that corresponds most closely to each parameter. See discussion for details and intuition on these correspondences.

There are eight parameters estimated using the simulated method of moments to match eight data moments. Table 1 lists these eight parameters and their estimated values while Table 2 lists the eight targeted moments and their values in both the data and the model. The parameters fall into two rough categories — those corresponding closely to worker-level moments (above the dividing line in Tables 1 and 2) and those corresponding closely to firm-level moments (below the line).

Worker moments: The data for the worker-level moments come from two sources. The proportion of individuals engaged in wage work and the aggregate unemployment rate are measured using the 2018-2019 wave of the Ethiopia Living Standard and Measurement Survery (LSMS), limited to individuals in Addis Ababa.[17] The wage sector premium is estimated on the same data by including a dummy vari-

[17] While the other data sources used in estimation are from 2014-2015, the 2018 wave of the Ethiopia LSMS was the first wave capable of providing representative estimates for Addis Ababa (previous waves were not representative at a sub-national level). For this reason, I opt to use the data from 2018 rather than the 2015 wave, which would otherwise line up better with the other datasets temporally.

Table 2: Moments Targeted using the Simulated Method of Moments

Moment	Source	Data	Model
% wage work	LSMS	30%	29%
Unemployment rate	LSMS	10%	12%
Wage sector premium	LSMS	39%	39%
% of expenditure on search	Abebe et al. (2021)	15%	16%
Control wage emp. after 16 weeks	Abebe et al. (2021)	12%	13%
Cost to hire as % of wage	Abebe et al. (2017)	120%	120%
Elas. of avg. wage to output per worker	World Bank ES	25%	25%
Avg. growth rate	World Bank ES	4.4%	4.4%

Note: This table displays the moments targeted in the simulated method of moments estimation, their source, and their values in both the data and model. See the discussion for details.

able indicating whether an individual is employed in a permanent wage job (vs self-employment or temporary) work in an otherwise standard Mincer regression of (log) earnings on age, as well as some controls (rural/urban, region, and sector fixed effects).[18]

The two remaining household moments come from the aforementioned weekly data on job seekers from Abebe et al. (2021). The first is average expenditure on job search (for weeks in which an individual searches) as a percentage of total expenditure. This is calculated directly via survey responses (i.e. individuals are asked directly how much they spent on search and in total). The second, labeled "control wage emp. after 16 weeks", reflects the proportion of individuals in the experimental control group with wage employment 16 weeks after baseline. Although I discuss the experiment in more detail in the next sub-section, it is important to note here that only data from the experimental control group is used in model estimation while data from the treatment group is reserved for model validation.

Together, these five moments pin down the five parameters above the dividing

[18]The Ethiopian Productive Safety Net Programme (PSNP), a relatively new "workfare" program administered by the Government of Ethiopia, presents a potential complication. The program provides temporary employment and was present in some regions of Addis Ababa during the 2018 LSMS survey. It is unclear whether earnings from the PSNP should be included in estimation. Fortunately, dropping these earnings from the analysis changes the estimate by less than one percentage point, rendering the issue quantitatively moot. I default to including all earnings from temporary employment, including those from the PSNP.

line in Table 1. The risk aversion parameter σ and the job separation rate λ are disciplined (mostly) by the size of the wage sector and the unemployment rate. While the link between the separation rate and the unemployment rate is clear, the link between risk aversion and the size of the wage sector arises from the fact that, for the worker, search is the higher-risk, higher-return option (relative to self-employment). Thus individuals' risk tolerance ends up being a primary determinant of the size of the wage sector.

The earnings premium in the wage sector naturally pins down the productivity of the self-employment technology, as a more productive technology shrinks the earnings gap between the two sectors. Similarly, the percent of total expenditure that goes towards search costs almost mechanically pins down the goods cost of search. The final moment, the employment rate of control group job seekers after 16 weeks, conceptually pins down the (weekly) job-finding rate. The parameter most directly linked to this equilibrium object is the initial size of a newborn entrepreneur (given by $M\underline{f}$, which are not separately identified) — if newborn entrepreneurs are larger, they will end up posting more vacancies, directly impacting the job-finding rate.

Firm moments: The remaining three parameters — the vacancy posting cost c, the wage bargaining parameter χ, and the upper bound of firm productivity $\bar{z}$ — are estimated to match firm-level moments. Abebe et al. (2017) survey firms in Addis about hiring practices and find that the average cost to a firm of making one additional hire is equal to 120 percent of the average wage. This moment directly pins down the vacancy posting cost.

The bargaining parameter χ, via the wage bargaining equation (11), is pinned down by the relationship between firm-level average wages and output per worker. I estimate this elasticity to be 25 percent (meaning, a firm with 100 percent higher output per worker pays its workers on average 25 percent more) in World Bank Enterprise Survey data and use this as the target in estimation. The final model object to be pinned down is the distribution of entrepreneur productivity. I choose an upper-truncated Pareto distribution with tail parameter 2.1 (as close to Zipf's law as possible while maintaining well-defined variance). The truncation point $\bar{z}$ is disciplined by the average (self-reported) annual growth rate for firms in the Enterprise Survey as a higher $\bar{z}$ directly corresponds to a higher average growth rate due to the fact that more productive firms grow faster (at least in the model).

4.3. Model Validation and the Experiment of Abebe et al. (2021)

To validate the model, I replicate an experiment performed by Abebe et al. (2021) in the model and compare the model outcomes to the experimentally estimated outcomes. As mentioned above, it is important to note that while control outcomes from the experiment are used to estimate the model, treatment outcomes and data are not. Thus comparing the model's predicted treatment effects to those estimated in the experiment represents an "out-of-sample" test of the model.

This experiment took place in 2014-2015 and evaluated the effects of providing a cash subsidy covering some of the costs of job search to prospective searchers in Addis Ababa, Ethiopia. In the context of Addis Ababa, the majority of job search takes place in person in the city center. Thus the cost of travel (typically by minibus) to the city center represents a large and salient cost of job search.

The experiment sampled young individuals who "(i) were between 18 and 29 years of age; (ii) had completed high school; (iii) were available to start working in the next three months; and (iv) were not currently working in a permanent job or enrolled in full time education." (Abebe et al. 2021) and randomly offered some individuals cash that could be collected in person at the city center up to three times each week. While not literally a job search subsidy as individuals could theoretically travel to the city center, collect the cash, and leave without searching, doing so would be ineffective as the cost of the subsidy is not large enough to cover the full round-trip journey.[19] Thus collecting the cash only makes sense if the individual intended to travel to the city center for other purposes (presumably job search). The cash was available for 16 weeks after which treated individuals were 3.4 percentage points ($p<0.1$) more likely to be employed in a permanent job.

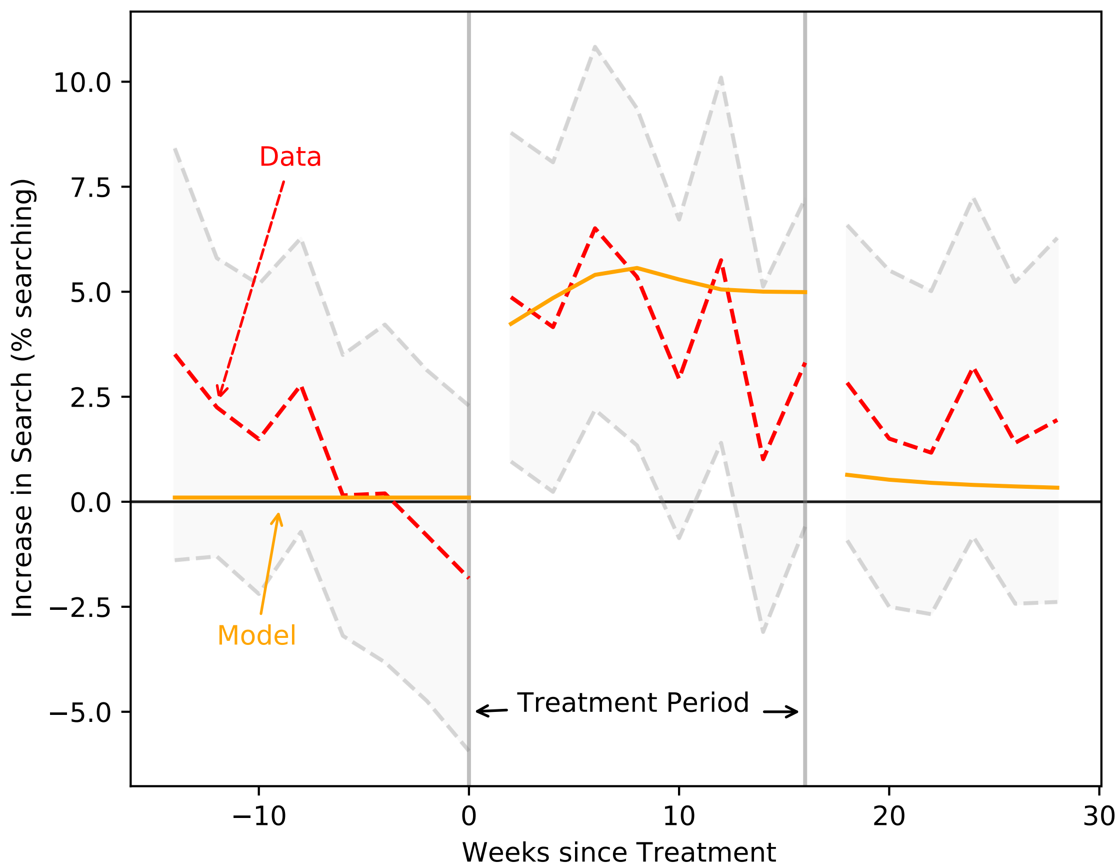
To replicate the experiment in the model, I select a representative but small (measure 0) subset of individuals not employed in the wage sector from the steady-state distribution of individuals. In this sense, the outcomes of sampled individuals do not affect equilibrium outcomes, and the experiment happens in "partial equilibrium". The sample is divided equally into treatment and control groups, and the cost of search parameter b is reduced by two-thirds (the median subsidy offered in the experiment) for the treatment group for 16 periods.

Experimental outcomes can then be observed by simulating the behavior of the

[19] In fact, the authors make sure of this by varying the subsidy offered to each individual based on the location, and thus minibus ticket cost, of the individual's home. However, I abstract from this heterogeneity and model the subsidy as uniform at the median value of subsidy offered.

treatment and control groups forward over time, and comparisons of means between the two groups correspond to Average Treatment Effects estimated by the experiment. For treatment households, I treat the experiment as an unanticipated MIT shock; households do not know ahead of time that they have been selected for treatment and cannot alter their behavior in response to such information (and are also fully aware that it will end after 16 periods). Thus differences between treatment and control groups before the treatment occurs are zero by construction.

Figure 2: Treatment Effect on Search Behavior over Time: Data and Model



This figure displays the treatment effect on search behavior as a function of "weeks since treatment" in both the data and estimated model.

Model vs Data: Figure 2 compares the model's predictions for the increase in search behavior as a result of the subsidy to those observed in the data. The solid orange line depicts model predictions and the dotted red line depicts the experimentally estimated effects along with the associated 95 percent confidence interval.

The model lines up with the experiment remarkably well. During the treatment period (between 0 and 16 weeks since treatment), treated individuals were roughly 5 percentage points more likely to search, a fact which is replicated in the model. There is a small decline in the point estimates in the last few weeks of treatment that is not quantitatively replicated by the model, but this decline is statistically insignificant, and the model continues to fall within the estimated 95 percent confidence interval.

The model also qualitatively replicates the fact that effects seem to persist for some weeks after treatment is ended, although the experimental point estimates here are noisy. The model's predictions are quantitatively smaller than these point estimates, but are well within the 95 confidence interval. One explanation for the model's underprediction of persistence is that increased search results in some sort of learning or habit formation, leading treated individuals to search more often even after the end of treatment, that is not captured in the model.

Even if the model accurately matches the increase in search behavior due to treatment, it may not match the increase in wage employment if, for example, search within a short time period exhibits substantial diminishing returns (i.e. job seekers first go after opportunities they judge most promising). Reflecting the implicit assumption of constant returns, the model predicts a roughly 5 percentage point higher probability of being employed after 16 weeks, the same as the increase in search behavior. The experimental equivalent is 3.4 percentage points (90 percent confidence interval 0.3 to 6.3). This is slightly lower, but the model is still reasonably accurate, and the 90 percent confidence interval is not sufficient to rule out the assumption of constant returns.

5. Efficient Policy in the Estimated Model

With the estimated model in hand, we can now quantify the optimal feasible policy and investigate the relative sizes and contributions of the four externalities. The are many ways to approach this, but the simplest is to directly solve for the optimal policy tax/subsidy rates on search and self-employment and then decompose the total impact of the policy across the four externalities.[20] Here it is important to note that while equations (17) – (20) are written under the simplifi-

[20]This is much more straightforward than approaches centered around recomputing optimal policies with and without the presence of certain channels due to the fact that steady-state equilibria often become degenerate (i.e. no self-employment or no wage work) or fail to exist when certain channels are shut down.

cation of linear utility, the results in this section are computed using the estimated model which exhibits substantial curvature ($\sigma = 5.2$).

5.1. How to Decompose the Impact of the Externalities

The task of decomposing the impact of the policy across the externalities is not entirely straightforward. The Congestion, Capital Shallowing, and Allocative Efficiency externalities all operate through the adjustment of particular equilibrium values ($\theta p(\theta)$, $w(z)$, and $H(z)$; see equation 19). The impact of each channel, then, can be decomposed by examining how the policy's impact on welfare changes when the change in the appropriate equilibrium value — and thus the impact of the externality — is marginally reduced (using a marginal calculation, rather than an average, reflects the fact that the optimal policy balances the marginal impacts of each channel).

Less clear, however, is the Firm Size externality which at first glance does not appear to depend on changes in any equilibrium values, instead depending only on the level of firms' growth rates $g(z)$. The necessary insight stems from noting that the firm size externality can only occur through an implicit change in the number of vacancies. An additional searcher today generates $\theta p(\theta)$ workers tomorrow (i.e. by finding a job) which generate $(\theta p(\theta))g(z)$ workers the next day and, as a matter of accounting, $(\theta p(\theta))\frac{g(z)}{p(\theta)}$ vacancies (and $(\theta p(\theta))\frac{g(z)^2}{p(\theta)}$ vacancies the next day and so on). These additional vacancies turn out to be the equilibrium object through which the Firm Size externality occurs. Once this is know, this externality can be treated symmetrical with the other three (Appendix D.1 provides a complete but dry description of exactly how this is computed).

5.2. The Efficient Policy

With these computational details set aside, Table 3 reports the tax rates of the optimal policy (which, recall, consists of taxes on search and self-employment earnings subject to a balanced-budget constraint) as well as the impact of the policy on welfare and the size of the wage sector. Although there is some minor variation along the transition path after the policy is implemented, these rates correspond to the eventual rates in the post-policy steady-state.

The most surprising result is that the optimal tax rate on on search is positive and large at 101 percent of baseline search costs. In other words, the competitive equilibrium exhibits too much search, and the planner finds it necessary to discourage this through a tax that shrinks the size of the wage sector from 30 percent

Table 3: Results of the Efficient Policy

Optimal Rates			
Search Tax:	101%	Self-emp. Tax:	-2.0%
Welfare:	+1.5%	Size of Wage Sector:	-6.0pp
(Self-employed):	+1.6%	(Pre-policy):	30%
(Employed):	+1.4%	(Post-policy):	24%

Note: This tables displays the search tax and self-employment subsidy rates that make up the optimal feasible policy in the estimated model, as well as the impact on welfare and the size of the wage sector that occurs when these rates are implemented. See text for details.

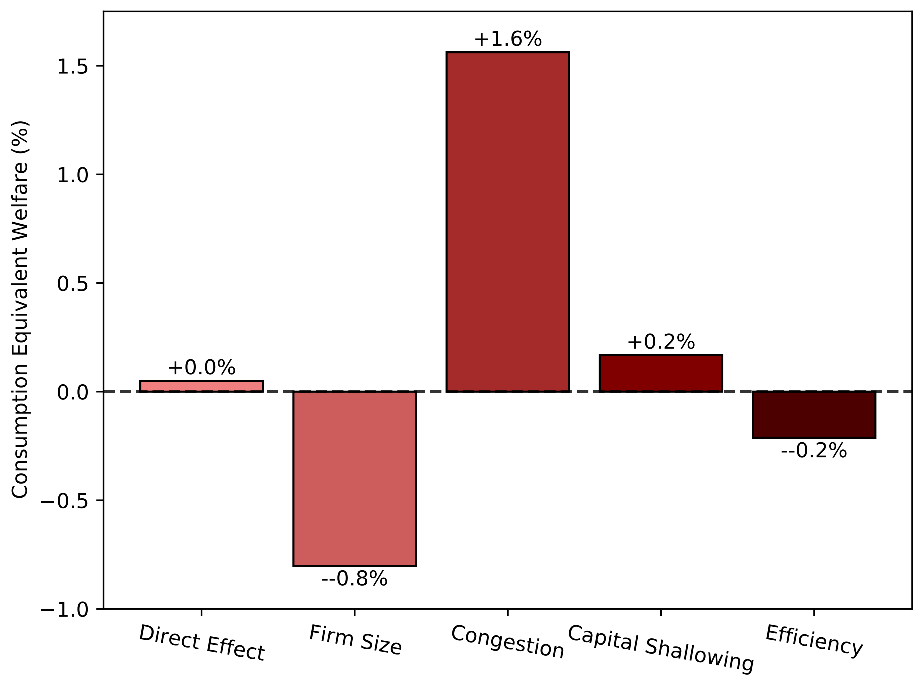
(of workers) to 24 percent. This contradicts the intuition of many policymakers and economists that barriers to search represent a substantial problem for (potential) workers in developing countries. At least from the perspective of the model, the crowd-out externalities dominate the crowd-in externalities, and more barriers need to be erected.

The positive tax rate on search is mirrored by the negative tax rate on self-employment earnings (i.e. a subsidy) to comply with the balanced-budget constraint. This subsidy is moderate in size at 2 percent of earnings. The difference in magnitude between the tax on search and the subsidy to self-employment partially reflects the difference in size between the cost of search and average earnings (without taxes, the search cost is roughly 20 percent of average self-employment earnings) and partially reflects the difference in popularity of the two activities (there are approximately 10 times as many self-employed individuals as there are searching individuals.)

The overall impact on welfare of the policy is substantial. Average welfare increases by 1.5 percent of consumption. This impact is not particularly regressive or progressive as the average impact among the self-employed and the average impact among the employed are similar (1.6 percent and 1.4 percent respectively). Although the employed pay more of the search cost, as they are more likely to search in the near future, they also reap more of the benefits from shrinking the crowd-out externalities. On the other hand, the self-employed largely benefit through higher earnings from the subsidy.

Figure 3 displays the contribution of each of the four externalities to the overall effect, as well as the gains from the direct impact of the policy on budgets (refer to equation 20). Because the optimal policy is a tax that reduces search, the *positive* search externalities (Firm Size and Efficiency) take on negative values as a reduction in search shrinks the size of these externalities which contributes negatively to welfare. Similarly, the negative externalities of search (Congestion and Capital Shallowing) take on positive values.

Figure 3: Sources of Welfare Gains from Optimal Policy



Note: This figure displays the (marginal) contribution of each of the four externalities to the overall welfare impact of the optimal feasible policy. Refer to subsection 5.1 for details on how this decomposition is performed.

Both the Firm Size and Congestion externalities contribute substantially to the welfare impact of the policy, accounting for -0.8 percent and +1.6 percent of the change respectively. Meanwhile, the Capital Shallowing and Allocative Efficiency externalities make much smaller contributions, accouting for +0.2 and -0.2 percent. Finally, the Direct Effect of the policy on budgets accounts for a positive but negligibly small (0.01 pecent) portion of the welfare gains, suggesting that the optimal policy under the restricted set of tax instruments is very close to what would be achieved without restrictions.[21]

This decomposition yields two main takeaways. The first is that the search externalities impacting employment (Firm Size and Congestion) are quantitatively much larger than the externalities impacting average wages (Capital Shallowing And Efficiency). This occurs (somewhat mechanically) due to the fact that the impact of the policy on wages is small — they increase by 0.8 percent as a result of Efficiency and decline by 0.2 percent as a result of Capital Shallowing for a net increase of only 0.6 percent. The second takeaway is the quantitative dominance of the Congestion externality; the tax results in a net increase in the job finding probability, which is highly valued by individuals.

Driving Forces: What features of the data drive the model to these conclusions? Put another way, what does the model suggest are the key features in look for in a particular labor market in order to determine which externalities are important? [To be filled out more]

6. Policy Analysis of Search Subsidies

The results of the previous section indicate that the level of search is too high and that the optimal policy is a tax on search, rather than a subsidy. Although this policy improves average welfare, it also shrinks the wage sector, which may be contradictory to the goals of many policymakers who would prefer to grow the size of the wage sector even at the expense of efficiency. Although such a goal is dubious from the perspective of the model, it may stem from practical concerns extending beyond the model's scope, such as the need to attract Foreign Direct Investment or to cement a nascent industrial base.

To accommodate this, this final section briefly pivots from normative to positive analysis and uses the estimated model to understand the impact of implementing a subsidy to job search. Even for a policymaker who values only the size of the wage sector, the crowd-in and crowd-out externalities remain important as they account for the difference between the impact of the subsidy when evaluated in partial equilibrium (i.e. on a small experimental sample) and the impact of the subsidy in general equilibrium (i.e. implemented for the entire labor market). This section, then, follows the spirit of the literature on interpreting experimental results

[21] Here it is worth noting that the fact that the contribution of each channel is computed by marginally, rather than totally, reducing the impact of that channel is the primary reason that summing that channels' contributions (+0.8 percent) does not yield the total welfare impact of the policy (+1.5 percent).

through macroeconomic models (e.g. Brooks, Donovan & Johnson 2020, Fujimoto, Lagakos & VanVuren 2023, Lagakos, Mobarak & Waugh 2023).

The policy analyzed is a cash transfer to searchers, effectively reducing the cost of search b, as in the experiment used to estimate the model. The size of the transfers is also chosen to be the same as that used in estimation and is equal to about two-thirds of the cost of search. In the baseline evaluation, the subsidy is funded through a proportional tax on the self-employed (keeping in line with the previous section).

Table 4 shows the impact of the subsidy on the size of the wage sector, as well as welfare, in both partial and general equilibrium. The general equilibrium results correspond to the case where all equilibrium values (the job-finding probability $\theta p(\theta)$, the wage-productivity relationship $w(z)$, and the z-type match probability $H(z)$) are allowed to adjust to their new equilibrium level, and these results represent what occurs when the policy is implemented economy wide and available to all. In contrast, the partial equilibrium results correspond to a model where these equilibrium values are fixed at their pre-policy levels. These outcomes, then, correspond to what would be observed in an experimental evaluation of the policy (as the experimental sample is small and does not influence equilbrium outcomes). Importantly, because the four externalities occur through these equilibrium adjustments, these are shut off in partial equilibrium as well.

In partial equilibrium, the results of the policy seem very promising. When the externalities are shut off, the size of the wage sector increases by 17 percentage points to 47 percent with essentially no impact on welfare (mirroring the small "Direct Effect" of Figure 3). These results assume that individuals are treated with both the subsidy and the tax used to fund the policy (for comparability to the previous section). If, consistent with experimental procedures, the subsidy is funded via grants from funding organizations rather than a tax of the self-employment earnings of recipients (displayed in row "Only Subsidy, No Tax"), the outcome looks even better with a welfare increase of 1.3 percent.

In general equilibrium, the results are substantially more pessimistic. The policy increases the size of the wage sector by about half as much (10 percentage points) and carries a substantial welfare penalty of 2.5 percent. Relative to the efficiency policy (which increased welfare by 1.5 percent), welfare is 4.0 percent lower. As the results of the previous section suggest, this pessimism is largely explained by the Congestion externality — when the other three externalities are shut down ("Only

Table 4: Result of Search Subsidies

Model	Wage Sector	Welfare
Baseline	30%	–
w/ Subsidy (Partial Eq.)	47%	-0.0%
w/ Subsidy (General Eq.)	40%	-2.5%
Only Subsidy, No Tax (PE)	43%	+1.3%
Only Congestion Channel	27%	-4.6%
Tax on Employed (GE)	32%	+0.0%

Note: This table displays the impact of implementing a subsidy for job search on the size of the wage sector and average welfare in the estimated model. Refer to the text for details on the models represented by each row.

Congestion"), the policy actually shrinks the wage sector to 27 percent and leads to a massive 4.6 percent welfare loss..

Finally, although being consistent with the previous section required a baseline policy in which the search subsidy was funded by a tax on self-employed workers, one might argue that it is more natural to fund the subsidy through a tax on wage workers — moving income from a "good" state (employment) to a "bad" state (search) can increase welfare. This policy is evaluated in the row labeled "Tax on Employed (GE)". The additional gains in welfare from redistribution are barely enough to offset the losses from the decline in efficiency (welfare increases by less than a tenth of a percent). More notably, the policy ends up barely growing the wage sector at all (increasing to only 32 percent) as the tax on wage work rather the self-employment serves to further discourage search.

The upshot of this section is that promising experimental (partial equilibrium) results do not necessarily guarantee that a policy will be successful when scaled-up to a general equilibrium level, at least for policies aimed at expanding the wage sector. Such results do not provide enough information (at least directly) about the sizes of the various search externalities and the effects of the policy in general equilibrium, highlighting the role of search models like the one developed in this paper in aggregating these results.

7. Conclusion

Many policies and interventions aim to expand the wage sector by increasing the extent to which (potential) workers can search for jobs; however, frictional labor markets generate search externalities. This paper develops and estimates a model that incorporates key features of developing countries in order to understand the inefficiencies that arise in this setting. Contrary to the intuition that search should be encouraged, the estimated model suggests that the optimal policy is a substantial tax increasing the cost of search.

One broad takeaway of the model and ensuing quantitative analysis, relevant to policymakers and economists alike, is that policies aimed at assisting job seekers should be very careful to distinguish between the extent to which policies encourage search (i.e. increase in individual's incentive or ability to search) versus the extent to which they improve the effectiveness of search (i.e. improve the productivity of the matching function), as improvements in search efficiency are not subject to the concern of crowding-out. Because many policies represent a combination of these two effects (e.g. government subsidies for employment agencies, discussed in Wu & Wang 2023, may encourage search by lowering the price of this service but may also improve efficiency if agencies are able to effectively streamline the matching process), experimental evaluations of these policies can productively try to distinguish between their impact on each.

The quantitative conclusions of Sections 5 and 6 should be caveated by noting that the model is estimated to the specific setting of Addis Ababa. Although quantitative exploration reveals that it is fairly difficult (though not impossible) to overturn the conclusion that the optimal policy is a tax on search, the exact level of the optimal tax can vary substantially when the targeted moments are changed. Applying the model in different settings would require new data on these moments, which may be difficult to find depending on the setting.

References

Abebe, G., Caria, A. S., Fafchamps, M., Falco, P., Franklin, S. & Quinn, S. (2021), 'Anonymity or distance? job search and labour market exclusion in a growing african city', *The Review of Economic Studies* **88**(3), 1279–1310.

Abebe, G. T., Caria, S., Fafchamps, M., Falco, P., Franklin, S., Quinn, S. & Shilpi, F. (2017), 'Job fairs: Matching firms and workers in a field experiment in ethiopia', *World Bank Policy Research Working Paper* (8092).

Acemoglu, D. & Hawkins, W. B. (2014), 'Search with multi-worker firms', *Theoretical Economics* **9**(3), 583–628.

Aiyagari, S. R. (1994), 'Uninsured idiosyncratic risk and aggregate saving', *The Quarterly Journal of Economics* **109**(3), 659–684.

Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M. & Vitali, A. (2020), 'Tackling youth unemployment: Evidence from a labor market experiment in uganda', *Econometrica* **88**(6), 2369–2414.

Algan, Y., Crépon, B. & Glover, D. (2020), 'Are active labor market policies directed at firms effective? evidence from a randomized evaluation with local employment agencies', *J-PAL working paper* .

Banerjee, A., Basu, P. & Keller, E. (2021), Cross-country disparities in skill premium and skill acquisition, Technical report.

Banerjee, A., Karlan, D. & Zinman, J. (2015), 'Six randomized evaluations of microcredit: Introduction and further steps', *American Economic Journal: Applied Economics* **7**(1), 1–21.

Bewley, T. (1977), 'The permanent income hypothesis: A theoretical formulation', *Journal of Economic Theory* **16**(2), 252–292.

Brooks, W., Donovan, K. & Johnson, T. R. (2020), From micro to macro in an equilibrium diffusion model, Technical report, Working Paper.

Brügemann, B., Gautier, P. & Menzio, G. (2019), 'Intra firm bargaining and shapley values', *The Review of Economic Studies* **86**(2), 564–592.

Buera, F. J., Kaboski, J. P. & Shin, Y. (2011), 'Finance and development: A tale of two sectors', *American economic review* **101**(5), 1964–2002.

Buera, F. J., Kaboski, J. P. & Shin, Y. (2021), 'The macroeconomics of microfinance', *The Review of Economic Studies* **88**(1), 126–161.

Dávila, E. & Schaab, A. (2023), 'Optimal monetary policy with heterogeneous agents: Discretion, commitment, and timeless policy'.

Davila, J., Hong, J. H., Krusell, P. & Ríos-Rull, J.-V. (2012), 'Constrained efficiency in the neoclassical growth model with uninsurable idiosyncratic shocks', *Econometrica* **80**(6), 2431–2467.

De Mel, S., McKenzie, D. & Woodruff, C. (2019), 'Labor drops: Experimental evidence on the return to additional labor in microenterprises', *American Economic Journal: Applied Economics* **11**(1), 202–35.

Donovan, K. (2021), 'The equilibrium impact of agricultural risk on intermediate inputs and aggregate productivity', *The Review of Economic Studies* **88**(5), 2275–2307.

Donovan, K., Lu, W. J. & Schoellman, T. (2023), 'Labor market dynamics and development', *The Quarterly Journal of Economics* **138**(4), 2287–2325.

Duflo, E. & Banerjee, A. (2011), *Poor economics*, Vol. 619, PublicAffairs.

Feng, Y., Lagakos, D. & Rauch, J. E. (2018), Unemployment and development, Technical report, National bureau of economic research.

Feng, Y. & Ren, J. (2021), Skill bias, financial frictions, and selection into entrepreneurship, Technical report.

Franklin, S. (2018), 'Location, search costs and youth unemployment: experimental evidence from transport subsidies', *The Economic Journal* **128**(614), 2353–2379.

Fujimoto, J., Lagakos, D. & VanVuren, M. (2023), Aggregate and distributional effects of 'free'secondary schooling in the developing world, Technical report, National Bureau of Economic Research.

Hall, R. E. & Schulhofer-Wohl, S. (2018), 'Measuring job-finding rates and matching efficiency with heterogeneous job-seekers', *American Economic Journal: Macroeconomics* **10**(1), 1–32.

Herreño, J. & Ocampo, S. (2021), 'The macroeconomic consequences of subsistence self-employment'.

Hosios, A. J. (1990), 'On the efficiency of matching and related models of search and unemployment', *The Review of Economic Studies* **57**(2), 279–298.

Huggett, M. (1993), 'The risk-free rate in heterogeneous-agent incomplete-insurance economies', *Journal of economic Dynamics and Control* **17**(5-6), 953–969.

Itskhoki, O. & Moll, B. (2019), 'Optimal development policies with financial frictions', *Econometrica* **87**(1), 139–173.

Lagakos, D., Mobarak, A. M. & Waugh, M. E. (2023), 'The welfare effects of encouraging rural–urban migration', *Econometrica* **91**(3), 803–837.

Lagakos, D. & Waugh, M. E. (2013), 'Selection, agriculture, and cross-country productivity differences', *American Economic Review* **103**(2), 948–980.

Levinsohn, J., Rankin, N., Roberts, G., Schöer, V. et al. (2014), 'Wage subsidies and youth employment in south africa: Evidence from a randomised control trial', *Work. Pap* **2**, 14.

McKenzie, D. (2017), 'How effective are active labor market policies in developing countries? a critical review of recent evidence', *The World Bank Research Observer* **32**(2), 127–154.

Moll, B. (2014), 'Productivity losses from financial frictions: Can self-financing undo capital misallocation?', *American Economic Review* **104**(10), 3186–3221.

Porzio, T., Rossi, F. & Santangelo, G. V. (2021), The human side of structural transformation, Technical report, National Bureau of Economic Research.

Poschke, M. (2019), 'Wage employment, unemployment and self-employment across countries'.

Schoar, A. (2010), 'The divide between subsistence and transformational entrepreneurship', *Innovation policy and the economy* **10**(1), 57–81.

Smith, E. (1999), 'Search, concave production, and optimal firm size', *Review of Economic Dynamics* **2**(2), 456–471.

Stole, L. A. & Zwiebel, J. (1996), 'Intra-firm bargaining under non-binding contracts', *The Review of Economic Studies* **63**(3), 375–410.

Wu, D. Q. & Wang, S. Z. (2023), Desperate talents, concerned employers, and employment agencies in ethiopia, Technical report, Working Paper.

Appendix

A. Additional Tables and Figure

Table A.1: Effect of Search Subsidy on Labor Market Outcomes (Abebe et al. 2021)

Outcome	Control Mean	Effect of Subsidy
Any Work	0.526	0.037
		(0.029)
Hours Worked	26.18	0.183
		(1.543)
Monthly Wages	857.9	65.88
		(63.86)
Permanent Job	0.171	0.033*
		(0.018)
Formal Job	0.224	0.054**
		(0.019)
Job Satisfaction	0.237	-0.001
		(0.027)

This table reproduces the primary results of Abebe et al. (2021) and displays the control mean for a variety of labor market outcomes as well as the experimentally estimated treatment effect of a conditional cash transfer to job seekers.

B. Derivations and Proofs from Section 2.6

The first result to show is that the entrepreneur's optimal choice of f' and n' satisfy $\eta(z;X) = \frac{\gamma f'^*}{n'^*}$ for some function η depending only on z and X. Substituting in the wage determination equation (which the entrepreneur takes as given) and the vacancy posting constraint, the first-order condition for f' and n' can be combined with the envelope condition for f and n to generate

$$\beta\Delta\mu'\Bigg((1-\alpha)(1-\chi)z(\frac{\gamma f'}{n'})^{\alpha} - \big((1-\chi)\underline{w} - \frac{c}{p(\theta(X'))}(1-\lambda)\big)\Bigg) = \frac{c}{p(\theta(X))}\mu$$

$$\beta\Delta\mu'\Bigg(\gamma\alpha(1-\chi)z(\frac{\gamma f'}{n'})^{\alpha-1} + 1 - \gamma(r+\delta)\Bigg) = \mu$$

where μ is the Lagrange multiplier on the budget constraint and $\theta(X')$ is a price function. Combining these two equations, substituting in η, and defining A, $B(X')$, and $C(X')$ for clarity yields

$$Az\eta^{\alpha} + B(X, X')z\eta^{\alpha-1} + C(X, X') = 0 \tag{21}$$

which, for $0 < \alpha < 1$, can be shown to have a unique and positive solution for η for any value of z, X, and X'. Call this solution $\tilde{\eta}(z; X, X')$. Finally, substituting $X' = H(X)$ and defining $\eta(z; X) = \tilde{\eta}(z; X, H(X))$ completes the derivation.

The next result to show is that entrepreneurs' growth rates depend only on z and aggregate state variables. This follows almost directly from the previous result. Substituting $n = \frac{\gamma}{\tilde{\eta}(z;X)} f$ in to the budget constraint of the entrepreneur problem reveals that the RHS of the budget constraint is now linear in f and can be written

$$d + E(z, X)f' = D(z, X)f \tag{22}$$

for appropriately define functions $D(z, X)$ and $E(z, X)$ which depend only on z, X, and parameters. Because entrepreneurs possess CRRA utility, the entrepreneur problem looks similar to a cake-eating problem has the well-known solution of a constant growth rate in f depending on the values of D and E, implying that that $f' = g(z; X)f$ for some function g depending only on z, X, and parameters.

The final result is the proof of Proposition 1. By assumption, θ is now constant. Let $\hat{E}(z, \theta)$ and $\hat{D}(z, \theta)$ correspond to E and D with with $\theta(X)$ simply replaced by θ (this can be done because E and D both depend on X only through $\theta(X)$). Then we have the explicit solution[22]

$$\begin{aligned}
\hat{g}(z, \theta) &= \left(\beta\Delta\frac{\hat{D}(z, \theta)}{\hat{E}(z, \theta)}\right)^{\frac{1}{\sigma}} \\
&= \left(\beta\Delta\frac{((1-\chi)\gamma z\hat{\eta}(z;\theta)^{\alpha-1} - \left((1-\chi)\underline{w} - \frac{c}{p(\theta)}(1-\lambda)\right)\frac{\gamma}{\hat{\eta}(z;\theta)} + \left(1-\gamma(r+\delta)\right)}{\left(1 + \frac{c}{p(\theta)}\frac{\gamma}{\hat{\eta}(z;\theta)}\right)}\right)^{\frac{1}{\sigma}}
\end{aligned}$$

The chain rule yields $\frac{d\hat{g}}{d\theta} = \frac{\partial\hat{g}}{\partial c/p(\theta)}\frac{dc/p(\theta)}{d\theta} + \frac{\partial\hat{g}}{\partial\hat{\eta}}\frac{d\hat{\eta}}{dc/p(\theta)}\frac{dc/p(\theta)}{d\theta}$. Using either direct calculation of partial derivatives or implicit differentiation (in the case of $\frac{d\hat{\eta}}{dc/p(\theta)}$), we

[22] While this solution to a "generalized cake eating problem" is straightforward, I have been unable to locate this exact formulation of the problem anywhere. As such, a derivation is available upon request.

can express each individual piece as

$$\frac{\partial \hat{g}}{\partial c/p(\theta)} = -\frac{1}{\sigma}\hat{g}^{1-\sigma}\left(\frac{(\frac{\hat{g}}{\beta\Delta}-1)+\lambda}{\frac{\eta}{\gamma}+\frac{c}{p(\theta)}}\right) \leq 0$$

$$\frac{\partial \hat{g}}{\partial \hat{\eta}} = \frac{1}{\sigma}\hat{g}^{1-\sigma}\left(\frac{\frac{\beta\Delta}{\hat{g}}-\frac{\hat{g}}{\beta\Delta}}{\frac{\eta}{\gamma}+\frac{c}{p(\theta)}}\right) \leq 0$$

$$\frac{d\hat{\eta}}{dc/p(\theta)} = \frac{\gamma\big(\alpha(1-\chi)z\hat{\eta}^{\alpha-1}-(r+\delta)\big)+\lambda}{J(\theta)} > 0$$

where $J(\theta)$ is a placeholder for a complex but unambiguously positive expression (note that the second expression simplifies some terms using the first order condition for f').

It is worth commenting briefly on why the claimed inequalities hold. Both the first and second inequalities follow directly from the fact that an optimizing entrepreneur will ensure that $g \geq \beta\Delta$ (an entrepreneur can always choose to select $k = 0, n = 0$ and simply eat their cake, yielding $g = \beta\Delta$, so this acts as a lower bound on all growth rates). The final expression follows from the fact that the presence of a collateral constraint ensures that the marginal product of capital ($\alpha(1-\chi)z\hat{\eta}^{\alpha-1}$) is always larger than the marginal cost of capital ($r+\delta$).

Returning to the main results and noting that $\frac{dc/p(\theta)}{d\theta} > 0$ by assumption, combining these inequalities with the chain rule shows that $\frac{d\hat{g}}{d\theta} < 0$ and $\frac{d\hat{\eta}}{d\theta} > 0$. The result for $\frac{\partial \hat{g}}{\partial\theta\partial z}$ is straightforward. We have $\frac{\partial \hat{g}}{\partial z} = \frac{1}{\sigma}\hat{g}^{1-\sigma}\big(\frac{(1-\chi)\hat{\eta}^{\alpha}}{\frac{\hat{\eta}}{\gamma}+\frac{c}{p(\theta)}}\big)$ which is also clearly greater than zero and decreasing in θ. Although this result holds only for partial derivatives (i.e. with $\hat{\eta}$ being held constant), it can also be shown to hold for total derivatives in the case where $\hat{\eta} \geq \alpha(1+\frac{c}{p(\theta)}\gamma)$ by applying the chain rule as above and computing $\frac{d\hat{\eta}}{dz}$ using implicit differentiation.

C. Derivations and Proofs from Section 3 [Need to do pass to align notation and results across all sections]

First, I formally define the functions v and H introduced in equation 16.

$$v(m_t, \eta_t, \eta_{t+1}, g_t) = \frac{1}{p(\theta)} \int \left[g_t(z)\Delta \frac{\eta_t(z)}{\eta_{t+1}(z)} - (1-\tilde{\lambda})\right] \int m_t(a,z)da + \frac{\hat{D}(z,\theta_t,\eta_t(z))\gamma \underline{f}}{\eta_{t+1}(z)} h(z)\, dz \tag{23}$$

$$H(z, m_t, \eta_t, \eta_{t+1}, g_t) = \frac{\left[g_t(z)\Delta \frac{\eta_t(z)}{\eta_{t+1}(z)} - (1-\tilde{\lambda})\right] \int m_t(a,z)da + \frac{\hat{D}(z,\theta_t,\eta_t(z))\gamma \underline{f}}{\eta_{t+1}(z)} h(z)}{p(\theta) v(m_t, \eta_t, \eta_{t+1}, g_t)} \tag{24}$$

The numerator is the number of matches with a productivity z entrepreneur and the denominator is the total number of matches.

The problem of the constrained social planner is given sequentially by

$$\begin{aligned}
\max_{\{c_t, a'_t, s_t, \theta_t, m_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \int\int\int u(c_t) m_t(a,z) j(y) dy da dz \\
s.t.\ \ a'_t + c_t &= Ra + (1-s_t)y + s_t(w_t(z) - (1-z)b) \ \ \forall a, y, z \\
a_{t+1} &\geq 0 \\
s_t(a,z) &\in \{0,1\} \\
\frac{v(m_t, \eta_t, \eta_{t+1}, g_t)}{\theta_t} &= \int\int s_t(a,0) m_t(a,0) j(y) dy da \\
m_{t+1}(a'_t, 0) &= m_t(a,0) - \theta_t p(\theta_t) \int s_t(a,0) m_t(a,0) j(y) dy \\
m_{t+1}(a'_t, z) = (1-\tilde{\lambda}) m_t(a,z) &+ H(z, m_t, \eta_t, \eta_{t+1}, g_t) \theta_t p(\theta_t) \int s_t(a,0) m_t(a,0) j(y) dy
\end{aligned} \tag{25}$$

where the functions η_t and g_t arise from the slightly modified sequential problem of an entrepreneur:

$$\begin{aligned}
&\max_{\{d_t, f_{t+1}, k_t, n_t, v_t\}} \sum_{t=0}^{\infty} (\beta\Delta)^t \frac{c_t^{1-\sigma}}{1-\sigma} \\
s.t.\ \ d_t + f_{t+1} &= (1-\chi) z k_t^{\alpha} n_t^{1-\alpha} - (r+\delta)k_t - (1-\chi)\underline{w} n_t + f_t - c v_t \\
n_{t+1} &= (1-\lambda) n_t + p(\theta_t) v_t \qquad (26) \\
k_t &\leq \gamma f_t \\
f_0 &\in \mathbb{R}
\end{aligned}$$

so that $\eta_t = \frac{\gamma f_t}{n_t}$ and $g_t = \frac{f_{t+1}}{f_t}$.[23] Note that here I have suppressed the initial condition of the planner's problem and imposed the scale-invariance of the entrepreneurs optimal capital-labor ratio and growth rate by leaving the initial condition f_0 arbitrary.

In analysis of the problem of the social planner, it will be useful to note that while η_t and g_t are potentially functions of z and the entire sequence of labor market tightness $\{\theta\}_{t=0}^{\infty}$, solving the entrepreneur's problem reveals that they depend only on ability z and current and future tightness θ_t, θ_{t+1} and thus can be written as $\eta_t(z, \theta_t, \theta_{t+1})$ and $g(z, \theta_t, \theta_{t+1})$. The independence of entrepreneur policy functions from values of θ beyond period $t+1$ follows directly from the linearity of the hiring cost, combined with the parameter assumptions that ensure that any operating entrepreneur will choose $v_t > 0$ each period. While the continuation value of an entrepreneurs labor force depends in theory on the whole sequence of labor market tightness, the ability to re-optimize at linear cost tomorrow ensures that this continuation value is equal to the "liquidation value" of the workforce next period.

[23]Even here in the appendix I opt to write the planner's problem for the case of no autocorrelation in individuals' self-employment productivity (i.e. y is drawn from $j(y)$ each period). Including autocorrelation is conceptually simple and involves adjusting only the final two inequalities governing the evolution of the distribution m_t (and the integral in the objective function); however, doing so leads to prohibitively cumbersome notation and adds no additional insight.

C.1. Notes and Proof for Proposition 2

The dynamic terms in equation 19 are given by

$$
\begin{aligned}
&\text{Anticipation Terms} = \\
&\frac{S}{\bar{\theta}}\Bigg(\mu_{t-2}\big(\frac{\partial v_{t-2}}{\partial \eta_{t-1}}\frac{\partial \eta_{t-1}}{\partial \theta_t}\big) + \mu_{t-1}\big(\frac{\partial v_{t-1}}{\partial \eta_{t-1}}\frac{\partial \eta_{t-1}}{\partial \theta_t} + \frac{\partial v_{t-1}}{\partial \eta_t}\frac{\partial \eta_t}{\partial \theta_t}\big) + \mu_t\big(\frac{\partial v_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \theta_t}\big)\Bigg) + \\
&\bar{\theta}p(\bar{\theta})S\Bigg(\int_z \lambda_{t-2}(z)\big(\frac{\partial H_{t-2}}{\partial \eta_{t-1}}\frac{\partial \eta_{t-1}}{\partial \theta_t}\big)dz + \int_z \lambda_{t-1}(z)\big(\frac{\partial H_{t-1}}{\partial \eta_{t-1}}\frac{\partial \eta_{t-1}}{\partial \theta_t} + \frac{\partial H_{t-1}}{\partial \eta_t}\frac{\partial \eta_t}{\partial \theta_t}\big)dz + \\
&\int_z \lambda_t(z)\big(\frac{\partial H_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \theta_t}\big)dz\Bigg)
\end{aligned}
\tag{27}
$$

where μ_t and $\lambda_t(z)$ are the shadow prices associated with the constraints on aggregate labor market tightness and productivity-specific matching rates respectively. These terms essentially capture the welfare gains from anticipatory hiring when labor market tightness is changed. While the welfare changes from permanent changes in hiring are captures in the other terms of equation 19, this term captures the small gains that occur due to the fact that some of this hiring is done in anticipation of the change, shifting some hiring forward temporally.

Proof: The first step is to rewrite the planner's problem to eliminate the binary choice of s_t which complicates analysis. It's fairly straightforward to show that, for utility functions exhibiting diminishing marginal utility, the optimal choice of s_t takes the form of a cutoff rule in a above which individuals search and below which they do not (this fact arises directly from the fact that c_t^* is monotonically increasing a conditional on s_t and diminishing marginal utility). Thus we can rewrite the planner's problem as selecting an optimal cutoff s_t, which is differentiable. I also rewrite the planner's problem in recursive form to simplify analysis.

$$
\begin{aligned}
V(\theta_{-2},\theta_{-1},m) &= \max_{c,a',s,\theta,m'} \int\int\int u(c)m(a,z)j(y)dydadz + \beta V(\theta_{-1},\theta,m') \\
s.t.\;\; a'+c &= Ra + (1-\mathsf{St}(a-s))y + \mathsf{St}(a-s)(w(z)-(1-z)b) \;\; \forall a,y,z \\
a' &\geq 0 \qquad\qquad (28)\\
\frac{v(m,\eta,\eta',g)}{\theta} &= \int_s^\infty m(a,0)da \\
m'_e(a',0) &= \int \tilde{\lambda} m(a,z)dz \\
m'_u(a',0) &= m(a,0) - \mathsf{St}(a-s)\theta p(\theta) m(a,0) \\
m'_e(a',z) &= (1-\tilde{\lambda})m(a,z) \\
m'_u(a',z) &= H(z,m,\eta,\eta',g)\mathsf{St}(a-s)\theta p(\theta)m(a,0) \\
m'(x,0) &= m_e(x,0) + m_u(x,0) \\
m'(x,z) &= m_e(x,z) + m_u(x,z)
\end{aligned}
$$

where $\mathsf{St}(x)$ is the step function defined via the integral of Dirac's delta δ_x.

Because the state variable describing the distribution of agents across states m is a function $\mathbb{R}^2 \to \mathbb{R}$, the value function V is technically a functional and making progress requires dipping into functional analysis. I keep things relatively simple and try to align notation as closely as possible to what is typical in more standard situations. To this end, define the following shorthand to capture the notion of a "derivative of y with respect to the value of m at point (a,z)":

$$
\frac{dy}{dm(a,z)} \equiv \frac{d}{d\epsilon} y(m+\epsilon\delta_a\delta_z)\Big|_{\epsilon=0}
$$

With this defined, we can proceed.

The first order condition with respect to s yields

$$
\frac{\lambda(s,0)}{m(s,0)}(s+b) = \theta p(\theta)\big(\int \mu(s,z)H(z)dz - \mu(s,0)\big) + \tau \qquad (29)
$$

where λ and τ are the Lagrange multipliers on the budget and theta constraints respectively. We can then generate a pair of envelope conditions with respect to $m(s,z)$ and $m(s,0)$ (note that I have used the first order condition for s to eliminate

τ from both).

$$\frac{1}{m(s,z)}\frac{dV}{dm(s,z)} = \int u(c)j(y)dy + \big(g(z)\Delta\frac{\eta}{\eta'} - (1-\tilde{\lambda})\big)\Big(\int H(z)\mu(s,z)dz - \mu(s,0)\Big)$$
$$- \frac{\lambda(s,0)}{\theta p(\theta)m(s,0)}(s+b)\big(g(z)\Delta\frac{\eta}{\eta'} - (1-\tilde{\lambda})\big) + \tilde{\lambda}\mu(s,0) \tag{30}$$
$$+ (1-\tilde{\lambda})\mu(s,z) + \big(g(z)\Delta\frac{\eta}{\eta'} - (1-\tilde{\lambda})\big)(\tilde{\omega}_1 - \tilde{\omega}_2)$$

$$\frac{1}{m(s,0)}\frac{dV}{dm(s,0)} = \int u(c)j(y)dy + \mu(s,0) + \lambda(s,0)(s+b) \tag{31}$$

where $\mu(a,z)$ and $\mu(a,0)$ are the Lagrange multiplier on the constraints governing the evolution of m' and $(\tilde{\omega}_1, \tilde{\omega}_2)$ are defined in the discussion at the endof this section.

We can then use these conditions to generate an expression for $\int H(z)\frac{1}{m(s,z)}\frac{dV}{dm(s,z)}dz - \frac{1}{m(s,0)}\frac{dV}{dm(s,0)}$ which should be interpretted as the planner's increase in value from moving one (normalized) unit of workers into employment while obeying the constraint that fracion $H(z)$ of workers must be matched with an entrepreneur of productivity z.

We also have from the first order conditions on $m'(s,z)$ and $m'(s,0)$[24]

$$\int H(z)\mu(s,z)dz - \mu(s,0) = \beta\Bigg(\int H(z)\frac{1}{m(s,z)}\frac{dV}{dm'(s,z)}dz - \frac{1}{m(s,0)}\frac{dV}{dm'(s,0)}\Bigg) \tag{32}$$

Combining this expression with the expression for the RHS referenced above, restricting to steady-state, and solving for the desired quantity yields

$$\int H(z)\mu(s,z)dz - \mu(s,0) = \frac{\beta\int\int H(z)\big(u(c_z) - u(c_0)\big)j(y)dydz}{1 - \beta\int H(z)g(z)\Delta dz} + \text{Drift Terms} \tag{33}$$

where c_z and c_0 are notation-saving shorthand for $c(a,z,y)$ and $c(a,0,y)$ respectively, and the drift terms are discussed further below. This term can be plugged directly in to the first order condition with respect to s.

[24]This phrase should be interpreted as intuitive shorthand for the first order conditioned generated by examining a delta-perturbation of m' at (a,z) i.e. $\frac{d\mathcal{L}}{dm'(a,z)}$ in the shorthand defined above.

With the hard part done, all that remains is to use the first order condition for θ to find the following expression for τ:

$$\tau = \frac{\theta / \int_s^\infty m(a,0)da}{\frac{d\log v}{d\log\theta} - 1}\Bigg(\underbrace{(1 + \frac{d\log p}{d\log\theta})\int_s^\infty \big(\int H(z)\mu(a,z)dz - \mu(a,0)\big)m(a,0)p(\theta)da}_{\text{Congestion}}$$

$$+ \underbrace{\int\int\int \lambda(a,z,y)\frac{dw}{d\theta}j(y)dydzda}_{\text{Capital Shallowing}} + \underbrace{\theta p(\theta)\int_s^\infty \int \frac{dH(z)}{d\theta}\mu(a,z)dzm(a,0)da}_{\text{Composition of Jobs}}\Bigg) \quad (34)$$

$$+ \underbrace{\beta\frac{dV}{d\theta_{-1}}}_{\text{Anticipation}}$$

Finally, plugging everything in to the first order condition for s shows that the planner assigns an individual in state $(a,0)$ to search if and only if

$$u'(c_0)\big(y+b\big) \le \beta\theta p(\theta)\frac{\int\int H(z)\big(u(c_z) - u(c_0)\big)j(y)dydz}{1 - \beta\int H(z)g(z)\Delta dz} + \text{Drift Terms} + \tau \quad (35)$$

The exact formulation of the decision rule used in Proposition 2 can be found simply by letting $\sigma \to 0$ and noting that the drift terms collapse to zero in this limit, concluding the proof.

Discussion of Drift Terms: The drift terms in the planner's decision rule serve as adjustments for the fact that the marginal job-seeker has a different level of asset holdings than the average job-seeker and, similarly, that the marginal newly employed individual has different assets than the average employed individual. Essentially, they adjust for the fact that the asset level of searchers will "drift" away from s over time.

$$\text{Drift Terms} = \frac{\big(1 - \frac{\int_s^\infty m(a,0)da}{m(s,0)}\big)\lambda(s,0)(y+b) + \big(g(z)\Delta\frac{\eta}{\eta'} - (1-\tilde{\lambda})\big)(\tilde{\omega}_1 - \tilde{\omega}_2)}{1 - \beta\int H(z)g(z)\Delta dz}$$

$$\tilde{\omega_1} = \int \frac{\int m(a,z)\mu(a,z)\text{St}(a-s)m(a,0)da}{m(s,z)\mu(s,z)m(s,0)}dz$$

$$\tilde{\omega_2} = \int \frac{\int\int \big(g(x)\Delta - (1-\tilde{\lambda})\big)H(x)m(a,x)\mu(a,x)\text{St}(a-s)m(a,0)dadx}{\big(g(z)\Delta - (1-\tilde{\lambda})\big)m(s,z)\mu(s,z)m(s,0)}dz$$

To see this, note that the drift terms collapse to zero when the distribution of asset holdings among both the employed and unemployed are concentrated at s (i.e. $m(a,z) = \delta_s m$ and $m(a,0) = \delta_s(1-m)$). Further analysis of this term is possible but involves substantial technical complication (due to the necessity of tracking the evolution of assets over time) and provides very little additional insight.

No (Additional) Externalities in Savings Decision: Here I sketch the argument/proof of the fact that the presence of search does not induce an externality in individuals' savings decisions. That is, individuals facing a search tax/subsidy aligning their privately optimal search decision rule with that of the planner will choose the same savings policy function as the planner.

The approach follows that of Davila et al. (2012) and leverages a change of variables in the planner's objective function from time-space to individual-space for any finite (N period) optimization sub-problem. Consider the sub-problem of a planner facing a distribution of agents m and who has already settled on the two-period-ahead policy function a'' but must decide today's policy function a'. One could consider the maximization of the sum of today's utility (averaged over m) and tomorrow's utility (averaged over the appropriately defined m'); this is the period approach and is how the planner's problem in (25) is written. One could alternatively consider the maximization of the *two period* utility for all agents alive in the first period (i.e. averaged over m) — the individual approach. These two objects are different ways of computing the same quantity.

The approach above lets us consider the following optimization problem:

$$\begin{aligned}\max \int\int\Bigg(&\int u(Ra - \text{Inc}(y,a,s,z) - a')j(y)dy \\ &+ \beta\mathbb{E}\Big[\int u(Ra' - \text{Inc}(y,a',s',z') - a'')j(y)dy|s,z\Big]\Bigg)m(a,z)dadz \qquad (36)\\ \text{Inc} = (1-&\text{St}(a-s))y + \text{St}(a-s)(w(z)-(1-z)b)\end{aligned}$$

Note that all the transition dynamics across employment states z are implicit in the expectations operator (a rigorous proof would require fully specifying these details, but they can be ignored in a proof sketch).

Taking the first order condition for a' from this problem reveals that it is identical to that derived from the individual problem. Of course these first order conditions contain the policy function for s, but the assumption that the search subsidy/tax implements the planner's search policy in the decentralized economy ensures that these functions are identical. Thus the savings policies are identical, and the proof sketch is complete.

Individual's Search Decision Rule: As was the case for the planner's problem, it is easy to show that individual's search decision rule is monotonic in their assets and thus the binary search choice in the individual problem can be replaced by the choice of an asset cutoff s above which the individual will search and below which they will not. Restating the relevant portion of the individual problem (4) for convenience (with auto-correlation in y removed and the aggregate state X suppressed):

$$\begin{aligned}
V_u(a) = \max_{c,a',s} u(c) + \beta\Bigg(\big(1 - \text{St}(a-s)\theta p(\theta)\big)E_y[V_u(a')] + \text{St}(a-s)\theta p(\theta)\big(E_z[V_e(a',z)]\Bigg) \\
V_e(a,z) = \max_{c,a'} \;\; u(c) + \beta\Bigg((1-\tilde{\lambda})V_e(a',z) + \tilde{\lambda}E_y[V_u(a')]\Bigg) \\
s.t. \;\; a' + c =& (1+r)a + \big(1 - \text{St}(a-s)\big)y - \text{St}(a-s)b \text{ for } V_u \\
a' + c =& (1+r)a + w(z) \text{ for } V_e
\end{aligned}$$

The first-order condition for s then yields

$$u'(c_u)(y+b) = \beta\theta p(\theta)\big(E_z[V_e(a'_u, z)] - E_y[V_u(a'_u)]\big) \tag{37}$$

where c_u and a'_u denote the policy functions of the "unemployed" with their dependence on (a, y) suppressed. Plugging the policy functions into the value functions above and doing some careful rearranging to the RHS yields the policy rule for search.

$$u'(c_u)(y+b) \leq \beta\theta p(\theta)\frac{E_z[u(c_e)] - E_y[u(c_u)]}{1 - \beta(1-\tilde{\lambda})} + \text{Drift Terms}$$

$$\begin{aligned}
\text{Drift Terms} &= \beta(1-\tilde{\lambda})\Delta_{a'_e,a'_u} - \Delta_{s,a'_u} + \beta\big(E_y[V_u(a'_e)] - E_y[V_u(a'_u)]\big) \\
\Delta_{x,y} &= \big(E_z[V_e(x)] - E_y[V_u(x)]\big) - \big(E_z[V_e(y)] - E_y[V_u(y)]\big)
\end{aligned} \tag{38}$$

As in the planner's problem, the inclusion of curvature in the utility function induces some "drift terms" that account for the fact that individuals' savings drift away from the cutoff s over time.

It turns out that the drift terms in the privately optimal decision rule (38) and equivalent to the drift terms in the planner's decision rule (35) in the sense that both terms yield the same value when given the same policy function a'. This is not particularly surprising, as both terms simply exist to adjust for curvature in the utility function. The powerful implication of this fact is that the externalities contained in τ above (as well as the difference in discount rates) make up an exhaustive list of wedges between the privately and publicly optimal decision rules, even with curvature in the utility. Formally showing this equivalence is somewhat cumbersome; the quickest approach involves an awkward change-of-variables in the planner's problem (to make it look more like the individual's problem) and can be provided upon request.

C.2. Positive Weight on Entrepreneurs' Welfare

The results above all depend on the assumption that the planner places zero weight on the welfare of entrepreneurs. It is not too difficult to generalize this assumption and allow the planner to place arbitrary Pareto weight on entrepreneurs' utility. Doing so requires the planner to track the distribution of entrepreneurs over their individual states. When approaching this in a fully general manner (i.e. tracking this distribution via a pdf m_e), the fact that the model is written in discrete time and the fact that entrepreneurs face no uncertainty combine to generate some technical unpleasantness — the distribution of entrepreneurs becomes "concentrated" at many points (i.e. there are point masses of z-entrepreneurs with collateral $\underline{f}$ and at $g(z,\theta)\underline{f}$ with nothing in between).

While this can be handled via Dirac's delta, it is much simpler to directly impose the fact that the entrepreneur states evolve discretely. To that end, define $\{f_{t,\tau}(z)\}_{\tau=0}^{\infty}$ to be the collateral of a z-type entrepreneur born in period $t-\tau$ (which is identical for all z-types as their problem is deterministic conditional on survival) and $\{\hat{m}_{t,\tau}(z)\}_{\tau=0}^{\infty}$ to be the number of such entrepreneurs alive. From the planner's perspective, all necessary entrepreneur behavior can be inferred from these two sequences, making these the additional state variables required to generalize the problem.

The planner's objective function then becomes

$$\max \sum_{t=0}^{\infty} \beta^t \int\int\int u(c_t) m_t(a,z) j(y) dy da dz + \Lambda \sum_{t=0}^{\infty} \beta^t \int \sum_{\tau=0}^{\infty} u(d_{t,\tau}(z)) dz \qquad (39)$$

where Λ is the relative weight on entrepreneurs' welfare. This can be written recursively as

$$
\begin{aligned}
&V(\theta_{-2},\theta_{-1},m)+\Lambda V_E(\theta_{-2},\theta_{-1},f_\tau,\hat{m}_\tau)= \\
&\max \int\int\int u(c)m(a,z)j(y)dydadz+\Lambda\sum_{\tau=0}^{\infty}\int u(d_\tau(z))m_\tau(z)dz \\
&+\beta V(\theta_{-1},\theta,m')+\Lambda\beta V_E(\theta_{-1},\theta,f'_\tau,\hat{m}'_\tau)
\end{aligned}
\tag{40}
$$

The additional constraints are

$$d_\tau(z)=F(z,\eta,\theta)f_\tau(z) \tag{41}$$

$$f'_{\tau+1}(z)=g(z,\eta,\theta_{-1},\theta)f_\tau(z) \tag{42}$$

$$m_{\tau+1}(z)=\Delta m_\tau(z) \tag{43}$$

$$f'_0(z)=\left(\frac{1}{1+\frac{c}{p(\theta)}\frac{\gamma}{\eta_t(z,\theta_{-1},\theta)}}\right)\underline{f} \tag{44}$$

$$m'_0(z)=Mh(z) \tag{45}$$

where F is the multiple of collateral consumed by the entrepreneur each period ($F=(D-E(\beta\Delta\frac{D}{E})^{\frac{1}{\sigma}})$) for D,E as defined in (22).

Examining the new objective functions and constraints, it is clear that values of Λ greater than zero will the only change in the planner's optimal search policy through the appearance of new terms in the first-order condition for θ given in (34) as the search cutoff s appears nowhere in the new objective function or constraints. These new terms, which represent new search externalities, are given by

$$\sum_{\tau=0}^{\infty}\int\lambda_c(\tau,z)\frac{dF}{d\theta}f_\tau(z)dz+\sum_{\tau=0}^{\infty}\int\lambda_f(\tau,z)\frac{dg}{d\theta}f_\tau(z)dz+\int\lambda_f(0,z)\frac{dG}{d\theta}\underline{f}dz \tag{46}$$

$$\frac{1}{G}=1+\frac{c}{p(\theta)}\frac{\gamma}{\eta_t(z,\theta_{-1},\theta)} \tag{47}$$

where G is the (inverse) price of one "production unit" (i.e. one unit of collateral and $\frac{\gamma}{\eta}$ units of labor).

Using the first-order and envelope conditions arising from the new constraints and rearranging allows us to express these terms in steady-state as

$$\Lambda \sum_{\tau=0}^{\infty} \int \Bigg(\underbrace{u'(\bar{d}_\tau)\frac{dF}{d\theta}}_{\text{Current consumption change}} + \underbrace{\sum_{i=0}^{\infty} (\beta\bar{g})^i u'(\bar{d}_{\tau+i})\frac{dg}{d\theta}}_{\text{Future consumption changes}} \Bigg) \bar{f}_\tau \bar{m}_\tau dz + \tag{48}$$

$$+ \Lambda \underbrace{\beta \Bigg(\sum_{\tau=0}^{\infty} \int (\beta\bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau \Bigg) \frac{dG}{d\theta} \underline{f}}_{\text{Initial reoptimization of newborns}}$$

where I have suppressed the dependence of many outcomes on z for legibility. The first two terms correspond to changes in consumption for living entrepreneurs, both in the current period (due to changes in the multiple of collateral that is consumed F) and in future periods (due to changes in the growth rate g). The last term corresponds to the change in lifetime consumption for newborn entrepreneurs arising from changes in the price of one production unit (i.e. one unit of f and $\frac{\gamma}{\eta}$ units of labor) which impacts lifetime consumption by changing entrepreneurs' initial size.

These terms all depend on the response of the appropriate object (F, g, G) to the change in θ. With a slight abuse of notation we can use the chain rule to separate these changes into those that occur due to changes in the optimal capital-labor ratio η and those that do not via $\frac{dF}{d\theta} = \frac{\partial F}{\partial \theta} + \frac{\partial F}{\partial \eta}\frac{\partial \eta}{\partial \theta}$ where the first partial derivative is taken while holding η (which depends on θ) constant. The expression in (48) can then be split into two terms, one depending on θ and one depending on η.

$$\Lambda \underbrace{\Bigg(\sum_{\tau=0}^{\infty} \int \big(u'(\bar{d}_\tau)\frac{\partial F}{\partial \theta} + \sum_{i=0}^{\infty} (\beta\bar{g})^i u'(\bar{d}_{\tau+i})\frac{\partial g}{\partial \theta} \big) \bar{f}_\tau \bar{m}_\tau dz + \beta \big(\sum_{\tau=0}^{\infty} \int (\beta\bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau \big) \frac{\partial G}{\partial \theta} \underline{f} \Bigg)}_{\text{Market Thickness}} + \tag{49}$$

$$\Lambda \frac{\partial \eta}{\partial \theta} \underbrace{\Bigg(\sum_{\tau=0}^{\infty} \int \big(u'(\bar{d}_\tau)\frac{\partial F}{\partial \eta} + \sum_{i=0}^{\infty} (\beta\bar{g})^i u'(\bar{d}_{\tau+i})\frac{\partial g}{\partial \eta} \big) \bar{f}_\tau \bar{m}_\tau dz + \beta \big(\sum_{\tau=0}^{\infty} \int (\beta\bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau \big) \frac{\partial G}{\partial \eta} \underline{f} \Bigg)}_{\text{Capital Shallowing}} \tag{50}$$

This, finally, is the most intuitive expression of the additional externalities that emerge when putting positive weight on entrepreneurs' welfare.[25] The first term, labeled "Market Thickness", captures the logic of the textbook market thickness externality. When search increases (and thus θ decreases) and the cost of hiring a worker declines, entrepreneur resources are freed up which increase consumption through a variety of channels. Lower hiring costs mean that entrepreneurs can consume a larger multiple of their collateral each period as fewer profits must be allocated towards hiring for a given growth rate ($\frac{\partial F}{\partial \theta}$) and that entrepreneurs are able to grow their size and thus their consumption faster ($\frac{\partial g}{\partial \theta}$). Finally, lower hiring costs mean that newborn entrepreneurs are able to start at a higher initial size as fewer initial resources must be allocated towards hiring the initial workforce, increasing lifetime consumption ($\frac{\partial G}{\partial \theta}$).

The second term corresponds to the same effects that occur due to changes in the capital-labor ratio η arising from the change in θ. In particular, a decline in the capital-labor ratio as a result of search leads to an increase in the profit per unit of capital, and thus per unit of collateral. This higher profit then increases the multiple of collateral consumed, firm growth, and entrepreneurs' initial size, all of which lead to higher consumption. I refer to this as the "Capital Shallowing" externality, mirroring the similar externality arising due to changes in η impacting households' consumption. This further highlights the sense in which the capital shallowing externality is a so-called monopolist externality. Changes in η serve to reallocate income between households and entrepreneurs, and the size of the Pareto weight Λ determines the extent to which the planner wishes to act as a monopolist on behalf of household or firms.

The final step is to add these terms into the first-order condition for θ described

[25] Although in theory this term can be further simplified by leveraging the fact that CRRA utility allows the various summations to be expressed as geometric series, this obscures rather than illuminates the underlying intuition.

in (34) in order to write the full expression in the case of $\Lambda > 0$.

$$
\begin{aligned}
\tau_\Lambda = \frac{\theta / \int_s^\infty m(a,0)da}{\frac{d\log v}{d\log\theta} - 1}\Big[& \underbrace{(1 + \frac{d\log p}{d\log\theta})\int_s^\infty \big(\int H(z)\mu(a,z)dz - \mu(a,0)\big)m(a,0)p(\theta)da}_{\text{Congestion}} \\
& + \underbrace{\theta p(\theta)\int_s^\infty \int \frac{dH(z)}{d\theta}\mu(a,z)dzm(a,0)da}_{\text{Composition of Jobs}} \\
& + \frac{\partial\eta}{\partial\theta}\underbrace{\Big(\int\int\int \lambda(a,z,y)\frac{dw}{d\eta}j(y)dydzda}_{\text{Capital Shallowing (Workers)}} \qquad (51) \\
& + \underbrace{\Lambda\sum_{\tau=0}^{\infty}\int \big(u'(\bar{d}_\tau)\frac{\partial F}{\partial\eta} + \sum_{i=0}^{\infty}(\beta\bar{g})^i u'(\bar{d}_{\tau+i})\frac{\partial g}{\partial\eta}\big)\bar{f}_\tau\bar{m}_\tau dz + \beta\big(\sum_{\tau=0}^{\infty}\int(\beta\bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau\big)\frac{\partial G}{\partial\eta}\underline{f}\Big)}_{\text{Capital Shallowing (Entrepreneurs)}} \\
& + \underbrace{\Lambda\sum_{\tau=0}^{\infty}\int \big(u'(\bar{d}_\tau)\frac{\partial F}{\partial\theta} + \sum_{i=0}^{\infty}(\beta\bar{g})^i u'(\bar{d}_{\tau+i})\frac{\partial g}{\partial\theta}\big)\bar{f}_\tau\bar{m}_\tau dz + \beta\big(\sum_{\tau=0}^{\infty}\int(\beta\bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau\big)\frac{\partial G}{\partial\theta}\underline{f}}_{\text{Market Thickness}} \\
& + \underbrace{\beta\big(\frac{dV}{d\theta_{-1}} + \Lambda\frac{dV_e}{d\theta_{-1}}\big)}_{\text{Anticipation}}
\end{aligned}
$$

Plugging this expression in to the first order condition for s yields the an identical planner's optimal search rule but with the additional capital shallowing and market thickness externalities present.

C.3. Relationship to the Hosios (1990) Condition

Given that the main results in this paper have focused on characterizing the efficient level of search, it is natural to wonder if something similar to the famous Hosios (1990) condition holds and if there are (relatively simple) conditions on parameters that lead the competitive allocation to be efficient. This turns out not to be the case in general once all the model features are added; however, building from a simplified version of the model in which the Hosios (1990) condition (almost) holds to the baseline result presented in Proposition 2 is valuable as it provides some intuition for how the main results of this paper fit into the broader search literature.

Some parameters assumptions greatly reduce the model's complexity and eliminate some externalities. The first step is to choose $\sigma = 0$ so that workers and entrepreneurs exhibit linear utility and choosing a Pareto weight of 1 so that the planner values all consumption equally. I then take the capital share parameter α to be 0 (i.e. a linear-in-labor production technology) which eliminates the monopolist externalities related to the capital-labor ratio (refer to 51). Taking the distribution of entrepreneur productivity z to be a point mass eliminates the Allocative Efficiency externality (as entrepreneurs become homogeneous) and, further, letting the collateral constraint parameter $\gamma \to \infty$ allows entrepreneurs to expand immediately to any size they desire, eliminating the Firm Size externality. Finally, for simplicity I set the cost of search b to 0 and assume that there is no auto-correlation in self-employment earnings.

Because these parameter assumptions leave behind only the Congestion and Market Thickness externalities, it is intuitive that something similar to the Hosios condition may arise. To generate the Hosios condition exactly, we need to make two adjustments to the model.

First, we must adjust workers' outside option in the bargaining protocol. Rather than non-cooperating workers drawing from the distribution of self-employment productivities and engaging in self-employment for a period, we allow them to choose between engaging in self-employment for the period or engaging in job search and "selling" any jobs obtained to other (jobless) workers (with a price equal to the value of the job, something agreed upon by all workers due to linear utility). Although somewhat contrived, this small adjustment is necessary to align the outside option used in wage bargaining with the self-employment option available to unemployed workers (which otherwise differs slightly). The second adjustment is to allow the planner to dictate the policy functions of entrepreneurs (which, under these simplifications, boils down to the choice of how many vacancies to post) rather than being constrained to choose values consist with entrepreneur optimality.

Under these assumptions, the thresholds in income below which individuals opt to search are given by

$$\begin{aligned} s_c &= \beta\theta p(\theta)\frac{\chi(z-\tilde{y}_c)}{1-\beta(1-\tilde{\lambda})} \\ s_p &= \beta\theta p(\theta)\frac{-\epsilon_{p,\theta}(z-\tilde{y}_p)}{1-\beta(1-\tilde{\lambda})} \end{aligned} \tag{52}$$

$$\tilde{y}_x = \int_{y<s_x} yj(y)dy + s_x(1-J(s_x))$$

where s_c, s_p denote the competitive and planner policies respectively and $\epsilon_{p,\theta}$ is the elasticity of $p(\theta)$. The term $\tilde{y}_x$, defined for compactness, summarizes the value generated by an unemployed individual — y for those who draw incomes below the threshold and s for those who search.

Here, it is clear that the Hosios condition — equality between the bargaining parameter χ and (the negative of) the elasticity of the matching function $\epsilon_{p,\theta}$ — ensures that the competitive and planner search thresholds align.[26]

With this baseline established, we can undo our two adjustments to the model (i.e. modifying the outside option and giving the planner the ability to dictate the decisions of entrepreneurs) and recover the following thresholds for the unmodified model.

$$\begin{aligned} s_c &= \beta\theta p(\theta)\frac{\chi(z-\tilde{y}_c)-(1-\chi)(\tilde{y}_c-\bar{y})}{1-\beta(1-\tilde{\lambda})} \\ s_p &= \beta\theta p(\theta)\frac{\left(1+\frac{1+\epsilon_{p,\theta}}{\epsilon_{v,p}\epsilon_{p,\theta}-1}\right)(z-\tilde{y}_p)-\left(1+\frac{1}{\epsilon_{v,p}\epsilon_{p,\theta}-1}\right)\left(\lambda+\frac{1-\beta}{\beta}\right)\frac{c}{p(\theta)}}{1-\beta(1-\tilde{\lambda})} \end{aligned} \tag{53}$$

Relative to (52), two adjustments appear, one in the competitive cutoff (blue) and one in the planner's cutoff (red). These make clear why the previous model modifications were necessary to generate the Hosios condition exactly. The blue term appearing in the equation for s_c arises as a result of the bargaining protocol, which imposes outside options for workers that are different than the options available to the unemployed. This term adjusts for this fact, leading to a slightly

[26]The presence of χ and $\epsilon_{p,\theta}$ terms without additional terms of $1-\chi$ and $1-\epsilon_{p,\theta}$ (which appear in textbook models) arises from the fact adding a choice between search and self-employment leads to two first-order conditions (one for search and one for vacancies). The missing terms are contained in the first order condition for the optimal level of vacancies (not written here).

lower cutoff (i.e. $\tilde{y}_c - \bar{y}$ is positive).[27]

The second modification, appearing in red in the equation for s_p, arise from the fact that the planner is now only able to dictate the decisions of households and, consequently, is constrained by the fact that the number of vacancies posted responds to changes in the vacancy filling probability $p(\theta)$. As a result, the planner cares about the "net" elasticity of θ with respect to the number of searchers, given by $\frac{1}{\epsilon_{v,p}\epsilon p,\theta-1}$, which when multiplied by the elasticity of the job finding probability with respect to θ, $1+\epsilon_{p,\theta}$, yields the net change in the job finding probability due to a change in search. For the planner, the surplus generated by the marginal searcher is equal to the surplus generated by the average searcher ($z - \tilde{y}_p$) adjusted by this net congestion effect.

The remaining red terms involving $\frac{c}{p(\theta)}$ also arise as a direct result of the removal of the planner's ability to dictate vacancy postings. In the modified model of (52), the impact of vacancies on entrepreneur consumption was accounted for in the first-order condition for vacancies. Now that planner can only control θ through search decisions, this impact is incorporated into the planner's optimal decision rule for search (and arises via the same net elasticity $\frac{1}{\epsilon_{v,p}\epsilon p,\theta-1}$). This is the Market Thickness externality of search.

The final step in bridging the gap between the Hosios condition in (52) and the decision rule in the full model (i.e. 51) is to write (53) in wage terms using both the wage bargaining equation ($w = \chi z + (1-\chi)\bar{y}$) and the optimal vacancy posting condition ($z - w - (\lambda + \frac{1-\beta}{\beta})\frac{c}{p(\theta)} = 0$) and do some simple rearranging.

$$
\begin{aligned}
s_c &= \beta\theta p(\theta)\frac{w-\tilde{y}_c}{1-\beta(1-\tilde{\lambda})} \\
s_p &= \beta\theta p(\theta)\frac{w-\tilde{y}_p}{1-\beta(1-\tilde{\lambda})} + \tau
\end{aligned} \tag{54}
$$

$$
\tau = \frac{1}{\epsilon_{v,p}\epsilon_{p,\theta}-1}\left(\underbrace{\frac{(1+\epsilon_{p,\theta})(z-\tilde{y}_p)}{1-\beta(1-\tilde{\lambda})}}_{\text{Congestion}} - \underbrace{\frac{(\lambda+\frac{1-\beta}{\beta})\frac{c}{p(\theta)}}{1-\beta(1-\tilde{\lambda})}}_{\text{Market Thickness}}\right)
$$

This expression for the cutoffs is, finally, the simplified model's equivalent of the

[27] As we will see shortly, this term "disappears" once the cutoffs are written in wage terms but, for now, thinking in terms of productivity behooves comparison between (52) and (53).

search decision rule in Proposition 2 or (more completely) in equation (51). The expressions for the Congestion and Market Thickness externalities are simpler and the Firm Size, Capitial Shallowing, and Efficiency externalities are zero as a result of the simplifying parameter choices. We can then relax these choices one-by-one to add these externalities back to τ and eventually end up at the expression in (51) and, from there, set the Pareto weight on entrepreneurs' utility to 0 and apply linear utility ($\sigma = 0$) to end up with the expressions in Proposition 2.

D. Details on Model Estimation and Quantitative Exercises

Many model parameters are chosen to match values typical in the macroeconomics, are taken from external sources, or are estimated directly. These are displayed in Table D.1, along with their values and sources. The discount rate β is chosen to match an annual discount rate of 0.95. Because a model period corresponds to two weeks, this is corresponds to a value of $0.95^{\frac{1}{26}}$. The rate of return on worker's savings R is taken to be exogenously equal to $0.9^{\frac{1}{26}}$. The assumption that the return to savings is less than one is typical models of developing countries (see e.g. Donovan 2021, Fujimoto, Lagakos & VanVuren 2023) and representative of the fact that households in these countries lack access to formal investment with positive returns. The value of 0.9 matches an annual inflation rate of roughly 10 percent, roughly consistent with World Bank estimates of inflation in Ethiopia over the last few years; thus the model asset a most closely reflects cash holdings. The capital share of income is set at 0.33 as is standard.

For the matching function, I use a simple generalized urn-ball matching function so that $p(\theta) = \frac{1-e^{-\zeta\theta}}{\theta}$. This particularly choice of functional form is unimportant beyond the fact that it introduces a free parameter that can be used to target any desired elasticity for p in the model steady-state.[28] Absent detailed estimates of this elasticity in the context of Addis Ababa and lacking the necessary data to estimate it, I choose ζ to generate an elasticity of -0.3. This is a fairly typical value and roughly in line with the estimates of Hall & Schulhofer-Wohl (2018) for the United States.

The interest rate faced by entrepreneurs is disciplined using World Bank MIX Market data containing financial information on microcredit providers in Ethiopia. Yields on loans from microfinance institutions range from 20 percent to 30 percent

[28] A functional form exhibiting a constant elasticity, such as Cobb-Douglas, would be ideal, but the use of discrete time limits sensible choices for p to those that lead both $p(\theta)$ and $\theta p(\theta)$ to be bounded between zero and one.

with negligible loan loss rates (typically less than one percent). Combining this rough average of a 25 percent annual return with 8 percent depreciation yields a depreciation-inclusive user cost of capital of 33 percent annually. This value is high relative to developed countries but is fairly typical for developing countries (see e.g. Banerjee et al. 2015, who document similar values in multiple countries including Ethiopia).

Collateral constraints are measured directly using data from the Ethiopian portion of the World Bank Enterprise Survey for the year 2015. The average collateral requirement reported by firms is slightly larger than 350 percent of loan value, meaning that a firm that owned 350,000 Birr worth of capital could pledge this as collateral and finance a loan for an additional 100,000 Birr of capital. Thus the implied value for γ is $1 + \frac{1}{3.5} = 1.29$. The Enterprise Survey is also used to estimate the entrepreneur survival probability Δ. Because productivity is constant for the life of an entrepreneur, entrepreneur death is the only reason that firms will shutdown in steady state. Consequently, the steady-state distribution of firm ages is geometric with decay parameter Δ whose value can be recovered through the simple maximum likelihood estimation. In this case, the estimate for Δ is given by $1 - \frac{1}{\hat{\mu}}$ where $\hat{\mu}$ is the sample average firm age, yielding an annual value for Δ of 0.92.

The self employment productivity process also comes directly from data. This productivity is modeled as a simple binary Markov process, drawing on the fact that earnings for those without permanent wage jobs are highly bimodal at a fortnightly frequency (seen in the high-frequency data of Abebe et al. 2021, described below). Such bimodality seems to stem from the fact that opportunities for self employment (or, often in the case of Addis Ababa, temporary ”gig-style” labor that functions similarly to self employment), and many individuals report neither working nor searching in a given period, presumably earning very little.

One advantage on using a binary income process instead of a more typical AR(1) is that transitions in and out of this idle state can be observed and measured directly. Using fortnightly data on work and searcher activities (described in the next section), I estimate the transition probabilities from engaged in self employment or temporary work to idleness and back. Although there is no reason for these transitions probabilities to be identical, the estimated value for both is approximately 11 percent. While average self employment earnings (i.e. the productivity parameter A_s) are estimated using SMM, the ratio of earnings in the low productivity state to the high productivity state is chosen to match the standard deviation of self employment earnings observed in the data. In particular, I isolate the transitory, idiosyncratic variance of earnings by regressing (log) earnings on individual and week fixed effects and calculating the standard deviation of the residuals (similar to the process employed in Lagakos & Waugh 2013). Conditional on the transition probabilities, there is a one to one correspondence between the standard deviation of income and the ratio of interest.[29]. The estimated ratio is 0.38 corresponding to an estimated standard deviation of .48.

Finally, the distribution from which newborn entrepreneurs draw their productivity is chosen to be an upper-truncated Pareto distribution (truncated as a bounded support for productivity is required for steady-state equilibrium to exist in the model). I set the lower bound of the distribution to a small but arbitrary number; because entrepreneurs endogeneous shut down below a threshold productivity level and the truncated Pareto distribution is scale-invariant, the lower threshold has no impact on model outcomes as long as it is below the shutdown threshold. The tail parameter is set to unite. It is worth noting that because of upper truncation, the mean and variance of productivity remain finite. The upper bound $\bar{z}$ is included in the SMM estimation, described below.

D.1. Computing the Contribution of the Firm Size Externality

Consider the laws of motion for total employment at z-type firms in period $t+i$ (denoted m_{t+i}) as a function of the number of searchers in period in t (denoted S_t). Writing the model with $\sigma \to 0$ in order to stay consistent with Proposition 2 and simplify expressions yields

$$\begin{aligned} m_{t+1}(z;S_t) &= (1-\tilde{\lambda})m_t(z) + H(z)\theta_t p(\theta_t) S_t \\ m_{t+i+1}(z;S_t) &= \underbrace{(1-\tilde{\lambda})m_{t+i}(z;S_t)}_{\text{Continuing Hires}} + \underbrace{p(\theta_{t+i})v_{t+i}(S_t)}_{\text{New Hires}} \\ v_{t+i}(S_t) &= \frac{1}{p(\theta_{t+i})}\big(\Delta g(z) - (1-\tilde{\lambda})\big)m_{t+i}(z;S_t) \end{aligned} \tag{55}$$

where I have suppressed the dependence of m_{t+i+1} on the entire sequence $\{\theta_{t+n}\}_{n=t}^{i}$ to focus on the impact of S_t.

From equation (55), it is clear that total employment in $t+i+1$ depends on

[29]For a symmetric transition matrix, as is the case here, this correspondence is given simply by $\frac{y_l}{y_h} = e^{-2\sigma}$

Table D.1: Directly Estimated Parameters

Parameter	Value	Description	Source
β	$.95^{\frac{1}{26}}$	Discount rate	Standard value
R	$.9^{\frac{1}{26}}$	Return to savings	10% annual inflation
α	$.33$	Capital share	Standard value
r	$1.33^{\frac{1}{26}}-1$	Capital cost for entrepreneurs	MIX Market
γ	1.29	Collateral constraint	World Bank ES
Δ	$.92^{\frac{1}{26}}$	Entrepreneur death prob.	World Bank ES
$M(y)$	$\begin{bmatrix} .89 & .11 \\ .11 & .89 \end{bmatrix}$	High and low y trans.	Abebe et al. (2021)
$\frac{y_l}{y_h}$	$.38$	Ratio low to high productivity	

This table displays the model parameters that are estimated directly as well as their values and sources. To help comparisons to typical values, parameters are displayed in annual terms. See the discussion for details on each parameter.

search today both through the probability that the searcher today will still be employed in $t+i$ and the (average) effect that the searcher will have on future hiring; this is not surprising as this is precisely the firm size externality. The key is to note how the impact of a change in S_t differs when the "New Hires" term is and is not allowed to adjust. When allowed to adjust, the difference between the "Continuing Hires" and "New Hires" terms can be ignored, yielding the simple formula depending on $\big(\Delta g(z)\big)^i$ in (56) below. When the "New Hires" term is not allow to change with changes in S_t, by replacing the impact of changes in employment in period $t+i$ on vacancies $\frac{dv_{t+i}}{m_{t+1}}$ with 0, we instead get the effect depending on $1-\tilde{\lambda}$ written in (57) below. This substitution of $\Delta g(z)$ for $1-\tilde{\lambda}$ is familiar — it is exactly the difference between the planner's and individuals' valuation of the benefits of search in Proposition 2.

$$\text{Both channels:} \qquad \frac{dm_{t+i+1}}{dS_t} = \big(\Delta g(z)\big)^i H(z) \frac{d}{dS_t}\big(\theta_t p(\theta_t) S_t\big) \tag{56}$$

$$\text{Continuing hires only:} \qquad \frac{dm_{t+i+1}}{dS_t} = \big(1-\tilde{\lambda}\big)^i H(z) \frac{d}{dS_t}\big(\theta_t p(\theta_t) S_t\big) \tag{57}$$

This, at last, makes it clear how to isolate the effect of the firm size channel. Similar to the approach for other externalities, we can shut down the externality by preventing the corresponding adjustment from occurring. In this case, this amounts to shrinking the change in vacancies (and the resulting job-finding probability) in period $i > 1$ by a factor of $\frac{\sum_{t=1}^{i}\left(1-\tilde{\lambda}\right)^{t}}{\sum_{t=1}\left(\int \Delta g(z)H(z)dz\right)^{t}}$, effectively imposing an evolution according to (57) rather than (56).[30]

[30]Shrinking the change in vacancies, rather than the level, is done to remain consistent with the approach applied to the other externalities which similarly measures the impact of the changes in externalities rather than levels.