Optimal Labor Market Policy in Developing Countries: A General Equilibrium Analysis

Mitchell VanVuren (Yale)

October 2024

Please click here for the latest version

Abstract

Workers in developing countries face substantial constraints to job search. Many policies aim to lower search barriers and expand the wage sector, but the efficiency and optimality of these policies remain 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 substantially increases the cost of search, rather than a subsidy.

This paper was previously circulated under the title *Job Search in Developing Countries: Crowd-In and Crowd-Out Externalities*. For their helpful comments, I thank David Argente, Kevin Donovan, Niklas Engbom, Joe Kaboski, Hannes Malmberg, Guiseppe Moscarini, Joseph Mullins, David Lagakos, Remy Levin, Kristina Manysheva, Michael Peters, Tommaso Porzio, Valerie Ramey, Pascual Restrepo, Martin Rotemburg, Todd Schoellman, Jedediah Silver, Daniela Vidart, and various seminar participants at UCSD, University of Connecticut, Yale, the NEUDC, Midwest Macro, Y-RISE Annual Conference, and the NBER Summer Institute. I am particularly indebted to Alessandra Peter for her insightful discussion of the paper. Any errors remain my own. Mitchell VanVuren: Yale Research Initiative on Innovation and Scale (Y-RISE) email: mitchell.vanvuren@yale.edu

1. Introduction

For policymakers in developing countries, creating jobs is a top economic priority. Recent empirical work, however, has documented that workers face substantial constraints that prevent them from searching for work and often spend months or even years alternating between marginal employment and job search before finally finding long-term employment (Abebe, Caria, Fafchamps, Falco, Franklin & Quinn 2021a, Abebe, Caria & Ortiz-Ospina 2021b, Donovan, Lu & Schoellman 2023). Simultaneously, the difficulty of finding and hiring the right workers is often understood to be a substantial constraint to firm growth (WDR 2013). It is not surprising, then, that there has been significant interest in policies aimed towards reducing workers' barrier to search — these offer the potential win-win of improving outcomes for workers while also allowing firms to hire the workers needed for growth. Examples of such policies include subsidies to labor search, transport, and the first few months of wages (e.g. Levinsohn et al. 2014, Franklin 2018, De Mel et al. 2019, Abebe et al. 2021a, and many others).

Even with experimental results, however, it be can difficult to determine the impact of implementing such policies for an entire labor market. If treated individuals who receive the benefits of the policy are more likely to find work, does this come at the expense of other searchers who are not in the experimental sample at all? Or does it grow the wage sector, improving future outcomes for these out-of-sample workers? Frictional search models are the standard tools for answering these questions and understanding what they imply for optimal labor market policy, but they lack many features central to developing economies, including large shares of self-employment and substantial credit and financial frictions for households and firms. Accounting for the interaction of search with these fundamental distortions is necessary to realistically evaluate search policies in this context.

This paper develops a model of frictional labor markets — accounting for these common characteristics of developing countries — and examines and quantifies the externalities of search that arise in this setting, as well as the optimal level of search. Workers have access to a frictionless self-employment option and are savings-constrained, limiting their ability to fund job search. They 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). Thus the model reproduces the empirical fact that individuals frequently and stochastically shift between self employment and job search before finding wage work (Donovan, Lu & Schoellman 2023).

Firms are owned and operated by entrepreneurs, which are treated as distinct agents from workers (as in Itskhoki & Moll 2019). Entrepreneurs are heterogeneous in productivity and operate a constant returns to scale production technology. Consequently, they 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 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 and decrease firms' ability to hire workers and provide employment in future periods.

A worker's search generates employment for themself (with some probability) but also impacts the job-finding prospects of other workers. In the short run, the searcher creates additional competition for jobs, decreasing the likelihood that others find employment and making it harder for them to break out of the cycle of self-employment and search. Consequently, too many searchers can crowd each other out, and search should be discouraged. In the long-run, however, a successful searcher finds a job and generates additional output. A portion of this output is captured by firms and used to finance growth, leading to additional hiring and employment in the future. Thus a searcher can crowd in additional employment and, in contrast to the previous effect, search should be encouraged.

The optimal labor market policy balances these tradeoffs, and I formalize this idea using a constrained social planner. The planner's problem takes the position of a government policymaker or ministry focused exclusively on labor market policy and without the ability to adjust other frictions, as this is the most realistic margin for policy intervention. For example, in Ethiopia (where I estimate the model) the Ministry and Labor and Skills can direct labor market policy but has little ability to dictate firm or financial policies. The result is a dynamic Ramsey problem that maximizes average worker utility while respecting search-and-matching frictions as well as the various budget constraints of workers and firms. Comparing the socially optimal search decision rule to the privately optimal rule reveals four search externalities, three critical economics ones and one (small) technical externality common to these types of models.[1]

As in standard models, search carries a negative externality that embeds the short-run intuition above — an additional searcher lowers today's job finding probability for all other searchers — which I refer to as the Crowd Out effect.[2] Unlike in standard models, however, firm financial frictions and heterogeneity lead to two positive externalities that arise from firm growth and potentially offset this negative effect.

The first of these positive externalities (Firm Size) arises from the effect of search on the *level* of employment in the future. A searcher has a chance to create a job today, and they internalize this benefit. They do not, however, internalize that some portion of the output generated by the match is captured by entrepreneurs and used to finance firm growth, leading to larger firms and higher levels of employment and hiring tomorrow. In addition to this level effect, an additional positive effect (Allocative Efficiency) arises from changes in the *composition* of employment — jobs at more productive firms generate more output and, as a result, the increases growth and employment are larger at more productive firms (which pay more). Thus in addition to creating more jobs, additional search increases average wages by improving allocative efficiency (at least in the long run).

Whether search should be subsidized or taxed depends on whether the positive or negative externalities dominate. As neither can be shown to dominate in general, this is ultimately a quantitative question. To answer it, 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. 2021a). 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 search subsidy in the estimated model is negative and equal to 50 of average self-employment earnings — that is, the optimal policy is

[1] The technical externality arises from the fact that the planner values only worker welfare and does not care about the welfare of entrepreneurs. As a result, the planner essentially wishes to act as a monopolist on behalf of workers and restrict labor supply above its competitive level in order to boost wages. Similar effects arise in all models that treat workers and entrepreneurs as distinct agents (e.g. Itskhoki & Moll 2019).

[2] Although conceptually similar to a standard congestion externality, the presence of both the frictionless casual sector and credit frictions modify the effect enough to justify a distinct name.

actually a significant tax on search. The gains from implementing the optimal policy are substantial — on average, welfare increases by 1.8 percent of consumption. A basic decomposition exercise reveals that the Crowd-Out and Firm Size externalities account for 1.5 and -0.8 percent of the welfare gains respectively, while the Allocative Efficiency and Monopolist effects account for only -0.0 and 0.4 percent.[3] In essence, the Crowd Out effect quantitatively dominates the positive externalities.

Why is this the case? Or more specifically, what aspect of the data drives the conclusion that the Crowd Out effect is large? In the model, a searcher creates jobs both directly through a higher level of search — as in traditional search models — but also indirectly as firms respond in equilibrium to a looser labor market by posting additional vacancies (which then generate jobs). If this indirect effect is small and creates few jobs, the Crowd Out effect will be large and the optimal policy will be a tax on search. This turns out to be the case in the estimated model, driven mostly by the fact that the estimated cost of posting vacancies is small. Because vacancies (and thus hiring costs) are cheap, the reduction in hiring costs due to falling market tightness does little to impact firms' desired levels of employment, and firms respond little to additional searchers.

In model estimation, the vacancy posting cost is chosen to match firm's reported marginal cost of hiring (i.e. the cost needed to hire one additional worker) equal to 120 percent of average wages, and this moment ends up being the primary determinant of the optimal policy. An quantitative experiment confirms this to be the case. After increasing the marginal cost of hiring by a factor of 5 (to 600 percent of average wages) and re-estimating the model, the optimal policy changes from tax to a subsidy equal to 25 percent of average earnings.

In essence, the estimated model suggests that firms are not particularly constrained by hiring costs. As a result, the marginal searcher represents only a minor reduction in constraints to firm growth and, consequently, generates few additional jobs — the negative Crowd Out externality is large and the searcher substantially reduces the probability that other searchers find jobs. Because workers are estimated to be fairly risk averse (with coefficient of relative risk aversion of 3.6), this decline in job finding probabilities leads to a substantial reduction in welfare. The result is the model's conclusion that the marginal searcher contributes

[3] Due to the nature of the decomposition exercise, the sum of these individual impacts is not necessarily equal to the overall welfare impact of 1.9 percent.

negatively to social welfare, and search should be taxed.

The final section of this paper pivots from normative to positive analysis and examines the aggregate impact of a search subsidy. The purpose of this section, following in the spirit of the literature on using macroeconomic models to interpret experimental results (e.g. Brooks, Donovan & Johnson 2020, Lagakos, Mobarak & Waugh 2023), is to highlight the importance equilibrium adjustment in determining the impact of the optimal policy. A two-thirds subsidy to search costs appears to be a promising policy in partial equilibrium (i.e. without the impact of externalities) and both increases the size of the wage sector (from 30 to 48 percent) and improves welfare (by 2.4 percent). However, these gains are substantially muted once equilibrium adjusts — although the wage sector still grows (by 38 percent), welfare declines by about half of a percent. Although the subsidy substantially reduces the amount of that workers pay per period to search, the dominance of Crowd Out leads to a decline in job finding probabilities, extending the average required duration of search. On net, the total cost (including opportunity cost) of finding one job is essentially unchanged by the subsidy. Thus, although a subsidy is effective at reducing the cost of search in partial equilibrium, the dominance of Crowd Out substantially mutes this effect in general equilibrium.

Overall, the surprising conclusion of this paper is that labor markets in developing countries (at least those similar to the labor 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, run the risk of being 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 even 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. Franklin (2018) and Abebe et al. (2021a) both evaluate RCTs of search subsidies. 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.

2. Model

Time is discrete. There is measure one of individuals (workers) and an endogenous measure of entrepreneurs. Workers consume, save, and choose between working in self-employment or participating in the frictional labor market for wage jobs. 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 (jobs) is given by a homogeneous of degree 1 matching function $m(S, V)$ where S is the measure of individuals searching for a job and V is the number of vacancies posted by firms. When convenient, $\theta = \frac{V}{S}$ is defined to be labor market tightness so that $p(\theta) \equiv m(\frac{1}{\theta}, 1) = \frac{m(S,V)}{V}$ is the probability that any vacancy is filled and $\theta p(\theta) =$

$\frac{m(S,V)}{S}$ 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, representing the idea that worker or firm needs may exogenous change in a way that causes either to terminate the match.

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.[4]

Labor Decisions and Earnings: Any individual can engage in self-employment and operate a linear-in-labor self-employment technology given by $y_t = A_s l_t$. By assumption, self-employment uses only an individual's own labor and does not involve hiring workers from outside the household. For simplicity, I normalize A_s to unity and assume that a workers' "effective labor" l_t follows Markov process M so that self-employment earnings are effectively exogenous, given by the process

$$y_{t+1} \sim M(y_t) \tag{2}$$

Instead of engaging in self-employment, an individual can choose to pay the 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 engages in wage work until they are hit by an exogenous separation shock (in equilibrium, no workers will endogenously quit). Wages are determined via bargaining (discussed later) and depend on the productivity of the entrepreneur with whom the individual is matched, given by z_t.[5]

[4] Quantitative experiments with allowing workers to choose any mixture of work and search, rather than being restricted to committing to a single choice for an entire period, 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. Thus the assumption that time is supplied indivisibly is of little quantitative consequence.

[5] Section 2.4 shows that the bargained wage depends only on the productivity of the entrepreneur and not on other entrepreneur or individual state variables.

In total, a worker's earnings are given by

$$\text{Earnings}_t = (1-e_t)\big((1-s_t)y_t - s_t b\big) + e_t w_t(z_t) \tag{3}$$

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.

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. $a_t \geq 0$), and their budget constraint is given by

$$a_{t+1} + c_t = (1+r)a_t + \text{Earnings}_t \tag{4}$$

Search: Search is undirected, and every vacancy has an equal probability of being filled. Conditional on matching with any job, an individual's probability of being matched with job that will pay $w(z)$ is denoted $H(z;X)$ (where X is a vector of aggregate state variables) and is equal to the share of vacancies posted by z-type entrepreneurs. Note here the implicit restriction that w and H depend only on z and not on other worker or entrepreneur state variables. This restriction will be justified in a later section

Separation: Employed workers are separated from their jobs with exogenous probability λ, representing the idea that worker or firm needs may change over time, resulting in the termination of matches. Workers can also lose their jobs is 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 and assumed throughout the rest of the paper), it can be shown that downsizing never occurs in equilibrium. 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

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

$$V_e(a,z;X) = \max_{c,a'} \frac{c^{1-\sigma}}{1-\sigma} + \beta\Bigg((1-\tilde{\lambda})V_e(a',z;X') + \tilde{\lambda}E_{y'}[V_u(a',y';X')]\Bigg) \tag{5}$$

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

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.

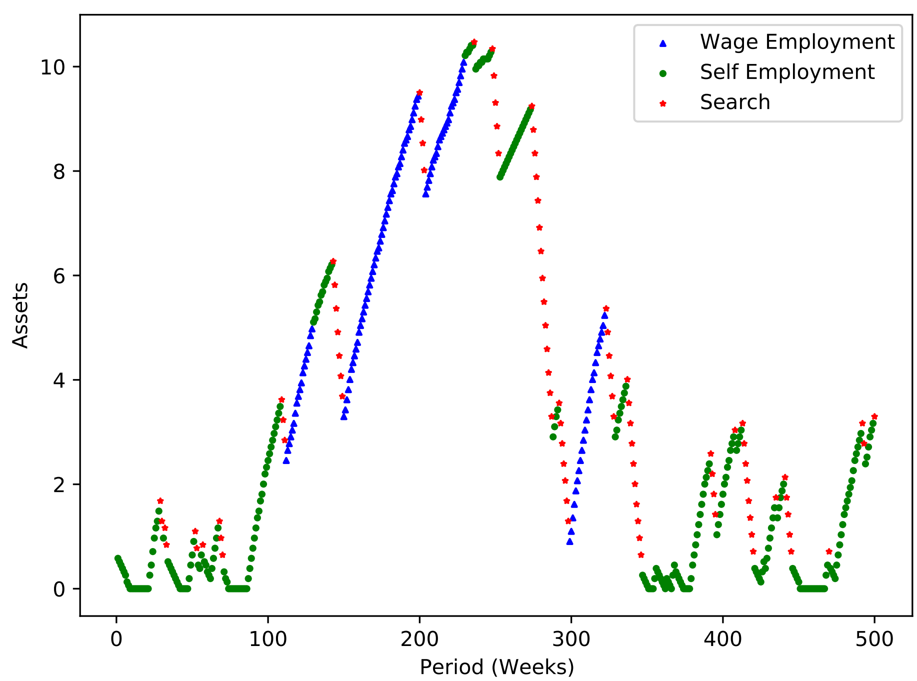
2.3. Worker Behavior

Workers weigh the costs of search, the explicit cost b and the opportunity cost of forgone self-employment earnings, against the benefit, the probability of finding high-paying wage work. This benefit, however, is highly uncertain, particularly if the probability of finding a job is small, as it is in many developing countries. It has a fairly small chance of yielding a very large benefit (a successful match) but high probability of yielding no benefit (no match) and leaving the worker substantial worse off than if they opted for the lower but guaranteed income of self-employment.

Because workers are risk-averse, face incomplete markets, and have no way to insure against risk other than self-insurance via asset accumulation, only those with enough assets to maintain a reasonable level of consumption in the likely situation that they fail to find a job will opt for search. Those without much self-insurance will opt for the guaranteed income in self-employment. For those that search, the search cost quickly depletes their savings and reduces their self-insurance, eventually driving them to self-employment until they can re-accumulate assets.

The result is that workers 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. The exact cutoff in savings above which households decide to search depends on their self-employment income y_t which is stochastic, leading to unpredictability in the exact timing of switches.

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.

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

2.4. Entrepreneurs

Entrepreneurs operate wage-sector firms and employ workers. Including entrepreneurs as distinct agents instead of an occupational choice for workers (as in Buera, Kaboski & Shin 2021) reflects the qualitative difference between "subsistence self-employment" and productive entrepreneurship with the potential to grow and employ many workers (e.g. Schoar 2010), 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 $N \to \infty$.[6] At the end of a period, entrepreneurs die with probability 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

$$\sum_{t=0}^{\infty} (\beta\Delta)^t \frac{d_t^{1-\sigma}}{1-\sigma} \tag{6}$$

Each entrepreneur operates a Cobb-Douglas production technology that depends on their ability, rents capital from the international capital market at an exogenous rental cost $(r+\delta)$, and pays workers wage w_t (determined by bargaining)

[6] The assumption that there are an infinite number of atomic entrepreneurs rather than a measure 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.

so that period profits are given by

$$\pi_t(z, k_t, n_t) = z k_t^{\alpha} n_t^{1-\alpha} - (r+\delta)k_t - w_t n_t \tag{7}$$

Entrepreneurs face financial frictions and must use their own assets f_t as collateral to finance capital which restricts their size despite constant returns to scale. Their collateral constraint is given by

$$k_t \leq \gamma f_t \tag{8}$$

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.[7]

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

$$n_{t+1} = (1-\lambda)n_t + p(\theta)v_t \tag{9}$$

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.

Entrepreneurs split their profits between consumption, savings (which are used as collateral tomorrow), and hiring so that their budget constraint is given by

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

Wage Bargaining: Each period, entrepreneurs and workers bargain over wages. Because capital acts as a fixed factor of production within a period (the collateral constraint always binds in equilibrium), firm output exhibits decreasing returns to scale in labor. To accommodate this, I follow previous literature (Stole & Zwiebel 1996, Smith 1999, Acemoglu & Hawkins 2014) and model production as a cooperative game between workers and entrepreneurs in which each agent is paid their

[7] 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).

Shapley value.[8]

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 and each earns (in expectation) average self employment earnings $\bar{y}$. If the entrepreneur and x of their n workers choose to cooperate, they operate the entrepreneur's production technology, and produce $zk^{\alpha}x^{1-\alpha}$.

Each agent is paid their Shapley value arising from this game, so that the wage per worker is given by

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

where χ is a parameter governing the bargaining power of the entrepreneur relative to workers.[9] This wage determination equation is straightforward; workers are simply paid some linear combination of the average product of labor and their outside option $\bar{y}$, with the weight determined by bargaining power.

Bellman Equation: Combining equations (6) - (10) and the wage bargaining equation (11), the entrepreneur's problem can be written recursively as

$$\begin{aligned} V(z,f,n;X) = \max_{f',n',k,v,d} \; & \frac{c^{1-\sigma}}{1-\sigma} + \beta\Delta V(z,f',n';X) \\ s.t. \;\; d + f' =& (1-\chi)zk^{\alpha}n^{1-\alpha} - (r+\delta)k - (1-\chi)\bar{y}n + f - cv \\ n' =& (1-\lambda)n + p(\theta)v \\ k \leq& \gamma f \\ v \geq& 0 \\ X' =& J(X) \end{aligned} \tag{12}$$

where X is a vector of aggregate state variables and J is the entrepreneur's perceptions function for the evolution of the aggregate state. Notably the composition of

[8]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.

[9]At a technical level, the game is between an atomistic entrepreneur and a continuum of workers so that the parameter χ is the relative size of the atomistic entrepreneur.

the entrepreneur's work force does not need to be tracked as a state variable due to the fact that bargained wages depend only on z. The full definition of recursive competitive equilibrium is relegated to the appendix.

2.5. Entrepreneur Behavior

While the entrepreneur's problem is complex, it results in fairly simple steady-state behavior (largely due to the fact that an entrepreneur's cost structure is entirely linear) which is summarized in 1 below.

Proposition 1 *Consider the entrepreneur's problem (12) in steady-state recursive competitive equilibrium, and let $\bar{\theta}$ denote the steady-state value of labor market tightness.*

Then the entrepreneur's optimal policy functions for capital k^*, next period savings f'^*), and next period labor n'^* satisfy

$$k^* = \gamma f \tag{13}$$

$$f'^* = g(z, \bar{\theta}) f \tag{14}$$

$$\frac{\gamma f'^*}{n'^*} = \eta(z, \bar{\theta}) \tag{15}$$

*where g and η depend only on an entrepreneur's productivity z and steady-state tightness $\bar{\theta}$. In particular, they do **not** depend on firm size through either f or n.*

Proof: Refer to Appendix B.

In essence, Proposition 1 says that entrepreneurs pick constant capital-labor ratios (η) and growth rates (g) that do not change over the course of the entrepreneur's life. An entrepreneur is born, observes their productivity z and labor market conditions $\bar{\theta}$, chooses a capital-labor ratio and a growth rate, and sticks to those for their entire life. Together, the two functions η and g are sufficient to fully characterize entrepreneur behavior in steady-state.

In addition to lending conceptual clarity, these results are important in keeping the model tractable. Although the bargained wage $w(z)$ could depend in general on all entrepreneur and worker state variables, combining the wage bargaining equation (11) with the unchanging capital-labor ratio (15) which depends only on entrepreneur productivity z yields the result that wages also depend solely on z. Consequently, equilibrium can be computed by solving for a one-dimensional wage function, rather than a four- or five- dimensional one.

To facilitate discussion of efficiency and optimal policy, it will be useful to note a couple properties of η and g. In particular, we have that

$$\frac{d\hat{\eta}}{d\bar{\theta}} > 0 \tag{16}$$

$$\frac{d\hat{g}}{d\bar{\theta}} < 0 \tag{17}$$

The first inequality ($\frac{d\hat{\eta}}{d\bar{\theta}} > 0$) says that capital-labor ratios are increasing in labor market tightness — a tighter labor market raises hiring costs and increases the price of labor relative to capital and, as a result, entrepreneurs opt to tilt their input mix more towards capital. The second ($\frac{d\hat{g}}{d\bar{\theta}} < 0$) tells us that growth rates are decreasing in tightness — paying more for hiring diverts income that otherwise would have been saved and used to finance expansion, capturing the idea that poor labor market conditions may serve as an impediment to firm growth.

3. Efficiency and Optimal Policy

As in many models of frictional search, there is a tension between the private benefit of search that accrues to an individual worker and the public/social benefit of search. This tension stems from a disconnect between the number of jobs that workers perceive as being created by their decision to search and the number of jobs that are actually created.

Short-run Job Creation: From a worker's perspective, the number of jobs created in the immediate short-run by a period of search is $\theta p(\theta)$ — the probability of finding one job in one period of search. In reality, however, total job creation differs from the worker's perception, and a hypothetical social planner would value search differently. The marginal searcher generates $\frac{\partial m}{\partial S}$ new matches/jobs directly which can be reexpressed as $\varepsilon_{m,S}\theta p(\theta)$ (where $\varepsilon_{m,S}$ is the elasticity of the matching function with respect to the number of searchers).[10] Noting that $\varepsilon_{m,S}$ is typically less than one by assumption, this suggests that workers perceive their direct impact on job creation ($\theta p(\theta)$ new jobs) to be larger than it is in reality ($\varepsilon_{m,S}\theta p(\theta)$ new jobs), a common effect in frictional search models.

In this model, however, there is an additional source of short-run job creation not perceived by workers. Jobs are created directly through their search but are also created indirectly due to the fact that firms respond to higher levels of search

[10] Although this elasticity is not necessarily constant and, in general, depends on the number of searchers and vacancies, treating it as a constant parameter here facilitates discussion without loss of generality.

(and thus looser labor markets and lower hiring costs) by posting additional vacancies (recall equation 16) which also generate jobs. Slightly abusing notation to let $\frac{d\log V}{d\log S}$ representing ”the percent change in aggregate equilibrium vacancy postings in response to a one percent increase in the aggregate number of searchers”, we differentiate the total number of jobs created each period $m(S, V(S))$ by S to obtain the total short-run job creation by the marginal searcher

$$\underbrace{\frac{dm(S,V(S))}{S}}_{\text{Short-run Job Creation}} = \theta p(\theta)\big(\underbrace{\varepsilon_{m,S}}_{\text{Direct}} + \underbrace{(1-\varepsilon_{m,S})\frac{d\log V}{d\log S}}_{\text{Indirect}}\big) \tag{18}$$

where I have made use of the fact that $\frac{\partial \log m}{\partial \log V} = 1 - \varepsilon_{m,S}$ due to constant returns to scale.

This expression for job creation makes it clear that the aggregate response of vacancies V to an additional searcher plays a central role in governing the extent to which workers overestimate the amount of jobs created by to search. Consider, for example, the case of no collateral constraints, $\gamma \to \infty$, which corresponds to a typical free-entry-of-firms assumption. In this case, $\frac{d\log V}{d\log S}$ is equal to one as free entry pins down the value of θ. Then total job creation of the marginal searcher is equal to $\theta p(\theta)$, just as workers believe, and there is no inefficiency at all. Conversely, the difference between workers' perceptions and truth, and thus inefficiency, is maximized when $\frac{d\log V}{d\log S}$ is zero. I return to this fact in Section 4.

Long-run Job Creation: In addition to disagreement about how many jobs are created by search in the short-run, a worker and a hypothetical social planner also disagree on long-run job creation. From the worker's perspective, searching and successfully finding a job today "generates" $(1-\tilde{\lambda})$ jobs (on average) tomorrow — the probability that the worker is not hit with a separation shock and continues their employment. More generally, finding a job in period t generates $(1-\tilde{\lambda})^\tau$ jobs in period $t+\tau$.

In reality, however, a new job generated by a searcher today leads to more than $(1-\tilde{\lambda})^\tau$ jobs τ periods in the future. This fact arises from the presence of entrepreneur collateral constraints — an employed worker today leads to more employment tomorrow not only through the continuation of their own job, but also because some of the output produced by the worker is captured by the entrepreneur (via bargaining) and can be used to fund firm expansion, including hiring. In particular, combining equations (14) and (15) from Proposition 1 yields a firm-level employment evolution equation of

$$n'^{*} = g(z; \bar{\theta})n \tag{19}$$

which suggests that one additional employee (n) today will lead to g additional employees tomorrow. Adjusting for the firm survival probability (Δ), we see that a new job at a z-type firm in period t generates $\left(\Delta g(z; \bar{\theta})\right)^{\tau}$ jobs in period $t + \tau$.

3.1. The Constrained Planner's Problem

Formalizing the intuition above requires defining a social planner's problem in order to make these externalities/disagreements explicit, but it is not immediately clear what the appropriate problem is. As in much of the 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 which teaches us nothing about search externalities.

Instead, I consider the practical problem of a government agency or ministry tasked with the goal of improving welfare by implementing labor market policy and concerned only with worker welfare.[11] In particular, I assume that the planner does not have the ability to adjust or fix other market failures or frictions. Like the constrained social planner in typical search models, they must respect the search-and-matching technology. Further, I assume that the planner cannot overcome workers' credit frictions and must respect worker budget and borrowing constraints (as in Davila et al. 2012).

We also wish to tie the planner's hands in a way that prevents them from overcoming entrepreneurs' financial constraints and improving outcomes simply by reducing misallocation. If the planner had the ability to dictate vacancy postings, for example, they could force unproductive entrepreneurs to post no vacancies and allocate employment to only the most productive entrepreneurs. This is an improvement, to be sure, but not one that stems from improving search decisions. The most natural way to enforce this constraint is to simply require that the planner's decisions for entrepreneur vacancy postings, savings, and consumption be consistent with entrepreneur optimality. Intuitively, this means that the planner

[11]Introducing positive weight on entrepreneur welfare can be done, as I show in Appendix C.2, but adds very little additional insight.

can only dictate the decision of workers and cannot dictate the decisions of entrepreneurs, preventing them from improving outcomes simply by "picking winners".

Allocations satisfying these constraints make up the set of feasible allocations for the planner, leading to the following definition:

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 tightnesses $\{\theta_t\}_{t=0}^{\infty}$ 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

$$\begin{aligned} a'_t + c_t &= Ra + w_t(z_t, \theta_t) && \forall a, y, t \text{ when } z \geq 0 \\ a'_t + c_t &= Ra + (1 - s_t)y + s_t b && \forall a, y, t \text{ when } z = 0 \\ a'_t &\geq 0 \end{aligned} \tag{20}$$

2. It respects the labor market matching technology

$$\begin{aligned} \frac{v(m_t, \eta_t, \eta_{t+1}, g_t)}{\theta_t} &= \int\int s_t(a,0) m_t(a,0) j(y) dy da \\ m_{t+1}(a', z) &= (1 - \tilde{\lambda}) m_t(a,z) + H(z, m_t, \eta_t, \eta_{t+1}, g_t) p(\theta_t) v(m_t, \eta_t, \eta_{t+1}, g_t) \end{aligned} \tag{21}$$

   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.[12]

3. The entrepreneur policy functions $\{g_t(z), \eta_t(z)\}_{t=0}^{\infty}$ solve the entrepreneurs' problem (Appendix equation 30), conditional on θ_{-1} and $\{\theta_t\}_{t=0}^{\infty}$.

The task of the social planner is to maximize average worker 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

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

in individuals' self-employment productivity y (i.e. the distributions m_t are only defined over (a, z)). This simplification substantially reduces notation, improves readability, and does not change any of the underlying results or economics; I maintain it throughout the rest of the paper. 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.

3.2. Privately- vs Socially- Optimal Search Decision Rules

With the planner's problem specified, Proposition 2 formalizes the externalities that were discussed above by characterizing both the privately optimal search decision rule (Individual) and the planner's socially optimal search decision rule (Planner). Differences between the two represent search externalities that are not internalized by workers in competitive equilibrium. For clarity, I make 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.[13]

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 s_c and s_p (respectively) which are characterized by the following*

[13]The 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 in the same ways. However, the assumption of linear utility substantially improves the readability of equations (22)-(24).

equations:

$$\textit{Individual: } s_c + b = \bar{\theta}p(\bar{\theta}) \int_z \frac{\beta E(z, s_c)}{1 - \beta(1 - \tilde{\lambda})} \bar{H}(z) dz \tag{22}$$

$$\textit{Planner: } s_p + b = \bar{\theta}p(\bar{\theta}) \int_z \frac{\beta E(z, s_p)}{1 - \beta \underbrace{\Delta g(z, \bar{\theta})}_{\text{Firm Size}}} \bar{H}(z) - \underbrace{\frac{(1 - \varepsilon_{m,S})}{1 + \int \frac{\partial \log v^*}{\partial \log S} dz} \bar{\lambda}(z)}_{\text{Crowd Out}} \bar{H}(z) dz + \tag{23}$$

$$+ \frac{1}{1 + \int \frac{\partial \log v^*}{\partial \log S} dz} \Bigg(\underbrace{\bar{S}\bar{\theta}p(\bar{\theta}) \int_z \bar{\lambda}(z) \frac{\partial H}{\partial g} \frac{\partial g}{\partial S} dz}_{\text{Allocative Efficiency}} + \underbrace{\int_z \frac{\partial w}{\partial \eta} \frac{\partial \eta}{\partial S} \bar{m}(z) dz}_{\text{Monopolist}} + \textit{Anticipation Term} \Bigg)$$

$$E(z, s) = w(z) - \big(\int_{y>s} y j(y) dy + J(s)(s - b) \big) \tag{24}$$

where $E(z, s)$ represents the excess value (over self-employment) generated by a job at a z-type firm, j and J are the PDF and CDF of the distribution of y respectively, and $\lambda(z)$ is the planner's Lagrange multiplier denoting the marginal value of an additional worker being matched with a productivity z entrepreneur. The anticipation term is described further in the appendix.

The privately optimal search rule is fairly straightforward. Workers weigh the opportunity-cost-inclusive cost of search ($y + b$) against the expected benefits (the probability of getting a job next period $\bar{\theta}p(\bar{\theta})$ multiplied by the expected benefit of the job, which is given by the per-period excess earnings E and then discounted according to the expected job duration). The level of income at which workers are indifferent between searching or not, s_c, occurs when these two values are equal. Below this level, workers will search while workers above this level will engage in self-employment.

Search Externalities: Relative to the privately-optimal rule, the planner's socially-optimal rule differs in four ways that are highlighted and labeled in equation (23), and I discuss each in turn.

First, the "Firm Size" term captures the disagreement between workers and the planner over long-run job creation discussed via equation (19) above. Although workers only internalize the fact that employment today increases their own employment prospects in the future (by $(1-\tilde{\lambda})^\tau$ for period $t+\tau$), the planner accounts for the fact that entrepreneurs capture some portion of the income generated by employment today and use this to fund future growth and hiring. As a result, the planner values not only the jobs created by search today but also the jobs created in the future, which amounts to increasing the discount rate used for excess earnings by $\Delta g(z,\bar{\theta}) - (1-\tilde{\lambda})$, the amount of future employment generated by a worker each period in excess of their own continuation probability.

The "Crowd Out" term captures the disagreement over short-run job creation discussed and summarized in equation (18). Although the worker believes that a period of search creates $\bar{\theta}p(\bar{\theta})$ jobs, the planner adjusts this number downwards according to equation (18).[14] Because the marginal searcher creates fewer than $\bar{\theta}p(\bar{\theta})$ jobs, they lower the overall job finding probability. Thus an alternative and equally valid interpretation of this term can be seen by noting that $(1-\varepsilon_{m,S})$ can be re-expressed as $\frac{\partial \log \theta p(\theta)}{\partial \log S}$, the change in the overall job finding probability due to the marginal searcher. From this perspective, this term is simply the planner internalizing the overall decline in the job finding probability induced by the marginal searcher (adjusted by $\frac{1}{1+\int \frac{\partial \log v^*}{\partial \log S} dz}$ to account for the fact that some of this decline is mitigated by increased vacancy postings due to lower market tightness).

Interestingly, an additional third term, "Allocative Efficiency", related to long-run job creation appears. While the Firm Size effect accounts for the fact that a portion of the output generated by a job is used to finance hiring tomorrow and increases the *level* of employment, it turns out that there is also a *compositional* effect. An additional searcher puts downward pressure on labor market tightness and lowers hiring costs, increasing firm growth rates. However, because more productive firms wish to grow faster and thus are more constrained by hiring costs, this results in a larger increase in growth and employment among high-productivity firms relative to low-productivity firms, increasing productive firms' share of resources and improving allocative efficiency. From the planner's perspective, this manifests as higher worker welfare through an increase in the probability that workers match with highly productive firms that pay higher wages ($\frac{\partial H}{\partial g}$).

[14]The exact equivalence is easier to see by noting that the aggregate elasticity in equation (18), $\frac{d \log V}{d \log S}$, is related to the individual elasticity in equation (23), $\frac{\partial \log v^*}{\partial \log S}$, via $\frac{d \log V}{d \log S} = \frac{\frac{\partial \log v^*}{\partial \log S}}{1+\frac{\partial \log v^*}{\partial \log S}}$. This expression can be obtained by differentiating the equilibrium condition $V = \int v^*(z,S,V)dz$ with respect to S and noting that constant returns to scale of the matching function imply that $\frac{\partial \log v^*}{\partial \log V} = -\frac{\partial \log v^*}{\partial \log S}$.

The final term, "Monopolist", is a technical externality rather than an economic one. Due to the assumption that the planner values only worker welfare, the planner has an additional incentive to act on a monopolist on behalf of workers.[15] Here, it takes the form of limiting search in order to push up the capital-labor ratio and, consequently, wages ($\frac{\partial w}{\partial \eta}\frac{\partial \eta}{\partial S}$). Fortunately, this technical externality ends up being quantitatively small in the estimated model and contributing little to the optimal policy.

3.3. Implementing the Planner's Allocation

The problem of selecting the welfare maximizing path subject to feasibility conditions is similar to other Ramsey-type problems often found in the literature dealing with welfare and efficiency in heterogeneous agent models (e.g. Davila et al. 2012, Itskhoki & Moll 2019, Dávila & Schaab 2023). Like all Ramsey problems, the planner's allocation can be decentralized and implemented in competitive equilibrium using a sufficiently rich set of tax and subsidy instruments.

Implementing the planner's allocation in this way, however, turns out not to be feasible in practice. Proposition 2 seems to suggests straightforward implementation — a subsidy or tax that aligns the competitive search cutoff s_c with the planner's search cutoff s_p should be sufficient. While this does succeed in aligning workers' search decisions with those of the planner, the subsidy also alters workers' budget constraints and, as a consequence, the resulting allocation violates workers' budget constraints in condition (20) and is not feasible. Thus it cannot solve planner's problem by construction. Instead, a complex set of state-contingent lump-sum transfers, well beyond the capabilities of any developing country government, are needed to restore feasibility.

Fortunately, this issue (which is admittedly more technical than economic) is of little consequence. Recasting the planner's problem as an optimal subsidy/tax problem under a restricted set of simpler tax instruments reveals little difference between the unrestricted planner's optimal search cutoff and the optimal search cutoff under these restrictions (refer to Appendix C) — the latter looks almost identical to (23) and simply carries an extra term to adjust for the changes in budgets that arise. To minimize the impact of this term, a natural choice of instruments is a single subsidy (or tax) to search funded by a single proportional tax (or sub-

[15] This type of externality is common in models that treat workers and entrepreneurs as distinct agents with distinct utility functions (see e.g. Itskhoki & Moll 2019).

sidy) on self-employment earnings as this minimizes the extent to which the policy generates (or reduces) welfare simply by moving income between labor market states (compared to e.g. a tax on wages) and overcoming worker credit constraints. Under these restrictions, the difference between the optimal restricted ("feasible") policy and the planner's full solution turns out to be quantitatively negligible. As a result, I focus on the feasible policy for the remainder of the paper.

4. Model Estimation

Neither the positive (Firm Size, Allocative Efficiency) nor the negative (Crowd Out) externalities of the previous section can be shown to dominate in general. Consequently, determining the optimal policy — whether search should be subsidized or taxed — is a quantitative question that requires bringing the model to data. This, in turn, requires narrowing our focus to 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 an experiment useful for model estimation happened to be conducted there.

I divide model parameters into two categories. The first are parameters that can be computed directly 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 experiment, 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 directly estimated parameters whose values are important.[16]

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 fortnightly frequency, this corresponds to a value of $0.9^{\frac{1}{26}}$. 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, Fujimoto, Lagakos & VanVuren 2023). Here, encoding this assumption is impor-

[16] 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.

tant as the difficulty of maintaining a cushion of savings is an oft-cited justification for search subsidies.

For similar reasons, the income process of the self-employed is also important. This is measured directly using fortnightly data on workers and job seekers collected by Abebe et al. (2021a) 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 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 so $p(\theta) = \frac{1-e^{-\zeta\theta}}{\theta}$, with an efficiency parameter of ζ that can be adjusted to target any elasticity $\varepsilon_{m,S} = -\frac{d\log p}{d\log\theta}$ in the model steady-state.[17] 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 $\varepsilon_{m,S}$ 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 (quantitative experiments in Section 5 below show that this choice ends up having little consequence for optimal policy).

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.[18] 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

[17]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.

[18]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	3.6	% wage work
λ	0.05	Unemployment rate
χ	0.50	Wage sector premium
b	0.06	% of expenditure on search
$M\underline{f}$	.002	Control wage employment after 16 weeks
c	0.48	Cost to hire as % of wage
$\bar{z}$	0.39	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 seven parameters estimated using the simulated method of moments to match seven data moments. Table 1 lists these parameters and their estimated values while Table 2 lists the 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.[19] The wage sector premium is estimated on the same data by including a dummy variable indicating whether an individual is employed in a permanent wage job (vs self-employment or temporary work) in an otherwise standard Mincer regression

[19]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%	13%
Wage sector premium	LSMS	39%	39%
% of expenditure on search	Abebe et al. (2021a)	15%	16%
Control wage emp. after 16 weeks	Abebe et al. (2021a)	12%	11%
Cost to hire as % of wage	Abebe et al. (2017)	120%	120%
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.

of (log) earnings on age, as well as some controls (rural/urban, region, and sector fixed effects).[20]

The two remaining household moments come from the aforementioned data on job seekers from Abebe et al. (2021a). 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 permanent employment 16 weeks after baseline. Although I discuss the experiment in more detail in the next subsection, 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 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,

[20]The Ethiopian Productive Safety Net Programme (PSNP), a relatively new (circa 2018) "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 individuals employed by the PSNP from the analysis changes the estimated wage premium by less than one percentage point, rendering the issue quantitatively moot. I default to including all earnings from temporary employment.

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 level of search and, consequently, the size of the wage sector.

The earnings premium in the wage sector naturally pins down the bargaining power parameter, as higher value (more bargaining power for workers) shifts the bargained wage towards the average product of labor and away from the outside option. 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) which determines the "size" of entrepreneurs relative to workers — if entrepreneurs are larger, they will post more vacancies, leading to higher job-finding rates.

Firm moments: The remaining parameters — the vacancy posting cost c and the distribution of firm productivity — 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. For the distribution of firm productivity, I choose an upper-truncated Pareto distribution with a tail parameter of unity (Zipf's law, note that upper-truncation ensures that the mean and variance of productivity remain finite). The truncation point $\bar{z}$ is disciplined by the average (self-reported) annual growth rate for firms in the Enterprise Survey — higher $\bar{z}$ directly corresponds to a higher average growth rate due to the fact that more productive firms grow faster.

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

One possibility, absent from discussion so far, is that the intuition embedded in the model that credit constraints are a substantial driver of low levels of search is simply not true. If this were the case, it would be difficult to put any stock in the model's conclusions for optimal policy. To address this and test whether the model can explain observed search behavior, I replicate an experiment performed by Abebe et al. (2021a) in the model and perform model validation by comparing 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 during model estimation, treatment outcomes and data are not. Thus comparing the model's predicted treatment effects to those estimated in the experiment represents a valid "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. 2021a) 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.[21] 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.

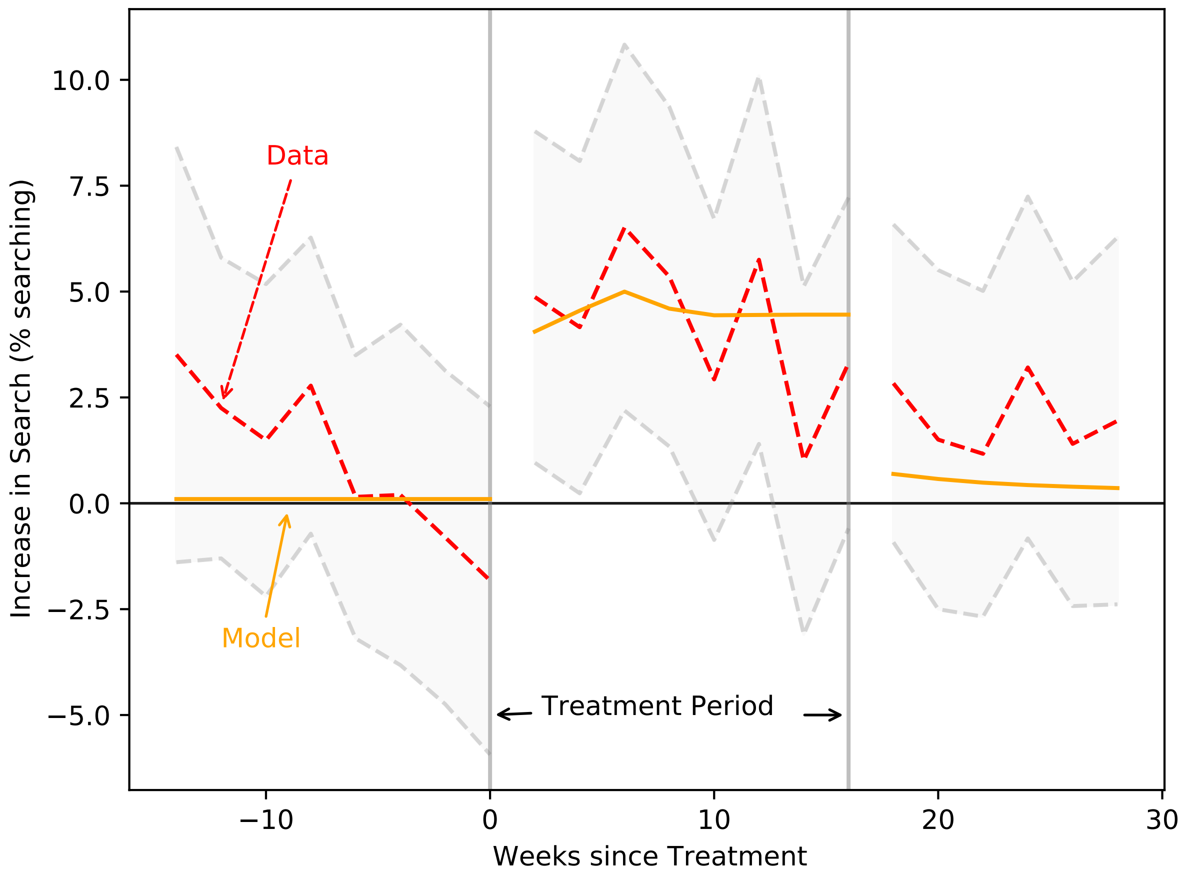
To replicate the experiment in the model, I select a representative but small (measure 0) subset of individuals not employed in the wage sector from the steady-state distribution of individuals. In this sense, the outcomes of sampled individuals do not affect equilibrium outcomes, and the experiment 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 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

[21] 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.

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.

I interpret the success of the model in predicting the response of worker search behavior to a search subsidy as evidence supporting (or, at least, failing to reject) the core idea that credit constraints are a substantial driver of search behavior.

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 various externalities laid out in Section 3. The simplest way to accomplish this is to directly solve for the optimal policy tax/subsidy rates on search and self-employment, compute the welfare gains from the optimal policy, and then decompose the total impact of the policy across the various channels. Here it is important to note that while equations (22) – (24) in Proposition 2 were expressed under the simplification of linear utility (for clarity), the results in this section are computed using the estimated model which exhibits substantial curvature ($\sigma = 3.6$).

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.

Table 3: Results of the Efficient Policy

Optimal Subsidies/Taxes			
Search Tax:	50%	Self-emp. Tax:	-2.5%
Welfare:	+1.9%	Size of Wage Sector:	-13pp
(Self-employed):	+1.9%	(Pre-policy):	29%
(Employed):	+1.7%	(Post-policy):	16%

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. For ease of comparison, the "Search Tax" and "Self-emp. Tax" values are displayed in common units of "Percentage of Average Self-Employment Earnings". See text for details.

The most surprising result is that the optimal tax rate on on search is positive and large, increasing the cost of search by 50 percent of average self-employment earnings. 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 29 percent (of workers) to 16 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 externality dominates, 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 about 2.5 percent of average earnings.

The overall impact on welfare of the policy is substantial. Average welfare increases by 1.9 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 (1.9 percent and 1.7 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 externality. On the other hand, the self-employed largely benefit through higher earnings due to the subsidy.

5.1. Decomposing the Impact of the Externalities

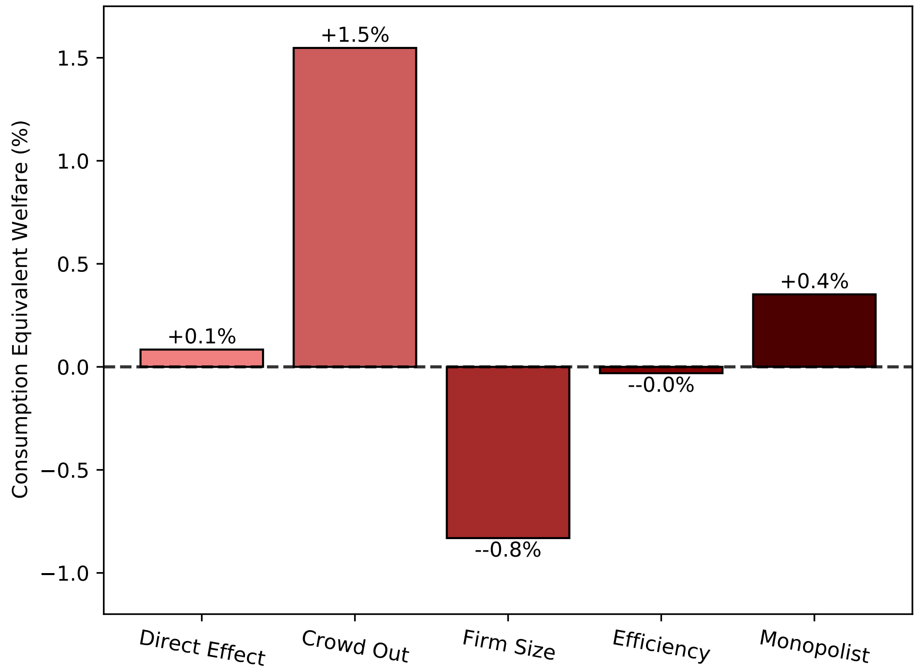
A natural way to examine an externality's individual contribution to the optimal policy is to marginally shrink the impact of that externality and re-examine what happens as a result of the policy.[22] For example, one could reduce the impact of Crowd Out by 10 percent by calculating how many more jobs would be created in the short-run if worker's private perceptions were correct and artificially increasing the number of jobs created by 10 percent of this. Computing the impact of the optimal policy under this adjustment and comparing to the policy's impact in the full model reveals Crowd Out's contribution. That is, if the optimal tax leads to a 1.9 percent increase in welfare in the baseline model and a 1.8 percent increase after this adjustment is made, we can say that 1.0 percentage points (1.9 less 1.8 divided by 10 percent) of the policy's welfare gains are attributable to Crowd Out (Appendix D.1 provides details on exactly how this computation is done for each channel).

Figure 3 displays the contribution of each of the externalities to the overall policy impact implied by this calculation. It also displays the change in welfare that occurs due to the fact that the optimal policy leads to minor changes in budget constraints (labeled "Direct Effect"; refer to subsection 3.3 for discussion). Because the optimal policy is a tax that reduces search, *negative* search externalities take on positive values, as they are the reason that the tax results in welfare gains. Similarly, positive externalities take on negative values.

Both the Crowd Out and Firm Size externalities contribute substantially, accounting for +1.5 percent and -0.8 percent of the welfare gains respectively. Meanwhile, the Allocative Efficiency and Monopolist channels make much smaller contributions, accouting for -0.0 and +0.4 percent. Finally, the Direct Effect of the policy on budgets accounts for a positive but small portion of the welfare gains, sug-

[22]Computing contributions by marginally reducing a channel (rather than entirely eliminating it) is keeping in line with the fact that the planner in Proposition 2 equates the marginal costs and benefits of search and thus cares about the marginal impact of each externality.

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.

gesting that the optimal policy under the restricted set of tax instruments is very close to what would be achieved without restrictions.[23]

The upshot of this decomposition is that the result that the optimal policy is a tax on search, rather than a subsidy, is primarily driven by the large size of the Crowd Out effect. The marginal searcher generates substantially fewer jobs than they take (in the short run), and the increases in hiring and employment in the long run stemming from the relaxation of entrepreneurs' financial constraints is not large enough to offset this effect. As a result, the competitive equilibrium exhibits too much search, and optimal policy corrects this.

[23]The fact that the contribution of each channel is computed by marginally, rather than totally, is the primary reason that summing that channels' contributions (+1.2 percent) does not yield the total welfare impact of the policy (+1.9 percent).

5.2. Driving Forces

What features of the data drive the model to the conclusion that Crowd Out is so large? The answer comes from examining the equation for short-run job creation (18), rewritten here for convenience.

$$\underbrace{\frac{dm(S,V(S))}{S}}_{\text{Short-run Job Creation}} = \theta p(\theta)\big(\underbrace{\varepsilon_{m,S}}_{\text{Direct}} + \underbrace{(1-\varepsilon_{m,S})\frac{d\log V}{d\log S}}_{\text{Indirect}}\big) \tag{25}$$

This equation makes it clear that the aggregate response of firms' vacancy postings to an increase in search ($\frac{d\log V}{d\log S}$) is the primary driver of the size of Crowd Out. A large value (close to the upper bound of one) reflects a world in which firms are highly responsive to increases in the number of searchers and the resulting declines in hiring costs. In this world, the marginal searcher creates many jobs as they induce a large increase in the number of vacancies (which create jobs) and Crowd Out is small. In contrast, a small value (close to the lower bound of zero) reflects a world where firms respond little to changes in the number of searchers and, as a result, the marginal searcher creates few jobs. In this world, Crowd Out is large.

In the model, the key determinant of firms' responsiveness turns out to be the vacancy posting cost c. If c is high, then hiring costs are large and the reduction in labor market tightness due to the marginal searcher represents a substantial reduction in a firm's total costs and frees up a substantial amount of resources that can now be allocated towards growth and expansion. As a result, firm hiring and thus vacancy postings increase substantially. On the other hand, if c is small, this reduction only represents a small portion of total costs, and there is not much expansion.

In model estimation, c is estimated to match the fact that firms report that hiring an additional worker would cost roughly 120 percent of the average (weekly) wage paid by the firm. We can test the intuition above quantitatively by artificially increasing this value by 5x to 600 percent, re-estimating the model to match this new moment, and seeing how the optimal policy changes.[24]

Performing this exercise confirms our intuition — in the re-estimated model, the optimal policy transforms from a 50 percent tax on search to a search subsidy equal to about 25 percent of average self-employment earnings. In this alternative model,

[24] An increase of 5x to 600 percent is chosen as this is roughly the largest value at which the model can maintain a reasonable fit to the data.

firms are substantially more responsive to changes in labor market tightness and the Crowd Out effect is substantially reduced. I interpret these results to conclude that the fact that the model perceives hiring costs equal to 120 percent of average weekly wages to be "small" is the primary driver of its conclusion that the optimal policy is to tax search.

What about $\varepsilon_{m,S}$?: Another potentially important parameter, based on equation (18), is the matching function elasticity $\varepsilon_{m,S}$, which governs the relative importance of searchers and vacancies in the process of job creation. Even if firms do not respond to higher search by posting many additional vacancies, this may not lead to substantial Crowd Out if vacancies are unimportant in job creation (i.e. $\varepsilon_{m,S}$ is close to unity). Although the importance of this parameter is reminiscent of the classic congestion externality of Hosios (1990), the presence of the indirect short-run job creation term, as well as the long-run Firm Size and Allocative Efficiency effects, means that comparing this elasticity to workers' bargaining share is no longer enough to determine whether the positive or negative externalities dominate.

We can do a quantitative experiment similar to the one we did for c above to test if our choice of value for $\varepsilon_{m,S}$ is important in driving our results. This is particularly important as the aggregate matching elasticity is a hard-to-pin-down-parameter, even in developed countries where there is substantially more data available (particularly on vacancy postings). In fact, the value of 0.3 used here is simply lifted from Hall & Schulhofer-Wohl (2018) who estimate the elasticity in the United States. If the elasticity in Addis Ababa is substantially different and the optimal policy depends heavily on the chosen value, this presents a threat to the validity of the quantitative results.

Fortunately, this concern appears to be unfounded. Even dramatically varying the value of $\varepsilon_{m,S}$ from 0.1 to 0.8 (from a baseline of 0.3) has little impact on the optimal policy, which remains a tax. The exact magnitude of the tax slightly varies from 60 percent (when $\varepsilon_{m,S} = 0.1$) to 30 percent (when $\varepsilon_{m,S} = 0.8$) but, generally speaking, remains very similar to the optimal tax of 50 percent of average self-employment earnings from the baseline parameterization. These results reinforce the conclusion that it is mostly firms' responsiveness to additional searchers, driven by whether hiring costs are large or small, that determines the optimal policy.

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. In this section, I pivot from normative to positive analysis and use the estimated model to examine what would happen if the optimal policy was ignored and search was subsidized anyways. I focus on both partial equilibrium results — where job finding probabilities and wages are fixed at their pre-policy values, eliminating the impact of search externalities — and general equilibrium results, where the externalities have impact. Thus this section follows the spirit of the literature on using using macroeconomic models to interpret partial equilibrium experimental results in a general equilibrium setting (e.g. Brooks, Donovan & Johnson 2020, Fujimoto, Lagakos & VanVuren 2023, Lagakos, Mobarak & Waugh 2023).

The policy considered is a subsidy to search that reduces search costs by two-thirds — the same policy implemented in the experiment used to estimate the model — funded by a tax on the self employed. Table 4 displays the results of this policy in both partial and general equilibrium.

Table 4: Result of Search Subsidies

Model	Wage Sector	Welfare
Baseline	30%	–
w/ Subsidy (Partial Eq.)	48%	2.4%
w/ Subsidy (General Eq.)	38%	-0.5%
Tax on Employed (Gen. Eeq.)	35%	+0.3%

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.

In partial equilibrium, where job finding rates and wages are fixed, the policy appears to be an extremely effective approach to growing the wage sector. The percent of the population engaged in wage work increases by 18 percentage points from 30 and 48 percent and, as a result, there is a dramatic increase in welfare equal to 2.4 percent of consumption. These large increases stem from the promising experimental results replicated by the model — the wage sector exhibits a substantial earnings premium and a search subsidy leads to a significant increase in search behavior, suggesting that workers who would otherwise substantially benefit from formal employment are credit-constrained away from search. As a result, when these constraints are reduced, employment and welfare substantially increase.

General equilibrium, however, tells a different story. Here, the policy looks much less effective. While it does increase the size of the wage sector by 8 percentage points (from 30 to 38 percent), it does so at substantial cost to welfare, which falls by about half of a percent. The reason, of course, is that allowing job finding probabilities and wages to change in response to the policy introduces search externalities and, as demonstrated above, the negative externalities dominate.

It is somewhat surprising, given the partial equilibrium results that suggest a large portion of workers are substantially credit-constrained, that the general equilibrium results are so pessimistic. Further investigating this provides some insight into the inner-workings of the model, and the key insight comes from examining the average opportunity-cost inclusive cost of finding a job — that is, the amount of consumption that a worker would forgo during an average-length job search which depends on both the search cost b (which determines the consumption given up each period) and the job-finding probability (which determines the length of search). Although in partial equilibrium this total cost falls by roughly 18 percent (due to the subsidy), the decline is only about 6 percent in general equilibrium due to a lower job finding rate.[25] Thus the subsidy ends up being a fairly ineffective way to actually reduce job-finding costs, at least once search externalities are accounted for.

As a final exercise, I consider the same subsidy (a two-thirds reduction in b) funded by a tax on wage workers, rather than a tax on the self-employed. Although taxing the self-employed makes sense from the perspective of Section 3 where one is trying to aligning the implemented policy as closely as possible to the social planner, taxing the self-employed is difficult (both operationally and politically) and a realistic policy may involve taxing wage workers instead. The results of this policy are also displayed in Table 4 and look fairly similar in general equilibrium. Wage sector growth remains small (at 5 percent) and, although the change

[25] Although this is still a decline on average, the variance in the total job finding cost increases substantially as a result of the lower job-finding probability, leading to the 0.5 percent decline in welfare overall.

in welfare is now positive (due to redistribution), the gains are still substantially smaller than what the partial equilibrium results suggest.

Overall, the result of this section is that promising experimental (partial equilibrium) results for search subsidies do not guarantee that a policy will be successful when scaled-up to a general equilibrium level. Interestingly, these results also provide further insight into why the optimal policy within the model is to tax search. In particular, it is *not* because the model concludes that there are few workers who would gain substantially from wage work but are constrained and unable to search — the partial equilibrium results show that there are many such workers. Instead, it is a result of the fact that a search subsidy is ineffective at reducing these constraints due to the dominance of the Crowd Out effect. Although a subsidy does reduce the per-period cost of search, this is mostly offset by a decline in the job-finding probability. As a result, the average cost of finding a job changes very little.

7. Conclusion

Many policies and interventions aim to expand the wage sector by increasing the extent to which (potential) workers can search for jobs. This paper develops and estimates a model that incorporates key features of developing countries in order to understand and quantify the search externalities 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. The primary reason for this conclusion is that hiring costs do not appear to be a substantial constraint to firm growth and consequently, firms do not substantially increase vacancy postings in response to higher levels of search. As a result, the negative Crowd Out externality of search is large — the marginal searcher generates few jobs and lowers the probability that other searchers find work.

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. (2021a), '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., Caria, A. S. & Ortiz-Ospina, E. (2021b), 'The selection of talent: Experimental and structural evidence from ethiopia', *American Economic Review* **111**(6), 1757–1806.

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.

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.

WDR (2013), *World development report 2013: Jobs*, The World Bank.

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. 2021a)

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. (2021a) 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 2.5

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

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

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

$$Az\eta^{\alpha} + B(X,X')z\eta^{\alpha-1} + C(X,X') = 0 \tag{26}$$

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

$$d + E(z,X)f' = D(z,X)f \tag{27}$$

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[26]

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

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

[26]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 21.

$$v(m_t, \eta_t, \eta_{t+1}, g_t) = \frac{1}{p(\theta)} \int \left[g_t(z)\Delta \frac{\eta_t(z)}{\eta_{t+1}(z)} - (1-\tilde{\lambda})\right] \int m_t(a,z)da + \frac{\hat{D}(z,\theta_t,\eta_t(z))\gamma \underline{f}}{\eta_{t+1}(z)} h(z)\, dz \tag{28}$$

$$H(z, m_t, \eta_t, \eta_{t+1}, g_t) = \frac{\left[g_t(z)\Delta \frac{\eta_t(z)}{\eta_{t+1}(z)} - (1-\tilde{\lambda})\right] \int m_t(a,z)da + \frac{\hat{D}(z,\theta_t,\eta_t(z))\gamma \underline{f}}{\eta_{t+1}(z)} h(z)}{p(\theta) v(m_t, \eta_t, \eta_{t+1}, g_t)} \tag{29}$$

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

$$\begin{aligned}
\max_{\{c_t, a'_t, s_t, \theta_t, m_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t & \int \int \int u(c_t) m_t(a,z) j(y) dy da dz \\
s.t.\ \ a'_t + c_t &= Ra + (1-s_t)y + s_t(w_t(z) - (1-z)b) \ \ \forall a, y, z \\
a_{t+1} &\geq 0 \\
s_t(a,z) &\in \{0,1\} \\
\frac{v(m_t, \eta_t, \eta_{t+1}, g_t)}{\theta_t} &= \int \int s_t(a,0) m_t(a,0) j(y) dy da \\
m_{t+1}(a'_t, 0) &= m_t(a,0) - \theta_t p(\theta_t) \int s_t(a,0) m_t(a,0) j(y) dy \\
m_{t+1}(a'_t, z) = (1-\tilde{\lambda}) m_t(a,z) + & H(z, m_t, \eta_t, \eta_{t+1}, g_t) \theta_t p(\theta_t) \int s_t(a,0) m_t(a,0) j(y) dy
\end{aligned} \tag{30}$$

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

$$
\begin{aligned}
&\max_{\{d_t, f_{t+1}, k_t, n_t, v_t\}} \sum_{t=0}^{\infty} (\beta\Delta)^t \frac{c_t^{1-\sigma}}{1-\sigma} \\
s.t.\ \ & d_t + f_{t+1} = (1-\chi) z k_t^{\alpha} n_t^{1-\alpha} - (r+\delta)k_t - (1-\chi)\underline{w} n_t + f_t - c v_t \\
& n_{t+1} = (1-\lambda)n_t + p(\theta_t) v_t \\
& k_t \leq \gamma f_t \\
& f_0 \in \mathbb{R}
\end{aligned}
\tag{31}
$$

so that $\eta_t = \frac{\gamma f_t}{n_t}$ and $g_t = \frac{f_{t+1}}{f_t}$.[27] 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.

[27] 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 **??** are given by

$$
\begin{aligned}
&\text{Anticipation Terms} = \\
&\frac{S}{\bar{\theta}}\Bigg(\mu_{t-2}\big(\frac{\partial v_{t-2}}{\partial \eta_{t-1}}\frac{\partial \eta_{t-1}}{\partial \theta_t}\big) + \mu_{t-1}\big(\frac{\partial v_{t-1}}{\partial \eta_{t-1}}\frac{\partial \eta_{t-1}}{\partial \theta_t} + \frac{\partial v_{t-1}}{\partial \eta_t}\frac{\partial \eta_t}{\partial \theta_t}\big) + \mu_t\big(\frac{\partial v_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \theta_t}\big)\Bigg)+ \\
&\bar{\theta}p(\bar{\theta})S\Bigg(\int_z \lambda_{t-2}(z)\big(\frac{\partial H_{t-2}}{\partial \eta_{t-1}}\frac{\partial \eta_{t-1}}{\partial \theta_t}\big)dz + \int_z \lambda_{t-1}(z)\big(\frac{\partial H_{t-1}}{\partial \eta_{t-1}}\frac{\partial \eta_{t-1}}{\partial \theta_t} + \frac{\partial H_{t-1}}{\partial \eta_t}\frac{\partial \eta_t}{\partial \theta_t}\big)dz + \\
&\int_z \lambda_t(z)\big(\frac{\partial H_t}{\partial \eta_t}\frac{\partial \eta_t}{\partial \theta_t}\big)dz\Bigg)
\end{aligned} \tag{32}
$$

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 **??**, 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.

$$
\begin{aligned}
V(\theta_{-2}, \theta_{-1}, m) &= \max_{c, a', s, \theta, m'} \int\int\int u(c) m(a, z) j(y) dy da dz + \beta V(\theta_{-1}, \theta, m') \\
s.t. \;\; a' + c &= Ra + (1 - \mathsf{St}(a - s))y + \mathsf{St}(a - s)(w(z) - (1 - z)b) \;\; \forall a, y, z \\
a' &\geq 0 \qquad (33) \\
\frac{v(m, \eta, \eta', g)}{\theta} &= \int_s^\infty m(a, 0) da \\
m'_e(a', 0) &= \int \tilde{\lambda} m(a, z) dz \\
m'_u(a', 0) &= m(a, 0) - \mathsf{St}(a - s)\theta p(\theta) m(a, 0) \\
m'_e(a', z) &= (1 - \tilde{\lambda}) m(a, z) \\
m'_u(a', z) &= H(z, m, \eta, \eta', g)\mathsf{St}(a - s)\theta p(\theta) m(a, 0) \\
m'(x, 0) &= m_e(x, 0) + m_u(x, 0) \\
m'(x, z) &= m_e(x, z) + m_u(x, z)
\end{aligned}
$$

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)":

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

With this defined, we can proceed.

The first order condition with respect to s yields

$$
\frac{\lambda(s, 0)}{m(s, 0)}(s + b) = \theta p(\theta)\big(\int \mu(s, z) H(z) dz - \mu(s, 0)\big) + \tau \qquad (34)
$$

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).

$$\frac{1}{m(s,z)}\frac{dV}{dm(s,z)} = \int u(c)j(y)dy + \big(g(z)\Delta\frac{\eta}{\eta'} - (1-\tilde{\lambda})\big)\Big(\int H(z)\mu(s,z)dz - \mu(s,0)\Big)$$
$$- \frac{\lambda(s,0)}{\theta p(\theta)m(s,0)}(s+b)\big(g(z)\Delta\frac{\eta}{\eta'} - (1-\tilde{\lambda})\big) + \tilde{\lambda}\mu(s,0) \tag{35}$$
$$+ (1-\tilde{\lambda})\mu(s,z) + \big(g(z)\Delta\frac{\eta}{\eta'} - (1-\tilde{\lambda})\big)(\tilde{\omega}_1 - \tilde{\omega}_2)$$

$$\frac{1}{m(s,0)}\frac{dV}{dm(s,0)} = \int u(c)j(y)dy + \mu(s,0) + \lambda(s,0)(s+b) \tag{36}$$

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 $m'(s,0)$[28]

$$\int H(z)\mu(s,z)dz - \mu(s,0) = \beta\bigg(\int H(z)\frac{1}{m(s,z)}\frac{dV}{dm'(s,z)}dz - \frac{1}{m(s,0)}\frac{dV}{dm'(s,0)}\bigg) \tag{37}$$

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

$$\int H(z)\mu(s,z)dz - \mu(s,0) = \frac{\beta\int\int H(z)\big(u(c_z) - u(c_0)\big)j(y)dydz}{1 - \beta\int H(z)g(z)\Delta dz} + \text{Drift Terms} \tag{38}$$

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.

[28]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:

$$\tau = \frac{\theta / \int_s^\infty m(a,0)da}{\frac{d\log v}{d\log\theta} - 1}\Bigg(\underbrace{(1 + \frac{d\log p}{d\log\theta}) \int_s^\infty \big(\int H(z)\mu(a,z)dz - \mu(a,0)\big) m(a,0)p(\theta)da}_{\text{Congestion}}$$
$$+ \underbrace{\int\int\int \lambda(a,z,y)\frac{dw}{d\theta}j(y)dydzda}_{\text{Capital Shallowing}} + \underbrace{\theta p(\theta)\int_s^\infty \int \frac{dH(z)}{d\theta}\mu(a,z)dzm(a,0)da}_{\text{Composition of Jobs}}\Bigg) \quad (39)$$
$$+ \underbrace{\beta\frac{dV}{d\theta_{-1}}}_{\text{Anticipation}}$$

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

$$u'(c_0)\big(y+b\big) \leq \beta\theta p(\theta)\frac{\int\int H(z)\big(u(c_z) - u(c_0)\big)j(y)dydz}{1 - \beta\int H(z)g(z)\Delta dz} + \text{Drift Terms} + \tau \quad (40)$$

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 (30) 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:

$$
\begin{aligned}
\max \int\int \Bigg(\int & u(Ra - \text{Inc}(y,a,s,z) - a')j(y)dy \\
& + \beta\mathbb{E}\Big[\int u(Ra' - \text{Inc}(y,a',s',z') - a'')j(y)dy|s,z\Big]\Bigg) m(a,z)dadz \qquad (41) \\
\text{Inc} &= (1 - \text{St}(a-s))y + \text{St}(a-s)(w(z) - (1-z)b)
\end{aligned}
$$

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 (5) for convenience (with auto-correlation in y removed and the aggregate state X suppressed):

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

The first-order condition for s then yields

$$u'(c_u)(y+b) = \beta\theta p(\theta)\big(E_z[V_e(a_u',z)] - E_y[V_u(a_u')]\big) \tag{42}$$

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.

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

$$\begin{aligned} \text{Drift Terms} &= \beta(1-\tilde{\lambda})\Delta_{a_e',a_u'} - \Delta_{s,a_u'} + \beta\big(E_y[V_u(a_e')] - E_y[V_u(a_u')]\big) \\ \Delta_{x,y} &= \big(E_z[V_e(x)] - E_y[V_u(x)]\big) - \big(E_z[V_e(y)] - E_y[V_u(y)]\big) \end{aligned} \tag{43}$$

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 (43) and equivalent to the drift terms in the planner's decision rule (40) 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

$$\max \sum_{t=0}^{\infty} \beta^t \int \int \int u(c_t) m_t(a,z) j(y) dy da dz + \Lambda \sum_{t=0}^{\infty} \beta^t \int \sum_{\tau=0}^{\infty} u(d_{t,\tau}(z)) dz \tag{44}$$

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

$$
\begin{aligned}
&V(\theta_{-2},\theta_{-1},m)+\Lambda V_E(\theta_{-2},\theta_{-1},f_\tau,\hat{m}_\tau)= \\
&\max \int\int\int u(c)m(a,z)j(y)dydadz+\Lambda\sum_{\tau=0}^{\infty}\int u(d_\tau(z))m_\tau(z)dz \\
&+\beta V(\theta_{-1},\theta,m')+\Lambda\beta V_E(\theta_{-1},\theta,f'_\tau,\hat{m}'_\tau)
\end{aligned}
\tag{45}
$$

The additional constraints are

$$d_\tau(z)=F(z,\eta,\theta)f_\tau(z) \tag{46}$$

$$f'_{\tau+1}(z)=g(z,\eta,\theta_{-1},\theta)f_\tau(z) \tag{47}$$

$$m_{\tau+1}(z)=\Delta m_\tau(z) \tag{48}$$

$$f'_0(z)=\left(\frac{1}{1+\frac{c}{p(\theta)}\frac{\gamma}{\eta_t(z,\theta_{-1},\theta)}}\right)\underline{f} \tag{49}$$

$$m'_0(z)=Mh(z) \tag{50}$$

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 (27).

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 (39) 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

$$\sum_{\tau=0}^{\infty}\int\lambda_c(\tau,z)\frac{dF}{d\theta}f_\tau(z)dz+\sum_{\tau=0}^{\infty}\int\lambda_f(\tau,z)\frac{dg}{d\theta}f_\tau(z)dz+\int\lambda_f(0,z)\frac{dG}{d\theta}\underline{f}dz \tag{51}$$

$$\frac{1}{G}=1+\frac{c}{p(\theta)}\frac{\gamma}{\eta_t(z,\theta_{-1},\theta)} \tag{52}$$

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

$$\Lambda \sum_{\tau=0}^{\infty} \int \Bigg(\underbrace{u'(\bar{d}_\tau)\frac{dF}{d\theta}}_{\text{Current consumption change}} + \underbrace{\sum_{i=0}^{\infty} (\beta \bar{g})^i u'(\bar{d}_{\tau+i})\frac{dg}{d\theta}}_{\text{Future consumption changes}} \Bigg) \bar{f}_\tau \bar{m}_\tau dz + \tag{53}$$

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

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 (53) can then be split into two terms, one depending on θ and one depending on η.

$$\Lambda \underbrace{\Bigg(\sum_{\tau=0}^{\infty} \int \big(u'(\bar{d}_\tau)\frac{\partial F}{\partial \theta} + \sum_{i=0}^{\infty} (\beta \bar{g})^i u'(\bar{d}_{\tau+i})\frac{\partial g}{\partial \theta}\big) \bar{f}_\tau \bar{m}_\tau dz + \beta \big(\sum_{\tau=0}^{\infty} \int (\beta \bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau \big) \frac{\partial G}{\partial \theta}\underline{f} \Bigg)}_{\text{Market Thickness}} + \tag{54}$$

$$\Lambda \frac{\partial \eta}{\partial \theta} \underbrace{\Bigg(\sum_{\tau=0}^{\infty} \int \big(u'(\bar{d}_\tau)\frac{\partial F}{\partial \eta} + \sum_{i=0}^{\infty} (\beta \bar{g})^i u'(\bar{d}_{\tau+i})\frac{\partial g}{\partial \eta}\big) \bar{f}_\tau \bar{m}_\tau dz + \beta \big(\sum_{\tau=0}^{\infty} \int (\beta \bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau \big) \frac{\partial G}{\partial \eta}\underline{f} \Bigg)}_{\text{Capital Shallowing}} \tag{55}$$

This, finally, is the most intuitive expression of the additional externalities that emerge when putting positive weight on entrepreneurs' welfare.[29] 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

[29] 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 (39) in order to write the full expression in the case of $\Lambda > 0$.

$$
\begin{aligned}
\tau_\Lambda = \frac{\theta / \int_s^\infty m(a,0)da}{\frac{d\log v}{d\log\theta} - 1} \Big[& \underbrace{(1 + \frac{d\log p}{d\log\theta}) \int_s^\infty \big(\int H(z)\mu(a,z)dz - \mu(a,0)\big) m(a,0)p(\theta)da}_{\text{Congestion}} \\
& + \underbrace{\theta p(\theta) \int_s^\infty \int \frac{dH(z)}{d\theta}\mu(a,z)dzm(a,0)da}_{\text{Composition of Jobs}} \\
& + \frac{\partial\eta}{\partial\theta} \underbrace{\Big(\int\int\int \lambda(a,z,y)\frac{dw}{d\eta} j(y)dydzda}_{\text{Capital Shallowing (Workers)}} \qquad (56) \\
& + \underbrace{\Lambda \sum_{\tau=0}^{\infty} \int \big(u'(\bar{d}_\tau)\frac{\partial F}{\partial\eta} + \sum_{i=0}^{\infty}(\beta\bar{g})^i u'(\bar{d}_{\tau+i})\frac{\partial g}{\partial\eta}\big) \bar{f}_\tau \bar{m}_\tau dz + \beta\big(\sum_{\tau=0}^{\infty} \int (\beta\bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau \big) \frac{\partial G}{\partial\eta} \underline{f} \Big)}_{\text{Capital Shallowing (Entrepreneurs)}} \\
& + \underbrace{\Lambda \sum_{\tau=0}^{\infty} \int \big(u'(\bar{d}_\tau)\frac{\partial F}{\partial\theta} + \sum_{i=0}^{\infty}(\beta\bar{g})^i u'(\bar{d}_{\tau+i})\frac{\partial g}{\partial\theta}\big) \bar{f}_\tau \bar{m}_\tau dz + \beta\big(\sum_{\tau=0}^{\infty} \int (\beta\bar{g})^\tau u'(\bar{d}_\tau)\bar{m}_\tau \big) \frac{\partial G}{\partial\theta} \underline{f}}_{\text{Market Thickness}} \\
& + \underbrace{\beta\big(\frac{dV}{d\theta_{-1}} + \Lambda\frac{dV_e}{d\theta_{-1}} \big)}_{\text{Anticipation}}
\end{aligned}
$$

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 to choose 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 56). 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 non-search 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

$$
\begin{aligned}
s_c &= \beta\theta p(\theta)\frac{\chi(z-\tilde{y}_c)}{1-\beta(1-\tilde{\lambda})} \\
s_p &= \beta\theta p(\theta)\frac{-\epsilon_{p,\theta}(z-\tilde{y}_p)}{1-\beta(1-\tilde{\lambda})}
\end{aligned} \tag{57}
$$

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

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.[30]

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.

$$
\begin{aligned}
s_c &= \beta\theta p(\theta)\frac{\chi(z-\tilde{y}_c)-(1-\chi)(\tilde{y}_c-\bar{y})}{1-\beta(1-\tilde{\lambda})} \\
s_p &= \beta\theta p(\theta)\frac{\left(1+\frac{1+\epsilon_{p,\theta}}{\epsilon_{v,p}\epsilon_{p,\theta}-1}\right)(z-\tilde{y}_p)-\left(1+\frac{1}{\epsilon_{v,p}\epsilon_{p,\theta}-1}\right)\left(\lambda+\frac{1-\beta}{\beta}\right)\frac{c}{p(\theta)}}{1-\beta(1-\tilde{\lambda})}
\end{aligned} \tag{58}
$$

Relative to (57), 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

[30]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).[31]

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 (57), 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 (57) and the decision rule in the full model (i.e. 56) is to write (58) 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.

$$
\begin{aligned}
s_c &= \beta\theta p(\theta)\frac{w-\tilde{y}_c}{1-\beta(1-\tilde{\lambda})} \\
s_p &= \beta\theta p(\theta)\frac{w-\tilde{y}_p}{1-\beta(1-\tilde{\lambda})} + \tau
\end{aligned}
\tag{59}
$$

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

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

[31] 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 (57) and (58).

search decision rule in Proposition 2 or (more completely) in equation (56). 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 (56) 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.[32] 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

[32] 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. 2021a, 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.[33]. 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 unity. 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 in the main text.

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

$$\begin{aligned} m_{t+1}(z;S_t) &= (1-\tilde{\lambda})m_t(z) + H(z)\theta_t p(\theta_t)S_t \\ m_{t+i+1}(z;S_t) &= \underbrace{(1-\tilde{\lambda})m_{t+i}(z;S_t)}_{\text{Continuing Hires}} + \underbrace{p(\theta_{t+i})v_{t+i}(S_t)}_{\text{New Hires}} \\ v_{t+i}(S_t) &= \frac{1}{p(\theta_{t+i})}\big(\Delta g(z) - (1-\tilde{\lambda})\big)m_{t+i}(z;S_t) \end{aligned} \tag{60}$$

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

[33] 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. (2021a)
$\frac{y_l}{y_h}$	.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 (61) 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 (62) 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.

$$\text{Both channels:} \qquad \frac{dm_{t+i+1}}{dS_t} = \big(\Delta g(z)\big)^i H(z)\frac{d}{dS_t}\big(\theta_t p(\theta_t) S_t\big) \tag{61}$$

$$\text{Continuing hires only:} \qquad \frac{dm_{t+i+1}}{dS_t} = \big(1-\tilde{\lambda}\big)^i H(z)\frac{d}{dS_t}\big(\theta_t p(\theta_t) S_t\big) \tag{62}$$

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 (62) rather than (61).[34]

[34]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.