Assortative Matching with Large Firms:

Span of Control over More versus Better Workers*

Jan Eeckhout† and Philipp Kircher‡

Preliminar Version - November 2011

Abstract

In large firms, management resolves a trade off between hiring more or better workers. The span of control or size is therefore intimately intertwined with the sorting pattern. We analyze firm size, assignment and wages when firms can adjust both the quantity and the quality margin. The sorting pattern between workers and firms is governed by a condition on the cross-margin-complementarity. There is positive sorting if the product of within-complementarities in the quality and quantity margin exceeds that of the between-complementarities across these margins. A simple system of two differential equations determines the equilibrium allocation, thus generating the distribution of firm size and of wages. More productive firms are larger if their advantage in employing more workers outweighs the additional resource demands of better workers. In the presence of frictional hiring, unemployment is inversely related to skills, while vacancies are vary ambiguously with firm productivity. In addition to the labor market, we highlight the relevance of the model for cross-country comparisons of productivity in international macro, and for applications in international trade.

Keywords. Span of Control. Sorting. Firm Size. Wage Distribution. Unemployment. Competitive Search Equilibrium. Two-Sided Matching. Supermodularity.

*We would like to thank numerous colleagues for insightful discussions and comments. Eeckhout gratefully acknowledges support from the ERC, grant 208068, and Kircher by the National Science Foundation, grant SES-0752076.

†University College London and ICREA-UPF, j.eeckhout@ucl.ac.uk.

‡Department of Economics, London School of Economics and University of Pennsylvania, P.Kircher@lse.ac.uk.

1 Introduction

Span of control – the number of workers under the control of management within a firm – attributes an essential role to the firm in economics. In the canonical macroeconomic context, firms predominantly make quantity decisions. Endowed with different management, technologies, or capital, companies choose the span of control accordingly, and this has important implications for the size of firms, as first pointed out by Lucas (1978). This labor factor intensity decision is both realistic and a convenient modeling device. It has been invoked to explain differences between countries (Restuccia-Rogerson, 2008), and to analyze technology adoption in evolving firms (Mortensen and Lentz, 2005, and Jovanovic, 1982).

Yet, in the real world firms typically face a more complex tradeoff. They simultaneously choose the *quality* of the workers as well as the quantity. Heterogeneity in skills and jobs is without doubt an important component of the labor market. The allocation process of differently skilled workers to jobs has extensively been analyzed, both with search frictions and without. In the standard frictionless matching model (Becker, 1973), each firm consists of exactly one job, just as in most of the matching models with search frictions.[1] This leads to sorting since the firm's choice is in effect about *which* worker to hire – the extensive margin –, rather than how many – the intensive margin.

The aim of this paper is to investigate sorting in an otherwise conventional macro environment where the firm simultaneously choose the quality as well as the quantity of the work force. This provides a much richer role for the firm and its span of control or size. For example, we shed light on why the high skilled upper management in firms like Walmart have an enormous span of control over relatively low skilled workers, while in mom-and-pop retail stores the span of control is small and skills of both managers and workers are average. Or, what are the consequences of information technology that improves the ability to manage many workers, such as monitoring and GPS tracking devices? These example illustrate that we can address general questions: Are more productive firms larger? Do they hire better workers; or both? How does this affect managerial compensation and firm profits? And how does it depend on the particular industry and country we are considering?

In the presence of heterogeneity, complementarity or supermodularity[2] between worker skills and job characteristics is the center piece of the firm's decision. Without complementarities (for example

[1] Burdett and Coles (1997), Shimer and Smith (2000), ...

[2] We will use the term complementarity and supermodularity interchangedly. For our purposes, it can best be thought of as the fact that the marginal contribution of higher types to output is higher when matched with other high types, i.e. there are synergies. In mathematical terms, the cross-partial of the output generated is positive (negative in the case of substitutes or submodularity).

when individuals only differ in efficiency units of labor) it does not matter for efficiency where each individual is working. Putting it stark: A CEO would add no more to the value of the economy cleaning offices than orchestrating mergers and acquisitions. In a competitive market she would earn no more in one activity than the other. Her productivity could be decomposed in an additive effect for the worker and the job she performs: output might be higher in some occupations such as orchestrating mergers and acquisitions, but it is the same for high-trained professionals as for untrained high-school dropouts. When analyzing

The paper contributes to existing work in four ways. First, we find a surprisingly simple condition for assortative matching in large firms that captures the quantity-quality tradeoff. This condition is new and compares the different degrees of complementarity. Where in the standard sorting model of Becker (1973), sorting is completely governed by the complementarity between worker and firm type, now the sorting pattern depends on the complementarity between all combinations of types and quantities. There are four complementarities. In addition to the (1) *type complementarity* just mentioned, there is the (2) complementarity in quantities of workers and resources, just as in the standard model with quantity choices only. There is the (3) span-of-control complementarity between the firm or manager type and the number of workers that features in Lucas (1978); how much do better managers have a higher marginal product of supervising more workers of a given skill? Finally, there is the (4) managerial resource complementarity, the complementarity between worker skills and managerial time: do better workers have a higher marginal product of receiving more supervision time? The tradeoff between these four forces determines the pattern of sorting, and therefore also the extent of the efficiency losses from misallocation.

Second, we can precisely pin down the composition of the work force within different firm types, i.e., how firms resolve the tradeoff of span of control over more versus better workers. The equilibrium allocation of types and quantities is entirely governed by a simple system of two differential equations. This also gives a prediction for the firm size distribution. For example under positive assortative matching, better management supervises larger groups (better firms are larger) provided the span of control complementarity (3) outweighs the managerial resource complementarity (4).

Revisiting our example of the retail industry, high productivity companies such as Walmart have high skilled managers and hire many mainly low skilled workers, compared to the smaller mom and pop stores. This indicates negative sorting together with a size distribution that exhibits a density of workers that is increasing in firm productivity. In the light of our theory, this is consistent with a type complementarity (1) that is small in this industry, whereas the span of control complementarity (3) is large: while the complementarity between managers and workers is small, at the margin management in better firms is much better at managing large groups than the low productivity firms. This is the case because cash registers and inventories are nearly trivial to operate, and they heavily invest in management and control tools that allow the supervision of many workers which allows them to get centralized information on performance on all registers and inventories. Whenever the managerial resource complementarity is positive, this leads not only to negative sorting, but also the a sharply increasing firm size in productivity, thus creating very large firms at the top. They do not need to spend much time with each employee because it is not worthwhile, and have the tools to supervise many. Instead, in management consulting, with strong complementarities in manager and subordinate skill ((1) large) but moderate span of control technologies ((3) moderate), there is positive sorting. Top firms are only larger than bottom firms if their span of control (3) outweighs the benefits from training and interacting with employees (4). Given that the two counteract, top consulting firms tend to be only moderately larger than other firms in the industry.

The Walmart example clearly illustrates that their way of doing business is very different from how the retail sector worked half a century ago. Over time, information technology and investment in knowledge dramatically changes the production process. We can therefore analyze how technological change affects the firm size distribution and the composition in the work force. Skill-biased technological change is usually viewed as a change that makes the complementarity between worker skill and firm technology larger. But much of technological change is in terms of information technology that changes the complementarity between manager skill and the amount of workers he supervises. In this model, increases in (3) change the sorting pattern, but in particular it spreads out the firm size distribution. Big firms become even bigger relative to the small firms.

This is particularly relevant for international trade, where trade changes the availability of factors of production and sometimes introduces new technologies. So far, such changes have been analyzed mainly for settings in which the size of each firm is limited by the extent of the demand. Typically, firms operate in Dixit-Stiglitz type markets where consumers have preference for variety (e.g., Costinot, 2010). Our framework is different and adds to those models of trade: output has decreasing returns because of scarce managerial resources, which limits the size of the firms. The advantage is that it can be studied without functional form assumptions. But it can also be integrated into a Dixit-Stiglitz type framework. Finally, the framework is easily extended to unemployment as well, which allows to study both the compensation and unemployment for workers of different skills. Again, this might be important in trade settings, where this has been studied recently by Helpman, Itskohki and Redding

(2011), yet in their setting workers are ex-ante identical and earn identical expected-payoffs, while in many trade settings we would like to start from a situation where workers of different types exist in the population.

In the debate on cross-country TFP differences, Restuccia-Rogerson (2008) and Hsieh-Klenow (2010) find that firm heterogeneity in conjunction with Lucas' span of control notion is important for explaining cross-country differences. By introducing working heterogeneity and sorting in this otherwise standard model, the debate can be illuminated taking into account differences in skill distributions across countries, as well as the size distribution across firms.

The paper contributes to existing work in a third way. We integrate labor market frictions into the model by means of directed search. The setup is sufficiently flexible to allow for us to introduce directed search frictions in the presence of sorting. We show that irrespective of the sorting pattern, unemployment rates are lower for more skilled workers, which is consistent with the empirically observed unemployment patterns. Instead, the vacancy rates across firm productivity levels is ambiguous. It depends on how firm size varies, which in turn is governed by the strength of the span of control complementarity relative to the managerial resource complementarity.

As a final contribution, we point out that our theory provides a unifying framework for previous models, most of which are special cases of ours. Clearly, Becker (1973) and Lucas (1978) are special cases. A general setup was also proposed in Rosen (1982), but solved only for a functional form that is a special case of our model, that of efficiency units of labor. Our setup also includes as special or limiting cases the functional forms of several existing models in this line of research such as Sattinger (1975), Garicano (2000), Antràs, Garicano and Rossi-Hansberg (2006), and Van Nieuwerburgh and Weill (2010). We can also adjust the setup to match the features of the Roy model (Heckman and Honore, 1990). Here we consider the competitive equilibrium outcome of the general model.

It should be noted that our setup is not restricted to the interpretation of span of control. For example, agricultural firms whose capital is land of different soil quality and who assign plots to workers of different endurance, or production firms whose capital is in its specific projects and who have to determine how many of these are handled by each of their workers.[3]

[3] Instead of projects, the capital may be their clients, and the decision is how many clients to assign to a member of the sales force of a given ability. Such capital is specific to the firm. In an extension we also accommodate generic physical capital on top of the differentiated firm-specific capital.

Related Literature

Our model relates mainly to the following four strands of literature: (1) assortative matching, (2) combinatorial assignment, (3) Roy-type models, and (4) frictional models with unemployment. Since the first is the closest, we devote most attention to it, and elaborate on the exact mathematical relationsships at the end in Section 6.

1. Relation to the assortative matching literature. The most common one-to-one matching models originating from Kantorovich (1942), Koopmans and Beckmann (1957), Shapley and Shubik (1971), Becker (1973), restrict attention to settings where agents have to be matched into pairs, with the obvious limitation that they do not provide insights into the size of the firm and the capital intensity. These models are captured in the specification that each worker needs exactly one unit of time, and those who do not obtain exactly one unit remain idle. This is a special case of our model. It captures that production is limited by either a scarcity of time or a scarcity of workers. It is a limiting case of our setup as production is only weakly increasing and convex the intensive margins, while most of our analysis proceeds assuming this holds strictly: more time strictly improves output but at a strictly diminishing rate. Becker's conditions arise when we take limits towards this Leontief specification. Notice that the matching models by Terviö (2008, and Gabaix and Landier (2007) to explain the changes of CEO compensation are of this kind. While they use firm size to determine the type of firm, only one worker (the CEO) is matched to one firm.

A number of both early and recent contributions have focussed on environments where managers can supervise more than one worker. An early model in this vain is presented in Sattinger (1975), where each employed worker type produces one unit of output, but requires supervision-time that depends on the manager type in a decreasing relation. A related structure arises in the span-of-control models of Garicano (2000) and Antràs, Garricano and Rossi-Hansberg (2006), where the supervision time is independent of the quality of the supervisor and refers to the number of problems that the subordinate cannot solve himself. The type of the supervisor remains important because it determines the number of problems she can solve. Both models have the feature that both the quantity and quality of workers play a role, but in a rather stark manner where additional supervision time above the minimum has no additional benefits. Their conditions again arise as limiting cases of our model.

In contrast to the assumption that a manager can supervise exactly up to some fixed number of workers but no more than that, most of the macro-economics literature has applied decreasing returns to labor in a smoother fashion to highlight the choice of the level of factor inputs. While examples abound, a classic is the widely adopted Lucas' (1978) span of control model.

Rosen (1982) proposed to a general setup with worker heterogeneity, where quantity and quality interact multiplicatively, which is a special case of ours. Rosen never solved his general setup, but assumed a functional form that guarantees perfect subsitutability, i.e., workers of a given type generate exactly the same output as twice as many workers of half that type. This is assumption is now commonly known as efficiency units of labor, and it is well-known that it generates no sorting implications. Any matching pattern is possible, and the main robust implication is that better workers obtain proportionally higher equilibrium wages.[4]

One of our contributions is to unify these strands of the assortative matching literature by analyzing a general function for output, to derive a general condition for sorting that approaches the conditions that exist in the literature as special cases, and to derive conditions that show when firm size is increasing and decreasing in manager type. The conditions are all transparent and easy to interpret, characterizing the trade-off in the number of workers that are supervised and the quality of these workers. It applies to any economy in which the output per worker can be decomposed into three components: his type, the type of the resource he is working with (here managerial quality, but it could also be the quality of land or of another physical capital that is allocated to him), and the amount of that resource that is devoted to him.

2. Relation to combinatorial matching. While the assortative matching literature has made rather specific assumptions for multi-worker firms that we attempt to generalize, the combinatorial matching and general equilibrium literature has stayed rather general but provided rather different sets of results. It mainly focussed on existence theorems. The classic example in the combinatorial matching literature is Kelso and Crawford (1982), who propose a many-to-one matching framework in a finite economy and allow for arbitrary production externalizes across workers in the same firm. While it is well-known that the stable equilibrium may not exist, they derive a sufficient condition for existence, that of gross substitutes. This condition means that adding another worker decreases the marginal value of each existing worker. This condition is satisfied in our setting since output is assumed to be concave in the number of workers. Gul and Stacchetti (1999) analyze the gross substitutes condition in the context of Walrasian equilibrium and show existence and the relation between the Walrasian price and the payment in the Vickrey-Clarke-Groves mechanism. In the context of auction design, Milgrom and Hatfield analyze package bidding as a model of many to one matching, generalize the Kelso and

[4]Rosen (1982) makes this assumption to focus on the question who becomes a manager and who becomes a worker, and even allows for more than one level of hierarchy, but never reconsiders the sorting conditions for a given set of managers and workers in more generality.

Crawford (1982) result, and propose an auction/matching algorithm that induces truthful reporting in dominant strategies. Rosen (1974) (product differentiation and hedonic prices) and Cole and Prescott (1997) (club formation) analyze general models of matching with intensive margins.

The drawback of allowing arbitrary externalizes among workers is that theorems are usually limited to existence proofs and auction-theoretic equilibrium constructions, while characterization results are essentially missing. Our work takes the alternative approach of restricting the environment such that output depends only the type of capital and the capital intensity that is assigned to each worker, which allows us to obtain analytic conditions for assortative matching, for changes in the size distribution, and generates differential equations for wages and employment that are easy to study in further applications.

3. Relation to the Roy model. Our model differs from settings such as the Roy (1951) model and its recent variants in e.g. Heckman and Honore (1990) where each firm (or sector) can absorb unbounded numbers of agents. In our setup marginal product decreases as any particular firm gets extensively large. Some models combine the Roy model with a demand by consumers that entials a constant elasticity of substitution (CES), which implies that the price falls when more workers produce output in a particular sector (see recently Costinot (2010)). The difference is that in such settings no agent internalizes the fact that the price falls when more output is produced. In our settings the firms understand that output falls when they produce more. In the final examples we also allow for a CES demand structure, but now this results in a model of imperfect competition similar to Dixit and Stiglitz (1977), only that now two-sided heterogeneity and an extensive margin are allowed.

4. Relation to frictional matching. Finally, the extension to search frictions is linked to recent developments on sorting in search markets by Shimer and Smith (2000), Shi (2001), Shimer (2005), Atakan (2006), and Eeckhout and Kircher (2010). All of these papers feature one-on-one matching. Still some of the techniques of the latter are helpful for our current setup. The real novelty of our approach here to is to provide general conditions for large firms that can acquire labor either competitively or, as we show in the later parts, through a competitive search channel.

2 The Model

We consider a static assignment problem in the tradition of Monge-Kantorovich but where the allocation is not limited to one-to-one matching.

Agents. The economy consists of heterogeneous firms and workers. Workers are indexed their skill $x \in X = [\underline{x}, \overline{x}]$, and $H_w(x)$ denotes the measure of workers with skills below x, with continuous non-zero density h_w. Also firms are heterogeneous in terms of some propriatory input into production that is exclusive to the firm, such as scarce managerial talent or particular propriatory capital goods. Firms are indexed by their productivity type $y \in Y = [\underline{y}, \overline{y}]$, where $H_f(y)$ denote the measure of firms with type below y, with non-zero continuous density h_f. It will be useful to think of the number of firms as small relative to the number of workers, which will allow each firm to hire a continuum of workers.[5]

Preferences and Production. Firms and workers are risk-neutral expected utility maximizers. If a firm of type y hires an amount of labor l_x of type x, it has to choose a fraction of its propriatory resources r_x that it dedicates to this worker type. This allows the firm to produce output

$$F(x, y, l_x, r_x)$$

with this worker type. In the production function the first two arguments (x, y) are quality variables regarding firm productivity and worker skill, while the latter two arguments (l_x, r_x) are quantity variables describing the level of inputs. The output is assumed to be increasing and twice differentiable in all arguments, and strictly concave in each of the quantity variables. For most analysis we also assume that output is constant returns to scale in the quantity variables, which turns it into a theory where the factor intensity r_x/l_x with which each worker is utilized becomes important. Other non-differentiated inputs into production can easily be accommodated, as we discuss at the end of this section.

The total output of a firm is the sum of the outputs accross all its worker types. A firm that produces with only one worker type generates output[6]

$$f(x, y, l) := F(x, y, l, 1).$$

This introduces an intensive margin into models that have traditionally focussed on pair-wise matching (see most literature following Becker (1973)), which is similar to imposing a one-unit labor force for all worker-firm pairs in this model. Since $f(x, y, l)$ is strictly decreasing in l, this theory provides a bridge to the literature on large firms with decreasing returns (see e.g., the literature following Stole and Zwiebel, 1996) that usually analyzes homogeneous workers and firms. The tractability arises because

[5] Although the set of individuals has the same cardinality as the set of firms, it is helpful to think of the set of firms as a closed interval in $[\underline{y}, \overline{y}] \subseteq \mathbb{R}$, and the set of workers as a two–dimensional subset $[\underline{x}, \overline{x}] \times [0, 1] \subseteq \mathbb{R}^2$. When both sets are endowed with the Lebesgue measure, an active firm employs a continuum of workers, albeit of mass zero.

[6] Given that $F(x, y, l, r)$ has constant returns to scale, we can write it as $F = rF(x, y, l/r, 1)$ and define $f(x, y, l/r) := F(x, y, l/r, 1)$ as the output per unit of resource. Or, as in the literature section, we can write it as $F = lF(x, y, 1, l/r)$, so that $g(x, y, r/l) = F(x, y, 1, l/r)$ represents the output per worker. In our exposition we work with the former as the firm's perspective is convenient in many derivations.

quality complementarities are concentrated on the worker-firm interaction rather than on intra-worker skill-complementarities.

Competitive Market Equilibrium. In equilibrium, workers of type x obtain some expected utility $w(x)$ that coincides with the expected wage that they are paid. Firms take this hedonic schedule as given when they make their hiring decision. We will first fix ideas by outlining the definitions of an intensive-margin hedonic pricing equilibrium without search frictions, and then handle the important case with search frictions and associated unemployment in an extension along the lines of the competitive search literature.

Firm optimality in a frictionless competitive market means that a firm of type y maximizes its output minus wage costs as follows:

$$\max_{l_x, r_x} \int [F(x, y, l_x, r_x) - w(x)l_x]dx \tag{1}$$

where r_x can be any probability density function over x. Factoring out r_x from the square bracket reveals that the interior depends only on the factor intensity $\theta = l_x/r_x$, which can be freely chosen at any level in $\Theta = \mathbb{R}_+$ by adjusting the labor input appropriately. Optimality requires that the firm places positive resources only on combinations of $x \in X$ and $\theta \in \Theta$ that solve[7]

$$\max_{x,\theta} f(x, y, \theta) - \theta w(x). \tag{2}$$

If there is only one such combination that solves this maximization problem, then the firm will hire only one worker type, allocate all resources to this type, and hire an amount of labor $l = \theta$.

Feasibility of the allocation implies that firms attempt to hire no more workers than there are in the population. Denote by $\mathcal{R}(x, y, \theta)$ the resource allocation in the economy, which describes the amount of resources that firms with a type below y devote to workers of a type below x that are employed with a factor intensity $l_x/r_x \leq \theta$. Let $\mathcal{R}(y|X, \Theta)$ denote the marginal over y when the other two variables can take any value in their type space. It denotes the amount of resources used by firms with type below y. Since the resources of each firm are normalized to one, this has to equal the amount of firms in the population, so feasibility requires $\mathcal{R}(y|X, \Theta) = H^f(y)$ for all y. Moreover, let $\mathcal{R}^\theta(\theta|x, Y)$ denote the marginal with respect to θ of the distribution conditional on a particular worker type x. It denotes the resources spent by all firms on workers of with type x employed with intensity less than θ. Feasibility

[7]Problem (1) is equivalent to $\max_{r_{(.)}} \int (r_x \max_{\theta_x} [F(x, y, \theta_x, 1) - w(x)\theta_x]) \, dx$, where $\theta_x = l_x/r_x$ can be adjusted through appropriate hiring of workers. Clearly, resources are only devoted to combinations of x and θ that maximize (2).

requires $\int \theta d\mathcal{R}(\theta|x, Y) \leq h_w(x)$. It states that the amount of workers of type x demanded across all firm types cannot exceed the number of such workers in the population (where the labor demand is the factor intensity times the amount of resources allocated at this factor intensity).

Definition 1 *An equilibrium is a tuple $(w,\mathcal{R})$ consisting of a non-negative hedonic wage schedule $w(\cdot)$ and a resource allocation $\mathcal{R}$ such that*

1. *Optimality: $(x, y, \theta) \in supp\mathcal{R}$ only if it satisfies (2).*

2. *Market Clearing: $\int \theta d\mathcal{R}(\theta|x, Y) \leq h_w(x)$, with equality if $w(x) > 0$.*

Assortative Matching: Let $\mathcal{R}(x, y|\Theta)$ be the marginal distribution of $\mathcal{R}$ over the firm and worker types at any level of intensity. It denotes the amount of resources devoted by firms with type below y to workers of skill below x. Matching is assortative if there exists a monotone function $\mu(x)$ and such that the support of $\mathcal{R}(x, y|\Theta)$ only includes points $(x, \mu(x))$. We call $\mu(x)$ the assignment function. Matching is (strictly) positive assortative if the assignment function has a (strictly) positive derivative, and it (strictly) negative assortative if the assignment function has a (strictly) negative derivative. As this definition makes clear, in the following we restrict attention to assortative matching that can be supported by a differentiable assignment function, even though all results can be proven with more technical involvement by considering non-differential monotone assignment functions.

Alternative interpretations of our setup. In our exposition we assume that firms own a unit measure of a scarce resource and allocate it to the different workers that they hire. Only the workers are traded in the market. Think about managers, each of whom has one unit of time for supervision, and who hire workers.

It might be worthwhile to note that there are alternative ways to set up our model that lead to identical results for sorting and factor prices. In our setup, we assumed that firms "buy" workers at wage $w(x)$. We could have chosen a different setup where workers buy resources for production at some endogenous price schedule $v(y)$. It turns out that our equilibrium profits according to (2) coincide with the equilibrium price $v(y)$ that arises in the alternative model where workers by resources.

Finally, we could assume that there are both resource owner and workers, and both workers and resources are traded in the market at endogenous prices $v(y)$ and $w(x)$, respectively. Both workers and resources can be put together to produce output. Anybody can set up a production entity and make profits

$$\max_{x,y,l,r} F(x, y, l, r) - lw(x) - rv(y),$$

which in equilibrium has to equal zero due to free entry, and demand has to equal supply of both workers and resources. Again, in equilibrium of this alternative model the wages are the same as in our equilibrium and the price of resources equals the firms' profits in our setup. In fact, this setup is identical to ours, only that we assumed that unit measure of resources are tied to a particular manager who runs the firm and reaps as profits the price of his resource.

Even within our exposition the production function can be interpreted in broader terms. First, we interpreted r as the fraction of the firm's resources, implicitly using a unit measure of resources for each firm. This is natural in the example of managerial time, but in many other settings firms differ in their endowments. It turns out that this is easily captures in our setting, since the unit restriction in terms of resources is a normalization. If firms of type y have $T(y)$ resources and produce $\tilde{F}(x, y, l, t)$ by using t units of them, we can express this in terms of the fraction r of their resources: $F(x, y, l, r) = \tilde{F}(x, y, l, rT(y))$.

Additionally, one might want to follow many macroeconomic models and include some kind of generic capital good that can be bought in the world market for price i per unit. If the firm buys k units generic capital and $\tilde{F}(x, y, l, t, k)$ is the corresponding output, then the production function we analyze is the induced production after optimal decisions on generic capital are made: $F(x, y, l, r) = \max_k \tilde{F}(x, y, l, rT(y), k) - ik$. We return to this extension in Section 7. In Section 4 we also cover the case where firms have to post vacancies in a frictional (competitive) search market, and the firm has to determine how many vacancies to post in order to attract the right level workers into production. This framework also allows us to capture unemployed workers in a large firm model with heterogeneity.

3 The Main Results

Models of assortative matching are in general difficult to characterize. Therefore, the literature has tried to identify conditions under which sorting is assortative. These conditions help our understanding of the underlying driving sources of sorting. In a setting like this where the welfare theorems hold, such conditions uncover the efficiency reasons behind the sorting pattners. And if the appropriate conditions are fullfilled, they substantially reduce the complecity of the assignment problem and allow further characterization of the equilibrium. In this section we derive necessary and sufficient conditions for assortative matching and characterize the assortative equilibrium.

3.1 Assortative Matching

Assume that the equilibrium is assortative, supported by some differentiable assignment function $\mu(x)$. Consider any $(x, \mu(x), \theta)$ in the support of the equilibrium allocation with $\theta > 0$, i.e., with positive amount of hiring. By (2) this means that (x, θ) are maximizers of the following problem for a firm of type $y = \mu(x)$:

$$\max_{x,\theta} f(x, y, \theta) - \theta w(x).$$

Assortative matching means that each firm only hires one type, and this problem can be understood as the problem of a firm that could choose any other worker type at any other quantity. The first order conditions for optimality are

$$\begin{aligned} f_\theta(x, \mu(x), \theta(x)) - w(x) &= 0 && (3) \\ f_x(x, \mu(x), \theta(x)) - \theta(x)w'(x) &= 0, && (4) \end{aligned}$$

where $\mu(x)$ and $l(x)$ are the equilibrium values. The second order condem requires the Hessian to be negative definite:

$$Hess = \begin{pmatrix} f_{\theta\theta} & f_{x\theta} - w'(x) \\ f_{x\theta} - w'(x) & f_{xx} - \theta w''(x) \end{pmatrix}.$$

This requires $f_{\theta\theta}$ to be negative and the determinant $|Hess|$ to be positive, or

$$f_{\theta\theta}[f_{xx} - \theta w''(x)] - (f_{x\theta} - w'(x))^2 \geq 0. \quad (5)$$

We can differentiate (3) and (4) with respect to the worker type to get

$$\begin{aligned} f_{x\theta} - w'(x) &= -\mu'(x)f_{y\theta} - \theta'(x)f_{\theta\theta} && (6) \\ f_{xx} - \theta(x)w''(x) &= -\mu'(x)f_{xy} - \theta'(x)\left[f_{x\theta} - w'(x)\right]. && (7) \end{aligned}$$

In the following three lines we successively substitute (6), (7) and then (4) into optimality condition

(5):

$$\begin{aligned} -\mu'(x) f_{\theta\theta} f_{xy} - \left[\theta'(x) f_{\theta\theta} + f_{x\theta} - w'(x)\right]\left[f_{x\theta} - w'(x)\right] &\geq 0 \\ -\mu'(x) f_{\theta\theta} f_{xy} + \mu'(x) f_{y\theta}\left[f_{x\theta} - w'(x)\right] &\geq 0 \\ -\mu'(x)[f_{\theta\theta} f_{xy} - f_{y\theta} f_{x\theta} + f_{y\theta} f_x/\theta] &\geq 0 \end{aligned}$$

For strictly positive assortative matching ($\mu'(x) > 0$) it has to hold the the term in square brackets is negative, for strictly negative assortative matching the term in square brackets need to be positive. Focussing on positive assortative matching, and using the relationsship in (4), we obtain the condition:

$$f_{\theta\theta} f_{xy} - f_{y\theta} f_{x\theta} + f_{y\theta} f_x/\theta \leq 0. \tag{8}$$

It turns out that this condition can more conveniently be summarized in terms of the original function $F(x, y, r, s)$, for which we know that $F(x, y, \theta, 1) = f(x, y, \theta)$. The following relationships will also prove useful. Homogeneity of F implies that $-F_{34} = \theta F_{33}$. Since F is constant returns, so is F_1.[8] A standard implication of constant returns it then $F_1(x, y, \theta, 1) = \theta F_{13} + F_{14}$. We can now rewrite (8) in terms of $F(x, y, \theta, 1)$ and rearrange to obtain the following cross-margin-complementarity condition:

$$F_{33}F_{12} - F_{23}\left[F_{13} - F_1/\theta\right] \leq 0 \tag{9}$$

$$\Leftrightarrow \quad F_{33}F_{12} + F_{23}F_{14}/\theta \leq 0$$

$$\Leftrightarrow \quad F_{12}F_{34} \geq F_{23}F_{14} \tag{10}$$

So the condition depends on the cross-partials in each dimension, relative to the cross-partials across the two dimensions. Only if the within-complementarities in extensive and intensive deminsion on the left hand side exceed the between-complementarities from extensive to intensive margin on the right hand side does positive assortative matching arise. We sum this finding up in the following proposition:

Proposition 1 *A necessary condition to have equilibria with positive assortative matching is that*

$$F_{12}F_{34} \geq F_{23}F_{14} \tag{11}$$

holds along the equilibrium path. If this inequality holds at all values strictly, then the only equilibria

[8] It holds that $F(x, y, r, s) = sF(x, y, r/s, 1)$, so differentiation implies that $F_1(x, y, r, s) = sF_1(x, y, r/s, 1)$.

that exists are positively assorted. In the same sense, the opposite inequality is necessary and sufficient for negative assortative matching.

Proof. Consider positive assortative matching. The proof of necessity along the equilibrium path is provided above. Similarly, it is clear that if the equality holds strictly at all points, than there cannot be any subset of the space where the assignment is locally negative assortative. To rule out non-differential assignments requires an additional argument that is available from the authors upon request. The prove for negative assortative matching proceeds along the same lines. ■

While the interpretation applies most generally, we like to think of the resources of a firm as the time spent by managers supervising other workers. Now one of the key determinants of positive/negative assortative matching is whether F_{14} is positive or negative. If managerial time is particularly productive when spent with high skilled types, then it is positive. If more managerial time is especially useful for the low skilled, then it is negative. Observe that this is a requirement for a given y and does not involve any variation in the managerial quality y. It is an empirical question whether it is positive or negative. To illustrate that F_{14} may be negative, a comparison can be insightful with the attention teachers pay to students. Typically, teachers will spend more time with the less gifted students rather than with the more gifted ones.

Interpreting this condition is relatively straightforward: On the left-hand side, a high cross-partial on the quality dimensions (F_{12}) means that higher types have ceteris paribus a higher marginal return for matching with higher types on the other side. This is reinforced by a higher cross-partial on the quality dimension, even though under constant returns to scale this can be viewed as a normalization. More importantly is the interpretation of the terms on the right-hand side. Consider the cross-partial F_{23}. If this is high, it means that we are in a setting where higher firms have a higher marginal valuation for the quantity of workers. That is, better firms value the number of "bodies" that work for them especially high. In this case better firms would like to employ many workers, which favors those where which do not require many resources. If F_{14} is negative these are the high skilled workers and the right hand side favors positive assortative matching, while F_{14} negative means that these are the low skilled workers which favors negative assorative matching.

The importance of the right hand side relies on the ability to substitute additional workers to make up for their lower quality. The following discussion reveals that as the elasticity of substition on the quantity dimension goes to zero in a way that agents can only be matched into pairs, the importance of the right hand side vanishes. It also dicusses other settings from the literature that arise as special cases.

3.2 Efficiency

One particular beauty of earlier work on one-on-one matching such as Becker (1973) is due to the fact that their condition for assortative matching can be understood by a simple efficiency consideration. If the production function is strictly supermodular but some agents are matched negatively assortative, a simple re-arrangement such that both high types and both low types are paired together increases efficiency. These efficiency gains induce assortative matching in the market game. Since our setisfies the first welfare theorem and we have quasi-linear preferences, it is easy to see that our condition has to relate to the efficiency of the underlying assignment problem. We highlight the connection by providing an analogue to the efficiency result in the previous literature. We present the result in terms of mass points of agents that match negatively assortative to make it as comparable to the previous literature as possible. Clearly it carries over to distributions with density when the efficiency increases at all the various points of negative sorting are added up.

Consider type distributions that allow for mass points, and therefore the distribution of resources allows for mass points. A feasible distribution $\mathcal{R}$ generates market surplus

$$S(\mathcal{R}) = \int F(x, y, \theta, 1) d\mathcal{R}.$$

Efficiency is achieved when the surplus is maximized over all feasible distributions of resources. The following result states that if the production function F fulfills strict cross-margin-supermodularity, the output is never maximized when $\mathcal{R}$ matches a positive measure of agents into combinations (x_1, y_1) and (x_2, y_2) that are negatively assortated.

Proposition 2 *Assume $F_{12}F_{34} > F_{23}F_{14}$ at all (x, y, l, r). Assume a feasible resource allocation $\mathcal{R}$ matches a measure $r_i > 0$ of resources at combination (x_i, y_i, θ_i) for $i \in \{1, 2\}$ where $x_1 > x_2$ but $y_1 < y_2$. Then there exists another feasible resource allocation $\mathcal{R}'$ that achieves higher output: $S(\mathcal{R}') > S(\mathcal{R})$.*

Proof. See Appendix. ■

In fact, an improvement in output can be achieved by positively assortative rematching only agents of types x_1, x_2, y_1 and y_2, while retaining the matching among all other agents. The main difficulty of the proof is to assign the right fraction of agents together, which is no longer necessarily one-to-one. This is indeed the key insight in this theory, which exploits the fact that output can be improved by improving the factor intensity, not just the matching pattern. An additional difficulty is that x_1 and x_2 are not necessarily close to each other, and neither are y_1 and y_2. While this is solved in one-to-one matching models by integrating the marginal gains over the cross-partial, this is more difficult in our setting where the condition involves not just one cross-partial. The appendix deals with both problems.

3.3 Equilibrium Characterization: Factor Intensity, Assignment and Wage Profile

In contrast to models with pairwise matching where assortativeness immediately implies who matches with whom (the best with the best, the second best with the second best, and so forth), this is not obvious in this framework as particular firms may hire more or less workers in equilibrium.

Proposition 3 *If matching is assortative, then the factor intensity, equilibrium assignment, and wages are determined by the following system of differential equations evaluated along the equilibrium allocation:*

$$\textit{PAM:} \qquad \theta'(x) = \frac{\mathcal{H}(x)F_{23} - F_{14}}{F_{34}}; \quad \mu'(x) = \frac{\mathcal{H}(x)}{\theta(x)}; \quad w'(x) = \frac{F_1}{\theta(x)}, \tag{12}$$

$$\textit{NAM:} \qquad \theta'(x) = -\frac{\mathcal{H}(x)F_{23} + F_{14}}{F_{34}}; \quad \mu'(x) = -\frac{\mathcal{H}(x)}{\theta(x)}; \quad w'(x) = \frac{F_1}{\theta(x)}, \tag{13}$$

where $\mathcal{H}(x) = \frac{h_w(x)}{h_f(x)}$.

Proof. Consider the case of PAM – the case of NAM can be derived in a similar way. The equilibrium condition for market clearing condition implies $H_w(\overline{x}) - H_w(x) = \int_{\mu(x)}^{\overline{y}} \theta(\tilde{x}) h_f(\tilde{x}) dx$. Differentiating with respect to x delivers the second differential equation in (12). The initial condition in the case of PAM is $\mu(\overline{x}) = \overline{y}$. From (4) we know that $w' = f_x/\theta$, which gives the third equation since $F_1 = f_x$. From the first-order condition in equation (3) we know that $f_\theta(x, \mu(x), \theta(x)) = w(x)$. Then from equation (6), after substituting for w' and μ' we obtain:

$$f_x/\theta = f_{x\theta} + f_{y\theta}/\theta + \theta' f_{\theta\theta}.$$

Using the same substitutions that we used in connection with equation (8) we obtain the first equation in (12). The initial condition for this differential equation obtains from running down the allocation from the top to the bottom and where the boundary condition holds either when the lowest type is attained or when the number of searchers goes to zero. An equilibrium allocation simultaneously solves the differential equation for μ' and θ' with the respective boundary conditions. ■

This gives an immediate propositon for the size of the different firms:

Proposition 4 *Under assortative matching, better firms hire more workers if and only if along the equilibrium path:*

1. $\mathcal{H}(x)F_{23} > F_{14}$ *under PAM,*

2. $-\mathcal{H}(x)F_{23} < F_{14}$ *under NAM.*

Proof. This follows readily from Proposition 3, once one realizes that under NAM $\theta'(x) < 0$. ■

Discuss F_{14} again

The result follows immediately from the proposition above once one observes that $F_{34} = -\theta F_{33}$ is strictly positive because of our concavity assumption, and that under NAM better workers are working for worse managers. The trade-off in the case of PAM is the following. Clearly, if better firms have a higher marginal value of hiring many workers (F_{23} large), this gives rise to better firms being large. Nevertheless, under assortative matching they also hire better workers. If these workers have a high marginal value from getting many resources of the firm (F_{14} large), then the firm will tend to be small. Clearly, if F_{14} is negative, meaning that better workers need less resources, this generates an even stronger force for firm growth. Under NAM, the first effect is the same, but now better firms are matched with worse workers. In this case, firms become exceptionally large if better workers need more resources, meaning that worse workers need less resources.

Interestingly, if matching is PAM and $F_{14} = F_{23}$, then the economy operates as in a one-to-one matching model: the ratio of workers to resources is always one, the assignment and the wages are as in Becker (1973). The reason is that the improvements of the firm in taking on more workers are exactly offset by the advantages of the workers to obtain more resources. Since the size distribution does not vary accross types, the renumeration also does not stray from the one that arises if we exogenously imposed a one-to-one matching ratio.

The importance of the general conditions in (11), (12) and (13) that allow for changes in the firm size due to differentials in the advantages between workers and firms is exactly to highlight the relevant sorting and assignment conditions when substitution between firm and worker inputs takes place and firm size varies across types.

We can use the equilibrium allocation to perform comparative statics. Calculating the firm distribution requires solving a system of differential equations of order two in $\theta(x)$ and $\mu(x)$. There are no solutions for general functional forms of F. However, we can explicitly solve the system of differential equation in particular cases and perform the comparative statics exercise.

4 Frictions and Involuntary Unemployment

Involuntary unemployment is not present in the model outlined so far. The model makes strong predictions on the wages that workers earn and their factor intensities, it makes no predictions on their probability of being employed. For some applications this might be rather limiting.

One simple frictional interpretation of our framework that takes partially care of this is the following: assume each firm has only a single job, but the number l of workers that it attracts constitute potential applicants who have to go through a matching function with the standard feature that the probability of filling the job goes up with the number of applicants. Since expected output is the product of the matching probability and the output produced when hiring, it is muliplicatively seperable and the sorting condition coincides with that in the second example in Section 6 (Eeckhout and Kircher (2010)).

The main drawback of this approach is that it does not capture the feature of true multi-worker firms: decreasing returns in production and actual choices of the number of jobs that are posted. The sorting framework that we laid out in the previous section is well-suited to capture true multi-worker firms with decreasing returns in production. In this section we embedd the previous setup in a costly recruiting and search process that has been used in other settings to capture the hiring behavior of large firms, albeit related work did not handle the two-sided heterogeneity. We will be able to derive predictions not only on the expected wages but also about the unemployment rate of workers of different skills. In particular, there exists a simple positive link between the worker's wages and their employment prospects. The following setup builds on the competitive search literature (e.g., Peters 1991; Acemoglu and Shimer 1999; Burdett, Shi and Wright 2001; Shi 2001; Shimer 2005; Eeckhout and Kircher 2009; Guerrieri, Shimer and Wright 2010) and its extensions to the analysis of multiworker firms (Menzio and Moen 2010; Garibaldi and Moen forthcoming; Kaas and Kircher 2011). We borrow the standard assumptions made in multi-worker firm models with search frictions, with the inoovation being the heterogeneity of both workers and firms.

Consider a situation where the workers are unemployed and can only hired by firms via a frictional hiring process. As part of this process, each firm decides how many vancancies v_x to post for each worker type x that it wants to hire. Posting v_x vacancies has a linear cost cv_x. It also decides to post wage ω_x for this worker type. Observing all vacancy postings, workers decide where to search for a job. Let q_x denote the "queue" of workers searching for a particular wage offer, defined as the number of workers per vacancy. Frictions in the hiring process make it impossible to fill a position for sure. Rather, the probability of filling a vacancy is a function of the number of workers queueing for this vacancy, denoted by $m(q_x)$, which is assumed to be strictly increasing and strictly concave.[9] Since there are q_x workers queueing per vacancy, the workers' job-finding rate for these workers is $m(q_x)/q_x$. The job finding rate is assumed to be strictly decreasing in the number of workers q_x queueing per vacancy. Firms can attract workers to their vacancies as long as these workers get in expectation their equilibrium utility, meaning that q_x adjusts depending on ω_x to satisfy: $\omega_x m(q_x)/q_x = w(x)$. Note the difference between the wage ω_x which is paid when a worker is actually hired, and the expected wage $w(x)$ of a queueing worker who does not yet know whether he will be hired or not. In equilibrium the firm takes the latter as given because this is the utility that workers can ensure themselves by searching for a job at other firms, while the former is the firm's choice variable with which it can affect how many workers will queue for its jobs. Therefore, a firm maximizes instead of (1) the new problem

$$\max_{r_x,\omega_x,v_x} \int [F(x,y,l_x,r_x) - l_x\omega_x - v_x c]\, dx \tag{14}$$
$$\text{s.t. } l_x = v_x m(q_x); \quad \text{and} \quad \omega_x m(q_x)/q_x = w(x)$$

and r_x integrates to unity. The first line simply takes into account that the firm has to pay the vacancy-creation cost, and that the number of hires depends on the amount of hiring per vacancy which is in turn related to the wage that it offers. There are two equivalent representations of this problem that substantially simplify the analysis. It can easily be verified that problem (14) is mathematically equivalent to both of the following two-step problems:

1. Let $G(x,y,s,r) = \max_v [F(x,y,vm(s/v),r) - vc]$, and solve $\max_{s_x,r_x} \int [G(x,y,s_x,r_x) - w(x)s_x]dx$ where r_x integrates to unity.

2. Let $C(l,x) = \min_{v,q}[cv + vqw(x)]$ s.t. $l = vm(q)$, and solve $\max_{s_x,r_x} \int [F(x,y,l_x,r_x) - C(l_x,x)]dx$ where r_x integrates to unity.

In the first equivalent formulation, the firm attracts "searchers" s_x, which queue up to get jobs at this firm. In order to entice them to do this, it has to offer wage $w(x)$ in expectation to them whether

[9] Careful elaborations how this queueing problem in a finite economy translates into matching probabilities as the population is exanded is given e.g. in Peters (1991) and Burdett, Shi and Wright (2001). It is based on the idea that workers approach vacancies unevenly due to coordination problems, which leads to excess applicants at some vacancies and to few vacancies at others.

or not they actually get hired. The definition of G then reflects the fact that the firm can still decide how many possible vacancies to create for these workers. If the firm creates more vacancies, searchers have an easier time finding a vacancy suitable to them, and this increases the amount of actual labor that is employed within the firm. In the second formulation the firm the output minus the costs of hiring the desired amount of labor. The costs include both the vacancy-creation costs as well as the wage costs, where again the expected wage has to be paid to all workers that are queueing for the jobs.

This has two direct consequences:

1. It has the beauty that G is fully determined by the primitives, and can be directly integrated into the framework we laid out in Section 2 (where now G replaces F). The firm looks as if it hires "searchers" which have to be paid their expected wage. Applying the machinery from the previous section allows us to assess whether sorting is assortative, and what the expected wages $w(x)$ are that are paid in equilibrium. We take this formulation embedded in the equilibrium definition of the previous section as the definition of a competitive search equilibrium with large firms.[10]

2. It then relates the expected wages $w(x)$ that were determined in the previous problem into job finding probabilties of the searchers. Substituting the constraint in Problem 2 into the objective function and taking first order conditions yields the main characterization of this section. It can best be expressed by writing the elasticity of the matching probability as $\eta(q) := qm'(q)/m(q)$ and by denoting the queue length that solves the minimization problem by $q(x)$. We then obtain

$$w(x)q(x) = \frac{\eta(q(x))}{1 - \eta(q(x))}c \tag{15}$$

The right hand side is related to the well-known Hosios condition (Hosios, 1990), which showed that efficient vacancy creation is related to the elasticity of the matching function. The condition becomes particularly tractable in commonly used settings in which the elasticity is constant. In this case the queue length that different workers face is inverse proportional to the expected utility that they obtain in equilibrium. Since better workers obtain higher expected utility $w(x)$ as determined in Problem 1 (otherwise a firm could higher better workers at equal cost), they face proportionally lower competition for each job and correspondingly higher job finding probabilities. This arises because the opportunity costs of having high skilled workers unsuccessfully queue for employment is higher, and therefore firms are more willing to create enough vacancies to enable most of these applicants to actually get hired for

[10]The same mathematical structure arises (after rearranging) when we start with an equilibrium definition in the natural way that is usually used in the competitive search literature, where firms compete in actual wages and not in terms of expected wage payments.

the job. The logic applies even if the elasticity is not constant:

Proposition 5 *In the competitive search equilibrium with large firms, higher skilled workers have face lower unemployment rates.*

Proof. The term $\eta(q)/[q(1-\eta(q))] = m'(q)/[m(q) - qm'(q)]$. This term is strictly decreasing in q, since the numerator is strictly decreasing and the denominator is strictly increasing in q. Since $w(x)$ is increasing in x in any equilibrium, implicit differentiation of (15) implies that $q(x)$ is decreasing, which in turn implies that the chances of finding employment are increasing in x. ■

Interestingly, this implies that under positive assortative matching the firm-size can be increasing in firm type even though the number of workers that apply for jobs is decreasing. This can be seens mathematically as follows. The amount of labor that is actually hired, $l(x)$, relates to the actual number of searchers and their queue per vacancy as $l(x) = s(x)m(q(x))/q(x)$, implying:

$$l'(x) = s'\frac{m}{q} + s\frac{m'q - m}{q^2}q'.$$

The change in the number of searchers (s') is determined by (12) under appropriate change of variables (θ and f replaced by s and g). Even if the number of workers that search for employment at better firms is not increasing, the number of hires might still be increasing because the second term is strictly positive. This may be due to the fact that high productivity firms put more resources into creating jobs for their high-skilled applicants. Recruiting of talented lawyers at a law firm is likely to involve more resources, either the direct cost or the opportunity cost of time, than what is spent on hiring low skilled labor at a fast food restaurant.

5 Simulations of Comparative Statics Results

TBD.

6 Special Cases

The following highlights how our model characterizes a number of existing setups that have been heavily used in the literature. It also highlights that it can capture new settings that have not been analyzed before. It also shows that it is not easy to start with certain separability assumptions (for example between the quality and quantity dimensions as in the second example below), because a lot of formulations in the literature have used different ways of interacting the variables that can be captured in our setup but not in more specialzied versions.

1. Efficiency units of labor. A particularly common assumption in the literature is the case of efficiency units of labor, where the output remains unchanged as long as the multiplicative term xl_x remains unchanged. In such a case workers of one type are completely replaceable by workers of half the skills as long as there are twice as many of them. Sorting is then essentially arbitrary: Each firm cares only about the right total amount of efficiency units, but not whether they are obtain by few high-type workers or many low-type workers. Our setup captures efficiency units of labor under production function $f(x, y, l) = \tilde{f}(y, xl)$. Taking cross-partials immediately reveals that we always obtain $F_{12}F_{34} = F_{23}F_{14}$ in this case.

2. Multiplicative separability. A particularly tractable case arises under multiplicative separability of the form $F(x, y, l, r) = A(x, y)B(l, r)$. In this case the condition (11) for positive assortative matching can be written as $[AA_{12}/(A_1A_2)][BB_{12}/(B_1B_2)] \geq 1$. If B has constant elasticity of substitution ε, we obtain an even simpler condition $AA_{12}/(A_1A_2) \geq \varepsilon$.[11]

3. Becker's one-on-one matching model as a limit case. Consider some output process $F(x, y, l, r)$. In the spirit of most of the sorting literaterature, we can now consider the restricted variant where only "paired" inputs can operate: every worker needs exactly one unit of resource and any resources needs exactly one worker, otherwise it is not used in production. The output can then be represented by $F(x, y, \min\{l, r\}, \min\{r, l\}) = F(x, y, 1, 1) \min\{l, r\}$, where the equality follows from constant returns to scale. This nicely corresponds to the multiplicatively separable setup discussed in the previous point. While our framework is build around the idea that more resources or more labor inputs improve production, this Leontief setup on the quantity dimension is exactly the limit case of a CES function with zero elasticity ($\varepsilon \to 0$). From the previous point we therefore know that sorting arises in this limit if $F_{12} \geq 0$, which is exactly the condition in Becker (1973).

4. Sattinger's and Garicano's span of control problem as a limit cases. One of the few contributions that provides clear conditions for sorting in a many-to-one matching model is presented in Sattinger (1975). His production function assumes that each worker produces the same, but a

[11]If ε is in the unit interval, this condition is equivalent to root-supermodularity, i.e., it is equivalent to $\sqrt[n]{A(x, y)}$ being supermodular with $n = (1 - \varepsilon)^{-1}$ as shown by Eeckhout and Kircher (2010) in a pairwise matching framework with directed search frictions. If $\varepsilon > 1$ this requires conditions on $A(x, y)$ that are stronger than log-supermodularity.

worker of type x needs $t(x,y)$ units of supervision time from manager of type y, where better types need to spend less time. The manager can only hire as many workers as he can supervise, so that $F(x,y,l,r) = \min\{r/t(x,y), l\}$, where the first terms in the minimization operator captures the number of workers that can be supervised and the second the number of workers hired. Our model allows for more flexibility in the substitution between inputs, but a CES extension that takes $r/t(x,y)$ and l as inputs again has the previous Liontieff specification as the inelastic limit.[12] Inspecting (11) and taking the inelastic limit reveals that positive sorting arises only if $t(x,y)$ is log-superodular. This exactly recovers the condition found by Sattinger.

Related is Garicano (2000) formulation where each worker can solve problems with difficulty equal to his type, and passes the remaining problems to the supervisor who need one unit of time to answer each problem passed to him. Here the supervision "time" $t(x)$ depends only on the worker type, but each manager can solve problems up to category y himself, leading to output $F = y\min\{r/t(x), l\}$. Approximating this with the appropriate CES highlights that better managers prefer to hire better workers to leverage their skills.

5. Extension of Lucas' Span of Control, and Rosen's general production: Lucas (1978) assumed a production function that is multiplicatively separable in the firm type and the amount of labor, where all labor is identical. Consider the following extension to heterogeneous labor: $F(x,y,l,r) = yg(x,l/r)r$, which boils down to $yg(x,l)$ in the case where all resources are spent on the same worker type. The new condition for assortative matching is $g_2g_{12} \geq g_1g_{22}$. If production is increasing in worker type and strictly concave, this means that sorting will be positive unless better workers types indeed dislike to work together because that limits the amount of resources they can obtain (g_{12} sufficiently negative).

This is related to Rosen's (1982) setup where $F = h(y)g(x, yl/r)r$ for the first level of supervision and the production workers, which has somewhat different separability assumptions.[13] He allows more flexibility in other parts by allowing for multiple layers of hirarchy and a choice on who performs on which layer. But he analyzes the model only for the case of linear homogeneity of g, which is equivalent to efficiency units of labor analyzed in point 1 above. Since it is a special case of our setup, our sorting conditions apply directly to this setting. One can easily write the conditions in terms of g, but does not get much additional insight above and beyond those we have discussed already for the general model.

[12]The function $F(x,y,l,r) = ([rg(x,y)]^{(\varepsilon-1)/\varepsilon} + l^{(\varepsilon-1)/\varepsilon})^{\varepsilon/(\varepsilon-1)}$ approaches $\min\{rg(x,y), l\}$ as $\varepsilon \to 0$.

[13]Rosen (1982) equation (1) for the output per worker can be written as $h(y)\xi(yr/l,x)$ for some functions h and ξ,so that total output is constant returns to scale. Output per resource is therefore $yg(yl/r,x)$ after appropriate transfromation (so that $g(y\theta,x) := \xi(y/\theta,x)/\theta$).

6. Spatial Sorting Within the Mono-centric City. The canonical model of the mono-centric city can explain how citizens locate across different locations, however there is no spatial sorting. All agents are identical and in equilibrium they are indifferent between living in the center or in the periphery by trading off commuting time for housing space and prices.[14] Let there be a continuum of locations y, each with housing stock $r(y)$. Let $y \in [0,1]$, where y is the center and y is the inverse of a measure of the distance from the center. Agents with budget x have preferences over consumption c and housing h represented by a quasi-linear utility function $u(c,h) = c + v(h)$. With consumption the numeraire good and $p_h(y)$ the price per unit of housing in location y, the budget constraint is $c + p_h(y)h = xg(y)$, where x is the worker skill and $g(y)$ is an increasing function representing the time at work rather than in commute. The closer to the center, the less time is spent on commuting and the more time is earned. Then we can write the individual citizen x's optimization problem as $xg(y) + v(h) - p_h(y)h$. The total supply of housing in location y is r and as a result, $l \cdot h = r$. Net of the transfers, the aggregate surplus for all l citizens is given by $F(x,y,l,r) = xg(y)l + v\left(\frac{r}{l}\right)l$. It is easily verified that $F_{12} = g'(y)l, F_{34} = -\frac{r}{l}v''\left(\frac{r}{l}\right), F_{14} = 0$ so that if $v(\cdot)$ is concave there is positive assortative matching of the high income earners into the center and the low income earners in the periphery. A similar functional form is used in Van Nieuwerburgh and Weill (2010) to consider differences between cities rather than within the city, where the term $xg(y)$ is replaced by a more agnostic worker-output $u(x,y)$ depending on worker skill x and city type y. Again, sorting is again fully determined by the cross-partial of x and y because $F_{14} = 0$.

The sorting conditions of the previous proposition establishes a positive relation between the productivity of the firm and the skills of the workers that it hires. It does not directly establish how many workers a given firm hires, and therefore it does not directly determine even under assortative matching who matches with whom.

7 Extensions

Our main example was the assignment of workers to firms, which can be viewed as the assignment of resources by firms to particular worker types. We expanded this leading example to the case of unemployment. The following gives extensions and other interpretations of our setup.

[14] Also Lucas and Rossi-Hansberg (2002) model the location of identical citizens but their model incorporates productive as well as residential land use. Though agents are identical, they earn different wages in different locations. The paper proves existence of a competitive equilibrium in this generalized location model which endogenously can generate multiple business centers.

1. Capital Investment. Consider a production process that not only takes as inputs the amount of labor and of proprietary firm resources, and creates output $\hat{F}(x,y,l,r,k)$. The generic capital k that can be bought on the world market at price i.[15] Optimal use of resources requires $F(x,y,l,r) = \max_k \left[\hat{F}(x,y,l,r,k) - ik\right]$, where F is constant returns in its last two arguments if $\hat{F}$ is constant returns in its last three arguments. Rewriting the cross-margin-complementarity condition (11) in terms of the new primitive yields the following condition for positive assortative matching: $\hat{F}_{12}\hat{F}_{34}\hat{F}_{55} - \hat{F}_{12}\hat{F}_{35}\hat{F}_{45} - \hat{F}_{15}\hat{F}_{25}\hat{F}_{34} \geq \hat{F}_{14}\hat{F}_{23}\hat{F}_{55} - \hat{F}_{14}\hat{F}_{25}\hat{F}_{35} - \hat{F}_{15}\hat{F}_{23}\hat{F}_{45}$.

2. Monopolistic Competition. In the previous sections, we analyzed the case where the firm's output is converted one-for-one into agents utility. Therefore, there are no consequences on the final output price of the good, which is normalized to one. An often used assumption in the trade literature concerns consumer preferences pioneered by Dixit and Stiglitz (1977) which are CES with elasticity of substitution $\rho \in (0,1)$ among the goods produced by different firms. For these preferences it is well-known that a firm that produces output $\tilde{f}$ has achieves a sales revenues $\chi\tilde{f}^{\rho}$, where χ is an equilibrium outcome that is viewed as constant from the perspective of the individual firm.[16] The difficulty in this setup is that, despite the fact that output is constant returns to scale in employment and firm resources, the revenue of the firm has decreasing returns to scale. Therefore, we cannot directly apply (11). But we can conjecture that there is assortative matching so that the firm employs only one worker type, in which case revenues are $f(x,y,l) = \chi\tilde{f}(x,y,l)^{\rho}$, and we can apply (8) directly. Rearranging and using $\tilde{F}(x,y,l,r) = r\tilde{f}(x,y,l/r)$ we get the condition for positive assortative matching

$$\begin{aligned}
&\left[\rho\tilde{F}_{12} + (1-\rho)(\tilde{F})\frac{\partial^2 \ln \tilde{F}}{\partial x \partial y}\right]\left[\rho\tilde{F}_{34} - (1-\rho)l\tilde{F}\frac{\partial^2 \ln \tilde{F}}{\partial l^2}\right] \\
\geq \quad &\left[\rho\tilde{F}_{23} + (1-\rho)\tilde{F}\frac{\partial^2 \ln \tilde{F}}{\partial y \partial l}\right]\left[\rho\tilde{F}_{14} + (1-\rho)\left(l\tilde{F}_{13} - l\tilde{F}\frac{\partial^2 \ln \tilde{F}}{\partial x \partial r}\right)\right].
\end{aligned}$$

Several points are note-worthy. First, the condition is independent of χ, and therefore can be checked before this term is computed as an outcome of the market interaction. Furthermore, for elastic preferences ($\delta \to 1$) the condition reduces to our original condition (11). In gereral, the condition relies not only on supermodularities in the production function, but also on log-supermodularities. This should

[15]This section expands on the short exposition in Footnote **??**.

[16]The underlying form for the utility function is $U = x_0^{1-\mu}\left(\int c(y)^{\rho}dy\right)^{\mu/\rho}$, where x_0 is a numeraire good and $c(y)$ is the amount of consumption of the good of producer y. Then one obtains $\chi = (\mu Y)^{1-\rho}P^{\rho}$ where Y is the aggregatve income, p_y denotes the price achieved by firm y through its equilibrium quantity, and $P = \left(\int p_y^{\rho/(1-\rho)}\right)^{\rho/(1-\rho)}$ represents the aggregate price index.

not be surprising. Even in the standard models supermodularity is the relevant condition when the marginal consumption value of output is normalized to one (Becker 1973), while sorting when output is CES-aggregated requires log-supermodularity. If $\tilde{F}$ is multiplicatively separable between quantity and quality dimension, and the quality dimension is CES, then as the quality dimension becomes increasingly inelastic it is easy to show that the condition reduces to log-supermodularity in x and y.

3. Optimal transport. Assume it costs $-r \cdot c(x, y)$ to move a r units of waste from production site x into destination storage y, and if one attempts to move more units r into any given amount l of storage then there is some probability of damage $d(r/l)$ that each unit that is stored gets destroyed. This leads to function $F(x, y, l, r) = -rc(x, y) - \alpha rd(r/l)$, where α represents the lost revenue because of destruction. Unlike in the standard Monge-Kantorovich transportation problem, here the allocation need not occur in fixed quantities.[17]

4. Frictional matching of men and women. Assume that there are different locations where men and women can meet. These are located in different distinct locations. If r men of type x search for s women of type y, then $M(r, s) \leq \min\{r, s\}$ matches are created, where M is a standard constant returns to scale matching function. Each match is worth $A(x, y)$ for the pair. Then output is given by $F(x, y, r, s) = M(r, s)A(x, y)$. This is essentially the setup in Eeckhout and Kircher (2010). Under the standard assumption that the search literature makes on M, they find a necessarily and sufficient condition for positive assortative matching is that A is root-supermodular. In a general production environment it is not possible to reduce the analysis to conditions on the quality dimension only, and a more general look at the cross-margin-complementarity is required.

5. Endogenous type distributions, technology choice, and team-work. One way to engodenize the type distribution is to assume that there is free entry of firms (free entry of resources in the model), but entry with type y costs $c(y)$. If output is increase in y, i.e., $F_2 > 0$, then it is crucial for a meaningful entry decision that $c(y)$ is strictly increasing. If c is increasing and differentiable, and our sorting condition is satisfied everywhere, it is not difficult to contruct an equilibrium where the profits of firms according to (2) equal the entry cost $c(y)$ for all active firms. In fact, this formulation is easier to construct: We know that the highest types match, so that $\mu(\bar{x}) = \bar{y}$. The problem is usually how to determine at which ratio they match, i.e., to find $\theta(\bar{x})$. But here it is given simply by requirement

[17]Observe that in the Monge-Kantorovich problem the allocation need not be in pure strategies. The optimal allocation may involve mixing or in large populations there may be a fraction of agents of a given type allocated to one location and the remainder to another location as long as the total measure of agents at any location does not exceed one. What differs here is the intensive margin. Any location is not restricted to taking on a given measure of agents.

that the profits of the highest firm equals the entry costs. Substituting the first order condition (3) into the objective function yields profit $f(\bar{x}, \mu(\bar{x}), \theta(\bar{x})) - \theta(\bar{x}) f_\theta(\bar{x}, \mu(\bar{x}), \theta(\bar{x}))$, which have to equal $c(\mu(\bar{x}))$. This can be then used together with the first order conditions and the differential equations in (3) to construct the type distribution after entry at all lower types.

More complicated is the analysis when one considers a common pool of workers, some of whom choose to be managers while others choose to remain workers. This is then a teamwork problem, where one team becomes the $y's$ and the other the $x's$. While interesting, we leave this analysis for further work.

6. Non-homotheticity: At the moment we have entertained the notion that each worker interacts with the resources to obtain output. Just envision briefly a production technology that is customized such that for each worker type, output can be produced easier when there are more workers and resources of that type in the sense that the output function F has increasing returns to scale. Clearly, in this the restriction that each firm has a unit-measure of resources becomes binding: firms would ideally like to merge to larger entities. While we do not want to stress this setup because it seems to us much less convincing than the setup under constant returns, in the appendix we can nevertheless show that our basic intuition up to equation (9) remains necessary. Only it is not longer possible to reformuate this as (10) which relied on constant returns.

8 Concluding Remarks

We have proposed a matching model that incorporates factor intensity and unemployment. We derive a simple condition for assortative matching and characterize the equilibrium firm size, unemployment level and unemployment by skills.

9 Appendix

Proof of Proposition 2

Proof. Strict cross-margin-supermodularity $F_{12}F_{34} > F_{14}F_{23}$ for all (x, y, l, r) is by (8) equivalent to $f_{\theta\theta}f_{xy} - f_{y\theta}f_{x\theta} + f_{y\theta}f_x/\theta < 0$ for all (x, y, θ). Assume a feasible resource allocation $\mathcal{R}$ matches a measure $r_i > 0$ of resources at combination (x_i, y_i, θ_i) for $i \in \{1, 2\}$ where $x_1 > x_2$ but $y_1 < y_2$. We will establish that output is strictly increased under a feasible variation yielding resource allocation $\mathcal{R}'$ that pairs some of the x_2 workers to some of the y_2 resources. We proceed in two steps. Step 1 has the key insight.

1. Establish the marginal benefit from assigning additional workers to some resource type:
Consider some (x, y, θ) such that r resources are deployed in this match (and are paired to θr workers). For the variational argument, we are interested in the marginal benefit of pairing an additional measure r' of resources of type y' with workers of type x. The optimal output is generated by withdrawing some optimal measure $\theta' r'$ of the workers that were supposed to be working to with resource y and reassigning them to work with resource y'. The joint output at (x, y) and (x, y') is given by

$$rf(x, y, r, \theta - \theta' r'/r) \; + \; r' f(x, y', \theta'). \tag{16}$$

Optimality of θ' requires according to the first order condition that $f_3(x, y, \theta - \theta' r'/r) = f_3(x, y', \theta')$, which shows that the optimal θ' is itsself a function of r'. Denote $\beta(y'; x, y, \theta)$ the marginal increase of (16) from increasing r', evaluated at $r' = 0$. It is given by

$$\beta(y'; x, y, \theta) = f(x, y', \theta') - \theta' f_3(x, y', \theta') \tag{17}$$

$$\text{where } \theta^{'} \text{ is determined by } \; f_3(x, y', \theta') = f_3(x, y, \theta). \tag{18}$$

The constrained (18) reiterates the optimality of θ' as a function of x, y, θ and y'. The cross-partial β_{12} of the marginal benefit in (17) with respect to x and y' is strictly positive, evaluated at $y' = y$, iff

$$f_{xy} > -\left[\theta f_{y\theta} f_{x\theta} + f_{y\theta} f_x\right] / \left[\theta f_{\theta\theta}\right],$$

i.e., exactly when our cross-margin condition holds. Therefore, it is optimal to assign higher buyers to higher sellers locally around (x, y). This is at the heart of the argument. The next step simply extends this logic to a global argument where y' might be far away from y.

2. Not PAM has strictly positive marginal benefits from matching the high types:

We started under the assumption that matching is not assortative since $x_1 > x_2$ but $y_1 < y_2$.. In particular, consider y_1 matched to x_2 at queue length λ_1 and y_2 matched to x_1 at queue λ_2, where $x_2 > x_1$ and $y_2 > y_1$. For (x_1, y_2) and (y_1, x_2) to be matched, optimality requires that the marginal benefit of types $y^v = y_1$ are higher when paired with x_2, while types $y^v = y_2$ yield higher benefit when paired with x_1 :

$$\beta(y_1; x_2, y_2, \theta_2) \leq \beta(y_1; x_1, y_1, \theta_1), \tag{19}$$

$$\beta(y_2; x_2, y_2, \theta_2) \geq \beta(y_2; x_1, y_1, \theta_1), \tag{20}$$

where $\beta(\cdot;\cdot,\cdot,\cdot)$ was defined in (16). We will show that if (19) holds, then (20) cannot hold, which yields the desired contradiction. We will show this by proving that the benefit $\beta(y'; x_1, y_1, \theta_1)$ on the right hand side of (19) and (20) always remains above the benefit $\beta(y'; x_2, y_2, \theta_2)$ on the left hand side. By (19) this has to be true at $y' = y_1$, and we will show that it remains true when we move to higher y'. The marginal increase of β with respect to its first argument y' is given by

$$\beta_1(y^v; x, y, \lambda) = f(x, y', \theta'), \tag{21}$$

where θ' is again determined as in (18). Assume there is some $y' \geq y_1$ such that marginal benefits are equalized, i.e., $\beta(y'; x_2, y_2, \theta_2) = \beta(y'; x_1, y_1, \theta_1)$. We have established the result when we can show that $\beta_1(y'; x_2, y_2, \theta_2) < \beta_1(y'; x_1, y_1, \theta_1)$.

By (21) this equivalent to showing that $f(x_2, y', \theta_2') < f(x_1, y', \theta_1')$, where $\theta_1' = \theta'(y'; x_1, y_2, \lambda_2)$ and $\theta_2' = \theta'(y'; x_2, y_1, \lambda_1)$ as in (18). To show this, define $\xi(x)$ for all x in resemblance of (17) by the following equality

$$f(x, y', \xi(x)) - \xi(x) f_3(x, y', \xi(x)) = \beta(y'; x_2, y_2, \theta_2),$$

which implies $\xi(x_2) = \theta_2'$ and $\xi(x_1) = \theta_1'$ by equality of the marginal benefits at y', i.e. by $\beta(y'; x_2, y_2, \theta_2) = \beta(y'; x_1, y_1, \theta_1)$. Differentiating $f(x, y', \xi(x))$ with respect to x reveals that it is strictly increasing exactly under our strict inequality $f_{\theta\theta} f_{xy} - f_{y\theta} f_{x\theta} + f_{y\theta} f_x / \theta < 0$. This in turn implies $f(x_2, y', \theta_2') < f(x_1, y', \theta_1')$. ■

The Non-Homogeneous Production Technology

Let output of the firm be $F(x, y, r, s)$, and the firm of type y chooses the worker type and the labor intensity l. As before, let the capital intensity r be given. Then the problem is

$$\max_{x,l} F(x, y, l, r) - lw(x) - rv(y).$$

The first order conditions for optimality are

$$\begin{aligned} F_1(x, \mu(x), l, r) - lw'(x) &= 0 \\ F_3(x, \mu(x), l, r) - w(x) &= 0 \end{aligned}$$

where $\mu(x)$ and l are the equilibrium values. The second order condition of this problem requires the Hessian H to be negative definite:

$$H = \begin{pmatrix} F_{11} - rw'' & F_{13} - w' \\ F_{13} - w' & F_{33} \end{pmatrix}$$

which requires that all the eigenvalues are negative or equivalently, $F_{11} - rw'' < 0$ (which follows from concavity in all the arguments (x, y, l, r)), and

$$\begin{vmatrix} F_{11} - rw'' & F_{13} - w' \\ F_{13} - w' & F_{33} \end{vmatrix} > 0.$$

After differentiating the two FOCs along the equilibrium allocation to substitute for $F_{11} - rw'' = -F_{12}\mu'$ and $F_{13} - w' = -F_{23}\mu'$ and also using the first FOC to rewrite $w' = F_1/r$ we get

$$\begin{vmatrix} -F_{12}\mu' & -F_{23}\mu' \\ F_{13} - w' & F_{33} \end{vmatrix} > 0$$

or $-F_{12}F_{33}\mu' + (F_{13} - F_1/r)F_{23}\mu' > 0$ and thus PAM requires (knowing that $F_{33} < 0$)

$$F_{12} > \frac{(F_1/r - F_{13})F_{23}}{|F_{33}|}.$$

Observe that this condition is similar to the one we obtained for the homogeneous case, only that now it depends on the marginal product F_1 and the concavity of F in l, F_{33}.

References

[1] Antras, Pol, Luis Garicano, and Esteban Rossi-Hansberg, "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, 2006, **1**, 31-77.

[2] Acemoglu, Daron, and Robert Shimer, "Efficient Unemployment Insurance," *Journal of Political Economy* **107**, 1999a, 893-928.

[3] Atakan, Alp, "Assortative Matching with Explicit Search Costs", *Econometrica* **74**, 2006, 667–680.

[4] Becker, Gary S., "A Theory of Marriage: Part I", *Journal of Political Economy* **81(4)**, 1973, 813-46.

[5] Cahuc, P., Postel-Vinay, F. and Robin, J.-M. (2006), "Wage Bargaining with On-the-Job Search: Theory and Evidence," *Econometrica*, 74: 323-364.

[6] Cole, H., and E. Prescott, "Valuation Equilibrium with Clubs," *Journal of Economic Theory*, **74**, 19-39, 1997.

[7] Costinot, Arnaud, and Jonathan Vogel, "Matching and Inequality in the World Economy," *Journal of Political Economy*, 118(4), 2010, 747-786.

[8] Davis, Steven J., R. Jason Faberman, and John Haltiwanger, "The Flow Approach to Labor Markets: New Data Sources and Micro–Macro Links," *Journal of Economic Perspectives* **20(3)**, 2006, 3–26.

[9] Dixit, A. K. and J. E. Stiglit, "Monopolistic competition and optimum product diversity," *American Economic Review*, 67(3), 1977, 297-308.

[10] Eeckhout, Jan and Philipp Kircher, "Sorting and Decentralized Price Competition", *Econometrica* **78(2)**, 2010, 539-574.

[11] Eeckhout, Jan and Roberto Pinheiro, "Diverse Organizations and the Competition for Talent", University of Pennsylvania mimeo, 2008.

[12] Garibaldi, Pietro, and Espen Moen, "Job-to-Job Movements in a Simple Search Model," *American Economic Review* Papers and Proceedings, forthcoming.

[13] Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 2000, **108 (5)**, 874-904.

[14] Guerrieri, Veronica, Robert Shimer and Randall Wright, "Adverse Selection in Competitive Search Equilibrium," *Econometrica*, 2010, **78 (6)**, 1823-1862.

[15] Gul, Faruk, and Ennio Stacchetti, "Walrasian Equilibrium with Gross Substitutes," *Journal of Economic Theory* **87**, 95-124, 1999.

[16] Heckman, James J., and Bo E. Honore, "The Empirical Content of the Roy Model," *Econometrica*, **58(5)**, 1121-1149.

[17] Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica* **50(3)**, 1982, 649-670

[18] Kaas, Leo and Philipp Kircher, "Efficient Firm Dynamics in a Frictional Labor Market", mimeo, 2011.

[19] Kantorovich, L.V., "On the Translocation of Masses", (Translated into English and reprinted in Management Science **5(1)**, 1-4 (1958)) *Comptes Rendus (Doklady) de l'Academie des Sciences de L'URSS* **37**, 199-201 (1942).

[20] Kelso, Alexander S., and Vincent P. Crawford, "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica* **50(6)**, 1982, 1483-1504.

[21] Koopmans, Tjalling C., and Martin Beckmann, "Assignment Problems and the Location of Economic Activities," *Econometrica* **25(1)**, 1957, 53-76.

[22] Lentz, R. and Mortensen, D. T., "An Empirical Model of Growth Through Product Innovation." *Econometrica,* **76**, 1317-1373.

[23] Lucas, Robert E., "On the Size Distribution of Business Firms," *The Bell Journal of Economics* **9(2)**, 1978, 508-523.

[24] Menzio, Guido, and Espen Moen, "Worker Replacement," *Journal of Monetary Economics*, **57**, 623-636.

[25] Milgrom, Paul R., and John W. Hatfield, "Matching with Contracts," *American Economic Review* **95(4)**, 913-935, 2005.

[26] Moen, Espen, "Competitive Search Equilibrium", *Journal of Political Economy* **105(2)**, 1997, 385-411.

[27] Peters, Michael, "Ex Ante Pricing in Matching Games: Non Steady States", *Econometrica* **59(5)**, 1991, 1425-1454.

[28] Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, **82**, 34–55, 1974.

[29] Rosen, S., "Authority, Control, and the Distribution of Earnings," *Bell Journal* 1982.

[30] Roy, A. D. "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 1951, **3(2)**, 135-146.

[31] Sattinger, Michael, "Comparative Advantage and the Distributions of Earnings and Abilities", Econometrica **43(3)**, 1975, 455–468.

[32] Shapley, Lloyd S. and Martin Shubik, "The Assignment Game I: The Core", *International Journal of Game Theory* **1(1)**, 1971, 111-130.

[33] Shi, Shouyong, "Frictional Assignment. 1. Efficiency," *Journal of Economic Theory*, 98, 2001, 232-260.

[34] Shimer, Robert, and Lones Smith, "Assortative Matching and Search," *Econometrica* **68**, 2000, 343–369.

[35] Van Nieuwerburgh, Stijn, and Pierre-Olivier Weill, "Why Has House Price Dispersion Gone Up?" *Review of Economic Studies*, 77, 2010, 1567-1606.