SORTING AND FACTOR INTENSITY:
PRODUCTION AND UNEMPLOYMENT ACROSS SKILLS*

JAN ECKHOUT† AND PHILIPP KIRCHER‡

– Preliminary Draft –
April 2011

Abstract

When firms choose the allocation of workers, they can adjust not only the type of worker, the extensive margin, but also the intensive margin, how many of those worker to employ. We propose a tractable matching model with such factor intensity. Positive sorting arises if a simple cross-margin-complementarity condition holds: within-complementarities in extensive and intensive margin have to exceed the between-complementarities across intensive and extensive margin. We characterize the equilibrium allocation, wages and factor intensities. In an immediate extension that allows for frictional hiring we also analyze the presence of unemployment across types. Unemployment is decreasing in skills, and we find conditions under which firm size is increasing in skill.


*We would like to thank numerous colleagues for insightful discussions and comments. Eckhout gratefully acknowledges support from the ERC, grant 208068, and Kircher by the National Science Foundation, grant SES-0752076.
†Department of Economics, ICREA-UPF Barcelona and UCL, j.eckhout@ucl.ac.uk.
‡Department of Economics, London School of Economics and University of Pennsylvania, P.Kircher@lse.ac.uk.
1 Introduction

We propose a theory of labor market matching where firms can adjust the factor intensity. When aggregate conditions change, for example due to business cycles, international competition or changes in factor endowments, firms react not only by adjusting the extensive margin (the skill of a worker that is right for the job), but also the intensive margin (how many of those workers to employ). The intensive margin indicates how labor productivity changes. For example, during a recession there can be both increased unemployment and growing labor productivity. Firms adjust capacity by laying off workers, which in the presence of a concave production technology raises the marginal product of the workers.

Key in understanding the interplay of productivity and the intensive margin is heterogeneity in skills. When workers differ in skills and ability, and firms employ technologies with different levels of productivity. In standard models of sorting the relevant choice is on the extensive margin by choosing the optimal allocation over different types of skills. Here, firms or technologies simultaneously operate two margins: not only do they optimally choose the skill of a worker, they also choose how many of those skilled workers to employ. The ensuing capital-labor ratio will as a result vary across skills and technologies. The combination of two-sided heterogeneity and endogenous factor intensity pins down the equilibrium outcome and determines wage formation, wage inequality and unemployment.

Unlike the standard Beckerian matching framework, firms are large and we can analyze changes in the firm size and composition. In the presence of an aggregate productivity shock for example, firms are able to adjust the skill quality as well as the number of those skilled workers. Even though our model is static, in a dynamic setting, the latter may turn out to be more easily adjusted since the adjustment is for marginal types only, rather than the entire pool.

Our setting also permits us to analyze equilibrium unemployment. In a directed search environment, firms choose their skill composition and size taking into account frictions in the hiring process. This permits us to analyze how unemployment changes across skills and firm size. Interestingly, in our framework long term unemployment arises naturally and it is determined characteristics of the production technology. In particular, firms might compete for higher skilled workers rather than substituting them with low skilled ones, thus leaving the low skilled permanently unemployed. Interestingly, aggregate productivity shocks will lead to the low skilled to move in and out the state of unemployability, while simultaneously equilibrium unemployment adjusts for higher-skilled workers.

Our main findings. First, we show a surprisingly simple condition for positive sorting. As in the one-to-one matching model, it requires a strong enough degree of complementarity between the type of
firm and that of the worker. Now the strength also depends on the marginal change of output in labor and technology intensity and how that varies across worker and firm types respectively. The relevant condition for sorting therefore has to involve changes in the quality dimensions (extensive margin) and in the quantity of factor inputs (intensive margin). Positive assortative matching arises if the the total degree of intensive and extensive complementarity exceed the cross-complementarities between the intensive and extensive margin. It might be interesting to note the latter become unimportant when the intensive margin is Lientieff, in which case standard supermodularity prevails. Second, even under assortative matching one still has to determine the exact allocation and the associated factor intensities. We derive these expressions, including conditions for them to be increasing. Finally, we integrate labor market frictions into the model and show how unemployment varies across worker types. Also, we show how the condition is extended when additional generic capital inputs can be bought in the world market, and when the goods have to be sold in an output market characterized by monopolistic competition.

**Relation to the literature.** While our exposition focuses foremost on the market equilibrium, we also show the equivalence to a planners problem following form the welfare theorem. The comparison with the literature is easiest in the following planner problem. Assume there is a continuum of workers and a continuum of resources (which in our main model are bundled inside firms). Workers differ by skill \( x \), and each skill is in fixed supply. Similarly, the resources of the firms have different quality \( y \), and each quality is in fixed supply.

For each type combination \((x, y)\) the planner can choose how many resources \( r \) to devote to this combination and how many workers \( l \) to employ in this combination. Clearly, he cannot assign more resources or workers than are present in the population.

Each combination produces some output \( F(x, y, r, l) \) based on the quality types and the quantity amounts of workers and resources, constant returns and increasing in the quantity dimensions. The planner maximizes the sum of outputs across all type combinations. Interpretations are manifold: The resources can be crop fields with different crops or different soil, employing workers of different strength or endurance. The overall output depends on how much land and of which type a worker ploughs. In modern economies the resources are better described as managerial time (in the spirit of Sattinger (1975)). We assume some smoothness on the quantity dimension and investigate the interplay between quantity and quality.

The main difference from the most common one-to-one matching model originating from Kan-
torovich (1942), Koopmans and Beckmann (1957), Shapley and Shubik (1971), Becker (1973), is that these earlier models restrict attention to settings where each agent on one side can only be matched with one agent on the other side. The formal requirement is that \( r = l \), which can be interpreted as excess resources or workers remaining idle: The total number of resources and workers that are actually used in production is \( n = \min\{r_{ij}, l_{ij}\} \) and so total output is \( F(x, y, n, n) \). Since the firms cannot adjust the amount of resources per worker, the assignment of workers to resources is determined by the supermodularity in their quality types.

Sattinger (1975) formulated an extensive margin of managerial span of control. The resources are managerial time. A manager of quality \( y \) needs \( t(x, y) \) units of time to supervise a subordinate of skill \( x \). Each subordinate produces one unit of output, but if a manager takes more subordinates than he can manage, these remain idle. When the manager devotes \( r \) units of time, the output is \( F = \min\{l, r/t(x, y)\} \) because any excess labor remains idle. Sattinger shows that better managers supervise better workers if the second argument is log-supermodular, where the difference from Becker (1973) arises because the number of workers per manager changes depending on the type.

The idea of modeling firms as many-to-one matchings has extensively been analyzed by Kelso and Crawford (1982). They propose a general, discrete agent framework. While it is well-known that the stable equilibrium of many-to-one matchings may not exist, Kelso and Crawford derive a sufficient condition for existence, that of gross substitutes. This condition basically means that adding one worker will not dramatically increase the productivity of all the other ones. This condition is satisfied in our setting since the intensive margin is concave: an additional worker is strictly less productive. While Kelso and Crawford allow arbitrary type-dependent production processes only restricted by the cross-substitute condition and show existence, characterization results under intensive margins are essentially missing. Our work is limited to the case where production depends only the type of capital and the capital intensity that is assigned to each worker. For our continuum economy this allows us to capture the intensive margin element while retaining a level of tractability that proves extremely useful in application.

Many-to-one matching has been analyzed in related but different contexts. Gul and Stacchetti (1999) analyze the gross substitutes condition in the context of Walrasian equilibrium. They 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 and generalize the Kelso and 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. Their focus is on setting up a general framework in general equilibrium. Here, the set up is related but our objective is to characterize the equilibrium allocation, in particular, who matches with whom and how much. We want to know how workers and firms interact: which firm hires what type of worker and how many; if there is unemployment, how does it change across firms.

The main difference of our approach to the standard matching models that feature in many recent applications originating from Kantorovich (1942), Koopmans and Beckmann (1957), Shapley and Shubik (1971), Becker (1973), is that these earlier models restrict attention to settings where each agent on one side can only be matched with one agent on the other side, and there are interaction in the amount that can be produced by each pair. Models following Sattinger (1975) do not have any complementarities in the output that a pair produces, but different workers need different amounts of supervision which limits the number of workers that a particular firm can hire at some maximum value. In both cases the condition for sorting remains simple (supermodularity in the former, log-supermodularity in the latter) but the ratio at which each firm can hire workers is very restrictive. In particular, the firm cannot substitute additional workers.

Using a continuum economy, we allow any ratio of agents from one side relative to agents on the other side. Restricting the output to Lintief retrieves the condition of these earlier models, but in general it reveals a richer condition that takes into account the extensive margin in the assignment.

While some settings such as the marriage market one-on-one matching clearly seems to be the relevant case, examples of intensive margin matching abound. Our leading example is the assignment of workers to firms, where the number of workers per firm does not need to be unity, but can be freely chosen. Another example is an extension to the classical transportation problem going back to Kantorovich (1942). He considered goods that a produced at various plants and are stored at various locations, and unit costs of transportation that depend on the production site and the storage location. In his setting, each good needs to be stored in exactly one unit of storage, while our setup allows for the case where more goods can be crammed into each unit of storage but damage increases with storage intensity. We discuss this example as well as other applications such as matching between teams of workers at the end, but will focus on our leading example in the main body.

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 used. Some models combine the
Roy model with a demand by consumers that entails a constant elasticity of substitution (CES), which implies that the price falls when more workers choose to work 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.

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), but the real novelty of our approach to is to provide general conditions for large firms that can acquire labor either competitively or through a competitive search channel.

2 The Model

Players. The economy consists of heterogeneous firms and workers. Workers are indexed their skill $x \in X = [x, \overline{x}]$, 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 proprietary input into production that is exclusive to the firm, such as scarce managerial talent or particular proprietary capital goods. Firms are indexed by their productivity type $y \in Y = [y, \overline{y}]$, 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.1

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 proprietary 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

---

1 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 $[y, \overline{y}] \subseteq \mathbb{R}$, and the set of workers as a two-dimensional subset $[x, \overline{x}] \times [0, 1] \subseteq \mathbb{R}^2$. When both sets are endowed with the Lebesgue measure, an active firm employs a continuum of workers, albeit of mass zero.
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 across all its worker types. A firm that
produces with only one worker type produces

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

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
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:

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

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\(^2\)

\(^2\)Problem (1) is equivalent to $\max_{\theta \in \Theta} \int (r_x \max_{\theta_x} [F(x, y, \theta_x, 1) - w(x)\theta_x]) dx$, 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 $\theta$ that maximize (2).
\[
\max_{x,\theta} f(x, y, \theta) - \theta w(x). 
\] (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 \( R(x, y, \theta) \) 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 \( R(y|X, \Theta) \) 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 \( R(y|X, \Theta) = H^f(y) \) for all \( y \). Moreover, let \( R^\theta(\theta|x, Y) \) denote the marginal with respect to \( \theta \) 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 \( \theta \). Feasibility requires \( \int \theta dR(\theta|x, Y) \leq h_w(x) \). 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 intensive-margin hedonic pricing equilibrium is a tuple \( (w, R) \) consisting of a non-negative hedonic wage schedule \( w(\cdot) \) and a resource allocation \( R \) such that

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

2. Market Clearing: \( \int \theta dR(\theta|x, Y) \leq h_w(x) \), with equality if \( w(x) > 0 \).

**Assortative Matching:** Let \( R(x, y|\Theta) \) be the marginal distribution of \( R \) 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 \( R(x, y|\Theta) \) only includes points \((x, \mu(x)) \). We call \( \mu \) the assignment function. Matching is (strictly) positive assortative if the assignment function is (strictly) increasing, and it is (strictly) negative assortative if the assignment function is (strictly) decreasing. In the following we will restrict attention to assortative matching that can be supported by a differentiable assignment function.

**Clarification of Market Structure and Integration of Different Resource Constraints, Generic Capital, and Competitive Search Frictions:** 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. As in other matching models, it turns out that exactly the same allocations arise if each worker would buy the resources he wants to use in production. In this case the workers production function would be $F(x, y, 1, r)$, and his return decreases as he buys additional resources. In this case the equilibrium price for a given resource equals the profit that firm obtains in our exposition. Similarly, if firms are not endowed with any resources but have to buy both resources and labor, the same allocation arises but firms make zero profits since all profits accrue to the owners of the scarce resources.

Even within our exposition the production function $F$ 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 just 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, our analysis focusses on proprietary resources such as land or managerial capital. Yet firms may also use some generic capital good. Assume this general capital can be obtained in the world market at unit price $i$, and a firm that produces with $k$ units of such capital achieves output $\tilde{F}(x, y, l, r, k)$. Then the relevant production function for our analysis is $F(x, y, l, r) = \max_k \tilde{F}(x, y, l, r, k) - ik$. That is, the production function we analyze is the induced production after optimal decisions on generic capital are made. We return to this extension in Section 4.

Finally, the search literature has emphasized the role matching frictions. Models like Caribaldi and Moen (forthcoming) assume that a firm posts $V$ vacancies at cost $cV$, and when $l$ workers try to get a job at these vacancies, only $lm(l/V)$ of them are hired while the remainder stay unemployed. Here $m$ captures the matching function, and this framework allows us to study frictional unemployment with large firms and two-sided heterogeneity. We show in the competitive search setting in Section 3: If the firm produces $\tilde{F}(x, y, h, r)$ when it hires $h$ workers, then the induced production function that we use for our general characterizations is $F(x, y, l, r) = \max_v \tilde{F}(x, y, lm(l/V), r) - cV$, i.e., the production that arises after optimal vacancy choices have been made.

## 3 Assortative Matching

Models or 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 partners. And if the appropriate conditions are fulfilled, they substantially reduce the complexity 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.

*Condition for 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 $(x, \theta)$ 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

$$f_\theta(x, \mu(x), \theta(x)) - w(x) = 0 \quad (3)$$

$$f_x(x, \mu(x), \theta(x)) - \theta(x)w'(x) = 0, \quad (4)$$

where $\mu(x)$ and $l(x)$ are the equilibrium values. The second order condition of this problem 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

$$f_{x \theta} - w'(x) = -\mu'(x)f_{\theta \theta} - \theta'(x)f_{\theta \theta} \quad (6)$$

$$f_{xx} - \theta(x)w''(x) = -\mu'(x)f_{x \theta} - \theta'(x) [f_{x \theta} - w'(x)]. \quad (7)$$

10
In the following three lines we successively substitute (6), (8) and then (4) into optimality condition (5):

\[-\mu'(x)f_{\theta\theta}f_{xy} - \left[\theta'(x)f_{\theta\theta} + f_{x\theta} - w'(x)\right] \left[f_{x\theta} - w'(x)\right] \geq 0\]

\[-\mu'(x)f_{\theta\theta}f_{xy} + \mu'(x)f_{y\theta} \left[f_{x\theta} - w'(x)\right] \geq 0\]

\[-\mu'(x)\left[f_{\theta\theta}f_{xy} - f_{y\theta}f_{x\theta} + f_{y\theta}f_{x}/\theta\right] \geq 0\]

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 relationship in (4), we obtain the condition:

\[f_{\theta\theta}f_{xy} - f_{y\theta}f_{x\theta} + f_{y\theta}f_{x}/\theta \leq 0.\]  (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 \(F_{1}.^{3}\)

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:

\[F_{33}F_{12} - F_{23} \left[F_{13} - F_{1}/\theta\right] \leq 0\]

\[\Leftrightarrow F_{33}F_{12} + F_{23}F_{14}/\theta \leq 0\]

\[\Leftrightarrow F_{12}F_{34} \geq F_{23}F_{14}\]

So the condition depends on the crosspartials in each dimension, relative to the crosspartials across the two dimensions. Only if the within-complementarities in extensive and intensive dimension 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

\[F_{12}F_{34} \geq F_{23}F_{14}\]  (9)

---

3It holds that \(F(x, y, r, s) = sF(x, y, r/s, 1)\), so differentiation implies that \(F_{1}(x, y, r, s) = sF_{1}(x, y, r/s, 1)\).
holds along the equilibrium path. If this inequality holds at all values strictly, then the only equilibria that exists are positively assorted. In the same sense, the opposite inequality is necessary and sufficient for negative assortative matching.

**Proof.** The proof of necessity is provided above. The proof for sufficiency is yet to be done. ■

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. It turns out that high quality workers are expensive, since they command higher wages. If the firms are predominantly interested in the number of bodies, they rather higher the less able workers but lots of them. These workers are cheaper, and overall profits go up.

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 substitution 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.

**Corollary 1** Condition (9) includes the following special cases:

**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 \( x_l \) 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-partially immediately reveals that we always obtain \( F_{12}F_{34} = F_{23}F_{14} \) in this case.

**Multiplicative separability of the quality and quantity dimensions.** 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 (9) for positive assortative matching can be written as \[ AA_{12}/(A_1 A_2)BB_{12}/(B_1 B_2) \geq 1. \] If \( B \) has constant elasticity of substitution \( \varepsilon \), we obtain an even simpler condition \( AA_{12}/(A_1 A_2) \geq \varepsilon. \)

**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 literature, we can now consider the restricted variant where only "paired" inputs can operate: every worker needs exactly one unit of resource and any resources need exactly one worker, and the remainder are idle. In this setting the true production is given 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 corresponds to the multiplicatively separable setup we discussed in the previous point. While our framework is built 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).

**Sattinger’s span of control problem as a limit case.** One of the few contributions that provides clear conditions for sorting in a many-to-one matching model is presented in Sattinger (1975). His model can be described as follows. The manager of each firm has one unit of time available. Each employed worker \( x \) produces one unit of output, but requires \( t(x, y) \) units of supervision-time from a manager type \( y \), where higher types need less time. This means that the manager can supervise \( g(x, y) = t(x, y)^{-1} \) workers per unit of time. Devoting \( r \) units of type allows him to supervise no more than \( rg(x, y) \) workers, and since the rest remain idle the production is with \( l \) workers is limited at \( F(x, y, l, r) = \min\{rg(x, y), l\} \). Our model allows for more flexibility in the substitution between inputs, but a CES extension that takes \( rg(x, y) \) and \( l \) as inputs again has the previous Leontief specification as the inelastic limit.\(^5\) Inspecting (9) and taking the inelastic limit reveals that positive sorting arises only if \( g(x, y) \) is log-supermodular. This exactly recovers the condition found by Sattinger.

**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.\(^6\) Let there be a continuum of locations \( y \),

\(^4\)If \( \varepsilon \) is in the unit interval, this condition is equivalent to root-supermodularity, i.e., it is equivalent to \( \sqrt{A(x, y)} \) being supermodular with \( n = (1 - \varepsilon)^{-1} \) as shown by Eckhout 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.

\(^5\)The function \( F(x, y, l, r) = (\|rg(x, y)\|^{(\varepsilon - 1)/\varepsilon} + l^{(\varepsilon - 1)/\varepsilon})^{(\varepsilon - 1)/\varepsilon} \) approaches \( \min\{rg(x, y), l\} \) as \( \varepsilon \to 0 \).

\(^6\)Also Lucas and Rossi-Hansberg (2002) model the location of identical citizens but their model incorporates productive
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 $F(x, y, l, r) = xg(y)l + v \left( \frac{r}{l} \right)$. It is easily verified that $F_{12} = g'(y)l, F_{34} = -\frac{r}{l^2}v'' \left( \frac{r}{l} \right), F_{14} = 0$ 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.

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 satisfies 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

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.
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 output $O(\mathcal{R}) = \int F(x, y, \theta, 1)d\mathcal{R}$. 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 $(x_i, y_i, \theta_i)$ 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: $O(\mathcal{R}') > O(\mathcal{R})$.

**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.

**Factor intensity, assignment, and wages in assortative equilibria:** In contrast to models with pair-wise 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. For the following characterization of the equilibrium factor intensity, assignment and wages it will be instructive to use up the one degree of freedom that we have in defining the economy by adopting the following normalization: Instead of referring to a particular worker by his type, we will from now on refer to him by the measure of workers that have a strictly lower type than him. That is, $x$ now denotes a worker such that exactly a measure $x$ of workers has a lower type than him. This still identifies the worker uniquely, but with the new convention the density of the distribution is unity. We will do the same for the firms. The main
advantage of this "relabeling" is that now the density on the new type space is equal to one. With this normalization, the following differential equations describe the equilibrium assignment, factor intensity and wages in this model.

**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 path:

\[ \begin{align*}
PAM: & \quad \theta'(x) = \frac{F_{23} - F_{14}}{F_{34}}; \quad \mu'(x) = \frac{1}{\theta(x)}; \quad w'(x) = \frac{F_1}{\theta(x)}, \\
NAM: & \quad \theta'(x) = \frac{F_{23} + F_{14}}{F_{34}}; \quad \mu'(x) = -\frac{1}{\theta(x)}; \quad w'(x) = \frac{F_1}{\theta(x)}.\end{align*} \tag{10, 11} \]

**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 \( H_w(\pi) - H_w(x) = \int_{\mu(x)} \theta(\hat{x}) h_f(\hat{x}) d\hat{x} \). Differentiating with respect to \( x \) and noting that under our normalization \( H_w(x) = x \) delivers the second differential equation in (10). The initial condition in the case of PAM is \( \mu(\pi) = \bar{\gamma} \). From (4) we know that

\[ w' = f_x / \theta, \]

which gives the third equation. 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 \( \mu' \) we obtain:

\[ f_x / \theta = f_x + f_y + \theta' f_{\theta y}. \]

Using the same substitutions that we used in connection with equation (8) we obtain the first equation in (10). 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 \( \mu' \) and \( \theta' \) with the respective boundary conditions. ■

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

**Corollary 2** Under assortative matching, better firms hire more workers if and only if along the equilibrium path

- \( F_{23} > F_{14} \) under PAM,
- \( F_{23} > -F_{14} \) under NAM.

---

\( ^8 \)With all arguments and with possibly non-normalized type distributions this condition reads in the case of PAM:

\[ \theta'(x) = \frac{F_{23}(x, \mu(x), \theta(x), 1) - \int_{\mu(x)} \theta(\hat{x}) h_f(\hat{x}) d\hat{x}}{F_{34}(x, \mu(x), \theta(x), 1)}; \quad \mu'(x) = \frac{h_w(x)}{\theta_f(\mu(x)\theta(x))}; \quad w'(x) = \frac{F_{23}(x, \mu(x), \theta(x), 1)}{\theta(x)}. \]
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 across types, the remuneration 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 (9), (10) and (11) 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.

**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 multiplicatively separable and the sorting condition coincides with that in the second example in Corollary 1 (Eeckhout and Kircher
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 embed 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 (Menzo 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 innovation 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 vacancies $v_x$ to post for each worker type $x$ that it wants to hire. Posting $v_x$ vacancies has a linear cost $c v_x$. It also decides to post wage $\omega_x$ 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 $\omega_x$ to satisfy: $\omega_x m(q_x)/q_x = w(x)$. Note the difference between the wage $\omega_x$ 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

\(^9\)Careful elaborations how this queueing problem in a finite economy translates into matching probabilities as the population is exausted 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.
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

\[
\max_{r_x, \omega_x, q_x} \int [F(x, y, l_x, r_x) - l_x \omega_x - v_x c] dx
\]

\[\text{s.t. } l_x = v_x m(q_x); \text{ and } \omega_x m(q_x)/q_x = w(x)\]

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 (12) is mathematically equivalent to both of the following two-step problems:

1. Let \(G(x, y, s, r) = \max_v [F(x, y, v m(s/v), r) - v c]\), 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 \(C(l, x) = \min_v [c v + v q w(x)]\) s.t. \(l = v m(q)\), 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 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:

**Problem 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 assertive, 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.\footnote{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.}

**Problem 2:** It then relates the expected wages $w(x)$ that were determined in the previous problem into job finding probabilities 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

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

(13)

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

**Proposition 4** 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 (13) 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 seen 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 (10) under appropriate change of variables \( (\theta \text{ and } f \text{ replaced by } s \text{ and } g) \). 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. The reason is that is high ability firms put more resources into creating jobs for their high-skilled applicants.

4 Additional Examples

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.

Example 1: Additional Capital Inputs. 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 \( i \).\(^\text{11}\) Optimal use of resources requires \( F(x, y, l, r) = \max_k \left[ \hat{F}(x, y, l, r, k) - i k \right] \), 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 (9) in terms of the new primitive yields the following condition for positive assortative matching:

\[
\hat{F}_{12} \hat{F}_{34} \hat{F}_{55} - \hat{F}_{12} \hat{F}_{35} \hat{F}_{54} - \hat{F}_{15} \hat{F}_{25} \hat{F}_{34} \geq \hat{F}_{14} \hat{F}_{23} \hat{F}_{55} - \hat{F}_{14} \hat{F}_{25} \hat{F}_{35} - \hat{F}_{15} \hat{F}_{23} \hat{F}_{45}.
\]

Example 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 \( \hat{f} \) has achieves a sales revenues \( \chi \hat{f}^\rho \), where \( \chi \) is an equilibrium outcome that is viewed as constant from the perspective of the individual firm.\(^\text{12}\) The

\(^{11}\)This section expands on the short exposition in Footnote ??.

\(^{12}\)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 aggregative income,
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 (9). 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

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

Several points are note-worthy. First, the condition is independent of \( \chi \), 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 (9). In general, the condition relies not only on supermodularities in the production function, but also on log-supermodularities. This should 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 \( y \).

**Example 3: Optimal transport.** Assume it costs \( -r \ast c(x, y) \) to move a \( r \) units of waist 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 \( \alpha \) represents the lost revenue because of destruction.

**Example 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 \( p_y \) 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.

22
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.

**Example 5: Matching of two teams of workers.** Rather than thinking about complementarities between the worker and the firm, one can equally well consider complementarities between teams of workers. One way to think about such complementarities may be the following. There is only one quality dimension \( Q \) and some distribution \( H^Q(q) \) of worker skills. But production needs two teams. So if one team has quality \( x \in Q \) and \( l \) team members while the other team has quality \( y \in Q \) and \( r \) team members, then output is \( F(x, y, l, r) \). Since the teams are drawn from the same base population, the natural requirement is symmetry of the form \( F(x, y, l, r) = F(y, x, r, l) \), at least in settings where the cooperation is symmetric. Then \( X = Y = Q \), and each side gets half the number of available workers \( H^f(x) = H^u(x) = H^Q(x)/2 \). This provides a stylized way to analyze patterns of teams that work together.

## 5 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.
6 Appendix

Proof. Proof of Proposition 2: Strict cross-margin-supermodularity \( F_{12}F_{34} > F_{14}F_{23} \) for all \( (x, y, l, r) \) is by (8) equivalent to \( f_{\theta \theta} f_{xy} - f_{y \theta} f_{x \theta} + f_{y \theta} f_{x \theta} / \theta < 0 \) for all \( (x, y, \theta) \). Assume a feasible resource allocation \( R \) matches a measure \( r_i > 0 \) of resources at combination \( (x_i, y_i, \theta_i) \) 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 \( R' \) 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 \( (x, y, \theta) \) such that \( r \) resources are deployed in this match (and are paired to \( \theta r \) 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

   \[
   r f(x, y, r, \theta - \theta' r' / r) + r' f(x, y', \theta').
   \]  

   Optimaly of \( \theta' \) 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 \( \theta' \) is itself a function of \( r' \). Denote \( \beta(y'; x, y, \theta) \) the marginal increase of (14) from increasing \( r' \), evaluated at \( r' = 0 \). It is given by

   \[
   \beta(y'; x, y, \theta) = f(x, y', \theta') - \theta' f_3(x, y', \theta')
   \]  

   where \( \theta' \) is determined by \( f_3(x, y', \theta') = f_3(x, y, \theta) \).

   The constrained (16) reiterates the optimality of \( \theta' \) as a function of \( x, y, \theta \) and \( y' \). The cross-partial \( \beta_{12} \) of the marginal benefit in (15) with respect to \( x \) and \( y' \) is strictly positive, evaluated at \( y' = y \), iff

   \[
   f_{xy} > -[\theta f_{y \theta} f_{x \theta} + f_{y \theta} f_{x \theta} / \theta f_{\theta \theta}],
   \]

   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 \( \lambda_1 \) and \( y_2 \) matched to \( x_1 \) at queue \( \lambda_2 \), 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 \( x_1 \):

\[
\beta(y_1; x_2, y_2, \theta_2) \leq \beta(y_1; x_1, y_1, \theta_1), \tag{17}
\]
\[
\beta(y_2; x_2, y_2, \theta_2) \geq \beta(y_2; x_1, y_1, \theta_1), \tag{18}
\]

where \( \beta(\cdot; \cdot, \cdot, \cdot) \) was defined in (14). We will show that if (17) holds, then (18) 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 (17) and (18) always remains above the benefit \( \beta(y'; x_2, y_2, \theta_2) \) on the left hand side. By (17) 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 \( \beta \) with respect to its first argument \( y' \) is given by

\[
\beta_1(y^v; x, y, \lambda) = f(x, y', \theta'), \tag{19}
\]

where \( \theta' \) is again determined as in (16). 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 (19) 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 (16). To show this, define \( \xi(x) \) for all \( x \) in resemblance of (15) 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 \( f(x_2, y', \theta'_2) < f(x_1, y', \theta'_1) \). \( \blacksquare \)
References


