DISCUSSION PAPER SERIES

DP18922
(v. 3)

WORKERS-TO-FIRMS MATCHING WHEN SKILLS ARE BUNDLED

Philippe Choné and Francis Kramarz

INDUSTRIAL ORGANIZATION AND LABOUR ECONOMICS

ISSN 0265-8003

WORKERS-TO-FIRMS MATCHING WHEN SKILLS ARE BUNDLED

Philippe Choné and Francis Kramarz

Discussion Paper DP18922
First Published 17 March 2024
This Revision 05 September 2024

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

WORKERS-TO-FIRMS MATCHING WHEN SKILLS ARE BUNDLED

Abstract

We study the workers-to-firms matching when firms' production technologies and worker's skills are both multidimensional. In this article, we provide fundamental welfare theorems when workers' skills are bundled (markets for stand-alone skills are missing). Sorting is then based on workers' comparative advantage. The bundling friction causes skills' prices to vary across firms, making the (unique) equilibrium wage schedule non-linear in skills. This non-linearity of wages, closely linked to the heterogeneity of workers' skills within firms, depends on the relative prevalence of specialist and generalist workers in the economy. In equilibrium, the wage is log-additive in worker quality and a worker-to-firm sorting effect that may reflect the firm's productivity. We provide descriptive evidence using Swedish matched employer-employee data providing us with direct measures of workers' cognitive and non-cognitive skills. Our companion paper, Choné, Gozlan, and Kramarz, 2024, examines the same questions when markets for stand-alone tasks open and the bundling friction gradually disappears.

JEL Classification: D20, D40, D51, J20, J24, J30

Keywords: Heterogeneous firms, Bundling, Sorting, Matching, Multidimensional skills

Philippe Choné - philippe.chone@ensae.fr
CREST - ENSAE and CEPR

Francis Kramarz - kramarz.francis@gmail.com
Innovation Lab, Collège de France and CEPR

Acknowledgements
We are grateful to Marco Cuturi, Xavier d'Haultfoeuille, Alfred Galichon, Nathael Gozlan, Gene Grossman, Philipp Kircher, Samuel Kortum, Cl\'emence Lenoir, Ilse Lindenlaub, Simon Mongey, Giuseppe Moscarini, Bérengère Patault, François-Pierre Paty, Torsten Persson, Roland Rathelot, Jean-Marc Robin, Jean-Charles Rochet, Bernard Salanié, David Schönholzer, Oskar Skans, and Bob Topel for extremely helpful conversations, remarks, and suggestions as well those of participants at multiple seminars and conferences. We thank Gaspard Choné-Ducasse and Pierre Krakovitch for excellent research assistance. Choné gratefully acknowledges support from the French National Research Agency (ANR), (Labex ECODEC No. ANR-11-LABEX-0047) and Kramarz from the European Research Council (ERC), advanced grant FIRMNET (project ID 741467).

Workers-to-Firms Matching When Skills Are Bundled*

Philippe Choné[†] and Francis Kramarz[‡]

[†] *CREST-ENSAE, Institut Polytechnique de Paris, France, and CEPR*

[‡] *Innovation Lab, Collège de France, Paris*

[‡] *Department of Economics, Uppsala University, and CEPR*

September 5, 2024

Abstract

We study the workers-to-firms matching when firms' production technologies and worker's skills are both multidimensional. In this article, we provide fundamental welfare theorems when workers' skills are bundled (markets for stand-alone skills are missing). Sorting is then based on workers' comparative advantage. The bundling friction causes skills' prices to vary across firms, making the (unique) equilibrium wage schedule non-linear in skills. This non-linearity of wages, closely linked to the heterogeneity of workers' skills within firms, depends on the relative prevalence of specialist and generalist workers in the economy. In equilibrium, the wage is log-additive in worker quality and a worker-to-firm sorting effect that may reflect the firm's productivity. We provide descriptive evidence using Swedish matched employer-employee data providing us with direct measures of workers' cognitive and non-cognitive skills. Our companion paper, Choné, Gozlan, and Kramarz (2024), examines the same questions when markets for stand-alone tasks open and the bundling friction gradually disappears.

JEL Codes: D20, D40, D51, J20, J24, J30
Keywords: bundling; multidimensional skills; matching; sorting; heterogeneous firms

*We are grateful to Marco Cuturi, Xavier d'Haultfoeuille, Alfred Galichon, Nathael Gozlan, Gene Grossman, Philipp Kircher, Samuel Kortum, Clémence Lenoir, Ilse Lindenlaub, Simon Mongey, Giuseppe Moscarini, Bérengère Patault, François-Pierre Paty, Torsten Persson, Roland Rathelot, Jean-Marc Robin, Jean-Charles Rochet, Bernard Salanié, David Schönholzer, Oskar Skans, and Bob Topel for extremely helpful conversations, remarks, and suggestions as well those of participants at multiple seminars and conferences. We thank Gaspard Choné-Ducasse and Pierre Krakovitch for excellent research assistance. Choné gratefully acknowledges support from the French National Research Agency (ANR), (Labex ECODEC No. ANR-11-LABEX-0047) and Kramarz from the European Research Council (ERC), advanced grant FIRMNET (project ID 741467).

1 Introduction

Uberization, the Gig Economy ... Words often used in the press to identify new forms of labor. Despite important work by Acemoglu and his co-authors on robots, see Acemoglu and Restrepo (2018), or by Autor (with co-authors) on skills, tasks, and technology (see Autor, 2015 and references therein and, recently, Autor, Chin, Salomons, and Seegmiller, forthcoming), a convincing theoretical framework to think about these new jobs has yet to emerge. To model how labor markets operate today, we believe that understanding how they operated in the past constitutes a necessary and preliminary step.

To characterize the past forms of labor, we build on Mandelbrot (1962), the first to note "the impossibility of renting the different factors to the different employers", as cited in Heckman and Scheinkman (1987) (HS, hereafter). Hence, firms are forced to hire workers endowed with their entire skill-set. HS use the word *Bundling* to name this constraint: the impossibility to unpack a worker's fixed package of skills in order to sell each such skill on a separate market.

To fully explicit what the bundling constraint implies, let us examine how skills are used in production. In a classical production function, with multiple inputs (e.g. skilled blue-collar workers vs unskilled blue-collar workers) and *without bundling*, any given worker belongs to a **unique** input category, never to each one (e.g. both). By contrast, in our model (following HS and Mandelbrot), the production function also comprises multiple inputs (factors, using Mandelbrot's words) but each and every worker contributes to **each and every** input category **as a direct consequence of *bundling***.

To characterize the new forms of labor, our companion paper, Choné, Gozlan, and Kramarz (2024), examines how labor markets are transformed when worker's skills supply stops to be fixed (becomes endogenous) and, even more importantly in the new world, when markets for individual skills open thanks to increased access to outsourcing, temp agencies, or platforms; a process we call *Unbundling*.

More precisely, in this first paper we study how workers are matched to firms in a bundled world and the resulting wage structure. The theoretical modelling inspired by our knowledge of labor markets and their economic environment, will be shown to be informative about the specificity of a market "trading" humans and to be in sharp contrast with markets where products are traded.

In this article, we build on Heckman and Scheinkman's theoretical insight. A *Bundle* will denote a set of skills *when it cannot be unpacked.* This bundle of (exogenously given) skills is what the employing firm may use when it hires a worker. There are k different skills (aggregated to produce a set of k different tasks by the firm) and a worker's endowment is denoted by the skill vector $s = (s_1, \ldots, s_j, \ldots, s_k)$, with j being the index for the skill-type. We follow Heckman and Scheinkman in assuming that *each firm's production function depends on its workers' (bundled) skills aggregated by skill-types*, $T = (T_1, \ldots, T_j, \ldots, T_k)$ with $T_j = \int s_j$ (the integral being taken over the measure of workers employed in the firm), to produce a bundle of k tasks *rather than each worker's (job) production aggregated over workers (jobs) employed at the firm.*[1]

Introducing bundling will allow us to examine questions that a uni-dimensional framework is unable (have difficulties, at least) to address within a competitive framework. To name but a few, the respective roles and positions of Generalists (endowed with multiple skills in balanced quantities) versus those of Specialists (endowed with a limited set of skills), the existence of a comparative advantage in some skills (rather than an absolute advantage), the emergence of polarization in a sorting of workers-to-firms equilibrium, departures from the law of one price in the wage schedule ...

Hence, to address these questions (among others), we model a world where both firms and workers display rich multidimensional heterogeneity. We examine the *matching* of workers to firms and the induced *sorting*. More precisely, we study how a continuum of workers, endowed with multidimensional (exogenously given) skills, match with a continuum of firms, also endowed with multidimensional (also exogenously given) heterogeneity (rather than a 2-sector model with a continuum of identical firms within each sector, the setup adopted by HS) within a General Equilibrium (GE) setup in which firms are allowed to choose their size, a concept precisely defined later in this article.

Let us summarize *our main results*, first from an economic viewpoint, second from a methodological perspective.

First, analyzing the properties of the "Primal" of our (general) equilibrium with Bundling, we exhibit the allocation of workers to firms and the *sorting* patterns displayed at this equilibrium. Under usual single-crossing conditions of the firm's technology, aggregate sorting obtains and firms hire their unique preferred mix of skill-types, say the ratio X_2/X_1 in a two-skills world, a phenomenon that we label "sorting in the horizontal dimension".[2] Depending on the skills supply prevailing in the economy, this preferred mix is obtained by either hiring workers with exactly that preferred mix or by hiring a combination of workers delivering the same exact preferred mix (a pattern

[1] In most of our analysis, tasks are obtained by a firm-level linear aggregation of skills. We show how this assumption can be relaxed to better capture the intuitive meaning of these two concepts in Section 4.

[2] This two-skill world seems to resemble the so-called "Roy model" but we discuss below why our model vastly differs from it.

we call "Bunching" in this paper). To give an intuition of this last result, consider a world with two skills, 1 and 2. In this world, let us assume that the supply is restricted to two types of workers with exactly $(s_1, 0)$ for type 1 and $(0, s_2)$ for type 2. A firm that needs both skills to produce will hire a mixture of workers of type 1 and type 2 so as to obtain its optimal mix X_2/X_1. In this example, no worker in the firm is endowed with the optimal mix, and the wage will be shown to be linear in the two skills, with one unique price per skill. By contrast, when most of the supply is situated away from the axes and closer to the 45 degree line of the (s_1, s_2) quadrant, at the equilibrium all workers in the firm are endowed with their employing firm's optimal mix. In this case, the wage is nonlinear, with the implicit price of each skill depending on the worker's employing firm.

Again at the equilibrium, the model delivers predictions about sorting patterns in the vertical (quality) dimension. First, a given firm does not necessarily employ workers of the same quality. For instance, in the absence of bunching, when supply is located away from the axes, the employees of a given firm have skill sets of the form $x = (\lambda X_1, \lambda X_2)$: while they are all endowed with the firm's optimal mix X_2/X_1, they may be heterogeneous in their quality, i.e., λ may vary within a subset of $\mathbb{R}_+$. Yet we demonstrate the *uniqueness* of the firm-aggregated vector of skills at any competitive equilibrium and show that high-productivity firms will employ a high-quality labor force (endowed with a high total amount of the different skills). Hence, a high-quality labor force, a well-defined firm-level concept, may stem from hiring many average workers or from hiring a smaller number of excellent workers. It follows that *conditional on employment* high-productivity firms employ high-quality individual workers.

Second, analyzing the "Dual" of our "bundling equilibrium", we prove that this equilibrium is decentralized by a unique wage schedule. This wage schedule is shown to be a) a homogenous function of degree one in the "quality" of the worker; b) a non-linear convex function in the bundle of skills.[3] Hence, in equilibrium, the implicit price of each skill-type varies across firms and the law of one price does not apply: *there is more than one price per type of skill*, potentially an infinite number of such prices.[4] This result is a direct consequence of the inefficiency – *constrained efficiency* – induced by bundling: the impossibility of unpacking a worker's multidimensional skills, each skill being used in a separate input.

Put differently, the wage schedule in this bundled world is shown to be log-additive in the worker's quality and in a "worker-to-firm sorting effect". The latter reflects the firm's production technology with the associated optimal mix derived from the sorting of those skills central to the firm-specific production function. This result holds exactly

[3]Rather than a linear function in skills with returns allowed to differ in each sector, again as in HS.

[4]Even though there is a unique price per bundle.

in the convex portions of the wage schedule. As mentioned above, however, supply together with demand conditions may yield an equilibrium in which firms must mix workers with skills that differ from the optimal mix. Bunching is shown to prevail in regions where the wage schedule in skills is linear. When those regions are "small" enough, the wage function is close to such log-additivity. Hence, in our bundled world – with multidimensional skills and firms with heterogeneous production functions – a wage equation that shares some features (but differs, in important respects discussed in the present article, and summarized just below) with the one studied in Abowd, Kramarz, and Margolis (1999), is *pervasive*.[5] First, and in stark contrast with AKM, workers endowed with identical skills are paid identical wages *within each labor market* (the locus where demand equals supply, e.g. an occupation within a municipality). This results from perfect sorting of workers to firms in the absence of bunching, again *within each labor market*. In addition, again because workers sort perfectly in our bundling world, no firm-to-firm mobility takes place in this static context. However, in Choné, Gozlan, and Kramarz (2024) firm-to-firm mobility takes place as a result of supply and demand shocks. Hence, our worker-to-firm sorting effect does **not** capture the wage counterpart (due to some prevailing firm-level wage policy) of workers' firm-to-firm mobility.[6] But, as we will see later, it may still be considered as a firm-specific component of pay, under some precise conditions.

Indeed, the properties of the production function have important consequences on the equilibrium outcomes. First, the first-order homogeneity of the wage function is a direct consequence of our (within-firm) aggregation of skills into tasks (Choné, Gozlan, and Kramarz, 2023). Second, when the production function becomes non-homothetic, we will show that the "firm-effect" in the wage schedule depends on the firm's total factor productivity, z, whereas it is independent of z under homotheticity. However, independently of the homotheticity of the production function, more productive firms are larger at the equilibrium. Finally, at our bundling equilibrium, the firm's labor share is decreasing in the firm's productivity when the firm's production function is non-homogenous (see also Simonovska, 2015 and Autor, Dorn, Katz, Patterson, and Van Reenen, 2020).

From a modeling perspective, we believe our article proposes a potential building block for different literatures. In a world of large firms and with a continuum of workers, we propose new results and new methods, when firms *have substance* – they aggregate their workers' skills to produce, decide their optimal size – rather than being a mere

[5]We will use the expression AKM-like in the following, a slight abuse admittedly.

[6]As those resulting from models analyzed in Card, Cardoso, Heining, and Kline (2018) or, more recently, by Wong (2023) where it captures an endogenous mixture of rents and of a monopsonistic markdown.

collection of jobs, for the case of an often neglected friction, workers' skills bundling. As a consequence, our model sheds light on, at least, one profound difference between the labor market and the product market: aggregation. In the latter, you cannot aggregate two product characteristics (say, two cars with five horsepower do not generate one car with ten horsepower) whereas in the former such an aggregation is obviously possible for production. As an example of the model versatility, we show that it is easily incorporated in a Dixit-Stiglitz framework (see Appendix C). We believe that it can also be embedded within a search environment or in other popular models of the literature. We also examine how technological change affects our Bundling equilibrium (Appendix D).

The limits to the model studied in the following are very clear: a "fixed, but bundled, exogenous labor supply". Hence, to circumvent some of these limits, we extend our model *in our companion paper*, Choné, Gozlan, and Kramarz (2024) in two ways. First, we introduce an "endogenous labor supply in a bundled environment", expliciting in particular how hours of work and their associated costs affect workers' skills supply. Second, we allow workers to sell their individual skills on markets. Combining these two possibilities will help us better contrast the old world, and the (new) world; a new world characterized by opening markets, through better technology, globalization, temp agencies, or, more recently, platforms. Hence, a world in which skills are easier, potentially at a cost, to alter and to unbundle.

In particular, we analyze the effect of increased market access on the matching of workers to firms. Workers' labor supply becomes endogenous: workers can choose how much skill to supply to their firm and how much skill to supply to the market. We also characterize the resulting wage schedule, the workers who benefit from this opening, and those harmed by it.

To take stock of the connections between what is indeed a very theoretical contribution and the labor market features that inspired our theory, we summarize the empirical content of our model in a separate Section. In particular, we briefly present results from two papers, Fredriksson, Hensvik, and Skans (2018) (FHS, hereafter) and Skans, Choné, and Kramarz (2022), co-written with Oskar Nordström Skans, who both use Swedish data on workers' skills, employers, and occupations. Both provide support to our model, in particular the role of comparative advantage in sorting and wages.

Connecting Literatures Our theoretical contribution incorporates four ingredients: 1) a continuum of heterogeneous workers with multidimensional skill-types; these skills being bundled in this paper; 2) a continuum of firms with heterogeneous and multidimensional production functions in which the (intermediary) inputs are tasks; 3) tasks are obtained by (type by type) aggregation of workers' skills employed at the firm *rather than by the aggregation of workers' individual production*; 4) an endogenous firm size.

We now examine the various articles that incorporate some (but not all) of these ingredients.

Bundling Multidimensional Skills: HS is the first paper (to the best of our knowledge) examining the consequences of skills' bundling. These authors were trying to understand whether the bundling of skills (first ingredient above) together with production obtained from an aggregation of workers' skills (third ingredient) could generate different returns to each skill in different sectors, in an economy with n sectors (and identical firms within each sector, the firms playing essentially no role). Their answer was positive: returns to skills could differ across sectors, in this Roy-style model.[7] Unfortunately, they did not provide general conditions for their result. Nor did they examine the structure of the matching between workers and firms (sectors) or the resulting wage schedule. By contrast, Lindenlaub (2017) focuses on sorting and provides a full characterization of positive assortative matching, PAM, or its negative counterpart, NAM, in a multidimensional framework with jobs but no aggregation of skills entering a firm-level production function (and, hence, no firm size). Lindenlaub and Postel-Vinay (2023) builds on Lindenlaub (2017) by adding random search to the initial sorting problem, a dimension that we do not examine here. Clearly, the search dimension brings important insights into skill-specific job ladders and the induced sorting of workers' skills bundles to jobs, rather than jobs. Because Lindenlaub (2017) is an important step in the study of the matching of workers to jobs in this multi-dimensional (with bundling) context, we will relate her results to ours directly within the body of our text.

Edmond and Mongey (2022) also examine bundling using a model with two tasks and two skills, adopting a macroeconomic perspective.[8] As in our approach, their workers are heterogeneous in their skill endowments. As in Murphy (1986) and HS, they have two firms in their economy (or, rather, two occupations). As we do here, each task (occupation, in their model) is produced from skills (using a CES function, in their model). Again, as we do, output is produced using the supply of both tasks as inputs. Because they have two occupations producing output, the question of sorting of workers to the two occupations is the one they ask (rather than across firms). In a recent contribution, Hernnäs (2021) studies the consequences of bundling in a world where tasks can be automated, using a framework close to that of Edmond and Mongey

[7] Because Roy (1951) is a stepping stone in this literature, we discuss extensively how our contribution relates to his in Subsection 3.6.

[8] They also study unbundling of skills (using this word as we do) in their contribution.

(2022). The paper shows that skill returns in the automated task decline if tasks are gross complements. More generally, Hernnäs (2021) allows to examine automation in a rich setting.

Consequences of skills-bundling have also been studied in International Trade (Ohnsorge and Trefler (2007)). There, workers have bundled skills and production relies on jobs rather than firms. Because the focus is on comparative advantage, we come back to this contribution just below.

Comparative Advantage in the Vertical Dimension: A macroeconomic literature studying trade, comparative advantage, and technical change has direct connections with our approach.

Costinot and Vogel (2010) combine a Roy-like assignment model and a Dixit-Stiglitz setting. Their workers are heterogeneous in a single dimension and, because there is a market for each task, full unbundling of tasks/skills prevails. Their high-skill workers have a comparative advantage in tasks with high-skill intensity – what we call the vertical dimension – resulting in pure sorting between skills and tasks. By contrast, in our approach, workers have bundled, multi-dimensional skills. And, in equilibrium, those workers with *a comparative advantage* in one skill – what we call the horizontal dimension – will work in a firm that values this exact skill more. In Appendix C, we study a Dixit-Stiglitz variant of our model, sharing features with Costinot and Vogel (2010).[9] It allows to clearly see how different environments (pure competition versus monopolistic competition, in this case) deliver similar effects (albeit based on different formulas) due to bundling, firm-specific aggregation of skills with real firm-level production functions, and endogenous size.

Ohnsorge and Trefler (2007) also have multidimensional skills (but no firms) and show that international differences in the distribution of workers' skill bundles, such as Japan's abundance of workers with a modest mix of both quantitative and teamwork skills, have important implications for international trade, industrial structure, and domestic income distribution.[10]

Connected to this trade literature, with a clear focus on comparative advantage in labor markets, two contributions must be mentioned. First, Teulings (2005) presents a theory of factor substitutability in a model with a continuum of worker and job (both uni-dimensional) types, with highly skilled workers having a comparative advantage in complex jobs. This model allows to generate patterns of substitutability between types that decline with their skill distance. Second, in a recent and very interesting article,

[9] We thank Sam Kortum for this suggestion.

[10] With Bundling, random search, and jobs rather than firms, the equilibrium in Lindenlaub and Postel-Vinay (2023) also displays elements of comparative advantage.

Haanwinckel (2023) contributes to this labor literature. His task-based production function requires combining tasks of different complexity levels, with task requirements depending on the good the firm decides to sell. As in Teulings (2005), the comparative advantage structure is uni-dimensional, corresponding to what we label the vertical dimension of skills. Interestingly, the firm assigns (optimally) each worker to tasks, resulting in within-firm heterogeneity in workers' types.

Giving Firms Substance: Our research is also inspired by a recent and important contribution, Eeckhout and Kircher (2018), in which assortative matching in so-called large firms is analyzed. Workers in their approach have one dimension of skills (hence, one type). However, to obtain firms that are more than a collection of jobs, Eeckhout and Kircher (2018) separate workers' quality from workers' quantity and assume constant returns to scale in those quantity variables. In addition, management decides the firm's span of control by setting the firm's "resources". This allows them to study rich patterns of sorting in which quality and quantity dimensions both play a role. The resulting sorting condition combines four types of complementarities. As a result of the constant returns assumptions (in particular), at the equilibrium a firm of quality y hires only one quality of worker x, hence the model generates no within-firm worker's heterogeneity. Unfortunately, very few other contributions address this firm's substance challenge. We mentioned above Haanwinckel (2023). A recent and interesting contribution is Boerma, Tsyvinski, and Zimin (2021) with firms of exogenous size (equal to two). Their model includes a team production function with bundling and heterogeneous firms (in productivity only, though). Their interest lies in the matching between such firms and workers.[11]

Connection to Hedonic Models: Following Rosen (1974), hedonic models applied to products markets focus on the matching of consumers and goods on the demand side, with goods and firms on the supply side. Matching models of the labor market are somewhat similar: workers' skills and tasks lie on the supply side when and tasks and firms lie on the demand side. At first glance, matching models look very much like hedonic models.[12] Our model, however, departs from an hedonic model of the labor market in two important dimensions. First , *tasks are not directly observed* by

[11]Firms also play a role in recent GE models of monopsonistic labor markets, such as Berger, Herkenhoff, and Mongey (2022) (see also references, therein). A finite number of firms in a market, each firm having an upward sloping labor supply curve, face workers endowed with different tastes for firms. The resulting equilibrium yields a markdown of wages. As mentioned earlier, our generalists – most constrained by bundling – also face a "markdown" within a purely competitive framework (except for bundling).

[12]Chiappori, McCann, and Nesheim (2010) express hedonic models as optimal transport problems. We discuss such connections below.

the researcher.[13] Hence, we model tasks as an unobserved function of skills. Second, the production process considered involves the aggregation of employees' skills within firms, whereas Rosen (1974) explicitly rules out such an aggregation,[14] something he calls buyer's arbitrage. Arbitrage in his context would consist in generating a new good by taking a linear combination of two goods' attributes which would force the price of the good to be linear (Rosen, 1974, page 37, last paragraph).

Optimal Transport and Matching: A growing strand of the literature leverages the insights of optimal transport theory to study the matching of agents in competitive markets.[15] Important papers there consider one-to-one matching, e.g. in the labor market (Lindenlaub, 2017) or in the marriage market (Galichon and Salanié, forthcoming). Boerma, Tsyvinski, and Zimin (2021), briefly presented above, use the multi-marginal version of optimal transport. By contrast, our analysis contributes to the literature that studies many-to-one matching with transferable utility. Because our model comprises a model of production in which skills are aggregated within firms, we are forced to use relatively new methods from OT theory, namely the so-called weak optimal transport (WOT) introduced by Gozlan, Roberto, Samson, and Tetali (2017). Because firms choose their size, we are again forced to rely on a very recent extension of WOT introduced by Choné, Gozlan, and Kramarz (2023). Appendix B presents the exact connections between competitive matching as examined here and optimal transport theory. Matching in labor markets was also examined in its discrete (game-theoretic) version[16]. We depart from this strand by working with a continuum of workers and a continuum of firms (the so-called "large firms" literature).

Bunching and bundling Using the literature on multidimensional optimal transport, Chiappori, McCann, and Pass (2016) connect their work to the multidimensional screening literature and argue that the bunching phenomenon, observed by Rochet and Choné (1998) in the monopoly context, does not occur in the competitive context. In the present paper, we find something akin to bunching in a competitive environment with multidimensional types where firms and workers have the same dimension of het-

[13] Indeed, we are not aware of any data source that would offer a comprehensive picture: workers' exact skills, the exact tasks each worker performs, together with the worker's employing firm. Often occupations are used as a proxy even though the tasks performed by the worker in her employing firm are virtually never measured. However, Bittarello, Kramarz, and Maitre (forthcoming) show the extent of dispersion in tasks within occupations.

[14] Two cars with 50 horsepower each are not equivalent to one with 100 horsepower is an obvious example. See also Lancaster (1966).

[15] See Villani (2009) for the mathematical theory, Galichon (2018) for applications to the economics of matching, and Peyré and Cuturi (2019) for computational optimal transport.

[16] Crawford (1991), Kelso and Crawford (1982), Hatfield and Milgrom (2005), and, more recently, Pycia (2012) and Pycia and Yenmez (2019) have contributed to this strand.

erogeneity. Indeed as explained above, in any bundling equilibrium, each firm has a preferred mix of skill-types that depends on its productive characteristics. And firms with different characteristics have different optimal mix (full sorting between firm-types and optimal mix of workers' types). However, in conditions of workers' supply of skill-types that we characterize, this optimal mix can **only** be achieved by combining workers endowed with different skill-types. In this precise situation, firms of different types optimally hire workers endowed with the exact same skill-type to achieve their (different) optimal mix; a phenomenon we call "bunching".

In the next Section, we present our model setup, when bundling prevails. Then, Section 3 examines how firms and workers are matched, again under bundling. Next, we discuss the empirical consequences of our model (Section 4). In the same Section, we very briefly mention empirical evidence based on a summary of a paper, co-written with Oskar Nordström Skans, Skans, Choné, and Kramarz (2022), in which we study aspects of the empirics of bundling (and unbundling) using Swedish data, a data set also used in Fredriksson, Hensvik, and Skans (2018). Section 5 concludes. All proofs are relegated to the Appendix.

2 Model Setup Under Bundling

The production process involves k intermediary inputs produced by workers, which we call tasks. Firms aggregate the tasks performed by their employees and transform them into final output. They are heterogeneous in their production technologies. Denoting by $T = (T_1, \dots, T_k)$ the aggregate vector of tasks produced by its employees, a firm of type ϕ produces final output $F(T; \phi)$, with F being nonnegative, continuous in (T, ϕ), concave in T, differentiable in T with partial derivatives continuous in (T, ϕ). We furthermore assume that the marginal productivities $\partial F \partial T_j(T; \phi)$ are continuous in ϕ. Firms' types are distributed according to a probability measure $H^f(\mathrm{d}\phi)$ on a compact support $\mathcal{X}^f \subset \mathbb{R}^k_+$.

Workers are indexed by vectors $s = (s_1, \dots, s_k)$, which we refer to as their "skills". The skill vectors s belong to a compact subset $\mathcal{X}^s$ of $\mathbb{R}^k_+$ and are distributed according to a probability measure $H^w(\mathrm{d}s)$ in the population of workers. We assume that there is no mass of workers with zero skills, $H^w(\{0\}) = 0$. We define the overall quality of a worker as the Euclidian norm $|s|$ of her skill vector and her skill profile as $\tilde{s} = s/|s|$. We refer to the former and latter respectively as to the vertical and horizontal dimensions of workers' heterogeneity. Skill profiles – the horizontal dimension of skills – reflects the workers' comparative advantage in any particular type of skill.

In most of this paper, we assume that a worker with skill set s produces task $t = s$.[17] In other words, we equate skills with tasks. In our companion paper, Choné, Gozlan, and Kramarz (2024), we endogenize the relationship between skills and tasks at the level of the workers, by modeling the production of tasks by workers and their choice of work hours. We thus check that the model presented here is a special case of a general model with endogenous labour supply.

Following in that Acemoglu and Autor (2011), our baseline specification assumes that firms aggregate the skills performed by their employees in a linear way. Given that an employee with skill vector s produces task t, the amount of task j produced in firms of type ϕ is

$$T_j^{\mathrm{D}}(\phi) = \int s_j \, n^{\mathrm{D}}(\mathrm{d}s; \phi), \tag{1}$$

where $n^{\mathrm{D}}(\mathrm{d}s; \phi)$ is a non-negative finite measure on $\mathcal{X}^s$ that represents the number of workers with skills s hired by a firm of type ϕ. We discuss alternative and rich aggregation schemes when analyzing the empirical content of our model (in Subsection 4.3).

An assignment of workers to firms is a family of a non-negative finite measures $n^{\mathrm{D}}(\mathrm{d}s; \phi)$ on $\mathcal{X}^s$. It is important to stress that we use "unnormalized" measures. The number of workers employed by a firm of type ϕ, which we denote by $N(\phi) = n^{\mathrm{D}}(\mathcal{X}; \phi)$, need not be one and the distribution $n^{\mathrm{D}}(\mathrm{d}s; \phi)$ need not be a probability measure. In fact, *the firms' sizes are endogenously determined in equilibrium.*

An assignment n^{D} "clears" the labor market if

$$\int n^{\mathrm{D}}(\mathrm{d}s; \phi) \, H^f(\mathrm{d}\phi) = H^w(\mathrm{d}s) \tag{2}$$

for H^w-almost all skill vectors $s \in \mathcal{X}^s$. In other words, market-clearing assignments "disintegrate" the skill distribution $H^w(\mathrm{d}s)$ and quantify the number of workers with skill s hired by firms of any type ϕ. Below, we often write the market clearing equation (2) in the shorter form $n^{\mathrm{D}} H^f = H^w$. Integrating this equation with respect to s shows that, for any market-clearing assignment n^{D}, the expected number of employees over all firms is one (w.l.o.g):

$$\int N(\phi) H^f(\mathrm{d}\phi) = 1. \tag{3}$$

It follows that the modified distribution of firms' types $\tilde{H}^f(\mathrm{d}\phi) = N(\phi) H^f(\mathrm{d}\phi)$ is a probability measure that is absolutely continuous with respect to the original distribution $H^f(\mathrm{d}\phi)$. Introducing $q(\mathrm{d}s; \phi) = n^{\mathrm{D}}(\mathrm{d}s; \phi)/N(\phi)$, a probability measure for any ϕ,

[17] We consider exogenous skills-to-tasks relationships of the form $t = g(s)$ in Subsection 4.3.

shows that the matching between workers' and firms' types

$$\pi(\mathrm{d}s, \mathrm{d}\phi) = n^{\mathrm{D}}(\mathrm{d}s;\phi)H^f(\mathrm{d}\phi) = q(\mathrm{d}s;\phi)\tilde{H}^f(\mathrm{d}\phi) \tag{4}$$

is a transport plan between the original skill distribution $H^w(\mathrm{d}s)$ and the modified firm distribution $\tilde{H}^f(\mathrm{d}\phi)$.

We say that *a market-clearing assignment* n^{D} *is optimal* if it maximizes total output in the economy, i.e., if it solves

$$\mathcal{J}^b(H^f, H^w) \stackrel{\mathrm{d}}{\equiv} \sup_{n^{\mathrm{D}} \mid n^{\mathrm{D}} H^f = H^w} \int F\left(\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi);\phi\right) H^f(\mathrm{d}\phi). \tag{5}$$

We show in Lemma A.1 that $\mathcal{J}^b(H^f, H^w) < \infty$. Whenever the production function F is nonlinear in the firm-aggregate vectors of tasks T, the total output in the economy is a nonlinear function of the assignment n^{D}. By contrast, if firms' production were just the sum of each of their employees' production – which is not what we do here – total output $\iint F(s;\phi)\, n^{\mathrm{D}}(\mathrm{d}s;\phi)H^f(\mathrm{d}\phi)$ would be linear in n^{D}.

Finally, we introduce the notion of competitive equilibrium. Under bundling, a worker's set of skills cannot be unpacked, hence firms must purchase her entire skill package $s = (s_1, \ldots, s_k)$. The workers' skills (or equivalently the tasks they perform) are observed by firms and are contractible. The wage of a worker with skill s is denoted by $w(s)$. The wage schedule $w(.)$ is therefore a map: $\mathcal{X}^s \to \mathbb{R}_+$. We rule out agency problems: a firm that hires a worker with skill s obtains the vector of tasks s in return for the paid wage $w(s)$. Given a wage schedule $w(.)$, the demand for skill is the assignment $n^{\mathrm{D}}(\mathrm{d}s;\phi)$ on $\mathcal{X}^s$ that maximizes the firms' profit:

$$\Pi(\phi;w) = \max_{\nu \in \mathcal{M}(\mathbb{R}^k_+)} F\left(\int s\,\nu(\mathrm{d}s);\phi\right) - \int w(s)\,\nu(\mathrm{d}s), \tag{6}$$

where $\mathcal{M}(\mathbb{R}^k_+)$ is the set of all positive measures on $\mathbb{R}^k_+$. A ***competitive equilibrium*** *is a pair* (n^{D}, w) *composed of a wage schedule* w *and a market-clearing assignment* n^{D} *such that* n^{D} *reflects the demand for skills under the wage* w*, i.e.,* $n^{\mathrm{D}}(\mathrm{d}s;\phi)$ *solves* (6) *for all firms' types* ϕ*.*

Of particular interest to us are the production functions of the form $F(T;\phi) = zF(T;\alpha)$ with the firms' types $\phi = (\alpha, z)$ having two components: z reflects total factor productivity and α reflects the relative importance of each task in the production process. We assume that the worker and firm heterogeneities have the same dimension, hence α lies in a space of dimension $k - 1$. Our leading example exhibits constant elasticity of substitution and decreasing returns to scale:

$$zF(T;\alpha) = (z/\eta)\ \left[\sum_{j=1}^{k} \alpha_j T_j^{\rho}\right]^{\eta/\rho}, \tag{7}$$

with $\sum_{j=1}^{k} \alpha_j = 1$, $0 < \eta < 1$, and $\rho < 1$. When $\rho < \eta$, the function displays increasing marginal productivities of aggregate skill types, $\partial^2 F/\partial T_j \partial T_k > 0$ for all $j \neq k$.[18] In other words, the marginal productivity of a worker in one skill increases with her co-workers' other skills. Under this specification, complementarities across workers result from complementarities across skill types.

Remark 1: The above CES example offers another opportunity to reiterate the nature of the bundling constraint. Each worker endowed with k skills provides each such skill to each skill type j, whereas in usual production functions, a worker contributes to a unique skill type (say skilled or unskilled blue-collar worker) which is then aggregated across workers.

Remark 2: The production functions considered in this paper, where output depends only on the within-firm aggregate vector, $\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)$, gives rise to a class of weak optimal transport problems, which Choné, Gozlan, and Kramarz (2023) refer to as "conical problems", see Appendix B. It is the simplest and most natural extension of usual production functions when incorporating the bundling friction. This conical class brings structure to the problem, as we see just below. We discuss in the Conclusion how the conical assumption can be relaxed to allow for a broader range of production functions.

3 Matching Workers and Firms Under Bundling

Under bundling, there are no markets for individual skills. Firms can acquire intermediary inputs only from their employees. Once hired, a firm can use the entirety of a worker's skills. In addition, we assume that a worker cannot be employed by more than one firm.[19]

In Subsection 3.1, we prove the existence of competitive equilibria using new insights from optimal transport theory. In Subsection 3.6, we connect our model with Roy (1951) and in Subsection 3.2, we examine how the firm-*aggregated* vectors of tasks depend on

[18] In this case, the aggregate skills are gross-complements, i.e., the demand for one skill is non-increasing in the prices of the other skills, see Theorem 4.D.3 in Takayama (1985).

[19] These assumptions are relaxed in our companion paper Choné, Gozlan, and Kramarz (2024).

the firms' technologies. Then, assuming homothetic production functions, we study the sorting of *individual* workers into firms. In Subsection 3.3, we focus on cases where pure sorting in the horizontal dimension obtains. In Subsection 3.4, we describe situations where, by contrast, skill profiles are heterogeneous within firms.

3.1 Competitive Equilibria and the Structure of Wages

Competitive equilibria will be shown to exist under the following assumptions:

Assumption 1. *For all $T \in \mathbb{R}_+^d$ and all $\phi \in \mathcal{X}^f$, $F(\mu T;\phi)/\mu$ tends to 0 as $\mu \to +\infty$.*

Assumption 1 holds for homogenous production functions with decreasing returns to scale, as is the case in our leading example (7). It also holds for the non-homothetic production function (29) used in Section 4.1.

The bundling environment is characterized by missing markets. Firms cannot purchase some amount of skills, separately for each skill type $j = 1, \ldots, k$. Proposition 1 below states new versions – to the best of our knowledge – of the two fundamental theorems of welfare economics that are adapted to this constrained environment. In particular the notion of optimality refers to the "Primal" Problem (5), which includes the constraints that only workers can be hired and that only skill-*vectors* can be traded.

Proposition 1 (The Fundamental Theorems Under Bundling)**.** *Suppose Assumption 1 holds. Then there exist optimal market-clearing assignments of workers to firms. Any such assignment can be decentralized by a wage schedule w. Conversely, any equilibrium assignment is optimal.*

In Appendix A.1, we use new tools in optimal transport theory to prove the above results. These tools allow us to overcome two challenges. First, contrary to the classic OT framework, the primal problem (5) is a nonlinear function of the assignment n^{D}. For this reason it belongs to the class of weak optimal transport problems introduced by Gozlan, Roberto, Samson, and Tetali (2017). Second, firms' sizes are endogenous in equilibrium, and thus $n^{\mathrm{D}}(\mathrm{d}s;\phi)$ is not necessarily a probability measure for all firms' types ϕ. In the terminology of Choné, Gozlan, and Kramarz (2023), assignments of workers to firms are *unnormalized* kernels. The primal problem can be rewritten as a transport problem over the set of probability distributions whose first marginal is absolutely continuous with respect to the distribution of firms' types, H^f, recall the discussion after (3). (Appendix B presents in greater detail the connection between competitive matching equilibria and optimal transport theory.)

The next proposition describes in more detail the structure of wages. It strongly relies on the conical nature of the production function discussed at the end of Section 2. We briefly discuss in Subsection 4.3 how wages are affected with more general aggregation schemes. Hereafter we denote by Φ^+_{nd} the set of all convex, positively one-homogenous, and non-decreasing functions $w : \mathbb{R}^k_+ \to \mathbb{R}_+$.

Proposition 2 (Structure of Wages)**.** *Suppose Assumption 1 holds. Then any optimal market-clearing assignment can be decentralized by a wage schedule $w \in \Phi^+_{\text{nd}}$ that is convex and homogenous of degree one.*

The convexity and homogeneity of the wage schedule come from the linear aggregation of skills within firms, given by equation (1). They guarantee the absence of arbitrage opportunities for firms. If these properties did not hold, firms could reduce their wage bill by replacing some workers with combinations of workers yielding the same aggregate skills.

Suppose for instance that there exist worker types s, s', and s'' such that $s'' = \nu s + (1-\nu)s'$ with $0 < \nu < 1$, $w(s) = w(s') = 1$, and $w(s'') > 1$. Then, no firm would want to hire type-s'' workers because a combination of type-s and type-s' workers would deliver the same amount of intermediary inputs in return for a lower wage bill. Specifically, diminishing demand $n^{\text{D}}(s'', \phi)$ by ε and increasing $n^{\text{D}}(s, \phi)$ by $\nu\varepsilon$ and $n^{\text{D}}(s', \phi)$ by $(1-\nu)\varepsilon$ leaves the firm-aggregated vector of skills (hence the firm-level vector of tasks) unchanged and reduces the wage bill.

To prove homogeneity, consider two workers with proportional skills s and μs for some $\mu > 0$. These workers have the same relative skill endowments but differ in their overall quality, embodied by the multiplicative factor μ. Assume, by contradiction, that $w(\mu s) < \mu w(s)$. Then no firm would hire worker type s as diminishing $N(s; \phi)$ by ε and increasing $N(\mu s; \phi)$ by ε/μ leaves the firm aggregate of skills unchanged while reducing the wage bill. It follows that the demand for worker s is zero, a contradiction. The reverse inequality, $w(\mu s) > \mu w(s)$, is ruled out by the same argument.

Proposition 2 implies that the wage is sub-additive, which has important economic implications. Let (e_i) be the canonical basis of $\mathbb{R}^k$, i.e., $e_i = (0, \ldots, 1, \ldots, 0)$, with 1 in the ith coordinate. Because w is convex and homogenous of degree one, it is sub-additive, hence

$$w(s) = w\left(\sum_{i=1}^{k} s_i e_i\right) \leq \sum_{i=1}^{k} w(e_i s_i) = \sum_{i=1}^{k} w(e_i) s_i. \tag{8}$$

Hereafter, we call a worker *specialist* if she is endowed with an unbalanced set of skills, with one dominating skill, and *generalist* if she is endowed with a balanced set of skills. The subadditivity property (8) expresses that it is less costly for firms to hire a generalist worker with skill set $s = (s_1, \ldots, s_k)$ than k specialist workers endowed with the corresponding amount s_i of skill in each dimension.

3.2 Aggregate Sorting

The firms' problem (6) can be broken down into two sub-problems that consist respectively in finding the firm-aggregated skill vector $T^{\mathrm{D}}(\phi) = \int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)$ and in achieving that aggregate vector in the most economical way. In this subsection, we study the properties of the aggregated skill vector $T^{\mathrm{D}}(\phi)$ and examine how it varies with the firms' technological characteristics ϕ.

Proposition 3 (Uniqueness of wage schedule and firm-aggregated skill vector)**.** *Suppose Assumption 1 holds and assume furthermore that $F(T;\phi)$ is* strictly *concave in T. Then the firm-aggregated skill vector $T^{\mathrm{D}}(\phi) = \int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)$ and the wage schedule $w \in \Phi^{+}_{\mathrm{nd}}$ are unique among all competitive equilibria (n^{D}, w). We have*

$$\Pi(\phi; w) = \max_{T} F(T;\phi) - w(T), \tag{9}$$

where w is any equilibrium wage schedule that is convex and homogenous of degree one.

Since F is concave and w is convex, the above problem is well-posed, with a unique solution characterized by

$$F_j(T^{\mathrm{D}}(\phi);\phi) = w_j(T^{\mathrm{D}}(\phi)). \tag{10}$$

At any competitive equilibrium, the productivity of each skill equals its marginal price. When the wage schedule is locally linear, i.e., is of the form $\sum \bar{w}_j s_j$, we are back to $F_j(T^{\mathrm{D}}(\phi);\phi) = \bar{w}_j$, i.e., the price of each skills equals its marginal productivity. More generally, the productivity of skill j equals its *implicit* price in the neighborhood of the aggregate skill T, i.e., the partial derivative $w_j = \partial w/\partial s_j$ evaluated at that point. Figure 1 shows the tangency of the firm's production isoquant and the iso-wage surface.

In the rest of this subsection, we study how the aggregate vector $T^{\mathrm{D}}(\phi)$ varies with the firm's type ϕ. We distinguish the (quality-adjusted) size of a firm and the aggregate profile of its employees. Specifically, we write the firm-aggregated skill vector of firm ϕ as $T^{\mathrm{D}}(\phi) = \Lambda^{\mathrm{D}}(\phi)\tilde{S}^{\mathrm{D}}(\phi)$, where $\Lambda^{\mathrm{D}}(\phi) = |T^{\mathrm{D}}(\phi)|$ is the total quality of the firm's employees and $\tilde{S}^{\mathrm{D}}(\phi)$ is their average skill profile.

Corollary 1 (Matching of aggregate skill profiles)**.** *Assume that production functions have homothetic isoquants. Then we have:*

$$\frac{F_j(\tilde{S}^{\mathrm{D}}(\phi);\phi)}{F_k(\tilde{S}^{\mathrm{D}}(\phi);\phi)} = \frac{w_j(\tilde{S}^{\mathrm{D}}(\phi))}{w_k(\tilde{S}^{\mathrm{D}}(\phi))}. \tag{11}$$

If a firm's technology is more intensive in skill j, the firm uses relatively more of that skill.

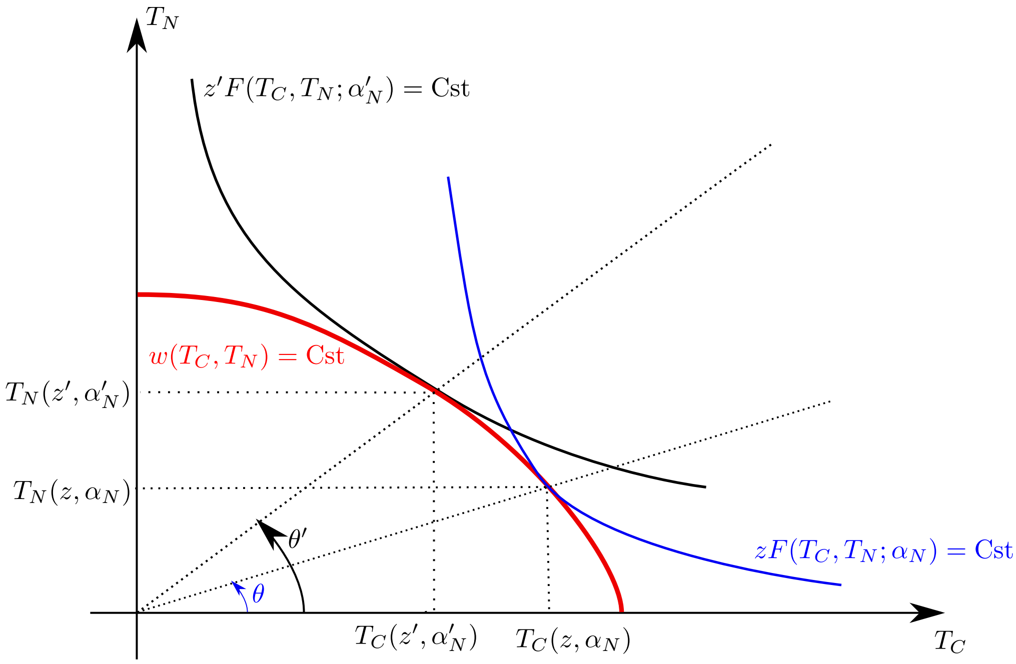


Figure 1: Matching in the horizontal (skill profile) dimension: Firm (α'_N, z') is more intensive in skill N than firm (α_N, z), with $\alpha'_N > \alpha_N$

The aggregate profile of the workers employed by a firm therefore depends on the marginal rates of technical substitution. When ϕ takes the form $\phi = (\alpha, z)$, where z reflects total factor productivity, i.e., $F(T, \phi) = zF(T, \alpha)$, these rates do not depend on TFP, z. As a consequence, the same is true for aggregate skill profile: $\tilde{S}^{\mathrm{D}}(\phi)$ depends only on the technological intensity parameters α that reflect the importance of each task for the firm. This is the case for instance in our leading example (7), for which $F_j/F_j = (\alpha_j/\alpha_k)(T_k/T_j)^{1-\rho}$.

Corollary 2 (Homogenous production functions and TFP)**.** *Assume furthermore that the production functions are homogenous of degree $\eta < 1$. Then the firm-aggregated intermediary input $T^{\mathrm{D}}(\phi)$, the firm's wage bill, and the firm's profits are proportional to $z^{1/(1-\eta)}$, where z denotes firm's total factor productivity.*

Two skills ($k = 2$): Suppose there are two (cognitive and non-cognitive) skills C and N. We may represent the firm-aggregated skill vector $T^{\mathrm{D}} = (\Lambda^{\mathrm{D}} \cos\theta^{\mathrm{D}}, \Lambda^{\mathrm{D}} \sin\theta^{\mathrm{D}})$ in polar coordinates, where Λ^{D} reflects the size of the firm defined as the total quality of its employees.

Proposition 4. *Assume that there are two skills/tasks and that the production $zF(T; \alpha_N)$ is concave in T. Then the total quality of the workers employed by a firm, $\Lambda^{\mathrm{D}}(\alpha_N, z)$, increases with the firm's total factor productivity z.*

Assume furthermore that the production functions have homothetic isoquants and that F_N/F_C increases with α_N. Then the firm-aggregated matching $(\theta^{\mathrm{D}}(\alpha_N,z),\Lambda^{\mathrm{D}}(\alpha_N,z))$ exhibits positive assortative matching in the sense of Lindenlaub (2017).

Hence, total quality Λ^{D} increases with TFP z. In addition, with homothetic isoquants, the aggregate workers-to-firms matching pattern exhibits positive assortative matching (PAM), in the sense that the Jacobian $D_{(\alpha_N,z)}(\theta^{\mathrm{D}},\Lambda^{\mathrm{D}})$ is a P-matrix, i.e., all the principal minors of the Jacobian are positive.[20] In contrast to Lindenlaub (2017), however, the above PAM property applies in our context to firms' *aggregates* rather than to individual workers' characteristics. At the individual level, two points are worth mentioning. First, even though the workers-to-firms matching is arbitrary in the vertical dimension (worker qualities), we explain in Section 4.1 that the monotonicity of the total quality of employees with the firms' total factor productivity does have testable implications. Second, regarding the horizontal dimension (worker profiles), workers' sorting patterns may be blurred by bunching, as we discuss in Section 3.4.

Two-skill CES example Denoting by C and N the cognitive and non-cognitive skills, we consider the production function (7):

$$zF(T_C,T_N;\alpha_N)=\frac{z}{\eta}\left[\alpha_C T_C^{\rho}+\alpha_N T_N^{\rho}\right]^{\eta/\rho}, \tag{12}$$

with $\alpha_C+\alpha_N=1$. Denoting $\tilde{S}^{\mathrm{D}}=(\cos\theta^{\mathrm{D}},\sin\theta^{\mathrm{D}})$ the average skill profile, the general workers-to-firms matching condition (11) writes

$$\left[\tan\theta^{\mathrm{D}}(\alpha)\right]^{1-\rho}=\frac{\alpha_N}{1-\alpha_N}\,\frac{w_C\left(\theta^{\mathrm{D}}(\alpha)\right)}{w_N\left(\theta^{\mathrm{D}}(\alpha)\right)}. \tag{13}$$

The matching between workers and firms is represented by the increasing function $\theta^{\mathrm{D}}(\alpha)$ implicitly defined by (13). The relative skill endowment in non-cognitive skills of the workers, $\theta^{\mathrm{D}}(\alpha)$, increases with the demand intensity in that skill, α_N, as illustrated on Figures 1 and 3. Equation (A.12) in the Appendix gives the aggregate quality of the firms' employees.

3.3 Pure Sorting in the Horizontal Dimension

We now examine the matching of worker skills s to firm technologies ϕ, which is represented by the transport plan π given by (4). In this subsection and the next, we focus on the horizontal dimension of skills, i.e., on the profiles $s/|s|$ of workers employed by

[20] In Appendix A.4, we provide a sufficient condition for PAM that does not require homothetic production isoquants, see inequality (A.18).

any given firm. To do this, we examine the second part of a firm-ϕ's problem, namely achieving the aggregated skill vector $T^{\mathrm{D}}(\phi)$ in the most economical way:

$$w(T^{\mathrm{D}}(\phi)) = \inf \left\{ \int w(s)\, n^{\mathrm{D}}(\mathrm{d}s) : \, n^{\mathrm{D}} \in \mathcal{M}(\mathcal{X}^s), \int s\, n^{\mathrm{D}}(\mathrm{d}s) = T^{\mathrm{D}}(\phi) \right\}, \tag{14}$$

where $\mathcal{M}(\mathcal{X}^s)$ is the set of all positive measure on $\mathcal{X}^s$ and w is convex and homogenous of degree one.

We start with the case where the iso-wage surfaces are strictly concave. Under this circumstance, the minimization of the wage bill at a given aggregate skill in (14) imposes that firm ϕ hires only workers with skill profile $\tilde{S}^{\mathrm{D}}(\phi) = T^{\mathrm{D}}(\phi)/\Lambda^{\mathrm{D}}(\phi)$. We can thus show that the equilibrium is characterized by a one-dimensional condition. To prove this result, we use the projection $s/w(s)$ onto the iso-wage surface $w = 1$. Because for any skill vector s, the wage earned by a worker of type $\tilde{s} = s/w(s)$ is equal to one, the integral $\int \lambda H^f(\mathrm{d}\lambda|\tilde{s})$ represents the total wage earned by workers with the same skill profile as $\tilde{s}$. More generally, for any distribution H on $\mathcal{X}^s$, we define the distribution $W_\# H$ as the push-forward of the positive measure $w(x)H(x)$ by the projection $s/w(s)$:[21]

$$W_\# H = \left(\frac{s}{w(s)}\right)_\# w(s) H. \tag{15}$$

The distribution $W_\# H$, which is supported on the iso-wage surface $w = 1$, places the mass $\int_0^\infty \lambda H(\mathrm{d}\lambda|\tilde{s})$ on any point $\tilde{s}$ with $w(\tilde{s}) = 1$. This mass, again, is nothing but the sum of the wages received by all the workers with skill profile $\tilde{s}$.

Proposition 5. *When the iso-wage schedule surfaces are strictly concave, all employees within the same firm share the same skill profile, i.e., the matching is pure in the horizontal dimension*

$$\textit{Support}\ \pi \subset \{(\tilde{S}^{\mathrm{D}}(\phi) \times \mathbb{R}_+ \,,\, \phi) \mid \phi \in \Phi\}. \tag{16}$$

In equilibrium, the aggregate skill vector $T^{\mathrm{D}}(\phi)$ and the wage schedule w satisfy

$$W_\# H^w = W_\# T^{\mathrm{D}}_\# H^f, \tag{17}$$

where W is given by (15).

When the iso-wage is strictly concave, any firm ϕ picks all its employees from the ray $\tilde{S}^{\mathrm{D}}(\phi) \times \mathbb{R}_+$ in $\mathcal{X}^s$. Intuitively, the equilibrium condition holds pointwise on the iso-wage surface, i.e., separately for each ray. The measure $T^{\mathrm{D}}_\# H^f$ represents the demand

[21] The push-forward operator is defined in Appendix A.5.

for skill vectors expressed by all firms in the economy.[22] The measure $W_\# T^{\mathrm{D}}_\# H^f$ reflects the wage bills paid by all these firms for each skill profile and can can be thought of as the demand for skill *profiles*. Similarly, $W_\# H^w$ represents the total wages earned by workers with each skill profile and can be thought of as the (wage-weighted) supply of skill profiles in the economy. The equilibrium condition (17) says that the demand and supply of skill profiles coincide. It translates into an ordinary differential equation for the matching map as we now illustrate in the case of two tasks.

Back to the two skills-tasks example: Assume that the production function is homogenous of degree $\eta < 1$ and F_N/F_C increases with α_N as in Proposition 4. As above, the firm-aggregated skill vector is represented as $T^{\mathrm{D}} = (T^{\mathrm{D}}_C, T^{\mathrm{D}}_N) = (\Lambda^{\mathrm{D}} \cos\theta^{\mathrm{D}}, \Lambda^{\mathrm{D}} \sin\theta^{\mathrm{D}})$, where Λ^{D} is the total quality of workers employed at firm ϕ. The workers-to-firms matching condition (11) can be written in this context

$$\frac{F_N\left(\cos\theta^{\mathrm{D}}(\alpha_N), \sin\theta^{\mathrm{D}}(\alpha_N); \alpha_N\right)}{F_C\left(\cos\theta^{\mathrm{D}}(\alpha_N), \sin\theta^{\mathrm{D}}(\alpha_N); \alpha_N\right)} = \frac{w_N\left(\theta^{\mathrm{D}}(\alpha_N)\right)}{w_C\left(\theta^{\mathrm{D}}(\alpha_N)\right)}, \tag{18}$$

which implicitly defines an increasing matching map $\theta^{\mathrm{D}}(\alpha_N)$. Setting $\tilde{w}(\theta) = w(\cos\theta, \sin\theta)$ as in (27), we can write the equilibrium condition for any α_N as

$$\int_0^{\theta^{\mathrm{D}}(\alpha_N)} \Lambda^{\mathrm{S}}(\theta) H^w(\mathrm{d}\theta) = \int_0^{\alpha_N} Z^f(\alpha)\Lambda^{\mathrm{D}}(\alpha_N; 1) H^f(\mathrm{d}\alpha), \tag{19}$$

where $\Lambda^{\mathrm{S}}(\theta) = \int_z \lambda H^w(\mathrm{d}\lambda|\theta)$ and $Z^f(\alpha) = \int_z z^{1/(1-\eta)} H^f(\mathrm{d}z|\alpha)$ are exogenous quantities that depend on the primitive distributions H^f and H^w, and the size $\Lambda^{\mathrm{D}}(\alpha_N; 1)$ is given by (A.12):

$$\Lambda^{\mathrm{D}}(\alpha_N; 1) = \left[\eta\, \frac{F(\cos\theta^{\mathrm{D}}(\alpha), \sin\theta^{\mathrm{D}}(\alpha); \alpha)}{\tilde{w}(\theta^{\mathrm{D}}(\alpha))}\right]^{1/(1-\eta)}.$$

The left-hand side of (19) represents the total quality of workers with skill profile below $\theta^{\mathrm{D}}(\alpha_N)$. The right-hand side represents the total quality of workers employed by firms with technological parameter below α_N.

Differentiating with respect to α_N yields the ordinary differential equation for the matching map $\theta^{\mathrm{D}}(\alpha_N)$:

$$\Lambda^{\mathrm{S}}(\theta^{\mathrm{D}}(\alpha_N))\, h^w(\theta^{\mathrm{D}}(\alpha_N))\, \frac{\mathrm{d}\theta^{\mathrm{D}}}{\mathrm{d}\alpha_N} = Z^f(\alpha_N) h^f(\alpha_N) \left[\eta\, \frac{F(\cos\theta^{\mathrm{D}}(\alpha_N), \sin\theta^{\mathrm{D}}(\alpha_N); \alpha_N)}{\tilde{w}(\theta^{\mathrm{D}}(\alpha_N))}\right]^{1/(1-\eta)}, \tag{20}$$

[22] $T^{\mathrm{D}}_\# H^f$ is the push-forward of the distribution of the firms' technological parameters H^f by their skill aggregate skill demand T^{D}, see Appendix A.5.

where h^f and h^w are the densities of the distributions of θ and α. Equation (20) relates the wage schedule and the matching map $\theta^{\mathrm{D}}(\alpha_N)$ implicitly given by (18) to the distributions of workers' skills and firms' technologies. The equilibrium conditions (15) and (20) are equivalent; they are expressed in terms of a one-dimensional outcome, respectively firms' sizes and wage bills.[23] In the next Subsection, we show that writing the equilibrium condition as a one-dimensional equality is no longer possible when wage iso-lines have linear parts and bunching prevails in equilibrium.

3.4 The Impact of Bunching

We now turn to situations in which different firm-types hire workers with similar skill-types (albeit never using the same combination because of the aggregate workers-to-firms matching condition). We refer to this phenomenon as bunching. First, we explain intuitively how bunching can arise in equilibrium, and how it is connected to the heterogeneity of skill profiles within firms. Next, we formally characterize equilibria with bunching.

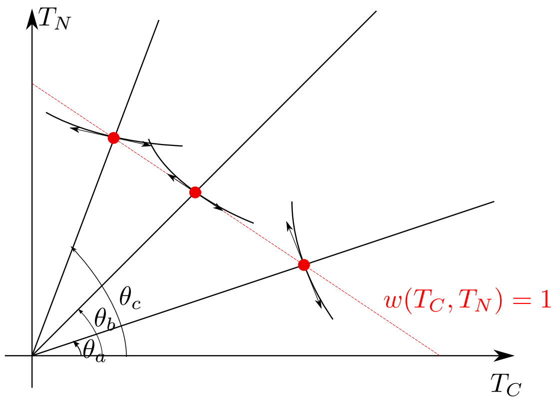
A simple economy with two tasks and three skill profiles: We start from an initial equilibrium without bunching for which the price schedule is linear, and from this equilibrium we change the distribution of skills in the economy. We first show that if we increase the relative number of "generalists" (workers with a balanced set of skills), their price falls and the wage schedule becomes nonlinear. We then show that if we decrease the relative number of generalists starting from this initial equilibrium, the wage schedule remains linear, the skill profiles of workers *within firms* become heterogeneous, in short, bunching emerges.

We illustrate the mechanism in a setting with two tasks and three skill profiles $\theta_a < \theta_b < \theta_c$, see Figure 2. Recall that $\tan\theta_i = s_{Ni}/s_{Ci}$ is the endowment of workers $i \in \{a, b, c\}$ in skill N relative to skill C, i.e., their comparative advantage in the non-cognitive skill. We pick any numbers $w_C > 0$ and $w_N > 0$ and construct distributions H^w and H^f for which the linear wage schedule $w(s_C, s_N) = w_C s_C + w_N s_N$ prevails in equilibrium. To achieve that, we choose three values for the technological intensities in skill N, $\alpha_N j$, $j \in \{a, b, c\}$, such that

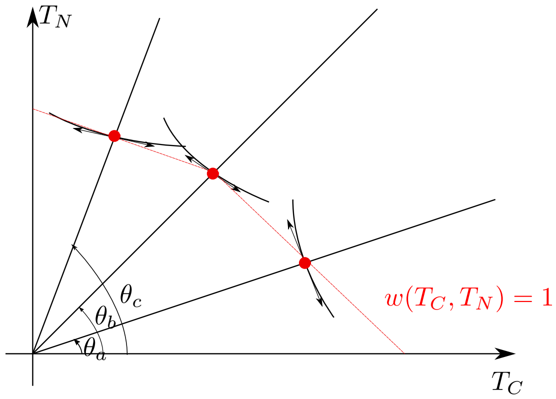
$$\frac{1-\alpha_{Nc}}{\alpha_{Nc}}\,(\tan\theta_c)^{1-\rho} < \frac{w_C}{w_N} = \frac{1-\alpha_{Nb}}{\alpha_{Nb}}\,(\tan\theta_b)^{1-\rho} < \frac{1-\alpha_{Na}}{\alpha_{Na}}\,(\tan\theta_a)^{1-\rho}.$$

Firms with technical intensity α_{Nj} hire workers with profile θ_j. Firms α_a would prefer workers endowed with more skill C relative to skill N, but no such workers are

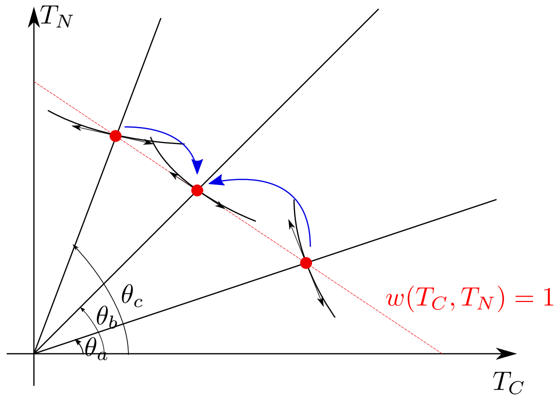
[23] Multiplying both sides of (20) by $\tilde{w}(\theta)$ yields in (15).



(a) Linear wage schedule



(b) More generalists make the schedule nonlinear



(c) Less generalists and more specialists create bunching

Figure 2: Equilibrium with three relative skill endowments in the economy

available in the economy. In this discrete setting, the equilibrium is achieved separately on each ray, i.e. for θ_a, θ_b and θ_c separately. Equation (20) takes the form

$$\Lambda^{\mathrm{S}}(\theta_i)\, h^w(\theta_i) = Z^f(\alpha_i) h^f(\alpha_i) \left[\frac{F\left(\cos\theta_i, \sin\theta_i; \alpha_i, 1\right)}{\tilde{w}(\theta_i)} \right]^{1/(1-\eta)}.$$

We choose $\Lambda^{\mathrm{S}}(\theta_i)\, h^w(\theta_i)$ and $Z^f(\alpha_i)h^f(\alpha_i)$ so that the above equation holds for all $i \in \{a, b, c\}$, i.e. so that Figure 2(a) represents the equilibrium configuration.

We now slightly increase the (quality-adjusted) number of generalist workers in the economy, $\Lambda^{\mathrm{S}}(\theta_b)\, h^w(\theta_b)$. To equalize the demand and the supply of generalists, we need to reduce their wage. The equilibrium configuration is modified as shown on Figure 2(b). The wages of the two specialist types a and c remain unchanged, as well as the behavior of α_{Na}-firms and α_{Nc}-firms. The wage schedule, however, has become nonlinear.

To generate bunching, we on the contrary decrease the number of generalist workers relative to the equilibrium of Figure 2(a). Specifically, we reduce $\Lambda^{\mathrm{S}}(\theta_b)\, h^w(\theta_b)$ by $\nu_b > 0$

and we define $\nu_a > 0$ and $\nu_c > 0$ by

$$\nu_b(\cos\theta_b, \sin\theta_b) = \nu_a(\cos\theta_a, \sin\theta_a) + \nu_c(\cos\theta_c, \sin\theta_c).$$

We raise the number of specialist workers $\Lambda^{\mathrm{S}}(\theta_a)\, h^w(\theta_a)$ and $\Lambda^{\mathrm{S}}(\theta_c)\, h^w(\theta_c)$ by ν_a and ν_c respectively. Figure 2(c) shows the new equilibrium configuration. Firms with technical intensities α_{Na} and α_{Nc} do not change their behavior. Firms with intensity α_{Nb} keep the same aggregate skill $T^{\mathrm{D}}(\phi)$ but obtain such an aggregate skill using a different composition of their workforce. They hire all workers with relative skill endowment θ_b, but also some workers of type θ_a and θ_c workers, specifically ν_a and ν_c efficiency units, respectively. Hence in equilibrium α_{Na}-firms and α_{Nb}-firms both hire some θ_a workers, while α_{Nb}-firms and α_{Nc}-firms both hire some θ_c workers. In the extreme case where $\nu_b = \Lambda^{\mathrm{S}}(\theta_b)\, h^w(\theta_b)$, there are no more θ_b workers in the economy, and firms with intensity α_{Nb} achieve their optimal aggregate skill θ_b by mixing θ_a and θ_c workers.

Remark: Our previous example should have made clear how we use the term *bunching*. Because there is always perfect separation in terms of the firm's aggregate skill mix – θ always increases with α – there is no bunching of the sort studied in goods consumption since there is full sorting. On the other hand, there is bunching in the sense that firms with different skills intensities, different α's, may hire workers of the same type to construct their optimal mix of skills, α.

Characterization of equilibrium under bunching: To characterize the bunching equilibria, we introduce the set of workers paid less than one dollar;

$$\mathcal{W} = \{s \,|\, w(s) \leq 1\}.$$

When the wage schedule is strictly concave as was assumed in Subsection 3.3, all the points of the iso-wage surfaces are extremal points of $\mathcal{W}$. Extremal points are degenerate faces of $\mathcal{W}$.[24] By contrast, when the schedule is locally linear, the set $\mathcal{W}$ has proper faces, i.e., faces that are neither a singleton nor the whole set $\mathcal{W}$ itself. For instance, on Figure 3, the segment $[AB]$ is a proper face of $\mathcal{W}$, while A is an extremal point.

Rockafellar (1970), Theorem 18.2., states that any convex set is the *disjoint* union of the relative interiors of all its faces. For any T, let $\mathcal{F}(T)$ be the (unique) face of $\mathcal{W}$ such that $T/w(T)$ belongs to the relative interior of $\mathcal{F}(T)$. The cone

$$\mathcal{C}\left(T^{\mathrm{D}}(\phi)\right) = \mathcal{F}\left(T^{\mathrm{D}}(\phi)\right) \times \mathbb{R}_+ \tag{21}$$

[24] A face $\mathcal{F}$ of a convex set $\mathcal{W}$ is a convex subset $\mathcal{F} \subset \mathcal{W}$ such that $W \setminus \mathcal{F}$ is convex.

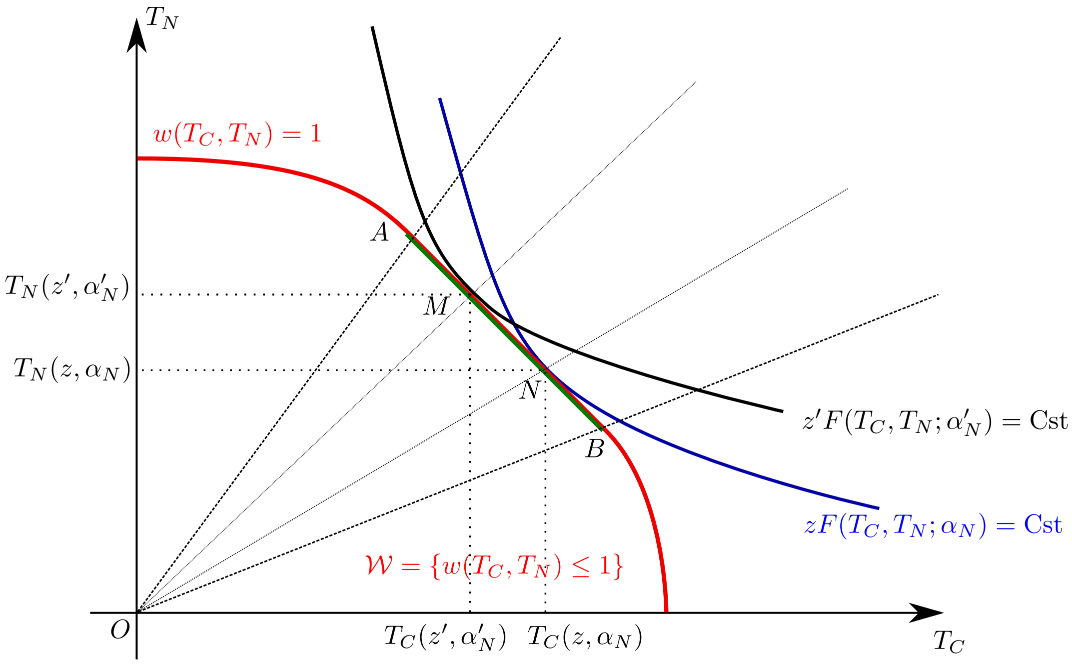


Figure 3: Matching is not pure. Firms $\phi = (\alpha_N, z)$ and $\phi' = (\alpha'_N, z')$, pick their employees in the cone generated by the face $[AB]$ of $\mathcal{W}$ in $\mathbb{R}^2_+$. Firm ϕ' is more intensive in skill N: $\alpha'_N > \alpha_N$ and $\theta^{\mathrm{D}}(\alpha'_N) > \theta^{\mathrm{D}}(\alpha_N)$.

is the largest set $\mathcal{C}$ in $\mathcal{X}$ such that (i) w is linear on $\mathcal{C}$; and (ii) the relative interior of $\mathcal{C}$ contains $\tilde{S}^{\mathrm{D}}(\phi)$, the average skill profile of workers employed by firm with type ϕ.

Figure 3 illustrates a case where w is linear on the non-degenerate cone lying between the rays (OA) and (OB). If $T/w(T)$ is an extremal point of $\mathcal{W}$ (such as point A on the figure), then $\mathcal{F}(X)$ is the singleton $\{T/w(T)\}$ and the cone is reduced to a ray (the ray containing A in the example). By contrast, for the firms of type ϕ and ϕ', the sets $\mathcal{F}\left(T^{\mathrm{D}}(\phi)\right)$ and $\mathcal{F}\left(T(\phi')\right)$ are the whole segment $[AB]$, with the cone $\mathcal{C}\left(T^{\mathrm{D}}(\phi)\right) = \mathcal{C}\left(T(\phi')\right)$ being (AOB).

When the wage schedule w is locally linear, the minimization of the wage bill, problem (14), is compatible with a firm hiring employees with different skill profiles. To minimize the firm's wage bill, the support of the assignment measure $n^{\mathrm{D}}(\mathrm{d}s;\phi)$ must be included in $\mathcal{C}\left(T^{\mathrm{D}}(\phi)\right)$. Because the wage schedule w is linear on that cone, we have

$$\int w(s)\; n^{\mathrm{D}}(\mathrm{d}s;\phi) = w\left(\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)\right) = w(T^{\mathrm{D}}(\phi)).$$

For instance, firms with type ϕ on Figure 3, rather than picking employees with skills proportional to $\tilde{S}^{\mathrm{D}}(\phi)$, i.e., along the half-line $[OM)$, can use skills located in the entire cone AOB.

Proposition 6. *When the equilibrium wage schedule is locally linear, the matching is not pure in the horizontal dimension*

$$\textit{Support}\ \pi \subset \{\, \mathcal{C}\left(T^{\mathrm{D}}(\phi)\right)\ ,\ \phi) \mid \phi \in \Phi \,\}, \tag{22}$$

where $\mathcal{C}\left(T^{\mathrm{D}}(\phi)\right)$ *is the cone given by* (21). *In equilibrium, the measure* $W_{\#}T_{\#}H^f$ *is dominated by* $W_{\#}H^w$ *in the convex order:*

$$W_{\#}H^w \succeq_C W_{\#}T^{\mathrm{D}}_{\#}H^f \tag{23}$$

where the operator W *is given by* (15).

When bunching prevails, it is no longer true that the total value of efficiency units of labor supplied by workers and demanded by firms coincide for each skill profile, i.e., that the distributions $W_{\#}T^{\mathrm{D}}_{\#}H^f$ and $W_{\#}H^w$ are equal. Recall that a measure μ_1 is dominated by a measure μ_2 in the convex order if and only if $\mu_2 h \geq \mu_1 h$ for all convex functions h.[25] The condition (23), which is weaker than (17), expresses that there is a local excess supply of specialist workers and an excess demand for generalist ones. In terms of efficiency units of labor (valued by wage), the distribution of workers' skills H^w lies closer to the boundary of the cone than the demand distribution $T^{\mathrm{D}}_{\#}H^f$. For instance, on Figures 3 and 4, the supply of skills is more concentrated along the rays OA and OB, while the demand is more concentrated in the interior of the cone.

Bunching in the horizontal dimension leads to many-to-many matching as illustrated on Figure 4. Firms with different types hire workers with the same skill profile, and workers with the same type may be employed by firms with different technologies. For instance, firms F and F' on the figure, which have different technological intensities in skill α, both hire workers with skills in the cone (AOB). In the extreme case where workers' skill are located only along the two rays (OA) and (OB), firms F and F' both hire workers with skill profiles A and B, but in different proportions to achieve their aggregate demand.[26]

To conclude this section, we connect our primal problem (5) to the classic optimal transport (OT) framework, used for instance in Lindenlaub (2017)'s study of worker-to-job matching. To fully understand how OT is connected to our contribution, a small detour is required. Our approach requires to account for workers' skill aggregation within firms and for endogenous firm size. These requirements demand a new mathematical framework, developed in Choné, Gozlan, and Kramarz (2023). In particular,

[25] It means that μ_2 is "riskier" than μ_1.

[26] In the absence of bunching, when the equilibrium wage schedule is strictly convex, cones are degenerate, i.e., coincide with rays.

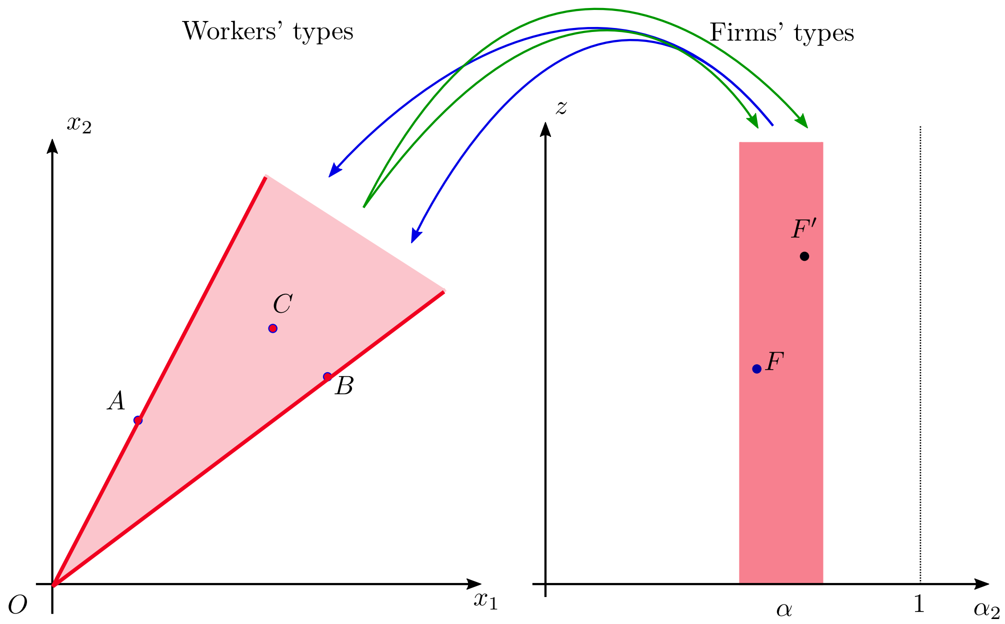


Figure 4: Sorting with bunching: Within-firm heterogeneity in skill profiles

it allows us to define precisely when one distribution is more "generalist" than another. Intuitively, in a two-skill world, it means that there are more generalists than specialists. Indeed, and back to our problem, we show in Appendix A.7 that the distribution of firm-aggregated skill vectors, $T^{\mathrm{D}}_{\#}H^f$, is more "generalist" than the original distribution of workers' skills in the economy, H^w, in the sense that $\int h(s)T^{\mathrm{D}}_{\#}H^f(\mathrm{d}s) \leq \int h(s)H^w(\mathrm{d}s)$ for all positively 1-homogenous convex functions h. When this property holds, Choné, Gozlan, and Kramarz (2023) say that $T^{\mathrm{D}}_{\#}H^f$ is dominated by H^w in the positively 1-homogenous convex order, something we denote by $T^{\mathrm{D}}_{\#}H^f \leq_{\mathrm{phc}} H^w$.

Proposition 7. *For any given map $T : \Phi \to \mathbb{R}^n_+$, the two properties are equivalent:*

1. *There exists a market-clearing assignment n^{D} such that $T(\phi)$ is the firm-aggregated skill vector $T(\phi) = T^{\mathrm{D}}(\phi) = \int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)$;*

2. *The probability distributions $T_{\#}H^f$ and H^w satisfy: $T_{\#}H^f \leq_{\mathrm{phc}} H^w$.*

Furthermore, if n^{D} is an optimal market-clearing assignment, $T_{\#}H^f$ is solution to

$$\mathcal{J}^b(H^f, H^w) = \int F(T^{\mathrm{D}}(\phi);\phi)H^f(\mathrm{d}\phi) = \max_{\gamma \leq_{\mathrm{phc}} H^w} \max_{\pi \in \mathcal{P}(\gamma, H^f)} \int F(s;\phi)\,\pi(\mathrm{d}s, \mathrm{d}\phi), \quad (24)$$

where $\mathcal{P}(\gamma, H^f)$ denotes the set of all couplings between γ and H^f.

The first part of Proposition 7 states that the ordering $T_\# H^f \leq_{\text{phc}} H^w$ is not only necessary but also sufficient for T being generated by a skill-aggregation process. The second part, namely equation (24), expresses that the optimal output under bundling, see (5), is the maximal output that can obtained *without* skill-aggregation *i.e., with classic OT*, among all skill distributions that are "more generalist" than the original distribution H^w.

Hence, when there are enough generalist workers in the economy, there is no bunching: $\gamma = T^{\text{D}}_\# H^f =_{\text{phc}} H^w$ as in Proposition 5.[27] If, on the contrary, $T^{\text{D}}_\# H^f$ is strictly dominated by H^w in the convex positively homogenous order – for instance if there are mostly specialist workers in the economy – then $T^{\text{D}}(\phi)$ is obtained by using workers with different skill profiles as in Proposition 6. In the latter case, there is within-firm heterogeneity in *skill profiles*. As a consequence, in equilibrium, complementarities across workers within the firm materialize. Hence, the productivity of workers endowed with (mostly) one skill and deprived of the other skills is enhanced by the presence of co-workers endowed with the other, complementary, skills.

An immediate consequence of (24) is that for any pair of skill distributions (H^w_1, H^w_2) such that $H^w_1 \leq_{\text{phc}} H^w_2$, i.e., such that H^w_2 is "more specialist" than H^w_1, total output is greater for H^w_2 than for H^w_1.[28] The unbundling process studied in our companion paper changes the initial skill distribution into a new, unbundled distribution of skills that is "more specialist" and hence increases total welfare. This process thus gets the economy closer to full efficiency, which is achieved if unbundling entails no costs for firms or workers, see Choné, Gozlan, and Kramarz (2024).

3.5 Numerical illustration

To illustrate the above propositions (with or without bunching), we now examine how the supply of skills in the economy affects equilibrium outcomes. Using the mirror descent algorithm presented in Paty, Choné, and Kramarz (2022), we contrast two polar cases in our two-skill environment. In Scenario (A) the skill supply comprises essentially generalists, whereas in Scenario (B) it comprises essentially specialists, i.e., workers are mostly endowed with either cognitive skills or non-cognitive skills. In both scenarios, the production technology is CES of the form (12), with the technical intensities α_C and α_N being uniformly distributed on $[0, 1]$. All firms have the same total factor productivity $z = 1$ and all workers have the same quality $\lambda = 1$. In Scenario (A) skill profiles $\theta = \arctan s_N/s_C$ are distributed as a Beta(6,6) random variable, whereas in Scenario (B) they are distributed as a Beta(.8,.8) variable.

[27] The projections of the distributions $T^{\text{D}}_\# H^f$ and H^w onto the iso-wage surface coincide.
[28] This is because the set $\{\gamma : \gamma \leq_{\text{phc}} H^w_2\}$ is larger than the set $\{\gamma : \gamma \leq_{\text{phc}} H^w_1\}$.

In Figure 5, we present for these two scenarios: (a) the matching patterns $\theta^{\mathrm{D}}(\alpha_N)$; (b) the wages $\tilde{w}(\theta) = w(\cos\theta, \sin\theta)$; (c) the implicit prices (marginal product of each skill) $w_C(\theta^{\mathrm{D}}(\alpha))$ and $w_N(\theta^{\mathrm{D}}(\alpha))$ as a function of α; (d) the wage isolines $w(s_C, s_N) = 1$, i.e., the quantity of each skill available for one dollar; (e) the firms' sizes $\Lambda^d(\alpha; 1)$. As explained at length, Scenario (A) with many generalists is one that leads to a nonlinear wage schedule, i.e., to the absence of bunching, see Figure 5(d). In this scenario, implicit prices differ across firms, see Figure 5(c), and specialist firms (with either low or large technical parameters αs) are forced to use workers that possess too much of the other specialist skill, see Figure 5(a). Since specialist workers are in "short" supply, their wage is high relative to generalists, who thus face a markdown, see Figure 5(b). Specialist firms cannot attain a large size because their favored employees are expensive.

In stark contrast, bunching is pervasive under Scenario (B). The wage schedule is linear, see the flat isolines on Figure 5(d), with the implicit price of each skill being constant across firms and skill profiles. As can be seen on Figure 5(b), the large supply of specialists decreases their wage, thus allowing specialist firms to increase their size, see Figure 5(e). Finally, the matching map that connects firms (α) and workers (θ) lies closer to the 45-degree line than under Scenario (A). Because more specialist workers are available in Scenario (B), specialist firms have access to skills that are more aligned with their technologies.

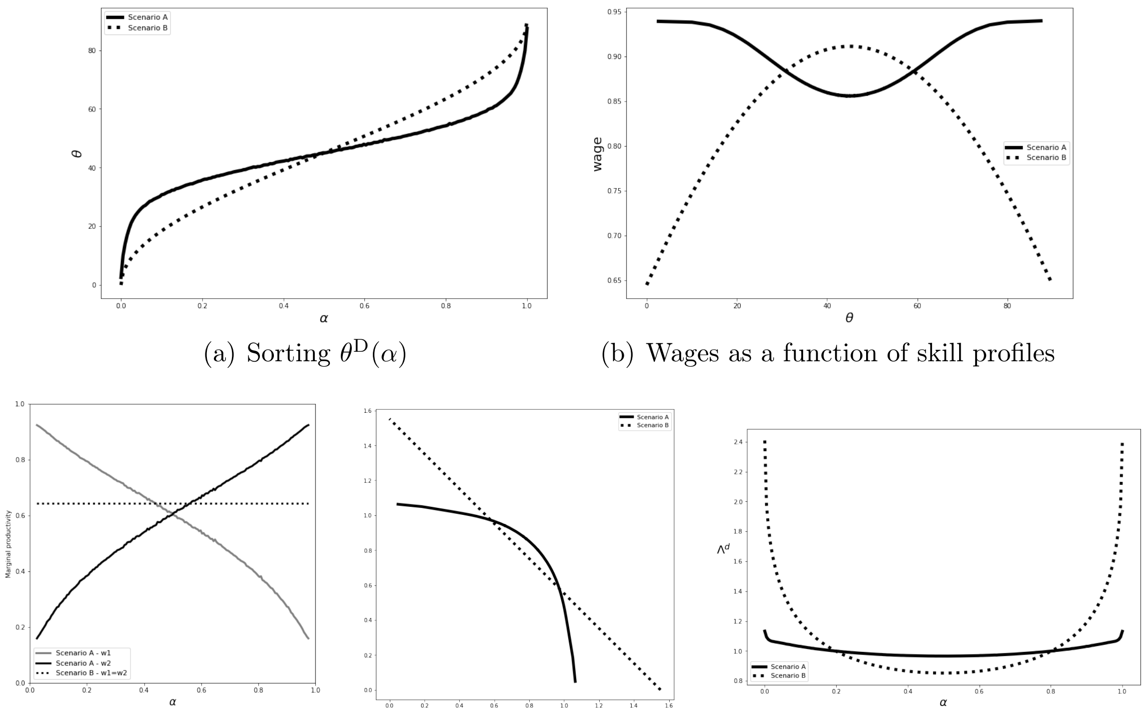
3.6 Connections to the Roy-Model

We now connect our model to Roy (1951). In doing so, we describe in more detail the structure of the convex and homogenous wage schedules. First, we need to define the *implicit price* of skill i for workers of type s as $w_i(s) = \partial w/\partial s_i$. These implicit prices are homogenous of degree zero, and as such depend on skill profiles $\tilde{s} = s/|s|$ but not on workers' qualities $|s|$. Using Euler's homogenous function theorem and the convexity of wages, we get, for all skill vectors s and s'

$$w(s) = \sum_{i=1}^{k} w_i(s)s_i \geq w(s') + \sum_{i=1}^{k} w_i(s')(s_i - s'_i) = \sum_{i=1}^{k} w_i(s'_i)s_i. \tag{25}$$

Using the above property, we show in Appendix A.3 that if a worker with skill vector s is employed by a firm of type ϕ, his wage is $w(s) = s \cdot \nabla w(T^{\mathrm{D}}(\phi))$ and

$$w(s) = \sup_{\phi'} s \cdot \nabla w(T^{\mathrm{D}}(\phi')). \tag{26}$$



(a) Sorting $\theta^{\mathrm{D}}(\alpha)$ (b) Wages as a function of skill profiles

(c) Implicit prices (d) Wage isolines (e) Firms' sizes $\Lambda^{\mathrm{D}}(\alpha, 1)$

Figure 5: The Effect of Skills Supply: Mostly Generalists (A) vs Mostly Specialists (B)

In a Roy-like assignment model, workers would decide to self-select into their preferred option among the menu of linear wage schedules $\sum_{i=1}^{k} w_i(s'_i)s_i$ indexed by s' or into their preferred firm. In such a context, Equations (25) would be thought of as an incentive constraint expressing that a worker with skills s prefers the linear schedule "designed for her", i.e., chooses $s' = s$.[29] Similarly, Equation (26) would represent the choice of firms by workers. By contrast, the present paper's modeling framework involves no supply-side decisions on workers' side; the above two equations are purely demand-driven: it results from the structure of our production function, in particular from the aggregation of skills within firms.

We shall repeatedly illustrate our results in the case of two skills, $k = 2$. To be consistent with the empirical analysis presented in Subsection 4.2, we shall refer to them as cognitive and non-cognitive skills, C and N. The worker's skill profiles $\tilde{s} = (s_C/|s|, s_N/|s|)$ can be parameterized as $\tilde{s} = (\cos\theta, \sin\theta)$, where θ belongs to $[0, \pi/2]$. For brevity, we often refer to θ as the worker's skill profile. The worker's comparative advantage in skill N over skill C is simply $s_N/s_C = \tan\theta$. The implicit prices of the two skills, $w_C(\theta)$ and $w_N(\theta)$, depend only on the profile θ. Equation (25)

[29] The empirical results of Section 4 illustrate that a worker is paid less if she deviates from $s' = s$ and in this sense is not well "matched".

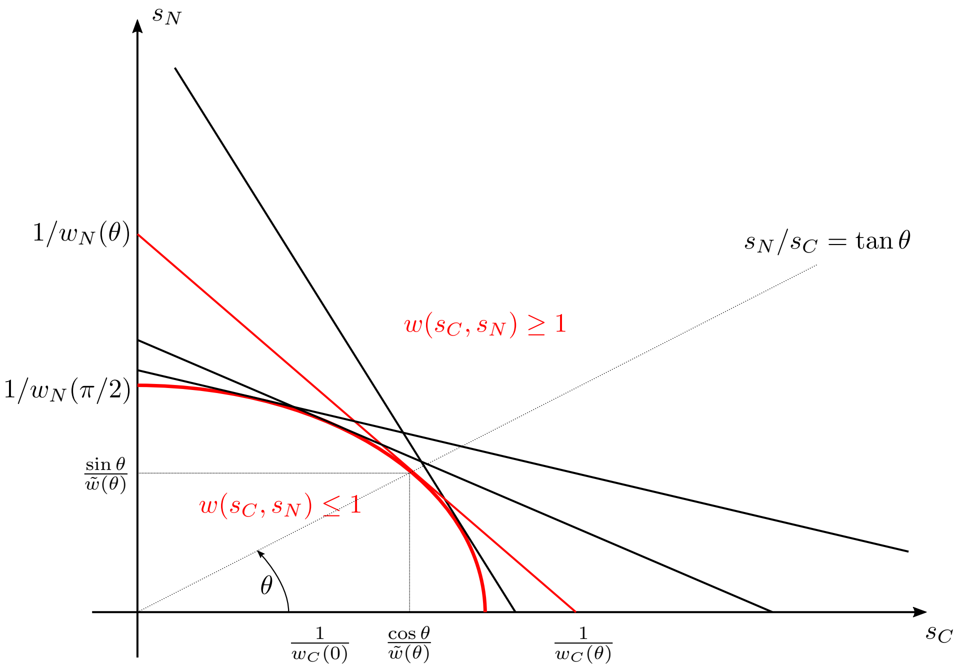


Figure 6: The set of workers paid less than one dollar is convex. The implicit prices of skills C and N for workers with skill profile θ are $w_C(\theta)$ and $w_N(\theta)$

can be rewritten here as:

$$\tilde{w}(\theta) \stackrel{\mathrm{d}}{=} w(\cos\theta, \sin\theta) = \max_{\theta'} w_C(\theta') \cos\theta + w_N(\theta') \sin\theta, \tag{27}$$

with the maximum being achieved for $\theta' = \theta$. Geometrically, convex and homogenous wage schedules are entirely determined by the associated iso-wage surface $w(s) = 1$, i.e., the sets of skill types that firms can obtain in return for one dollar. Figure 6 shows that the iso-wage surfaces are the envelopes of their tangents, a familiar property in the nonlinear pricing literature.[30]

The classical Roy literature (see for instance the presentation of Heckman and Honore, 1990) assumes two skills and two sectors, with each sector using only one skill and each skill being priced separately. Workers choosing one sector are paid for the skill valued in that sector, their other skill being left unused. The solid line in Figure 7 shows the wage isoline $w = 1$ for the Roy wage schedule $w(s_C, s_N) = \max(p_C s_C, p_N s_N)$, where p_C and p_N are the prices in the two sectors. This correspond to the special case where all firms use only one of the two skills; in our baseline example, all firms have $(\alpha_C, \alpha_N = 1)$ equal to either $(1, 0)$ or $(0, 1)$. Heckman and Scheinkman (1987), Edmond and Mongey (2022), assumes special forms for the family of linear tariffs. For instance, in the case

[30] See Wilson (1993) and Laffont and Martimort (2009). In the case depicted on Figure 6, the iso-wage curve $w = 1$ can be parameterized as $(s_C(\theta), s_N(\theta))$, with $s_C(\theta) = \cos\theta/\tilde{w}(\theta)$ and $s_N(\theta) = \sin\theta/\tilde{w}(\theta)$. It is the envelope of the family of straight lines $w_C(\theta')s_C + w_N(\theta')s_N = 1$ indexed by θ'.

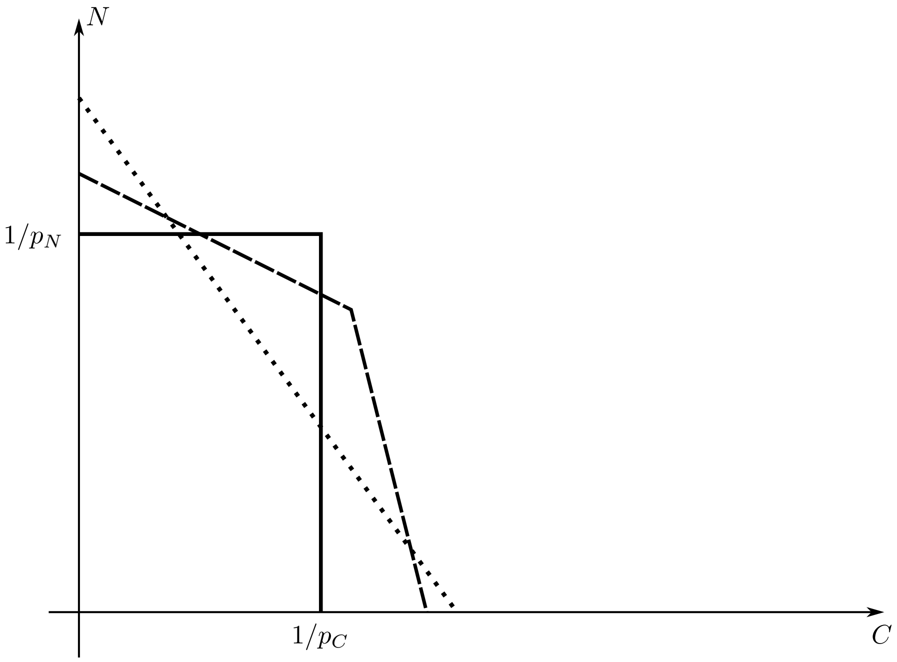


Figure 7: Isowage lines: Roy (1951): $w(s_C, s_N) = \max(p_C s_C, p_N s_N)$ (solid line); Edmond and Mongey (2022) or Heckman and Scheinkman (1987) (dashed); Katz and Murphy (1992) (dotted line)

of two skills, both of these papers assume two sectors with homogenous firms within each sector and a sector-specific wage schedule, in other words they restrict attention to two-part wage schedules. Katz and Murphy (1992) assume the law of one-price, i.e., one price per each separate skill, which corresponds to linear wage isolines.[31]

4 From the Model to the Data

In this Section, we discuss the main empirical predictions of our theory. We also mention the type of data needed to test such predictions and how data relate to potential identification. The Swedish data on workers' skills and their employing firms used in Fredriksson, Hensvik, and Skans (2018) as well as in our paper Skans, Choné, and Kramarz (2022) (SCK, hereafter) is one such data source. We briefly summarize both papers' main results. Finally, we discuss the connections between skills and tasks, even though *no data source* containing the nature of tasks, measured at the worker-job level over time and across employers, appears to exist. Indeed, occupations are often used as a proxy even though the tasks performed by the worker in her employing firm are

[31]Linearity also obtains under full unbundling. Considering a gradual process of skill unbundling, Choné, Gozlan, and Kramarz (2024) explain how the wage schedule changes as market for stand-alone skills open.

virtually never measured. However, Bittarello, Kramarz, and Maitre (forthcoming) show the extent of dispersion in tasks within occupations.

4.1 Our Model's Main Empirical Consequences

We summarize the main empirical consequences derived from our theory. All such empirical consequences should be understood as applying occupation by occupation (nurses, computer scientists, etc.) with potentially diverse skills, employed in a restricted set of firms, with a demand for skills and for the ensuing tasks that may vary from firm to firm. For ease of exposition, we assume hereafter that $k = 2$, skill s_C comprises all *Cognitive* skills and skill s_N comprises all *Non-Cognitive* skills, as is measured in the Swedish data source used in Skans, Choné, and Kramarz (2022) that we present later in this Section. In what follows, we define the skill profile θ of a worker with skill vector (s_C, s_N) by $\tan\theta = s_N/s_C$.

Firm-aggregated Workers' Qualities and Profiles: Proposition 3 proves the uniqueness of the firm-aggregated skill vector $T^{\mathrm{D}}(\phi) = \int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)$. Furthermore, by writing this skill vector as $T^{\mathrm{D}}(\phi) = \Lambda^{\mathrm{D}}(\phi)\tilde{S}^{\mathrm{D}}(\phi)$, where $\Lambda^{\mathrm{D}}(\phi) = |T^{\mathrm{D}}(\phi)|$ is the total quality of the firm's employees and $\tilde{S}^{\mathrm{D}}(\phi)$ is their average skill profile, we have shown that $\Lambda^{\mathrm{D}}(\phi)$ increases with total factor productivity z (see Appendix A.4 for detail).

Given the linear structure of our aggregation from skills to tasks, high-z firms, which are also high-Λ^d, can achieve their high total quality using either a large number of employees of average quality or a smaller number of high quality workers.

Furthermore, and assuming that the marginal rate of technical substitution F_C/F_N, evaluated at the firm-agregated skill vector $(\Lambda^{\mathrm{D}}\cos\theta, \Lambda^{\mathrm{D}}\sin\theta)$, increases with Λ^{D}, the equality $F_C/F_N = w_C/w_N$ implies that θ, our measure of sorting, decreases with z (see Appendix A.4 for detail).

All these firm-level variables have direct counterparts in the Swedish data sources using workers' skills measures mentioned just above as well as proxies for z, also available in the Swedish data.

Wages under Bundling: From Proposition 2 we know that the wage schedule is homogenous of degree one. Hence, he log-wage of a worker with skills (s_C, s_N) writes as

$$\ln w(s_C, s_N) = \ln\lambda + \ln\tilde{w}(\theta), \tag{28}$$

where $\lambda = |(s_C, s_N)|$ and θ are, respectively, worker's quality (the norm of the skill vector) and skill profile. In the absence of bunching, there is pure sorting in the horizontal dimension, recall Section 3.3, meaning that θ depends only on the technology (α_N, z)

of the worker's employing firm. This property is reminiscent of the additive decomposition of the log-wage into a person and a firm effect contained in Abowd, Kramarz, and Margolis (1999) (AKM, hereafter). We discuss econometric identification of this equation below, comparing our context with that analyzed in AKM. We also discuss there how the properties of the production function impact our results. In particular, we contrast below the case with homothetic isoquants (with an implied "worker-to-firm sorting effect" in our AKM-like decomposition independent of z, the firm's total factor productivity) with the non-homothetic case (with an implied "worker-to-firm sorting effect" depending on z showing in this setting why using "firm effect" to characterize this component of pay has much more than a grain of truth).

We now examine various features (sorting, prices, wages, ...) of the equilibrium when the production function is non-homothetic.

Some Implications of Non-Homotheticity: We consider now the following (non-homogenous, non-homothetic) production function:

$$F(T;\alpha,z)=\frac{z}{\eta}\left([\alpha_C(T_C+A)^\rho+\alpha_N(T_N+B)^\rho]^{\eta/\rho}-(\alpha_C A^\rho+\alpha_N B^\rho)^{\eta/\rho}\right), \tag{29}$$

where the technical intensities in the cognitive and non-cognitive skills, α_C and α_N, satisfy $\alpha_C+\alpha_N=1$. Setting $(T_C^{\mathrm{D}},T_N^{\mathrm{D}})=(\Lambda^{\mathrm{D}}\cos\theta^{\mathrm{D}},\Lambda^{\mathrm{D}}\sin\theta^{\mathrm{D}})$ and using the first-order conditions (10) as well as the 1-homogeneity of the wage, we get

$$\frac{F_C(\Lambda^{\mathrm{D}},\theta^{\mathrm{D}};\alpha_N)}{F_N(\Lambda^{\mathrm{D}},\theta^{\mathrm{D}};\alpha_N)}=\frac{\alpha_C}{\alpha_N}\left[\frac{\Lambda^{\mathrm{D}}\sin\theta^{\mathrm{D}}+B}{\Lambda^{\mathrm{D}}\cos\theta^{\mathrm{D}}+A}\right]^{1-\rho}=\frac{w_C(\theta^{\mathrm{D}})}{w_N(\theta^{\mathrm{D}})}. \tag{30}$$

The marginal product of T_C versus T_N, F_C/F_N, increases with Λ for skill profiles $\theta_{N/C}$ such that $\tan\theta_{N/C}>B/A$. In this case, the marginal productivity of *Cognitive* skills relative to that of *Non-Cognitive* skills increases with the size of firms, hence big firms use relatively more *Cognitive* skills, implying that θ decreases with z. The "worker-to-firm sorting effect" now becomes linked to the firm's productivity z and becomes closer to a true "firm-effect". Furthermore, under non-homotheticity, the "worker-to-firm sorting effect" (which captures the intensity of the relative use of the two skills) and the total quality of the firms' workers will be correlated. Indeed, in our model, the strength of the correlation between individual worker quality and the "worker-to-firm sorting effect" will vary: zero under homotheticity whereas, under non-homotheticity, this individual-level correlation will be positive and small when the productive firms employ many average workers but positive and large when the productive firms employ a small number of very high-quality workers.

Our central set of simulations (based on algorithms presented in Paty, Choné, and Kramarz (2022)) imposes $A = 1.0$ and $B = 0.3$ in the production function (affecting firms' demand). Firms are uniformly distributed with two productivity levels ($z = 1$ and $z = 2$) when the skills' supply follows a Beta(7,4) distribution with relatively few "Cognitive" workers. The distribution of firms and the workers' supply are presented in Figure 8. In this example, the equilibrium condition can be written as

$$\Lambda^{\mathrm{S}}(\theta)h^w(\theta) = \sum_{l=1}^{2} h^f(\alpha_N(\theta;z_l))\,\Lambda^{\mathrm{D}}(\alpha_N(\theta;z_l);z_l)\frac{\partial \alpha_N}{\partial \theta}(\theta;z_l), \tag{31}$$

where $z_1 = 1$ and $z_2 = 2$, $\Lambda^{\mathrm{S}}(\theta) = 1$, $h^w(\theta)$ is the density of the Beta(7,4) law, and $h^f(\alpha_N) = 1$. The sorting functions $\theta(\alpha_N;z_l)$ for $l = 1,2$ and the firms' demand for quality Λ^{D} jointly maximize $zF(\Lambda\theta;\alpha_N) - \Lambda\tilde{w}(\theta)$.[32]

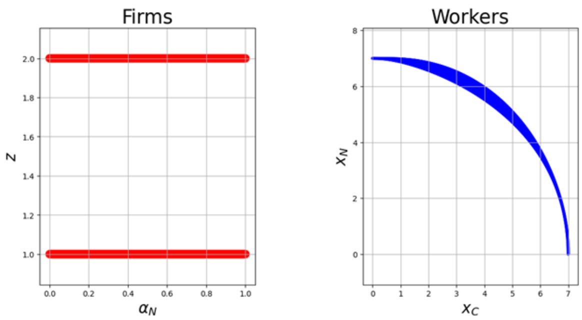


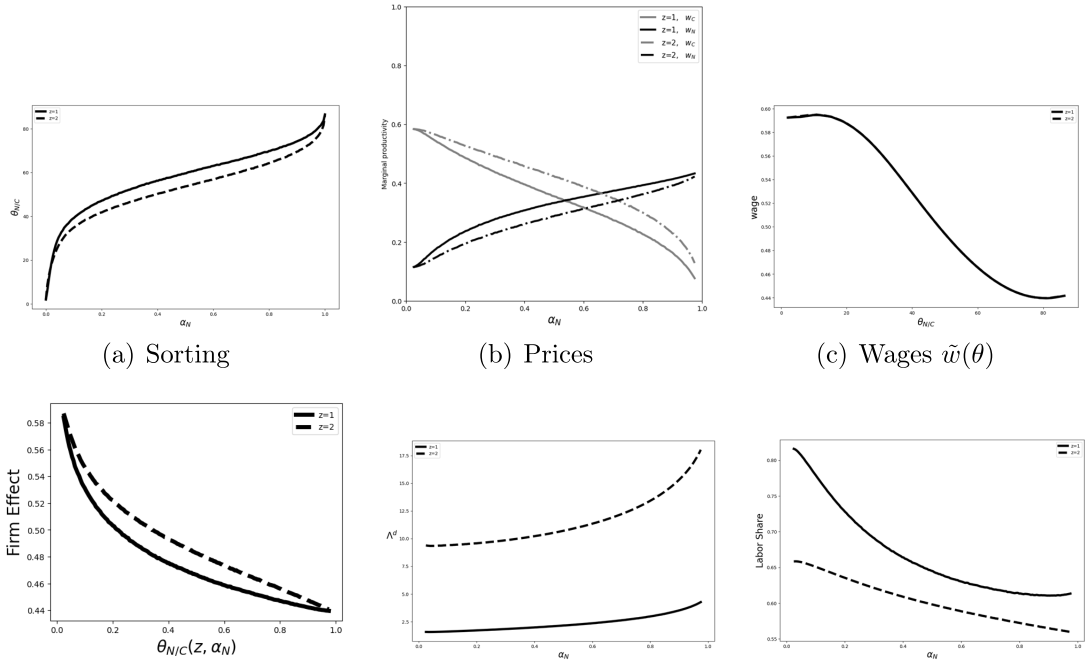
Figure 8: Firms (uniform with $z = 1, 2$); Skill Supply (Beta(7,4) i.e. "shortage" of skill C)

Figure 9 presents the resulting Sorting patterns ($\theta_{N/C}$ as a function of α_N), Prices (as a function of α_N), and Wages (as a function of $\theta_{N/C}$) in the top panel, when "worker-to-firm sorting effects" (as a function of $\theta_{N/C}(\alpha_N, z)$), Firm size, and Labor Shares[33] (both as functions of α_N) are presented in the bottom panel. The deficit in supply of cognitive workers tends to make the wage decreasing in the comparative advantage in N, i.e., in the skill profile $\theta_{N/C}$. As predicted by Proposition 4, firms' sizes Λ^{D} increase with their TFP parameter z. The relative productivity of the cognitive skill increases with the firm size Λ^{D} (and hence with the TFP z) when $\tan\theta_{N/C} > .3$ or $\theta_{N/C}$ above 17 degrees. In this region, more productive firms (dashed line) use relatively more cognitive (and high-wages) workers than less productive firms (solid line), see Figure 9(a). As a result, in this configuration, "worker-to-firm sorting effects" (W-to-F sorting effects, hereafter) become "firm-effects": they depend on z and are larger for more productive firms ($z = 2$), see Figure 9(d). The high wages commanded by

[32]The sorting maps $\theta(\alpha_N;z_l)$ for $l = 1,2$ are thus determined by (30). We denote their inverses by $\alpha_N(\theta;z_l)$ in (31).

[33]Labor shares are defined as the payroll-to-sales ratio, wL/F.

the C-specialists explain why *Cognitive* firms, *conditional* on z, are smaller than *Non-Cognitive* firms and also have a lower labor share (again, given z). Larger and more productive firms have a lower labor share.



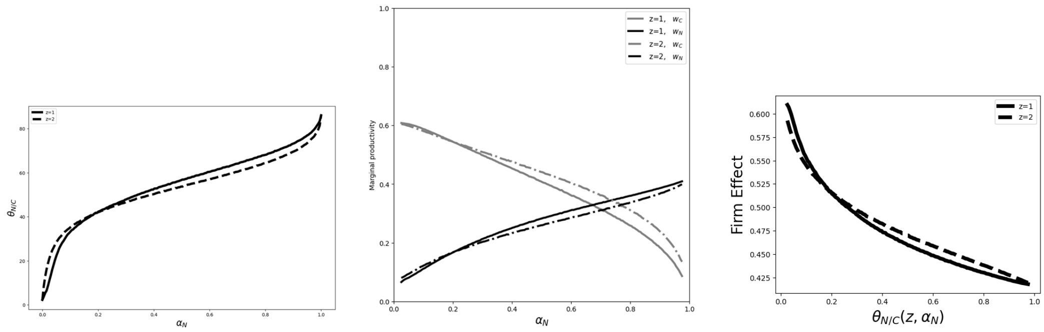
(a) Sorting (b) Prices (c) Wages $\tilde{w}(\theta)$

(d) W-to-F Sorting effects (e) Size (f) Labor shares

Figure 9: Wages and sorting with non-homothetic production function (29)

Figure 10 presents two simulations that highlight the role of asymmetries in the production function and in the supply of skills in the economy. The top panel shows Sorting, Wages, and W-to-F sorting effects for an almost symmetric production function ($A = 1.0, B = 0.9$) and for the same asymmetric supply as above (deficit of cognitive specialists). The bottom panel shows Wages, W-to-F sorting effects, and Labor shares for a symmetric supply of workers (namely skill profiles distributed as a Beta(6,6) random variable) and for the same production function as in Figure 9 ($A = 1.0, B = 0.3$).

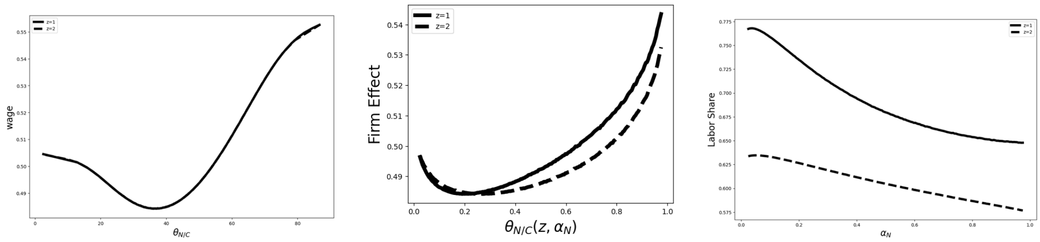
The asymmetry of the production function affects the sorting, the implicit prices and the W-to-F sorting effects. Indeed, when it is almost symmetric in the two skills ($A = 1.0, B = 0.9$), we observe that for all three variables the curves for low and high-productivity firms cross at the low end of the technical intensity in non-cognitive skills (α_N), with $z = 2$ firms having larger W-to-F sorting effects above the crossing. A visual inspection of Figures 9(a), 9(b), 9(d), 10(a), 10(b) and 10(c) shows that a larger value of B (from 0.3 to 0.9) reduces the difference between low-z and high-z firms. The reason is that when A and B are both large, their effect on the ratio of productivities



(a) Sorting; (A,B=1.0,0.9); Asy. Supply

(b) Prices; (A,B=1.0,0.9); Asy. Supply

(c) W-to-F sorting effects; (A,B=1.0,0.9); Asy. Supply



(d) Wage; Sym. Supply

(e) W-to-F Sorting effects; Sym. Supply

(f) Labor sh.; Sym. Supply

Figure 10: Non homothetic production function: Demand vs Supply

in (30) tends to dominate that of the skill aggregates T_C and T_N, and hence to weaken the role of the TFP parameter z.

The asymmetry of the skills supply, namely the relative scarcity of cognitive workers, also plays an important role in the results of Figure 9. It indeed causes the wage to decrease with the skill profile $\theta_{N|C}$, recall Figure 9(c). When on the contrary the distribution is symmetric in the two skills, wages are driven by the firms' demand and are locally increasing in $\theta_{N|C}$, see Figure 10(d). Hence, the W-to-F sorting effects are now larger for low-z firms. As before, because of the non-homogeneity of the production function, the labor share stays smaller for productive firms but, for a given z, larger for the non-cognitive firms which pay a high price for their favored specialists.

Comparing Figures 9(d), 10(c), and 10(e), we see that the variations of the Firm effects in the TFP parameter z result from complex general equilibrium effects.

Firm-Level Workers' Qualities and Profiles with Bunching: The above analysis has assumed bunching away. Under certain skills-supply and skills-demand conditions, all employees in a given firm share the same skill profile and differ only in their individual quality. When the supply of specialists w.r.t. generalists within a region of the skills space increases, however, (local) bunching emerge at equilibrium. A firm, to achieve its optimal mix of skill types, must hire workers situated between the two edges of the face that includes this optimal mix. As a result, within-firm heterogeneity in workers' profiles and qualities will increase (something that can be directly measured with the data at hand).

To assess the potential extent of bunching and the associated within-firm heterogeneity in workers' profiles, a first, fully reduced-form, approach is to examine the extent to which the sorting of workers within employing firms is mostly driven by such skill profiles. The leave-out regression analysis mentioned in the following subsection (Subsection 4.2) with full results in SCK) provides a first answer. A second empirical consequence of local bunching and the ensuing within-firm heterogeneity in skill profiles pertains to wages; the wage is linear in skills in zones (faces) where bunching takes place. The firm's optimal mix is comprised between the two extremal points of the cone. Assuming that the face is "small" enough, then the difference between worker's individual (log-) wage and her (log-) quality should be close to the (log-) firm-effect as measured at the optimal mix. However, when the (linear) face of the equilibrium wage schedule is large enough, the AKM-like property is likely to be lost. We defer the study of bunching in wages to a follow-up article, based upon structural modeling of our problem.

Identification of the wage schedule and production functions' parameters: As we have repeatedly mentioned, wages in our framework have an AKM-like property. But, as we will explain in this paragraph, this property has different foundations than those found in AKM, and, hence, very different interpretations and consequences.

We can rewrite equation (28) to include the person and the firm indices, i and j respectively, and their explicit characteristics, λ_i, θ_i, and $\alpha_{N,j}$, z_j.

$$\ln w(s_{C,i}, s_{N,i}) = \ln \lambda_i + \ln \tilde{w}(\theta(\alpha_{N,j}, z_j)), \tag{32}$$

where $\lambda_i = |(s_{C,i}, s_{N,i})|$ and $\theta_i = \theta(\alpha_{N,j}, z_j)$ are, respectively, worker's quality (the norm of the skill vector) and worker's skill profile. Indeed, in the absence of bunching, there is pure sorting in the horizontal dimension, meaning that θ_i depends only on the technology $(\alpha_{N,j}, z_j)$ of the worker's employing firm j.

To better see the differences with AKM, we first compare the person-effect as defined in AKM with the person component coming from the above equation. Let us assume that the econometrician perfectly observes the workers' skill *profiles* $\tilde{s}_i$ but observes the workers' skill vectors only up to a multiplicative factor. In other words, for an individual worker with skill vector $\hat{s}_i$ in the data, the true skill profile is $\tilde{s}_i = \hat{s}_i/|\hat{s}_i|$, and the true skill vector that enters the production (i.e., the worker's contribution to the aggregate task within the firm) is $s_i = \lambda_i^* \tilde{s}_i$, where λ_i^* is the unknown-to-the-econometrician worker quality. When including this component, equation (32) now comprises a true person effect, similar to that included in AKM.

Within the pure bundling framework adopted here, there is no mobility. Hence, there is no way – in contrast to what is required for identification in AKM – to differentiate out the person effect and be left with firm effects. Indeed, in the AKM framework, the latter manifest themselves when workers move from firm to firm. In addition, as mentioned multiple times, when there is a single market where supply and demand are equated, *similar* workers (endowed with identical skills including those unobserved by the econometrician but visible to the firms) are paid the *same* wage (due to perfect competition, except for the bundling constraint). As a consequence, there is no high-wage firms – firms that will pay every worker higher wages – as in AKM, in Card, Cardoso, Heining, and Kline (2018) or, more recently, Wong (2023). The W-to-F sorting effects are not AKM firm-effects.

We continue our discussion of identification of the wage schedule in Choné, Gozlan, and Kramarz (2024). Indeed, when workers alter their labor supply in response to shocks, or when markets open, multiple workers will move from firm-to-firm in response to the changing equilibrium. The induced mobility is likely to help identifying elements of the above AKM-like wage equation by differencing out the "person-effect". However, the resulting "firm-effect" may not capture the pay advantage (or loss) of moving into a particular firm (in stark contrast with the classic AKM framework) but elements specific to sorting.

We now briefly discuss identification of the production function parameters (z and α). Because the (Swedish) data allow us to measure wages, the matching of workers to firms, and the exact supply of skills, equations (18) and (20) show that we can recover the distribution of firms' technological parameters from the equilibrium matching and wage schedule. It follows that for any homogenous wage schedule $w(s)$ with strictly concave iso-wage curves, any homogenous production functions $zF(.; \alpha_N)$ such that F_N/F_C increases with α_N, and any skill distribution H^w, there exist distributions of the firms' technological parameters ϕ for which w is the equilibrium wage. Such distributions H^f are not uniquely identified as Equation (20) only determines (for any α_N) the quantity $Z^f(\alpha_N)h^f(\alpha_N)$ that drives the demand for workers with skill profile $\theta^{\mathrm{D}}(\alpha_N)$ by firms with intensity α_N in the non-cognitive skill N. Admittedly, the present discussion relies on the maintained assumption that all the endogenous objects are generated by our model of bundling, that workers' skills are perfectly observed by the analyst, and that bunching does not occur in equilibrium, even though the analysis can be extended in the presence of bunching.[34] And from a computational standpoint, the shape of the wage schedule and the firm-aggregated skill vector can be extremely precisely (numerically) approximated using results from Paty, Choné, and Kramarz (2022). The algorithms suggested they can be used to perform structural estimation of firms' technologies. Because the measured sorting appears to be far from the one predicted in our theory, any structural estimation will require the introduction of unobserved heterogeneity, among other elements. We leave such issues for further research (see also our final remarks in Section 5).

4.2 Some Empirical Evidence

We now provide a summary of the empirical evidence that Fredriksson, Hensvik, and Skans (2018) (FHS, hereafter) and Skans, Choné, and Kramarz (2022) (SCK, hereafter) produced. FHS's analysis, performed before our paper was written, provides evidence consistent with our model. SCK's descriptive analysis is directly inspired by our theory and precise knowledge of FHS's results.

Data Overview: Both FHS's and SCK's results rely on a data set measuring multidimensional skills of a large fraction of Swedish male workers. The data originate from the Swedish military conscription tests taken by most males born between 1952 and 1981.The tests were taken at age 18 and the data should therefore be understood as capturing pre-market abilities. There are two main components; *cognitive abilities*, henceforth denoted as C, measured through a set of written tests and *non-cognitive abilities*, henceforth denoted as N, measured during a structured interview with a specialized psychologist both on a 1 to 9 (non-parametric) scale. The data on employment cover the period 1996 to 2013 and include all workers with measured test results in ages 20 to 64. To examine sorting, the analysis examines each worker's co-workers rather than each worker's employing establishment and its characteristics (productivity for instance).

Some Lessons from FHS: FHS's analysis focuses on the "negative of sorting", i.e. on mismatch and its dynamics. If the talents measured before market entry are differentially productive across jobs[35], their allocation, in particular for experienced workers, is informative about sorting. "The results (...) imply substantial sorting on comparative advantages across jobs" (FHS, p. 3305). This sorting may take time, since workers learn and move between jobs, resulting in decreasing mismatch. More precisely,

[34] Although equilibrium conditions are more involved if bunching occurs (recall Section 3.4), most parameters remain identified.

[35] FHS define jobs as an occupation in an establishment and an entry year.

FHS show that workers are sorted into jobs on the basis of the specific types of skills they have (their Table 3), summarized as "The labor market is characterized by strong *horizontal* sorting on specific abilities across jobs within occupations." (p. 3312). To study sorting on wages, they compute the marginal job-specific returns of each skill (their Table 4). To do so, they run 60,500 wage regressions (i.e. one per job). They conclude that "Workers are found in jobs where the returns to their talents are higher than average, as suggested by Roy" (and fully consistent with our bundling framework, albeit without a Roy-style interpretation). In addition, a focus on tenured workers, a lesson from FHS, is incorporated in SCK.

Sorting: SCK classifies workers as *Generalists* or *Specialists* depending on the relationship between their two reported scores (trying to capture the skills ratio, s_C/s_N, defined in the theory Sections in the two skills case).[36]

Building on this worker-level classification, we classify establishments as a function of their workers' dominating type (*and not the employing firm's productivity since we examine workers' sorting rather than the workers-to-firms matching*) to inform us about α, i.e. the type of production function used by the establishment. SCK also classifies workers using their overall ability levels or "quality" (parameter λ in the theory).

SCK first define workers as low skilled if the sum of cognitive and non-cognitive abilities falls strictly below 9 and high-skilled if the sum is strictly above 11 whereas the mid-skilled are those in-between. Together with their types of skill, i.e. generalists, C and N-specialists, SCK creates 9 types of workers. Then, they run regressions where each of these 9 types is the outcome and the explanatory variables are the co-worker (leave-out) mean levels of these attributes.

Resulting estimates show that high-level N-specialists are employed together with high-level N-specialists. Similarly, high-level generalists and high-level C-specialists are employed with their peers. Similar patterns also appear for mid- and low-level workers although horizontal sorting appears to be stronger for the high total ability workers. Hence, workers are sorted into establishments where their co-workers are of a similar type, a result fully consistent with employers having heterogeneous production functions that differ in their productive values of N and C skills.

Skills and Wages: SCK examines whether market returns to each skill are higher in settings where the technology is likely to use more intensively this exact skill, as predicted by our theory. The type of employer is again based on the share of each type

[36] SCK heuristically define workers as *Generalists* if $|C_i - N_i| < 2$ and consequently define workers as *C-Specialists* if $C_i > N_i + 1$ and *N-Specialists* if $N_i > C_i + 1$. Notice that this distinction is absent from FHS.

of specialists that are employed by the establishment. SCK estimate an equation in which the type of the establishment is interacted with the specialization of the worker and estimate if the returns to being a C-intensive worker are higher if the employer uses a C-intensive technology (and conversely for N). Indeed, the results suggest that the wages in segments where employers rely intensively on C-skills also pay higher returns to these exact skills. Similarly, the results suggest a premium for N-skills in market segments dominated by N-intensive firms. These patterns are robust to controls for occupations, analyzing data at the job-level (other results with a similar flavor are given in Skans, Choné, and Kramarz, 2022).

4.3 From Workers' Skills to Firms' Tasks (and Back)

We now propose extensions of our baseline model in two directions. We examine (i) how the tasks performed by workers depend on their skills; (ii) how firms aggregate the tasks performed by their employees.

Link between skills and tasks at the worker level So far, we have equated skills and tasks, i.e., we have assumed $s = t$ for all workers. By allowing labor supply to respond to wages, Choné, Gozlan, and Kramarz (2024) endogenize the relationship between the workers' skills and the tasks they perform. For now, we restrict attention to skills-to-tasks relationships of the form $t = g(s)$, where the function g is exogenous and occupation-specific.[37] From the above analysis, we know that *in the space of tasks* the wage is convex and homogenous of degree one, i.e., $w(t_C, t_N)$ is convex and homogenous in (t_C, t_N). We also know that under the assumptions of Proposition 4, for instance with the production function (12), the firms-to-tasks matching is PAM, in the sense that there exists an increasing relationship between firms' intensities α_N and task profiles t_N/t_C. However, as already mentioned, tasks are unobserved in existing data; only workers' skills are. It is therefore important to know (i) whether the observed wage schedule $w(g(s))$, i.e., the wage as a function of the observed skills, inherits the properties of w; and (ii) whether the firms-to-skills matching inherits the properties of the firms-to-tasks matching.

To answer this question, we start from the simplest and most intuitive way to characterize the above relationship between skills and tasks and assume that each task uses each of the worker's skills in fixed quantities:

$$(t_C, t_N) = g(s_C, s_N) = (d_{CC}s_C + d_{CN}s_N, d_{NC}s_C + d_{NN}s_N), \tag{33}$$

[37] This specification happens to be a special case of the model with endogenous labor supply studied in Choné, Gozlan, and Kramarz (2024).

with $d_{ij} \geq 0$. For such a linear relationship $t = Ds$, it is straightforward to check that wages are convex and homogenous in skills. Moreover, if $\det D > 0$ there is an increasing relationship between workers' skill profiles s_N/s_C and task profiles t_N/t_C. The same is true when the skills-to-tasks relationship is homogenous of degree $\gamma > 0$, as in $(t_C, t_N) = g(s_C, s_N) = (s_C^\gamma, s_N^\gamma)$, so that in these two cases firms-to-tasks PAM translates into firms-to-skills PAM.[38] In Choné, Gozlan, and Kramarz (2024), we establish similar results with endogenous supply of skills.

Aggregation of workers' tasks within firms One may consider aggregation technologies that are not additively separable in the employees' tasks. An often used aggregation scheme is CES:

$$T = \left(\left[\int t_C^\gamma \, n^{\mathrm{D}}(\mathrm{d}t) \right]^{1/\gamma}, \left[\int t_N^\gamma \, n^{\mathrm{D}}(\mathrm{d}t) \right]^{1/\gamma} \right). \tag{34}$$

In our leading example (7) where the production function $F(T_C, T_N)$ is itself CES, such a skills-aggregation scheme leads to a two-level nested CES. For our theoretical results to apply, we need $F(T)$ to be concave in the assignment n^{D}, i.e., we need the modified production function $\tilde{F}(T_C, T_N) = F(T_C^{1/\gamma}, T_N^{1/\gamma})$ to be concave in T, which obtains if $\gamma > \max(\rho, \eta)$.[39] We can thus allow for some degree of imperfect substitution between co-workers' tasks.

Finally, the number of skills needs not be equal to the number of tasks. Suppose two types of cognitive skills and two types of non-cognitive skills are used to produce two tasks according to

$$T = (T_1, T_2) = (\mathrm{CES}(C_1, N_1; \beta_1), \mathrm{CES}(C_2, N_2; \beta_2)).$$

Task 1 uses skills C_1 and N_1, with β_1 representing the (potentially firm-specific) technical intensity in C_1 with an equivalent formulation holding for Task 2. The final output is then produced by combining the two tasks according to $F(T; \alpha)$. As above, the production function can be rewritten as $\tilde{F}(C_1; C_2, N_1, N_2; \alpha, \beta_1, \beta_2)$, where capital letters represent firm-aggregated quantities (for instance $C_1 = \int s_{C1} \, n^{\mathrm{D}}(\mathrm{d}s_{C1})$) and $(\alpha, \beta_1, \beta_2)$ is a *firm-specific* vector of technical parameters.

[38] Indeed, the linear case of (33) is not econometrically different from $t = s$ since, given data, g and the distribution of α cannot be separately identified. This is not so when g is, for instance, homogenous of degree 2, which would translate into a wage function which would be homogenous of degree $1/2$ in s, something potentially observable in the data.

[39] The new production function $\tilde{F}$ is quasi-concave ($\rho/\gamma < 1$) and homogenous of degree $\eta/\gamma < 1$.

A distinctive feature of the setup examined in the present paper is that firm-specific parameters interact only with firm-aggregated quantities. Using the first-order condition (11), aggregate sorting properties can be derived as in Proposition 4 noticing that $\tilde{F}_{C_i}/\tilde{F}_{N_i}$ increases with β_i and $\tilde{F}_{T_1}/\tilde{F}_{T_2}$ increases with α. This class of production functions strikingly differs from settings where a firm's set of technical characteristics interact with *individual* workers' characteristics, which pushes to individual rather than aggregate sorting.[40] The above formulation which connects skills and tasks, when compared with Haanwinckel (2023) or Teulings (2005), offers more between-firms heterogeneity or, when compared with Eeckhout and Kircher (2018), possesses a clear within-firm aggregation scheme.

5 Conclusion

This paper, albeit theoretical, has an applied motivation. It starts from empirical questions on the deep structure of labor markets as they operated until recently.

To recap some of our results, under bundling, the law of one price virtually never obtains: the implicit prices paid to workers for their skills vary across employing firms. Wages have an AKM-like log-additive structure. Sorting of workers across firms is based on comparative – rather than absolute – advantage and naturally generates within-firm worker heterogeneity, in a competitive framework but for the bundling friction.

Our modeling approach highlights the productive role of workers by having clear within-firm aggregation schemes of skills to tasks, the inputs of the firms' production functions. Our framework accommodates a lot of between-firms heterogeneity and is highly versatile. For instance, it may be embedded into a Dixit-Stiglitz framework with little changes in its principles.

Our companion paper, Choné, Gozlan, and Kramarz (2024), studies how such markets are being transformed today, with the trade of stand-alone skills being facilitated by new markets and intermediaries. The structure of wages provides a striking example of contrasts between the old and the new world. Associated with markets opening and unbundling, our companion paper demonstrates a "flattening" of wage schedules, inducing a potential attenuation of the workers-to-firms sorting effects, reminiscent of the disappearance of what the literature calls, after AKM, firm effects. As outsourcing markets gradually open, the distribution of skills on the labor market becomes more polarized and firms use more specialized skills, adhering more closely to their technological characteristics, thus improving their comparative advantage. Such effects will be empirically studied in Choné, Kramarz, and Skans (2024) using Swedish data.

[40] Choné and Kramarz (2022) considers the case where tasks are produced by interacting individual firm and worker characteristics and are then aggregated within each firm.

Beyond unbundling and its associated empirical consequences, there are at least two directions to expand our research agenda. First, incorporating bundling into other classic models such as Random Search or Bargaining à la Nash should enlarge considerably its scope and interest for various scholars. Second, both this paper and Choné, Gozlan, and Kramarz (2024) focus on conical aggregation schemes where production depends on the sum of skill vectors within firms. However, Choné, Gozlan, and Kramarz (2023) also provides existence and duality results for *non-conical* problems, allowing the analyst to examine a much broader class of production functions. In the spirit of the literature that incorporates the costs and benefits of within-firm workers' diversity (see e.g. Kremer, 1993, Garicano and Rossi-Hansberg, 2006), the above mathematical results provide other routes to penalize diversity (giving generalist workers a productive advantage over the corresponding combination of specialists) within a general equilibrium framework. How the cost of within-firm diversity affects the qualitative properties of the competitive wage schedule is left for future research.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67(2), 251–333.

Acemoglu, D., and D. Autor (2011): "Skills, tasks and technologies: Implications for employment and earnings," in *Handbook of labor economics*, vol. 4, pp. 1043–1171. Elsevier.

Acemoglu, D., and P. Restrepo (2018): "The race between man and machine: Implications of technology for growth, factor shares, and employment," *American Economic Review*, 108(6), 1488–1542.

Autor, D., C. Chin, A. Salomons, and B. Seegmiller (forthcoming): "New Frontiers: The Origins and Content of New Work, 1940-2018," *The Quarterly Journal of Economics*.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, 135(2), 645–709.

Autor, D. H. (2015): "Why are there still so many jobs? The history and future of workplace automation," *Journal of Economic Perspectives*, 29(3), 3–30.

Berger, D., K. Herkenhoff, and S. Mongey (2022): "Labor Market Power," *American Economic Review*, 112(4), 1147–1193.

BITTARELLO, L., F. KRAMARZ, AND A. MAITRE (forthcoming): "The Task Content of Occupations," *Revue Economique.*

BOERMA, J., A. TSYVINSKI, AND A. ZIMIN (2021): "Sorting with Team Formation," *Working Paper.*

CARD, D., A. R. CARDOSO, J. HEINING, AND P. KLINE (2018): "Firms and labor market inequality: evidence and some theory," *Journal of Labor Economics*, 36(S1), S13–S70.

CARD, D., AND J. E. DINARDO (2002): "Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *The Journal of Labor Economics*, 20(4), 733–783.

CHIAPPORI, P.-A., R. MCCANN, AND B. PASS (2016): "Multidimensional matching," *arXiv preprint arXiv:1604.05771.*

CHIAPPORI, P.-A., R. J. MCCANN, AND L. P. NESHEIM (2010): "Hedonic price equilibria, stable matching, and optimal transport: equivalence, topology, and uniqueness," *Economic Theory*, 42(2), 317–354.

CHONÉ, P., N. GOZLAN, AND F. KRAMARZ (2023): "Weak optimal transport with unnormalized kernels," *SIAM Journal of Mathematical Analysis*, 55(6), 6039–6092.

——— (2024): "Workers-to-Firms Matching When Skills Become Unbundled," *CEPR Discussion Paper*, DP19438.

CHONÉ, P., AND F. KRAMARZ (2022): "Transporting superstars," On-going work.

CHONÉ, P., F. KRAMARZ, AND O. N. SKANS (2024): "From Bundling to Unbundling: The Evolution of Labor Markets Over the Last Thirty Years," in progress.

COSTINOT, A., AND J. VOGEL (2010): "Matching and inequality in the world economy," *Journal of Political Economy*, 118(4), 747–786.

CRAWFORD, V. P. (1991): "Comparative statics in matching markets," *Journal of Economic Theory*, 54(2), 389–400.

EDMOND, C., AND S. MONGEY (2022): "Unbundling labor," Discussion paper.

EECKHOUT, J., AND P. KIRCHER (2018): "Assortative Matching With Large Firms," *Econometrica*, 86(1), 85–132.

Fredriksson, P., L. Hensvik, and O. N. Skans (2018): "Mismatch of talent: Evidence on match quality, entry wages, and job mobility," *American Economic Review*, 108(11), 3303–38.

Galichon, A. (2018): *Optimal transport methods in economics*. Princeton University Press.

Galichon, A., and B. Salanié (forthcoming): "Cupid's invisible hand: Social surplus and identification in matching models," *Review of Economic Studies*.

Garicano, L., and E. Rossi-Hansberg (2006): "Organization and inequality in a knowledge economy," *The Quarterly journal of economics*, 121(4), 1383–1435.

Gozlan, N., C. Roberto, P.-M. Samson, and P. Tetali (2017): "Kantorovich duality for general transport costs and applications," *Journal of Functional Analysis*, 273(11), 3327–3405.

Haanwinckel, D. (2023): "Supply, demand, institutions, and firms: A theory of labor market sorting and the wage distribution," *Working paper*.

Hatfield, J. W., and P. R. Milgrom (2005): "Matching with contracts," *American Economic Review*, 95(4), 913–935.

Heckman, J., and J. Scheinkman (1987): "The importance of bundling in a Gorman-Lancaster model of earnings," *The Review of Economic Studies*, 54(2), 243–255.

Heckman, J. J., and B. E. Honore (1990): "The empirical content of the Roy model," *Econometrica: Journal of the Econometric Society*, pp. 1121–1149.

Hernnäs, S. (2021): "Automation, Task Content, and Occupations when Skills and Tasks are Bundled," Discussion paper.

Jung, J. W., I. Simonovska, and A. Weinberger (2019): "Exporter heterogeneity and price discrimination: A quantitative view," *Journal of International Economics*, 116, 103–124.

Katz, L. F., and K. M. Murphy (1992): "Changes in relative wages, 1963–1987: supply and demand factors," *The Quarterly Journal of Economics*, 107(1), 35–78.

Kelso, A. S., and V. P. Crawford (1982): "Job matching, coalition formation, and gross substitutes," *Econometrica*, pp. 1483–1504.

Kremer, M. (1993): "The O-ring theory of economic development," *The quarterly journal of economics*, 108(3), 551–575.

Laffont, J.-J., and D. Martimort (2009): *The theory of incentives: the principal-agent model.* Princeton university press.

Lancaster, K. J. (1966): "A new approach to consumer theory," *Journal of Political Economy*, 74(2), 132–157.

Lemieux, T. (2006): "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill?," *American Economic Review*, 96(3), 461–498.

Lindenlaub, I. (2017): "Sorting multidimensional types: Theory and application," *The Review of Economic Studies*, 84(2), 718–789.

Lindenlaub, I., and F. Postel-Vinay (2023): "Multidimensional Sorting under Random Search," *Working Paper*.

Mandelbrot, B. (1962): "Paretian distributions and income maximization," *The Quarterly Journal of Economics*, 76(1), 57–85.

Murphy, K. M. (1986): "Specialization and Human Capital," Ph.D. thesis, University of Chicago.

Ohnsorge, F., and D. Trefler (2007): "Sorting It Out: International Trade with Heterogeneous Workers," *The Journal of Political Economy*, 115(5), 868–892.

Paty, F.-P., P. Choné, and F. Kramarz (2022): "Algorithms for Weak Optimal Transport with an Application to Economics," *arXiv preprint arXiv:2205.09825*.

Peyré, G., and M. Cuturi (2019): "Computational optimal transport," *Foundations and Trends® in Machine Learning*, 11(5-6), 355–607.

Pycia, M. (2012): "Stability and preference alignment in matching and coalition formation," *Econometrica*, 80(1), 323–362.

Pycia, M., and M. B. Yenmez (2019): "Matching with externalities," .

Rochet, J.-C., and P. Choné (1998): "Ironing, sweeping, and multidimensional screening," *Econometrica*, pp. 783–826.

Rockafellar, R. T. (1970): *Convex analysis*, no. 28. Princeton University Press.

Rosen, S. (1974): "Hedonic prices and implicit markets: product differentiation in pure competition," *Journal of Political Economy*, 82(1), 34–55.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings," *Oxford economic papers*, 3(2), 135–146.

Sato, R. (1977): "Homothetic and non-homothetic CES production functions," *The American Economic Review*, 67(4), 559–569.

Simonovska, I. (2015): "Income differences and prices of tradables: Insights from an online retailer," *The Review of Economic Studies*, 82(4), 1612–1656.

Skans, O. N., P. Choné, and F. Kramarz (2022): "The Empirics of Labor Market Sorting When Workers Have Bundles of Skills and Jobs Are Heterogeneous," IFAU Working Paper 2022:6.

Takayama, A. (1985): *Mathematical economics*. Cambridge university press.

Teulings, C. N. (2005): "Comparative advantage, relative wages, and the accumulation of human capital," *Journal of Political Economy*, 113(2), 425–461.

Villani, C. (2009): *Optimal transport: old and new*, vol. 338 of *Grundlehren der mathematischen Wissenschaften*. Springer Science, New York.

Wilson, R. B. (1993): *Nonlinear pricing*. Oxford University Press.

Wong, H. C. (2023): "Understanding High-Wage Firms," Discussion paper.

A Appendix

A.1 Proof of Proposition 1

Existence of an optimal market-clearing assignment To prove existence, we extend Theorem 6.4 of Choné, Gozlan, and Kramarz (2023), hereafter CGK. Specifically, we relax their assumption that $\inf_{\phi\in\mathcal{X}^f, s\in\mathcal{X}^s} F(\mu s;\phi)$ tends to $+\infty$ as $\mu\to+\infty$. This assumption indeed does not hold if some workers have zero skill in some dimensions (i.e., the support of H^w intersects the boundary of $\mathbb{R}^d_+$) and the skills are complements as is the case in our baseline CES example (7) with $\rho<0$.

Lemma A.1. *Under the regularity assumption made at the beginning of Section 2, the production function F satisfies Condition (A') of CGK, i.e., there exists sequences $(u_n(\phi))_{n\geq 0}$ and $(v_n(\phi))_{n\geq 0}$ of continuous functions on $\mathcal{X}^f$ and taking values respectively in $\mathbb{R}$ and $\mathbb{R}^d$ such that*

$$F(s;\phi) = \inf_n u_n(\phi) + v_n(\phi)\, s. \tag{A.1}$$

Furthermore, the value of the primal problem is finite: $\mathcal{J}^b(H^f, H^w) < \infty$.

Proof. Take $(s_n)_{n\geq 0}$ a dense subset of $\mathcal{X}^s$. For each n and all $s\in\mathcal{X}^s$, the concavity of F yields

$$F(s;\phi) \leq F(s_n;\phi) + \nabla_s F(s_n;\phi)(s-s_n).$$

By assumption, the functions u_n and v_n defined as

$$u_n(\phi) = F(s_n;\phi) - \nabla_s F(s_n;\phi)s_n \;\text{ and } v_n(\phi) = \nabla_s F(s_n;\phi)$$

are continuous in ϕ. For any ϕ, the function $\inf_n u_n(\phi) + v_n(\phi)\, s$ is concave in s, is greater than or equal to $F(s;\phi)$, and coincides with F at all points s_n. This yields the result by continuity. Furthermore, assuming that the total skills in the economy are finite and that $\mathcal{X}^f$ is compact, we have for all market-clearing assignment n^{D}:

$$\int F\left(\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)\right) H^f(\mathrm{d}\phi) \leq \bar{U} + \bar{V}\int sH^w(\mathrm{d}s) < \infty,$$

with $\bar{U} = \sup_{\phi\in\mathcal{X}^f} u_0(\phi)$ and $\bar{V} = \sup_{\phi\in\mathcal{X}^f} v_0(\phi)$.

The following a priori estimate holds:

$$\begin{aligned}\int F\left(\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)\right) H^f(\mathrm{d}\phi) &\leq \int \left\{u_0(\phi) + v_0(\phi)\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)\right\} H^f(\mathrm{d}\phi)\\ &\leq \bar{U} + \bar{V}\iint s\, n^{\mathrm{D}}(\mathrm{d}s;\phi) H^f(\mathrm{d}\phi)\\ &= \bar{U} + \bar{V}\int sH^w(\mathrm{d}s),\end{aligned}$$

implying that $\mathcal{J}^b(H^f, H^w) < \infty$. □

As explained below Equation (3), one can associate to any market-clearing assignment n^{D} a probability distribution π on $\mathcal{X}^f \times \mathcal{X}^s$ whose first marginal $\tilde{H}^f$ is absolutely continuous with respect to the firms' distribution H^f. Disintegrating such a distribution as $\pi(\mathrm{d}s, \mathrm{d}\phi) = \pi_1(\phi)\, q(\mathrm{d}s;\phi)$, where the first marginal π_1 is absolutely continuous with respect to H^f, one can define the functional

$$\mathcal{I}(\pi; H^f) = \int F\left(\int \frac{\mathrm{d}\pi_1}{\mathrm{d}H^f}\int sq(\mathrm{d}s;\phi)\right) H^f(\mathrm{d}\phi).$$

CGK introduce the set of distributions whose first marginal $\tilde{H}^f$ is absolutely continuous with respect to H^f and whose second marginal is the distribution of skills H^w:

$$\Pi(\ll H^f, H^w) \stackrel{\mathrm{d}}{=} \{\pi \in \Pi(\tilde{H}^f, H^w)\,,\ \tilde{H}^f \in \mathcal{P}(\mathcal{X}^f), \tilde{H}^f \ll H^f\},$$

with $\Pi(\mu,\nu)$ denoting the set of all couplings between μ and ν. An assignment n^{D} is optimal if and only if the associated coupling π maximizes $\mathcal{I}(\pi; H^f)$ on $\Pi(\ll H^f, H^w)$. The latter set, however, is generally not closed, and for this reason the existence of an optimal assignment, which CGK call a "strong solution", is not guaranteed. Accordingly, they introduce a notion of weak solution. They say that a finite measure π on $\mathcal{X}_f \times \mathbb{R}^d_+$ is a weak solution of the optimal transport problem if there exists a sequence of market-clearing assignments n^{D}_n such that the total output $\int F\left(\int s\, n^{\mathrm{D}}_n(\mathrm{d}s;\phi);\phi\right) H^f(\mathrm{d}\phi)$ tends to $\mathcal{J}^b(H^f, H^w)$. They prove that weak solutions always exist (Proposition 2.4).

To investigate the existence of strong solutions, CGK show that the closure of $\Pi(\ll H^f, H^w)$ is the set $\Pi(\mathcal{X}^f; H^w)$ of all probability measures such that $\pi_1(\mathcal{X}^f) = 1$ and $\pi_2 = H^w$. Assuming that $F'_\infty(s;\phi) = \lim_{\mu\to\infty} F(\mu s;\phi)/\mu$ is a continuous function on $\mathcal{X}^f \times \mathbb{R}^k_+$, they introduce the functional

$$\bar{\mathcal{I}}(\pi; H^f) = \int F\left(\int \frac{\mathrm{d}\pi_1^{\mathrm{ac}}}{\mathrm{d}H^f}\int sq(\mathrm{d}s;\phi)\right) H^f(\mathrm{d}\phi) + \iint F'_\infty(s;\phi) q(\mathrm{d}s;\phi)\pi_1^{\mathrm{sing}}(\mathrm{d}\phi),$$

where $\pi(\mathrm{d}\phi, \mathrm{d}s) = \pi_1(\mathrm{d}\phi)\, q(\mathrm{d}s;\phi)$ and $\pi_1 = \pi_1^{\mathrm{ac}} + \pi_1^{\mathrm{sing}}$ is the decomposition of the first marginal π_1 into absolutely continuous and singular parts with respect to H^f. They show that $\bar{\mathcal{I}}$ is a lower semicontinuous extension of $\mathcal{I}$ on $\Pi(\mathcal{X}^f; H^w)$. Under Assumption (A') mentioned in Lemma A.1, Theorem 3.7 of CGK states that a measure π is a weak solution if and only if it maximizes $\bar{\mathcal{I}}(\pi; H^f)$ over $\pi \in \Pi(\mathcal{X}^f; H^w)$. Moreover, the value of the maximum is nothing else than $\mathcal{J}^b(H^f, H^w)$. Assumption 1 states that $F'_\infty = 0$, implying that the functional simplifies into

$$\bar{\mathcal{I}}(\pi; H^f) = \int F\left(\int \frac{\mathrm{d}\pi_1^{\mathrm{ac}}}{\mathrm{d}H^f} \int sq(\mathrm{d}s;\phi)\right) H^f(\mathrm{d}\phi).$$

We now extend the results of CGK to prove the existence of an optimal market-clearing assignment under weaker assumptions. We start from any weak solution $\bar{\pi} = \bar{\pi}(\mathrm{d}s;\phi)\bar{\pi}_1(\mathrm{d}\phi)$. This means that $\bar{\pi}$ maximizes $\bar{\mathcal{I}}(\bar{\pi}; H^f)$ on $\Pi(\mathcal{X}^f; H^w)$ and $\bar{\mathcal{I}}(\bar{\pi}; H^f) = \mathcal{J}^b(H^f, H^w)$.

We set $\gamma(\mathrm{d}s) = \int \bar{\pi}^\phi(\mathrm{d}s)\bar{\pi}^{\mathrm{sing}}(\mathrm{d}\phi)$ and consider the assignment n^{D} defined by

$$n^{\mathrm{D}}(\mathrm{d}s;\phi) = \frac{\mathrm{d}\bar{\pi}^{\mathrm{ac}}}{\mathrm{d}H^f}(\phi)\bar{\pi}^\phi(\mathrm{d}s) + \gamma(\mathrm{d}s).$$

The assignment n^{D} clears the market because

$$\begin{aligned}\int n^{\mathrm{D}}(\mathrm{d}s;\phi)H^f(\mathrm{d}\phi) &= \gamma(\mathrm{d}s) + \int \bar{\pi}^{\mathrm{ac}}(\mathrm{d}\phi)\bar{\pi}^\phi(\mathrm{d}s) \\ &= \int \left[\bar{\pi}^{\mathrm{ac}}(\mathrm{d}\phi) + \bar{\pi}^{\mathrm{sing}}(\mathrm{d}\phi)\right]\, \bar{\pi}^\phi(\mathrm{d}s) = \int \bar{\pi}(\mathrm{d}\phi)\bar{\pi}^\phi(\mathrm{d}s) = H^w(\mathrm{d}s).\end{aligned}$$

Since $F(.;\phi)$ is nondecreasing on $\mathbb{R}_+^k$ for all ϕ,[41] it holds

$$F\left(\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi);\phi\right) = F\left(\frac{\mathrm{d}\bar{\pi}^{ac}}{\mathrm{d}H^f}(\phi)\int s\bar{\pi}^\phi(\mathrm{d}s) + \int s\gamma(\mathrm{d}s);\phi\right) \geq F\left(\frac{\mathrm{d}\bar{\pi}^{ac}}{\mathrm{d}H^f}(\phi)\int s\bar{\pi}^\phi(\mathrm{d}s);\phi\right).$$

Integrating with respect to $H^f(\mathrm{d}\phi)$ and using the fact that $\bar{\pi}$ is a weak solution, we find

$$\int F\left(\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi);\phi\right) H^f(\mathrm{d}\phi) \geq F\left(\frac{\mathrm{d}\bar{\pi}^{ac}}{\mathrm{d}H^f}(\phi)\int s\bar{\pi}^\phi(\mathrm{d}s);\phi\right) H^f(\mathrm{d}\phi) = \mathcal{J}^b(H^f, H^w)$$

and hence that n^{D} an optimal market-clearing assignment.

[41] Any concave function $F : \mathbb{R}_+^k \to \mathbb{R}_+$ is nondecreasing. To see this, consider $T_1 \leq T_2$, use concavity to write: $F(T_2) = F(T_1 + T_2 - T_1) \geq tF(T_1/t) + (1-t)F((T_2 - T_1)/(1-t)) \geq tF(T_1)$, and let t go to 1.

Duality formula and dual attainment CGK introduce a dual version of the primal problem

$$I^* \overset{\mathrm{d}}{\equiv} \inf_{w \in \Phi^+_{\mathrm{nd}}} \int \bar{\Pi}(\phi; w) H^f(\mathrm{d}\phi) + \int w(s) H^w(\mathrm{d}s), \tag{A.2}$$

where Φ^+_{nd} denotes the set of all convex, positively one-homogenous, and non-decreasing functions $w : \mathbb{R}^k_+ \to \mathbb{R}_+$, and $\bar{\Pi}(\phi; w)$ is given by

$$\bar{\Pi}(\phi; w) = \max_{T \in \mathbb{R}^k_+} F(T; \phi) - w(T). \tag{A.3}$$

Theorem 6.4 of the above paper establishes the duality formula $\mathcal{J}^b(H^f, H^w) = I^*$ as well as the existence of solutions to dual problems (5) and (A.2).

Connection between primal and dual solutions We now prove the Fundamental Welfare Theorems. On the one hand, there exists a unnormalized kernel, a family of nonnegative finite measures $n^{\mathrm{D}}(\mathrm{d}s, \phi)$ satisfying $n^{\mathrm{D}} H^f = H^w$ that achieves the upper bound in (5), hence the existence of an optimal assignment of workers to firms. On the other hand, there exists a convex, 1-homogeneous function w that achieves the lower bound in (A.2).

First, consider an equilibrium (n^{D}, w). Let us denote by T the firm-aggregated skill vector corresponding to the assignment n^{D}, i.e., $T^{\mathrm{D}}(\phi) = \int s\, n^{\mathrm{D}}(\mathrm{d}s; \phi)$. Using the market clearing condition $n^{\mathrm{D}} H^f = H^w$, we have

$$\begin{aligned} I^* &\leq \int \bar{\Pi}(\phi; w) H^f(\mathrm{d}\phi) + \int w(s) H^w(\mathrm{d}s) \\ &= \int F(T^{\mathrm{D}}(\phi); \phi) H^f(\mathrm{d}\phi) - \iint w(s)\, n^{\mathrm{D}}(\mathrm{d}s; \phi) H^f(\mathrm{d}\phi) + \int w(s) H^w(\mathrm{d}s) \\ &= \int F(T^{\mathrm{D}}(\phi); \phi) H^f(\mathrm{d}\phi) \leq \mathcal{J}^b(H^f, H^w). \end{aligned}$$

Because $\mathcal{J}^b(H^f, H^w) = I^*$, there is equality in the first and last inequalities above, implying that w is a dual optimizer (i.e., a solution of (A.2)) and that the equilibrium assignment n^{D} is optimal.

Conversely, consider an optimal market-clearing assignment n^{D}. As above, we denote by $T^{\mathrm{D}}(\phi)$ the corresponding firm-aggregated skill vector. Then, for any dual optimizer w, we have by definition of the profit function

$$F(T^{\mathrm{D}}(\phi); \phi) - \int w(s)\, n^{\mathrm{D}}(\mathrm{d}s; \phi) \leq \bar{\Pi}(\phi; w). \tag{A.4}$$

Integrating with respect to $H^f(\mathrm{d}\phi)$ and using $n^{\mathrm{D}}H^f = H^w$ yields

$$\mathcal{J}^b(H^f, H^w) = \int F(T^{\mathrm{D}}(\phi);\phi)H^f(\mathrm{d}\phi) \leq \int \bar{\Pi}(\phi;w)H^f(\mathrm{d}\phi) + \int w(s)H^w(\mathrm{d}s) = I^*. \tag{A.5}$$

The equality $\mathcal{J}^b(H^f, H^w) = I^*$ shows that we must have equality in (A.4) for H^f-almost every $\phi \in \Phi$, meaning that the optimal market-clearing assignment n^{D} is decentralized by the wage schedule w.

Skills equated with tasks With a general disutility cost function Ψ, Choné, Gozlan, and Kramarz (2024) show that the equilibrium price of tasks satisfies the following duality formula:

$$\mathcal{W}_\infty(H^f, H^w) = \inf_{p(.)\geq 0 \text{ convex 1-homog.}} \int \Pi(\phi;p)H^f(d\phi) + \int \bar{U}(s;p)H^w(\mathrm{d}s), \tag{A.6}$$

where workers' utility is given by

$$\bar{U}(s;w) = \max_{t\in\mathbb{R}^k_+} p(t) - \Psi(t;s). \tag{A.7}$$

They check that the price schedule can be assumed to be nondecreasing without any loss of generality. Now consider the nonnegative and nondecreasing cost function $\Psi(t;s)$ such that $\Psi(s;s) = 0$ and Ψ is infinite as soon as $t_j > s_j$ for some j. Because p is nondecreasing, $t = s$ is solution to (A.7).

A.2 Proof of Proposition 2

We explain here in more detail why we can restrict attention to convex, 1-homogeneous and non-decreasing wage schedules. The firms' problem (6) can be broken down into two subproblems that consist respectively in finding the firm-aggregated skill vector T and in achieving that aggregate vector in the most economical way. Formally

$$\Pi(\phi;w) = \max_{T\in\mathcal{Z}^s} F(T;\phi) - \bar{w}(T), \tag{A.8}$$

where $\mathcal{Z}^s$ is the conical hull of $\mathcal{X}^s$: $\mathcal{Z}^s = \left\{\sum_{j=1}^k a_j s_j, a_1, \dots a_k \in \mathbb{R}_+, s_1, \dots, s_k \in \mathcal{X}^s\right\}$ and

$$\bar{w}(T) = \inf\left\{\int w(s)\, n^{\mathrm{D}}(\mathrm{d}s) : n^{\mathrm{D}} \in \mathcal{M}(\mathcal{X}^s), \int s\, n^{\mathrm{D}}(\mathrm{d}s) = T\right\}. \tag{A.9}$$

It is easy to check that the function $\bar{w}$ defined in (A.9) is convex and homogenous of degree one. For any $s \in \mathcal{X}^s$, we can take the $n^{\mathrm{D}}(\mathrm{d}s)$ as the mass point at s, thus showing that $\bar{w}(s) \leq w(s)$. The map $\bar{w} : \mathcal{Z}^s \to \mathbb{R}_+$ is therefore the greatest convex and homogenous function such that $\bar{w} \leq w$ on $\mathcal{X}^s$.

By construction of $\bar{w}$, we have: $\Pi(\phi; w) = \bar{\Pi}(\phi; \bar{w})$. Moreover, because $\bar{w} \leq w$, we have: $\int \bar{w}(x) H^w(\mathrm{d}x) \leq \int w(x) H^w(\mathrm{d}x)$. It follows that if w is a dual optimizer, i.e., a solution of Problem (A.2), so is $\bar{w}$. Using $\bar{w}$ instead of w in (A.4) and (A.5) shows that the optimal market-clearing assignment n^{D} is decentralized by the convex and positively homogenous wage schedule $\bar{w}$.

Finally, we show that at any equilibrium the wage schedule can be assumed to nondecreasing. Starting from any convex, 1-homogeneous wage schedule p, we consider the modified schedule: $\bar{w}(t) = \inf_{t' \geq t} w(t')$. It is easy to see that $\bar{p}$ is convex, 1-homogeneous, and nondecreasing, and that $\bar{w} \leq w$. Under schedule w, because the production function $F(T; \phi)$ is nondecreasing in T, firms choose $T' > T$ if T' is less costly than T. So replacing w with $\bar{w}$ does not alter the firms' demand behavior, which implies: $\Pi(\phi; \bar{w}) = \Pi(\phi; w)$. It follows that $\bar{w} \leq w$ is solution to the dual problem (A.2).

Relation between skills and tasks We present the change of variables $t = g(x)$ mentioned in Subsection 4.3. We define the probability distribution over tasks: $\tilde{H}^w(\mathrm{d}t) = g_{\#} H^w(\mathrm{d}x)$. To any assignment $n^{\mathrm{D}}(\mathrm{d}x; \phi)$, we associate the assignment in the task space $M^d(\mathrm{d}t; \phi) = g_{\#}\, n^{\mathrm{D}}(\mathrm{d}x; \phi)$. Because g is one-to-one, the market clearing conditions $n^{\mathrm{D}} H^f = H^w$ and $M^d H^f = \tilde{H}^w$ are equivalent. The primal problem (5) that defines the optimal output under bundling can be rewritten as

$$\mathcal{J}^b(H^f, H^w) = \sup_{M^d | M^d H^f = \tilde{H}^w} \int F\left(\int t M^d(\mathrm{d}t; \phi)\right) H^f(\mathrm{d}\phi).$$

Starting from any wage schedule $w(x)$, we define the corresponding wage in the tasks space as $w^t(t) = w(g^{-1}(t))$ and rewrite the firms' profit (6) as

$$\tilde{\Pi}(\phi; p) = \max_{M^d(\mathrm{d}t; \phi)} F\left(\int t M^d(\mathrm{d}t; \phi)\right) - \int w^t(t) M^d(\mathrm{d}t; \phi).$$

We can also write the dual problem (A.2) as

$$I^* = \inf_{p \in \mathcal{C}_b(g(\mathcal{X}))} \int \tilde{\Pi}(\phi; p) H^f(\mathrm{d}\phi) + \int w^t(t) \tilde{H}^w(\mathrm{d}t).$$

We can thus apply the Fundamental Theorems in the tasks space $g(\mathcal{X})$ equipped with the probability measure $\tilde{H}^w(\mathrm{d}t)$ and the firm space Φ with the probability $H^f(\mathrm{d}\phi)$.

A.3 Proof of Proposition 3

Consider two competitive equilibria (n^{D}, w_1) and (n^{D}, w_2). Let $T_i^{\mathrm{D}} = \int x\, n_i^{\mathrm{D}}(\mathrm{d}x; \phi)$, $i = 1, 2$ denote the corresponding firm-aggregated skill vectors. We know from the proof of Proposition 1 that w_1 and w_2 are dual optimizers and that for any dual optimizer w the vectors $T_1^{\mathrm{D}}(\phi)$ and $T_2^{\mathrm{D}}(\phi)$ are solutions to Problem (9) for H^f-almost every ϕ. Because F is strictly concave and w is convex, this problem is strictly concave, which yields $T_1^{\mathrm{D}} = T_2^{\mathrm{D}} \stackrel{\mathrm{d}}{=} T^{\mathrm{D}}$. It follows that $\nabla F(T^{\mathrm{D}}; \phi) = \nabla w(T^{\mathrm{D}}) = \nabla w_1(T^{\mathrm{D}}) = \nabla w_2(T^{\mathrm{D}})$. By homogeneity of the wage (Euler's identity), we have $w(T^{\mathrm{D}}) = T^{\mathrm{D}} \cdot \nabla w(T^{\mathrm{D}})$. It follows that $w = w_1 = w_2$ on the image of T^{D}.

In the absence of bunching (Figure 1), all workers employed by a firm of type ϕ have the same skill profile, i.e., their skill vector is collinear to $T^{\mathrm{D}}(\phi)$. By homogeneity of the wage schedule, it follows that $w(s) = s.\nabla w(T^{\mathrm{D}}(\phi))$. In the presence of bunching (Figure 3), the skill vectors s of workers employed by firm ϕ belong to a linear part of the iso-wage line $w = 1$ and are paid the same implicit prices $w_i(T^{\mathrm{D}}(\phi))$. So $w(s) = s.\nabla w(T^{\mathrm{D}}(\phi))$ still holds if worker with skill vector s is employed by firm of type ϕ. Because w is convex and 1-homogenous, we have $w(s) \geq w(T^{\mathrm{D}}(\phi')) + \nabla w(T^{\mathrm{D}}(\phi')) \cdot (s - T^{\mathrm{D}}(\phi')) = s \cdot \nabla w(T^{\mathrm{D}}(\phi'))$ for all firm type ϕ', which yields

$$w(s) = \sup_{\phi'} s \cdot \nabla w(T^{\mathrm{D}}(\phi')). \tag{A.10}$$

Because $T^{\mathrm{D}}(\phi)$ and $\nabla w(T^{\mathrm{D}}(\phi)) = \nabla F(T^{\mathrm{D}}(\phi))$ are unique, so is the wage schedule w. $\square$

Notice that Proposition 3 holds for the CES production function (7) with complementary skills ($\rho < 0$) although it is not strictly concave (this function is zero if one skill is not used). This is because the maximizer T of $\bar{\Pi}(\phi; w)$ is unique for all non-trivial wage schedules w – the sole property of the production function used above.

Proof of Corollary 1 From (10), the marginal rate of technical substitution (MRTS) equals the ratio of implicit prices across skills:

$$\frac{F_j(T^{\mathrm{D}}(\phi); \alpha, z)}{F_j(T^{\mathrm{D}}(\phi); \alpha, z)} = \frac{w_j(T^{\mathrm{D}}(\phi))}{w_k(T^{\mathrm{D}}(\phi))},$$

where $T^{\mathrm{D}} = \Lambda^{\mathrm{D}} \tilde{S}^{\mathrm{D}}$, $\Lambda^{\mathrm{D}} > 0$. Because the wage schedule is positively homogenous, the wage isolines are homothetic, and the ratios w_j/w_k depend only on $\tilde{S}^{\mathrm{D}}$. If the production functions have homothetic isoquants, the same is true for the MRTS F_j/F_k.

Proof of Corollary 2 From Corollary 1, we know that the average skill profile $\tilde{S}^{\mathrm{D}}$ does not depend on z. The total quality of a firm ϕ's employees, $\Lambda^{\mathrm{D}}(\phi)$, is determined by maximizing its profit:

$$\Pi(\phi; w) = \max_{\Lambda} z\, F(\Lambda \tilde{S}^{\mathrm{D}}(\alpha); \alpha) - \Lambda w(\tilde{S}^{\mathrm{D}}(\alpha)). \tag{A.11}$$

Using that F is homogenous of degree $\eta < 1$, we find that the total quality of workers employed by firm $\phi = (\alpha, z)$:

$$\Lambda^{\mathrm{D}}(\alpha, z) = \left[\frac{\eta\, z\, F(\tilde{S}^{\mathrm{D}}(\alpha); \alpha)}{w(\tilde{S}^{\mathrm{D}}(\alpha))}\right]^{\frac{1}{1-\eta}}. \tag{A.12}$$

The firm's aggregate skill is $T^{\mathrm{D}}(\phi) = \Lambda^{\mathrm{D}}(\alpha, z)\tilde{S}^{\mathrm{D}}(\alpha)$. Using that F is homogenous of degree η, we can write its wage bill as

$$w(T^{\mathrm{D}}(\phi)) = \Lambda^{\mathrm{D}}(\alpha, z) w(\tilde{S}^{\mathrm{D}}(\alpha)) = \left[\eta\, z\, F\left(\frac{\tilde{S}^{\mathrm{D}}(\alpha)}{w(\tilde{S}^{\mathrm{D}}(\alpha))}; \alpha\right)\right]^{\frac{1}{1-\eta}}. \tag{A.13}$$

The firm's profit is

$$\begin{aligned}
\Pi(\phi; w) &= (1-\eta)\,(z\eta^{\eta})^{\frac{1}{1-\eta}} \left[F\left(\frac{\tilde{S}^{\mathrm{D}}(\alpha)}{w(\tilde{S}^{\mathrm{D}}(\alpha))}; \alpha\right)\right]^{\frac{1}{1-\eta}} \\
&= (1-\eta)\,(z\eta^{\eta})^{\frac{1}{1-\eta}}\, w(\tilde{S}^{\mathrm{D}}(\alpha)) \left[\frac{F\left(\tilde{S}^{\mathrm{D}}(\alpha); \alpha\right)}{w(\tilde{S}^{\mathrm{D}}(\alpha))}\right]^{\frac{1}{1-\eta}}. \qquad \text{(A.14)}
\end{aligned}$$

All the above quantities depend on the TFP parameter z through $z^{1/(1-\eta)}$.

A.4 Proof of Proposition 4

When there are two skills C and N, the average profile of the workers, θ^{D}, and their total quality, Λ^{D}, satisfy the first-order conditions

$$K_C(\theta^{\mathrm{D}}, \Lambda^{\mathrm{D}}) \stackrel{\mathrm{d}}{=} zF_C(\Lambda^{\mathrm{D}}\cos\theta^{\mathrm{D}}, \Lambda^{\mathrm{D}}\sin\theta^{\mathrm{D}}; \alpha_N) - w_C(\theta^{\mathrm{D}}) = 0 \tag{A.15}$$

$$K_N(\theta^{\mathrm{D}}, \Lambda^{\mathrm{D}}) \stackrel{\mathrm{d}}{=} zF_N(\Lambda^{\mathrm{D}}\cos\theta^{\mathrm{D}}, \Lambda^{\mathrm{D}}\sin\theta^{\mathrm{D}}; \alpha_N) - w_N(\theta^{\mathrm{D}}) = 0. \tag{A.16}$$

where K_C and K_N are the first derivatives of the firm's objective $F(T; \phi) - w(T)$. Differentiating the first-order conditions (A.15) and (A.16) and inverting the Jacobian of K yields

$$\begin{pmatrix} \dfrac{\partial \theta^{\mathrm{D}}}{\partial \alpha_N} & \dfrac{\partial \theta^{\mathrm{D}}}{\partial z} \\ \dfrac{\partial \Lambda^{\mathrm{D}}}{\partial \alpha_N} & \dfrac{\partial \Lambda^{\mathrm{D}}}{\partial z} \end{pmatrix} = -\frac{1}{d} \begin{pmatrix} z\dfrac{\partial F_N}{\partial \Lambda^{\mathrm{D}}} & -z\dfrac{\partial F_C}{\partial \Lambda^{\mathrm{D}}} \\ -\left(z\dfrac{\partial F_N}{\partial \theta^{\mathrm{D}}} - w_N'\right) & z\dfrac{\partial F_C}{\partial \theta^{\mathrm{D}}} - w_C' \end{pmatrix} \begin{pmatrix} z\dfrac{\partial F_C}{\partial \alpha_N} & F_C \\ z\dfrac{\partial F_N}{\partial \alpha_N} & F_N \end{pmatrix}, \tag{A.17}$$

where d is the determinant of the Jacobian of $K = (K_C, K_N)$ in polar coordinates, i.e., the determinant of

$$\begin{pmatrix} \dfrac{\partial K_C}{\partial \theta^{\mathrm{D}}} & \dfrac{\partial K_C}{\partial \Lambda^{\mathrm{D}}} \\ \dfrac{\partial K_N}{\partial \theta^{\mathrm{D}}} & \dfrac{\partial K_N}{\partial \Lambda^{\mathrm{D}}} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial K_C}{\partial s_C} & \dfrac{\partial K_C}{\partial s_N} \\ \dfrac{\partial K_N}{\partial s_C} & \dfrac{\partial K_N}{\partial s_N} \end{pmatrix} \begin{pmatrix} -\Lambda^{\mathrm{D}} \sin\theta^{\mathrm{D}} & \cos\theta^{\mathrm{D}} \\ \Lambda^{\mathrm{D}} \cos\theta^{\mathrm{D}} & \sin\theta^{\mathrm{D}} \end{pmatrix}.$$

By concavity of the firm's problem, the determinant of the first matrix on the right hand side is positive, hence $d < 0$.

To prove the first part of the proposition, we compute the derivative of total quality with respect to total factor productivity

$$\frac{\partial \Lambda^{\mathrm{D}}}{\partial z} = -\frac{1}{d} \left[F_N \left(z\frac{\partial F_C}{\partial \theta^{\mathrm{D}}} - w_C'\right) - F_C \left(z\frac{\partial F_N}{\partial \theta^{\mathrm{D}}} - w_N'\right)\right].$$

Consider the above bracketed terms. The first term $F_C w_N' - F_N w_C' = w_C w_N' - w_N w_C'$ is positive because the w_N/w_C increases with θ^{D} by concavity if the iso-wage curve. The second term $F_N \partial F_C/\partial\theta^{\mathrm{D}} - F_C \partial F_N/\partial\theta^{\mathrm{D}}$ is positive by convexity of the production isoquants. It follows that the bracketed terms is positive and hence that Λ^{D} increases with z.

To prove the second part – the PAM property –, we need to show that the determinant of the sorting matrix is positive and that θ^{D} increases with α_N. Regarding the former point, the determinant of the sorting matrix at left-hand side of (A.17) is positive because by concavity of the firm problem and the Assumption that F_N/F_C increases with α_N the two matrices on the right hand side have a negative determinant. Regarding the latter point, the derivative of the skill profile with respect to technological intensity is

$$\frac{\partial \theta^{\mathrm{D}}}{\partial \alpha_N} = -\frac{z^2}{d} \left[\frac{\partial F_C}{\partial \alpha_N}\frac{\partial F_N}{\partial \Lambda^{\mathrm{D}}} - \frac{\partial F_N}{\partial \alpha_N}\frac{\partial F_C}{\partial \Lambda^{\mathrm{D}}}\right].$$

Hence θ^{D} increases with α_N of and only if

$$\frac{\partial F_C}{\partial \alpha_N}\frac{\partial F_N}{\partial \Lambda^{\mathrm{D}}} - \frac{\partial F_N}{\partial \alpha_N}\frac{\partial F_C}{\partial \Lambda^{\mathrm{D}}} \geq 0. \tag{A.18}$$

It follows from the above analysis that (A.18), together with F_N/F_C increasing in α_N, is a sufficient condition for PAM. Condition (A.18) holds in particular if production isoquants are homothetic. Indeed, we have in this case that $F_C\partial F_N/\partial\Lambda^{\mathrm{D}} = F_N\partial F_C/\partial\Lambda^{\mathrm{D}}$ and hence $(\partial F_C/\partial\Lambda^{\mathrm{D}}, \partial F_N/\partial\Lambda^{\mathrm{D}}) = -\kappa(F_C, F_N)$ for some constant $\kappa > 0$, which, together with F_N/F_C increasing in α_N, guarantees that (A.18) holds.

Non-homothetic isoquants We now provide details about the sorting pattern when production isoquants are non-homothetic, see the discussion in Section 4. From (A.17), we have

$$\frac{\partial \theta^{\mathrm{D}}}{\partial z} = -(z/d)\left\{F_C\frac{\partial F_N}{\partial \Lambda^{\mathrm{D}}} - F_N\frac{\partial F_C}{\partial \Lambda^{\mathrm{D}}}\right\}.$$

where $d < 0$. It follows θ^{D} is independent of z when the production isoquants are homothetic and decreases with z if $\partial(F_C/F_N)/\partial\Lambda^{\mathrm{D}} > 0$. Adapting notations $F_C = F_C$ and $F_N = F_N$ yields the results announced in Subsection 4.1. The latter condition holds for instance for the CES function modified in the spirit of Simonovska (2015) and Jung, Simonovska, and Weinberger (2019):[42]

$$zF(T;\alpha_N) = (z/\eta)\ \left[\alpha_C(T_C+\bar{T}_C)^{\rho} + \alpha_N T_N^{\rho}\right]^{\eta/\rho} - K, \tag{A.19}$$

where $\alpha_C + \alpha_N = 1$ and $\bar{T}_C$ is a positive constant and the constant K ensures that $zF(0,\alpha_N) = 0$. Indeed here

$$\frac{F_C}{F_N} = \frac{\alpha_C}{\alpha_N}\left[\frac{T_N}{T_C+\bar{T}_C}\right]^{1-\rho}$$

and hence F_C/F_N evaluated at $(\Lambda^{\mathrm{D}}\cos\theta^{\mathrm{D}}, \Lambda^{\mathrm{D}}\sin\theta^{\mathrm{D}})$ increases with Λ^{D}.

A.5 Proof of Proposition 5

Let w be an equilibrium wage schedule that is convex and homogenous of degree one. We have, for any firm type ϕ

$$\frac{w\left(\int x\, n^{\mathrm{D}}(\mathrm{d}s;\phi)\right)}{\int w(x)\, n^{\mathrm{D}}(\mathrm{d}s;\phi)} = w\left(\frac{\int [x/w(x)]w(x)\, n^{\mathrm{D}}(\mathrm{d}s;\phi)}{\int w(x)\, n^{\mathrm{D}}(\mathrm{d}s;\phi)}\right) \leq \frac{\int w(x)\, n^{\mathrm{D}}(\mathrm{d}s;\phi)}{\int w(x)\, n^{\mathrm{D}}(\mathrm{d}s;\phi)} = 1. \tag{A.20}$$

[42]See Sato (1977) for a comprehensive study of non-homothetic CES function.

When the iso-wage surface $w = 1$ is strictly concave, the equality in (A.20) imposes that $x/w(x)$ is constant for n^{D}-almost every x, i.e., that all the workers employed by firms of type ϕ have the same skill profile.

Recall that for any measurable map $T : \mathcal{X} \to \mathcal{Y}$, the push-forward of a positive measure μ on $\mathcal{X}$ by T is the positive measure $T_{\#}\mu$ on $\mathcal{Y}$ that satisfies, for all continuous function h on $\mathcal{Y}$

$$(T_{\#}\mu)h = \int_{\mathcal{X}} h(T(x))\mathrm{d}\mu(x).$$

In the particular case of the operator W, we have

$$< W_{\#}H, h >= \int h\left(\frac{x}{w(x)}\right) w(x)\mathrm{d}H(x)$$

for any test function h. It follows that

$$\begin{aligned}
< W_{\#}T_{\#}H^f, h > &= \int_{\phi} h\left(\frac{T^{\mathrm{D}}(\phi)}{w(T^{\mathrm{D}}(\phi))}\right) w(T^{\mathrm{D}}(\phi))H^f(\mathrm{d}\phi) \\
&= \int_{\phi} h\left(\frac{T^{\mathrm{D}}(\phi)}{w(T^{\mathrm{D}}(\phi))}\right) \int_{x} w(x)\mathrm{d}\, n^{\mathrm{D}}(x;\phi)H^f(\mathrm{d}\phi) && \text{(A.21)} \\
&= \iint h\left(\frac{x}{w(x)}\right) w(x)\mathrm{d}\, n^{\mathrm{D}}(x;\phi)H^f(\mathrm{d}\phi) && \text{(A.22)} \\
&= \int_{x} h\left(\frac{x}{w(x)}\right) w(x)H^w(\mathrm{d}s) && \text{(A.23)} \\
&= < W_{\#}H^w, h > .
\end{aligned}$$

Equation (A.21) follows from the equality in (A.20). Equation (A.22) uses that $x/w(x) = T^{\mathrm{D}}(\phi)/w(T^{\mathrm{D}}(\phi))$ for all x in the support of $n^{\mathrm{D}}(\mathrm{d}s;\phi)$, i.e., for all x proportional to $\tilde{S}^{\mathrm{D}}(\alpha)$. Equation (A.23) uses the equilibrium condition (2).

A.6 Proof of Proposition 6

For any convex test function h, we have, using the equality in (A.20) for w and Jensen inequality for h

$$h\left(\frac{T^{\mathrm{D}}(\phi)}{w(T^{\mathrm{D}}(\phi))}\right) = h\left(\frac{\int [s/w(s)]w(s)\, n^{\mathrm{D}}(\mathrm{d}s;\phi)}{w(T^{\mathrm{D}}(\phi))}\right) \leq \frac{1}{w(T^{\mathrm{D}}(\phi))} \int h\left(\frac{s}{w(s)}\right) w(s)\, n^{\mathrm{D}}(\mathrm{d}s;\phi),$$

which yields

$$
\begin{aligned}
< W_\# T_\# H^f, h > &= \int_\phi h\left(\frac{T^{\mathrm{D}}(\phi)}{w(T^{\mathrm{D}}(\phi))}\right) w(T^{\mathrm{D}}(\phi)) H^f(\mathrm{d}\phi) \\
&\leq \iint_\phi h\left(\frac{s}{w(s)}\right) w(s) \mathrm{d}\, n^{\mathrm{D}}(s;\phi) H^f(\mathrm{d}\phi) \\
&= \int h\left(\frac{s}{w(s)}\right) w(s) H^w(\mathrm{d}s) \\
&= < W_\# H^w, h > .
\end{aligned}
$$

A.7 Proof of Proposition 7

Consider a market-clearing assignment n^{D} such that $T^{\mathrm{D}}(\phi)$ is the firm-aggregated skill vector $T^{\mathrm{D}}(\phi) = \int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)$. Because any convex and positively 1-homogenous function is sub-additive, we have

$$
\begin{aligned}
\int h(s) T_\# H^f(\mathrm{d}s) &= \int h(T^{\mathrm{D}}(\phi)) H^f(\mathrm{d}\phi) \\
&= \int h\left(\int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)\right) H^f(\mathrm{d}\phi) \\
&\leq \iint h(s)\, n^{\mathrm{D}}(\mathrm{d}s;\phi) H^f(\mathrm{d}\phi) = \int h(s) H^w(\mathrm{d}s),
\end{aligned}
$$

which proves $T_\# H^f \leq_{\mathrm{phc}} H^w$.

The converse property follows from the new variant of Strassen Theorem established by CGK. Theorem 5.2 in their paper establishes that for any distribution γ more "generalist" than H^w in the sense that $\gamma \leq_{\mathrm{phc}} H^w$, there exists a market-clearing assignment $n^{\mathrm{D}}(\mathrm{d}s;\phi)$ such that $n^{\mathrm{D}}\gamma = H^w$ and $y = \int s\, n^{\mathrm{D}}(\mathrm{d}s;\phi)$ for γ-almost every y. Applying this result to the distribution $\gamma = T_\# H^f$ yields the desired property. The equality (24) follows from Theorem 5.5 of CGK.

B Connection to optimal transport theory

In this section, we explain how our setup is related to optimal transport theory.

Weak optimal transport (WOT) Given two probability measures μ and ν, and a cost function $c(\phi, m)$ that is convex in m, Gozlan, Roberto, Samson, and Tetali (2017) consider the problem of minimizing

$$
\inf_{\pi \in \Pi(\mu,\nu)} \int c(\phi, p^\phi) \mu(\mathrm{d}\phi), \tag{B.1}
$$

where $\Pi(\mu, \nu)$ is the set of all couplings π of μ and ν (i.e., the set of probability measures over $\mathcal{X} \times \mathcal{Y}$ with marginals μ and ν) and p^ϕ is the (μ-almost surely unique) probability kernel such that

$$\pi(\mathrm{d}s, \mathrm{d}\phi) = p^\phi(\mathrm{d}s)\, \mu(\mathrm{d}\phi). \tag{B.2}$$

Gozlan, Roberto, Samson, and Tetali (2017) prove existence and duality results for Problem (B.1) under the main requirement that $c(\phi, m)$ is convex in m.

The problem of maximizing total output in the economy, which is given by (5), has the same form as (B.1), with $\mu = H^f$, $\nu = H^w$, and the transport cost defined (for any given $x_0 \in \mathcal{X}$) by

$$c(\phi,\, n^{\mathrm{D}}) = -F\left(\int s\; n^{\mathrm{D}}(\mathrm{d}s); \phi\right) + F(s_0; \phi) + \nabla_s F(s_0; \phi).\left(\int s\; n^{\mathrm{D}}(\mathrm{d}s) - s_0\right).$$

The above cost function is nonnegative by concavity of F in X. Under the equilibrium condition (2), minimizing (B.1) is equivalent to maximizing (5) because $\iint x\, p^\phi(\mathrm{d}s)\mu(\mathrm{d}\phi)$ equals $\int x\, \nu(\mathrm{d}s)$, which is a fixed and exogenous quantity.

Unnormalized kernels and endogenous firms' sizes As mentioned in Section 2, the framework developed in the present article has an important difference with the WOT problem described above. Specifically, we do not impose that the workers-to-firms assignments, $n^{\mathrm{D}}(\mathrm{d}s; \phi)$, are *probability* measures, as is required in the kernel disintegration (B.2). Accordingly, Choné, Gozlan, and Kramarz (2023) relax the assumption that the kernel p^ϕ in (B.1) is a *probability* measure. Denoting by $\mathcal{M}(\mathcal{Y})$ the set of nonnegative finite measures over $\mathcal{Y}$, they introduce the weak optimal transport problem with unnormalized kernel (WOTUK) as

$$\mathrm{WOTUK}(\mu, \nu) \overset{\mathrm{d}}{=} \sup_{\substack{q \in \mathcal{M}(\mathcal{Y})^{\mathcal{X}} \\ \int q^\phi\, \mu(\mathrm{d}\phi) = \nu}} \int_{\mathcal{X}} F(\phi, q^\phi)\, \mu(\mathrm{d}\phi), \tag{B.3}$$

where $\mathcal{F} : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$. The constraint $\int q^\phi\, \mu(\mathrm{d}\phi) = \nu$ expresses that the unnormalized kernel (q^ϕ) transports μ onto ν. CGK connect the WOTUK problem (B.3) to a WOT problem as follows. Letting

$$\Pi(\ll \mu, \nu) \overset{\mathrm{d}}{=} \{P \in \Pi(\eta, \nu)\,,\, \eta \in \mathcal{P}(\mathcal{X}), \eta \ll \mu\},$$

denote the set of probability measure over $\mathcal{X}$ that are absolutely continuous with respect to μ, CGK show that

$$\text{WOTUK}(\mu,\nu) = \sup_{\eta\in\Pi(\ll\mu,\nu)} \sup_{\pi\in\Pi(\eta,\nu)} \int F\left(\phi, \frac{d\eta}{d\mu}(\phi)\pi^{\phi}\right) \mu(\mathrm{d}\phi) \tag{B.4}$$

where $\pi^{\phi} \in \mathcal{P}(\mathcal{Y})$ is the unique disintegration of π with respect to η, *i.e.* such that $\pi(\mathrm{d}s, \mathrm{d}\phi) = \eta(\mathrm{d}\phi)\pi^{\phi}(\mathrm{d}s)$. At given η, we get a WOT problem. Instead of constraining the first marginal of π to be μ, the WOTUK problem only imposes that the first marginal is absolutely continuous with respect to μ.

In the economic setting of this paper, the kernel $q^{\phi}(\mathcal{Y})$ is denoted by $N(\phi)$ and represents the number of employees of firms with type ϕ. The modified firms' distribution is $\eta = \tilde{H}^f$, and $N(\phi)$ is the density of η with respect to μ. The constraint $\int q^{\phi}\,\mu(\mathrm{d}\phi) = \nu$ is the equilibrium condition (2). Allowing q^{ϕ} to be an unnormalized measure instead of a probability measure avoids having to assume that all firms have the same size.

Conical WOTUK problems The specification studied in the present paper corresponds to a special class of WOTUK problems, which CGK call conical WOTUK problems. It corresponds to the case where

$$\mathcal{F}(x,p) = F\left(x, \int_{\mathcal{Y}} y\, p(\mathrm{d}y)\right)$$

for some $F : \mathcal{X} \times \text{cone}(\mathcal{Y}) \to \mathbb{R}$, where the conical hull of $\mathcal{Y}$ is given by

$$\text{cone}(\mathcal{Y}) \overset{\mathrm{d}}{=} \left\{ \sum_{i=1}^{n} \lambda_i y_i\,,\ \lambda_C, \ldots, \lambda_n \in \mathbb{R}_+, y_C, \ldots, y_n \in \mathcal{Y}, n \geq 1 \right\}.$$

CGK establish the existence of solutions for the dual problem, which guarantee the existence of a competitive equilibria in our setting where a firm's output depends on the *conical* combination of its employees' types, $\int y\, dq_x(y)$. The combination is said to be "conical" because the mass of q_x is not necessarily equal to one. In other words, the aggregate skill of the workers hired by a firm is not their average skills as in the WOT setting, but their average skills *scaled by the positive factor* $q_x(\mathcal{Y})$ that represents the number of employees.

C A Dixit-Stiglitz Environment

In the paper, we assume that the price of the final good is exogenous and normalized to one, and quantities (output, labor demand, etc.) are determined by decreasing returns to scale. We now present a Dixit-Stiglitz environment where firms operate under constant returns to scale and quantities are set by monopolistic competition.[43] We then check that this environment is isomorphic to the model presented in the main text.

For simplicity of exposition, we assume in the following that the workers' skills are two-dimensional. Firms indexed by (α, z) produce a differentiated good under constant returns to scale. The production function takes the form $y(\alpha_N, z) = zF(T_C, T_N; \alpha_N)$, where F is homogenous of degree one. A representative consumer has income I and preferences over baskets $\mathbf{y} = (y(\alpha_N, z))$ given by

$$U(\mathbf{y}) = \left(\int y(\alpha_N, z)^{\frac{\sigma-1}{\sigma}} H^f(\mathrm{d}\alpha_N, \mathrm{d}z) \right)^{\frac{\sigma}{\sigma-1}},$$

with $\sigma > 1$. Let $p(\alpha_N, z)$ denote the price of good (α_N, z). The Marshallian demand is given by

$$y(\alpha_N, z) = I \frac{p(\alpha_N, z)^{-\sigma}}{\int [p(\alpha_N, z)]^{1-\sigma} H^f(\mathrm{d}\alpha_N, \mathrm{d}z)}. \tag{C.1}$$

As in the text, we parameterize aggregate skill vectors as $T^{\mathrm{D}} = (T_C, T_N) = \Lambda^{\mathrm{D}}(\cos\theta, \sin\theta)$, where θ is the firm's aggregate skill profile and Λ^{D} is the aggregate quality of its employees. The wage schedule, denoted $w(s_C, s_N)$, has the same properties as in the main text, namely convexity and one-homogeneity. There is monopolistic competition on the product market and firm (α_N, z) chooses its aggregate skill vector $T = \int s\, n^{\mathrm{D}}(\mathrm{d}s)$ to maximize its profit

$$\begin{aligned} p(\alpha_N, z)y - w(T) &= y \left[p(\alpha_N, z) - \frac{w(T)}{y} \right] \\ &= y \left[p(\alpha_N, z) - \frac{w(\cos\theta, \sin\theta)}{z\ F(\cos\theta; \sin\theta; \alpha_N)} \right] \end{aligned} \tag{C.2}$$

(To derive the last equality above, we use the one-homogeneity of w and F and eliminate Λ^d.) For any aggregate skill profile θ, the firm chooses its aggregate worker quality Λ^{D} (or equivalently its output y) under the demand equation, which yields the standard mark-up condition

$$p(\alpha_N, z) = c(\theta; \alpha_N, z)\ \frac{\sigma}{\sigma - 1}, \tag{C.3}$$

and the output

$$y(\alpha_N, z) = I\ \frac{\sigma - 1}{\sigma}\ \frac{c(\theta; \alpha_N, z)^{-\sigma}}{\int c(\theta; \alpha_N, z)^{1-\sigma} H^f(\mathrm{d}\alpha_N, \mathrm{d}z)}, \tag{C.4}$$

[43] See for instance Costinot and Vogel (2010) with one-dimensional skills.

where

$$c(\theta;\alpha_N,z)=\frac{w(\cos\theta,\sin\theta)}{z\ F(\cos\theta;\sin\theta;\alpha_N)}$$

is the firm's constant marginal cost. The firms' profit is decreasing in the unit cost, hence the aggregate skill profile θ is chosen to minimize the cost:

$$\tilde{c}(\alpha_N,z)=\min_{\theta} c(\theta;\alpha_N,z)=\min_{\theta}\frac{w(\cos\theta,\sin\theta)}{z\ F(\cos\theta;\sin\theta;\alpha_N)}. \tag{C.5}$$

From output (C.4), we obtain the resulting labor demand:

$$\begin{aligned}\Lambda^{\mathrm{D}}(\alpha_N,z) &= \frac{y(\alpha_N,z)}{zF(\cos\theta,\sin\theta;\alpha_N)}\\ &= I\frac{\sigma-1}{\sigma}\frac{1}{\int\tilde{c}(\alpha_N,z)^{1-\sigma}H^f(\mathrm{d}\alpha_N,\mathrm{d}z)}\frac{[zF(\cos\theta,\sin\theta;\alpha_N)]^{\sigma-1}}{w(\cos\theta,\sin\theta)^{\sigma}},\end{aligned} \tag{C.6}$$

and the wage bill:

$$\begin{aligned}w(T) &= \Lambda^{\mathrm{D}}(\alpha_N,z)\,w(\cos\theta,\sin\theta)\\ &= I\frac{\sigma-1}{\sigma}\frac{1}{\int\tilde{c}(\alpha_N,z)^{1-\sigma}H^f(\mathrm{d}\alpha_N,\mathrm{d}z)}\left[\frac{zF(\cos\theta,\sin\theta;\alpha_N)}{w(\cos\theta,\sin\theta)}\right]^{\sigma-1}.\end{aligned} \tag{C.7}$$

The two environments, the one presented in this Appendix and the competitive one from the main text, have deep similarities. In fact, the Dixit-Stiglitz variant is isomorphic to the model presented in the main text. Using the demand equation (C.1), we rewrite the firm's profit (C.2) as

$$p(\alpha_N,z)y-w(T)=C^{1/\sigma}\left[zF(T;\alpha_N)\right]^{1-1/\sigma}-w(T),$$

where C is a constant. We obtain expression (A.11) for the firms' profit used in our main model by replacing z with $z^{1-1/\sigma}$ and the one-homogenous function F with $C^{1/\sigma}F^{1-1/\sigma}$, which is homogenous of degree $\eta=1-1/\sigma$. Equations (C.6) and (C.7) correspond to Equations (A.12) and (A.13) after the above replacements of z and F. The labor demand elasticity is $\sigma>1$ here and $1/(1-\eta)>1$ in the main text. The equilibrium conditions (19) and (20) must be modified according to (C.6) and (C.7). Notice that the underlying primal problem maximizes pF, the gross revenue of firms to be shared with workers, and ignores the surplus of final consumers.

All results obtained under Bundling have their counterpart in the DS world despite differences in the resulting formulas. In particular, the logic of the horizontal matching (link between worker's skill profiles θ and firms' technological intensity α_N) is unchanged. The determination of the aggregate profile θ from (C.5) is the same as in the main text, see for instance Figure 1. The analysis of bunching is also similar as well as the AKM-like decomposition of the log-wage. Furthermore, the analysis of unbundling, presented in our companion paper, can also be carried out within the DS setting.

D Technological Shocks under Bundling

In this Appendix, we study the impact of a positive productivity shock that affects firms differently, depending on their technological profile (α).

Skill-Biased Technological Change (SBTC) was seen by many as an essential driving force of labor markets transformations (wage inequality, in particular) over the 80s and the 90s, in the context of an increased supply of educated workers in the US.[44] We study this eighties/nineties style environment faced with such technical change in the following.

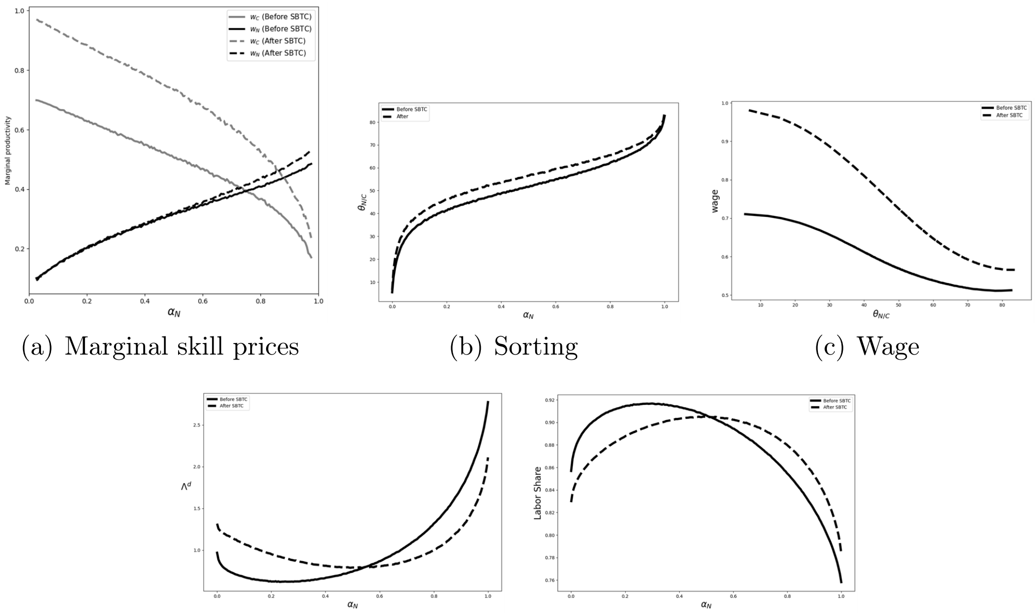
Workers supply their skills as a bundle, the quantity being fixed, but are faced with a SBTC shock affecting the productivity of cognitive tasks in their employing firms, in a labor market where cognitive skills are in (relative) short supply before the shock. Because of this SBTC shock, with or without an education shock, at the new equilibrium, prices (wages of each skill) change, hence the sorting of workers to firms changes, as well as the size of each firm. Following insights from Simonovska (2015), the production function we use to study SBTC is non-homogenous :

$$F(T;\alpha,z) = \frac{z}{\eta}\left[\left((\alpha_N T_N^{\rho} + \alpha_C T_C^{\rho})^{1/\rho} + A\right)^{\eta} - A^{\eta}\right], \tag{D.1}$$

with $A = 1$. The two skills are complements ($\rho = -1$) and returns to scale are decreasing ($\eta = .5$). We assume that the technical parameter α_N is uniformly distributed on $[0,1]$. In this Section, cognitive skills are relatively rare: workers' skill profiles $\theta = \arctan s^N/s^C$ are distributed as a Beta(7,4) random variable on $(0,\pi/2)$. There is no heterogeneity in workers' quality ($\lambda = 1$) or in firms' total factor productivity ($z = 1$). In the following numerical simulation, we use the mirror-descent algorithms developed in Paty, Choné, and Kramarz (2022).

We now consider a positive productivity shock that affects cognitive firms, starting from the environment described above. The firms' total factor productivity parameter

[44] See Katz and Murphy (1992) is an important view on SBTC, see Card and DiNardo (2002) for a thorough discussion, but Lemieux (2006) shows the role of the secular increase in education that took place there in explaining increasing inequality without a large role for SBTC. However, Bittarello, Kramarz, and Maitre (forthcoming) show that, in France, returns to cognitive skills decreased between 1990 and 2013 *because* of the increase in the supply of college-educated workers.



(a) Marginal skill prices (b) Sorting (c) Wage

(d) Firm's sizes (e) Labor share

Figure D.1: Skill-biased technical change. Cognitive firms experience a positive productivity shock

z increases from $z = 1$ to $z = 1 + (1 - \alpha_N)/2$. Purely non-cognitive firms, with $\alpha_N = 1$, are unaffected. Purely cognitive firms ($\alpha_N = 0$) have $z = 1.5$ after the shock.

The first-order effect of this SBTC shock is, obviously, an increase in the price of cognitive skills paid by all firms, in particular those most in demand for such skills (left side of Figure D.1(a)). As the price of cognitive tasks increase, firms use and demand more non-cognitive tasks, affecting sorting (Figure D.1(b)). This induced increase in demand for non-cognitive tasks in turn impacts the price of non-cognitive skills (see Figure D.1(a), right part). Wages are an outcome of both sorting and skills' prices. Indeed, Figure D.1(c) shows that following the SBTC all wages – including those of non-cognitive workers – increase because firms not only use and demand more cognitive tasks but also non-cognitive ones. Importantly, our SBTC shock increases wage inequality. Cognitive firms increase in size, as shown in Figure D.1(d) even though the price of cognitive skills increases because the productivity (direct) effect of the increases in the TFP parameter, z, dominates and outweighs the price effect. Similarly, the labor share decreases for the most cognitive firms (Figure D.1(e)) even though the corresponding wage increases: essentially because firms produce more (are more efficient), i.e., the direct effect of the increased productivity, z, which is included within the denominator of the labor share – wL/F – dominates.

As we just saw, SBTC affects the sorting of workers to firms. Hence, to accomodate this changing sorting, workers will move from firm-to-firm. More precisely, workers with a skill profile θ will move from firms with $\alpha^{-1}_{before}(\theta)$ to firms with $\alpha^{-1}_{after}(\theta)$. Indeed, workers' mobility is a feature of all shocks studied in the Choné, Gozlan, and Kramarz (2024) whenever the sorting of workers to firms is altered. Adding mobility costs will dampen the associated magnitude.[45]

A long line of research, briefly mentioned above, tries to identify how SBTC affects labor market outcomes, resulting potentially in increased wage inequality, in the context of a growing supply of college-educated workers, as happened in the US over the 80s or in France over the 90s and 2000s. In most studies, such as Katz and Murphy (1992) or Lemieux (2006), the potential role of skills-bundling was never evoked. Recently, Lindenlaub (2017) filled this gap by studying how bundling alters previous conclusions. Indeed, and fully in line with what we do here, Lindenlaub (2017) focuses on Task-Biased Technological Change (TBTC, using her words) within a Quadratic-Gaussian model with bi-linear production function, delivering a closed-form solution to her problem. Whereas we increase the productivity of firms having a "preference" for cognitive tasks, she models TBTC by decreasing the relative efficiency of the manual task w.r.t. the cognitive task. In this environment, she presents sufficient conditions for wage inequality to increase. Her work is important for us because it shows how bundling (for jobs in her approach) has implications that widely differ from those obtained in a fully unbundled world. Our analysis differs from hers because we define SBTC (or TBTC) in a world with firms (rather than jobs). Indeed, we are able to predict changes in firms' sizes and labor shares after SBTC.

We could also contrast SBTC without and with an increase in the supply of college-educated workers (as took place in the US or in France, as examined in Bittarello, Kramarz, and Maitre, forthcoming), increasing as a result the supply of cognitive skills. Hence, in comparison with our previous Figures, the price of cognitive skills will increase less. As a consequence, all the effects described will still exist but will be attenuated, potentially resulting in an absence of impact on wage inequality (as suggested in Lemieux, 2006).[46]

[45]Mobility is likely to be too large without such costs when compared to year-to-year realized moves seen in data. But estimated costs in structural analyses tend to be unrealistically large too.

[46]It is hard, not to say impossible, to go beyond this limited observation without structurally estimating our model, an endeavor beyond the purpose of the present article.