Bunching and Taxing Multidimensional Skills*

Job Boerma
University of Wisconsin-Madison

Aleh Tsyvinski
Yale University

Alexander P. Zimin
Broad Institute

September 2026

Abstract

We characterize optimal policy in a multidimensional nonlinear taxation model with bunching. We develop an empirically relevant model with cognitive and manual skills, firm heterogeneity, and labor market sorting. We first derive two conditions for the optimality of taxes that take into account bunching. The first condition − a stochastic dominance optimal tax condition − shows that at the optimum the schedule of benefits dominates the schedule of distortions in terms of second-order stochastic dominance. The second condition − a global optimal tax formula − provides a representation that balances the local costs and benefits of optimal taxation while explicitly accounting for global incentive constraints. Second, we use Legendre transformations to represent our problem as a linear program. This linearization allows us to solve the model quantitatively and to precisely characterize bunching. At an optimum, 10 percent of workers are bunched. We introduce two notions of bunching – blunt bunching and targeted bunching. Blunt bunching constitutes 30 percent of all bunching, occurs at the lowest regions of cognitive and manual skills, and lumps the allocations of these workers, resulting in a significant distortion. Targeted bunching constitutes 70 percent of all bunching and recognizes the workers' comparative advantage. The planner separates workers on their dominant skill and bunches them on their weaker skill, thus mitigating distortions along the dominant skill dimension.

*We thank Mark Aguiar, Dirk Bergemann, Hector Chade, George Mailath, Paolo Martellini, Christian Moser, Emmanuel Saez, Florian Scheuer, Andrew Shephard, Chris Taber and especially Jean-Charles Rochet for detailed insightful comments.

1 Introduction

We make significant progress in analyzing multidimensional optimal nonlinear income taxation problems with bunching. This is one of the important open questions in the theory and practice of optimal taxation. Our paper is the first to solve for optimal multidimensional taxation with bunching in an empirically relevant model of wage determination.

The key difficulty of analyzing multidimensional optimal tax problems lies in characterizing regions of bunching. Bunching occurs when workers of different types receive identical allocations. Kleven, Kreiner, and Saez (2009) establish the importance of bunching in a model of couples taxation in which one partner makes only an extensive margin labor supply decision. Little is known about optimal taxation and the nature of bunching in more general settings. At the same time, a large literature in labor economics emphasizes the importance of multidimensional skills and labor market sorting in determining wage dispersion.

We develop an empirically relevant model that incorporates three important elements of wage dispersion. Workers differ in both manual and in cognitive skills, firms differ in productivity, and workers' output depends on the firms they work for and coworkers they work with. We characterize equilibrium for the positive model in closed form and use this closed-form solution to identify the underlying multidimensional skill distribution.

The characterization of optimal taxes in our model is based on two main theoretical insights. First, we derive two conditions for optimal taxes that take into account bunching. The first condition − a stochastic dominance optimal taxation condition − shows that at the optimum the benefits and the costs are not necessarily equated for each skill level but rather the entire schedule of benefits dominates the entire schedule of distortions in terms of second-order stochastic dominance. The second condition − a global optimal tax formula − provides a representation that balances the local costs and benefits of optimal taxation while explicitly accounting for global incentive constraints. These optimal tax conditions generalize the classic unidimensional optimal taxation conditions in Mirrlees (1971), Diamond (1998), and Saez (2001) to a multidimensional optimal taxation problem, accounting for global incentive constraints and bunching. Second, we use Legendre transforms to represent our problem as a linear program. Legendre transforms are a powerful tool from convex analysis that allows us to represent a convex function by a family of its tangent lines. This linearization enables us to fully solve the model quantitatively and, in particular, precisely characterize the patterns of bunching.

We find that 10 percent of all workers are bunched at the optimal allocation for our empirically

estimated economy. We show that workers are bunched with other workers who are better in one dimension but worse in the other dimension. Moreover, a sizable portion of bunching is nonlocal.

We introduce two notions of bunching: blunt bunching and targeted bunching. Blunt bunching occurs at the lowest regions of cognitive and manual skills. The planner does not distinguish workers' cognitive skills from their manual skills and lumps their allocations together by an index of their skills. This is a blunt tool for providing incentives as it creates significant distortions leading to high marginal taxes − 30 percent of all bunching is blunt. Targeted bunching recognizes the workers' comparative advantage. The planner separates workers on their dominant skill while bunching them on their weaker skill − 70 percent of all bunching is targeted.

We now discuss our model and results in more detail. We consider the optimal policy problem which we formulate as a mechanism design problem (Mirrlees, 1971). The planner chooses consumption allocations, allocations of cognitive and manual tasks, and the assignment of workers to coworkers and firms subject to incentive constraints that workers truthfully report their type. The primary difficulty in analyzing multidimensional optimal taxation problems is in characterizing regions of bunching. Rochet and Choné (1998) shows that bunching is generic in the multidimensional multiproduct monopolist problem that is closely related to the multidimensional optimal taxation problem. McCann, Rankin, and Zhang (2026) focuses on the free boundary separating bunched from unbunched consumers to provide the most complete characterization of the multidimensional multiproduct monopolist problem with a quadratic technology and types that are uniformly distributed on a square.[1] They further prove that as the support is moved away from the origin, targeted bunching appears before the blunt region is present. An important paper by Kleven, Kreiner, and Saez (2009) solves a multidimensional optimal taxation model under the restriction that one of the allocations is binary and argues for the importance of bunching in their setup. In our model, both the allocations of the manual and cognitive tasks are instead unrestricted. For this general setup, no characterization of optimal tax policy is known.

Our first theoretical result is the derivation of two optimality conditions that take into account the regions of bunching. First, we show that, at the optimum, the utility and revenue benefits from the entire schedule of taxes second-order stochastically dominate the costs of distortions they induce. Without bunching, this tradeoff is made locally and leads the planner to equalize the marginal benefits and costs of taxes as in a unidimensional problem. We show that, when

[1] McCann and Zhang (2025) shows that the solution to this problem features blunt and targeted bunching. Both papers credit our paper for introducing the notions of blunt and targeted bunching. The regularity of the boundary between the targeted bunching region and the non-bunched region for this setting was established in Chen, Figalli, and Zhang (2026) and McCann, Rankin, and Zhang (2026).

there is bunching, the planner no longer equates the benefits and the costs of the taxes at each worker skill level but instead requires that the entire schedule of the benefits of taxes second-order stochastically dominate the entire schedule of distortions, making this tradeoff nonlocal. Second, we derive an optimal tax formula for the multidimensional taxation problem that holds with equality. We show that the local tradeoffs have to be augmented with an additional term that accounts for global incentive constraints.[2] Both formulas are new to the literature on optimal taxation.

Our next main theoretical insight is to transform the planner problem to a linear program. This is a key step that enables computation of the bunching regions. Legendre transformations linearize a convex function by replacing it with the upper envelope of all its tangent lines. The Legendre transform allows us to translate the multidimensional optimal taxation problem into a linear program that can be analyzed quantitatively with high precision. Numerical precision is not merely a technical curiosity but is essential to identify the regions and nature of bunching.

A parallel starting point of our analysis is a characterization of the equilibrium in a positive economy. In our positive economy, workers choose the amount of cognitive and manual tasks to deliver, coworkers to work with, and firms to work for. This problem integrates endogenous labor supply decisions with the assignment of multidimensional workers to teams and to heterogeneous firms. Our first result for the positive economy is that workers sort with identical coworkers (self-sorting) and that better teams work on more valuable projects (positive sorting). The resulting assignment is qualitatively identical to the assignment under the optimal policy problem but the exact assignment differs because of differences in labor supply due to incentive constraints.

We use the dual formulation of the equilibrium assignment problem to characterize equilibrium wages. We show that wages are a convex function of an index of the worker's task inputs rather than a function of each task individually. We further establish an exact mapping between curvature in the wage schedule and the distribution of firm productivity. By choosing a parametric convex function, we can then infer a distribution of firm projects such that this convex function is the equilibrium wage schedule.

Having developed the theory to characterize both positive and optimal allocations, we bring our theory to the data. In order to quantify cognitive and manual skill heterogeneity in the U.S.

[2]When there is no bunching, a classic pointwise optimality condition holds that can be rewritten in terms of a multidimensional ABC taxation formula similar to the unidimensional tax formula in Diamond (1998) and Saez (2001). The absence of bunching is equivalent to the indirect utility function being strongly convex and the first-order approach being valid. Kleven, Kreiner, and Saez (2006, p. 23) derive such a multidimensional ABC formula without bunching.

population, we use earnings information for all workers between 2000 and 2019 in the American Community Survey (ACS). We combine the earnings data from the ACS with data from O*NET on the manual and cognitive task intensity for every occupation (Acemoglu and Autor, 2011).

We use our closed-form characterization for wages in the positive economy to pointwise identify the level of manual and cognitive tasks completed by each worker. We then use the worker's optimality condition for each task together with these inferred task levels to identify levels of cognitive and manual skill that deliver each worker's wage and relative task intensity in the cross-sectional data as a model outcome. For the unidimensional taxation problem, an important contribution of Saez (2001) was to infer the underlying productivity distribution using earnings data which then becomes a central input to determine optimal taxes. Our identification generalizes these findings and delivers the distribution of manual and cognitive skills in a model accounting for multiple drivers of earnings (multidimensional skills, coworkers, firms). Our identification resembles Boerma and Karabarbounis (2020, 2021) who use explicit solutions for home production models to identify productivity at home and market productivity using data on consumption, home and market hours.

We next turn to the quantitative characterization of the optimum using the inferred skill distribution. In order to understand our quantitative characterization, we first describe a benchmark without incentive constraints. Due to separability of preferences and technology over tasks, the efforts in a given task depend exclusively on the worker's skills in this task and, hence, there is no cross-dependence between tasks. Trivially, there is no bunching and there are no distortions.

In sharp contrast to the benchmark, optimal task intensity in our model depends positively on both of the worker's skills. Workers with high manual skills also deliver high levels of cognitive tasks. This codependence is lower at the top end of the skill distribution. In the limit, workers face zero distortion in their manual task allocation at the top of the manual skill distribution, meaning there is no dependence of their manual task intensity on their cognitive skills as in the benchmark. At the lower end of the skill distribution, the distortion from this positive codependence is high.

A central part of our contribution is to characterize patterns of bunching. We first show that 10.4 percent of all workers are bunched at the optimum. Workers bunch with other workers both near and afar. Moreover, workers exclusively bunch with workers that are better in one skill dimension but worse in another. Workers do not bunch with workers over whom they have an absolute advantage nor with workers who have an absolute advantage over them.

We introduce two notions of bunching: blunt bunching and targeted bunching. In the blunt

bunching region, the planner requires all workers with the same effective skill index to deliver identical cognitive and manual tasks, and thus bunches workers that vastly differ in their skills. This is a blunt way to provide incentives and comes at a cost of significant output distortions. In the targeted bunching region, the planner recognizes the increasing efficiency costs of distorting higher skill workers. When workers have a higher relative level of, for example, manual skills, they are separated along this dimension but are bunched on their relatively low cognitive skill. The planner thus separates according to workers' comparative advantage and bunches workers by comparative disadvantage. In contrast to the blunt bunching region, targeted bunching occurs with workers who are more similar in skills: not too far away yet still nonlocally. In the region without bunching, the planner distorts allocations as in the unidimensional case.

We summarize the bunching patterns by describing the tax wedges they induce. In particular, we find that the level of tax wedges is high in the bunching regions. The tax wedges are particularly high for low-skill workers who are bluntly bunched and are also high along the dimension of comparative disadvantage for the more skilled workers in the targeted bunching region. We further show that the optimum is implementable by a tax function that is only a function of earnings and line of work.

Literature. We now describe related literature. Kleven, Kreiner, and Saez (2009) is the first paper that analyzed optimal multidimensional taxation with bunching. They model a binary labor supply choice for the secondary earner along with continuous labor supply choice for the primary earner. Judd, Ma, Saunders, and Su (2017) consider numerically some cases of optimal taxation with multiple dimensions of heterogeneity (up to five dimensions of heterogeneity with five individual types) and find that some nonlocal constraints bind. The most ambitious attempt to date to solve a multidimensional policy problem with bunching is Moser and Olea de Souza e Silva (2019) for a model where workers are heterogeneous in two dimensions but only one dimension of heterogeneity enters the planner's objective. Their key ingredient is paternalistic preferences, which delivers bunching due to disagreement between the planner and workers. In their environment bunching is optimal and, in fact, an essential feature even for the unidimensional problem. The fact that the planner cares only about one dimension of heterogeneity significantly reduces the complexity of deviations patterns. They characterize the model theoretically with the continuous skill distributions and also compute the model with six impatient types in one dimension and a large number of types in the second dimension. In our paper and, more broadly, for multidimensional optimal nonlinear taxation problems the planner cares about heterogeneity in several

dimensions and, hence, the deviations and bunching patterns are significantly more complicated and nuanced, especially when a large number of types within each skill dimension is analyzed. Heathcote and Tsujiyama (2026) comprehensively analyze computational performance of different algorithms for unidimensional optimal taxation. They show that the number of skill types is not just a technical detail but has an important impact on policy prescriptions. In our settings, the need for fine skill differentiation in both dimensions of heterogeneity is additionally important to recover the nuanced patterns of bunching and deviating, especially in the regions of targeted bunching. More broadly, there is a vast literature on multidimensional mechanisms (e.g., McAfee and McMillan (1988), Armstrong (1996) and Rochet and Choné (1998)) that also emphasizes the complexity, as well as the central role, of bunching for the optimal solutions.

An important strand of papers in Scheuer (2014), Rothschild and Scheuer (2013, 2014, 2016) analyzes optimal taxation with multidimensional heterogeneity. These papers achieve tractability by transforming the multidimensional problem into a unidimensional screening problem with an endogenous wage distribution. Rothschild and Scheuer (2014, 2016) in the multidimensional case and Scheuer and Werning (2017) also emphasize the importance of labor market sorting. Lehmann, Renes, Spiritus, and Zoutman (2021) and Golosov and Krasikov (2025) use a first-order approach to theoretically and numerically study multidimensional optimal taxation for a specification where there is no bunching.[3]

A complementary approach is to analyze optimal policy in economies with multidimensional heterogeneity by restricting taxes to parametric families. The most comprehensive recent analysis using this approach is Blundell and Shephard (2012) on optimal taxation of low-income families and Gayle and Shephard (2019) on optimal taxation of couples. Notable papers that use such a parametric approach in a variety of other areas of optimal taxes are, for example, Benabou (2002), Conesa, Kitao, and Krueger (2009), Heathcote, Storesletten, and Violante (2017). Heathcote and Tsujiyama (2021) synthesize the Mirrleesian approach and the parametric approach to optimal taxation.

Our positive wage determination model relates to a growing literature in labor economics that adopts a task approach to understand the contribution of multidimensional skills to labor market outcomes. Recent prominent examples in this area include Yamaguchi (2012), Sanders and Taber (2012), Lindenlaub (2017), Deming (2017), Guvenen, Kuruscu, Tanaka, and Wiczer (2020),

[3]Golosov and Krasikov (2025) illustrates that there is no bunching if the multidimensional type distribution is independent, the Pareto weights are additively separable and the Lagrange multiplier on the government budget constraint is equal to one. Under these restrictions, the multidimensional optimal taxation problem separates into two independent unidimensional taxation problems (page 16 of Golosov and Krasikov (2025)).

Lise and Postel-Vinay (2020), Roys and Taber (2022), and Lindenlaub and Postel-Vinay (2023). Differently from these papers, we combine multidimensional skill heterogeneity with sorting into worker teams and sorting with heterogeneous firms.

2 Environment

We consider an economy with two-dimensional worker skill heterogeneity and heterogeneous firms. Worker output depends not only on own cognitive and manual efforts but also on the coworker they work with and the firm they work for − as emphasized in the modern literature on wage determination.

Workers. The economy is populated by a measure two of workers who differ in two unobservable characteristics. Workers are endowed with a bundle of cognitive and manual skills $\alpha = (\alpha_c, \alpha_m) \in A$. The distribution over types α is denoted by Φ.

Workers have preferences over consumption c and experience disutility from effort in cognitive and manual activities $\ell = (\ell_c, \ell_m)$:

$$U(c, \ell) = u(c) - v(\ell_c) - v(\ell_m), \tag{1}$$

where consumption c and effort ℓ are positive, utility is increasing and concave in consumption, and decreasing and strictly concave in cognitive and manual efforts. We further assume disutility has the form:

$$v(\ell) = \kappa \ell^{\rho}, \tag{2}$$

with $\rho > 2, \kappa > 0$.

Technology. Cognitive and manual production input $(x_c, x_m) \in \mathbf{X}$ for a worker are the product of their skills and their efforts:

$$x_s = \alpha_s \ell_s, \tag{3}$$

for all tasks $s \in S = \{c, m\}$. The worker's skill is given by α, while their effort is given by ℓ.

The economy is populated by a unit mass of heterogeneous firms that produce a single output by organizing two workers into a team to work on a project z. Firm production is represented by y. We use a bilinear team technology together with a multiplicative firm technology:[4]

$$y(x_1, x_2, z) = z\left(x_{1c}x_{2c} + x_{1m}x_{2m}\right). \tag{4}$$

[4]The bilinear technology is also used in Tinbergen (1956), Lindenlaub (2017), Lise and Postel-Vinay (2020), Lindenlaub and Postel-Vinay (2023), and Boerma, Ottolini, and Tsyvinski (2025), among others.

Assignment. An assignment pairs workers with coworkers and projects. Formally, an assignment is a probability measure γ over workers, coworkers, and firms. Given a distribution of worker inputs F_x, a distribution of coworker inputs F_x, and a distribution of firms F_z, the set of feasible assignments is $\Gamma := \Gamma(F_x, F_x, F_z)$. This is the set of probability measures on the product space $\mathbf{X} \times \mathbf{X} \times Z$ such that the marginal distributions of γ onto the set of workers and coworkers $\mathbf{X}$ are F_x, and the marginal distribution of γ onto the set of firms Z is F_z. The assignment function captures the measure of workers that work together on a project as $\gamma(x_1, x_2, z)$. Given a feasible assignment total output is $\int y(x_1, x_2, z)\mathrm{d}\gamma$.

Resources. Aggregate output and external resources R are allocated to workers to consume:

$$\int y(x_1, x_2, z)\mathrm{d}\gamma + R \geq \int c(\alpha)\mathrm{d}\Phi, \tag{5}$$

where $\int c(\alpha)\mathrm{d}\Phi$ is aggregate consumption.

3 Planning Problem

In this section, we formulate the planning problem. The planner problem is to choose an allocation $\{(c(\alpha), x_c(\alpha), x_m(\alpha))\}_{\alpha \in A}$, where consumption and task inputs are both positive, and an assignment $\gamma \in \Gamma$ to minimize the resource costs of providing welfare $\mathcal{U}$. In Appendix A.1, we characterize the optimal assignment of workers into teams and workers with projects. We show that given an optimal assignment, the planner chooses allocations to minimize:

$$\int \Big(c(\alpha) - \frac{1}{2} z(\alpha) \big(x_c^2(\alpha) + x_m^2(\alpha) \big) \Big) \mathrm{d}\Phi. \tag{6}$$

subject to incentive constraints for all workers $\alpha \in A$, so that workers do not gain by misreporting types to be $\hat{\alpha} = (\hat{\alpha}_c, \hat{\alpha}_m)$:

$$U(c(\alpha), x_c(\alpha)/\alpha_c, x_m(\alpha)/\alpha_m) = \max_{\hat{\alpha} \in A} U(c(\hat{\alpha}), x_c(\hat{\alpha})/\alpha_c, x_m(\hat{\alpha})/\alpha_m), \tag{7}$$

and the promise keeping condition:

$$\int U(c(\alpha), x_c(\alpha)/\alpha_c, x_m(\alpha)/\alpha_m)\mathrm{d}\Phi \geq \mathcal{U}, \tag{8}$$

which requires that aggregate welfare exceeds promised value $\mathcal{U}$.[5]

[5]The planning problem is equivalent to maximizing utilitarian welfare function subject to the resource constraint and the incentive constraints (7). There are no incentive constraints for firms since we assume firm output y and inputs x_1 and x_2 are observed by the planner. Hence, firm productivity z is not private information.

3.1 Utility Allocations

We next transform the planner problem from choosing consumption and task allocations to choosing consumption utility and labor disutility allocations.

For each task $s \in S$, we define the skill parameter $p_s = \kappa \alpha_s^{-\rho}$ so the skill parameter is inversely related to the underlying skill α_s. The implied distribution function for the skill parameter vector p is denoted π, and the corresponding project is denoted $z(p)$. We use this skill transformation to define a worker's utility from consumption as a function of their skill vector as $c(p) := u(c(\alpha))$. Following this definition, the resource cost of consumption utility is $\mathcal{C}(c(p)) = u^{-1}(c(p))$. Since the utility from consumption is strictly increasing and concave in the consumption allocation, the resource cost is strictly increasing and convex in consumption utility. Similarly, we define labor disutility in each task $s \in \mathcal{S}$ as a function of the transformed skill parameter p as $x_s(p) := x_s(\alpha)^\rho$. The resource cost of providing disutility $\mathcal{X}(x_s(p)) := -\frac{1}{2} x_s(p)^{\frac{2}{\rho}}$ is decreasing and strictly convex in labor disutility for $\rho > 2$.

Given the introduction of the skill parameter p and the transformation of the choice variables from allocations to utilities, the planner chooses $\{(c(p), x_c(p), x_m(p))\}$ to minimize the resource cost of providing welfare:

$$\int \big(\mathcal{C}(c(p)) + z(p)\big(\mathcal{X}(x_c(p)) + \mathcal{X}(x_m(p))\big)\big)\, \pi(p) \mathrm{d}p, \tag{9}$$

subject to linear incentive constraints:

$$c(p) - p_c x_c(p) - p_m x_m(p) \geq c(q) - p_c x_c(q) - p_m x_m(q), \tag{10}$$

for all workers (p, q), and a linear promise keeping condition:

$$\int (c(p) - p_c x_c(p) - p_m x_m(p)) \pi(p) \mathrm{d}p \geq \mathcal{U}. \tag{11}$$

3.2 Incentive Compatibility

We show that the indirect utility for workers is convex and decreasing in type $p = (p_c, p_m)$. The indirect utility is defined as:

$$u(p) = c(p) - p_c x_c(p) - p_m x_m(p), \tag{12}$$

which implies that for any incentive compatible allocation $\nabla u(p) = -x(p) = -(x_c(p), x_m(p))^T$ almost everywhere. Using the indirect utility function, the incentive constraints (10) are $u(p) \geq$

$u(q) - (p_c - q_c)x_c(q) - (p_m - q_m)x_m(q)$ or, equivalently in notation of scalar product $\langle \cdot, \cdot \rangle$:

$$u(p) - u(q) \geq \langle p - q, -x(q) \rangle = \langle p - q, \nabla u(q) \rangle, \tag{13}$$

for the incentive constraint where worker type p does not want to report to be of type q.

A differentiable function u on a convex domain is convex if and only if $u(p) \geq u(q) + \langle p - q, \nabla u(q) \rangle$. This implies that an incentive compatible indirect utility function is necessarily convex. Since the gradient of the indirect utility function is the negative of a worker's production disutility, and production disutility is positive, the indirect utility function decreases in p, or $\nabla u(p) = -x(p) \leq 0$.[6]

Lemma 1. Any indirect utility function (12) that is incentive compatible is convex and decreasing in worker type p.

We denote the set of utility allocations that satisfy the set of incentive constraints by $\mathcal{I}$, which we refer to as feasible allocations.

3.3 Bunching

We refer to bunching as different workers being assigned the same labor allocation x, and therefore the same consumption allocation c. We label the set of bunching types by $\mathcal{B}$.[7]

Definition. Worker p is bunched, $p \in \mathcal{B}$, if and only if in any neighborhood around this worker there exist two other workers p' and p'' with identical allocations $x(p') = x(p'')$.

We now state the notions of convexity and strong convexity. Assume that the indirect utility is twice continuously differentiable in the neighborhood of a type p. An indirect utility function u is convex if and only if the Hessian matrix $H(u)$ is positive semidefinite. The indirect utility function is strongly convex if $H(u) - \alpha_I I$ is positive semidefinite for some strictly positive α_I, where I is the identity matrix.

Lemma 2. If the indirect utility (12) is strongly convex, then there is no bunching. If the indirect utility is not strongly convex at all points in the neighborhood of type p, then worker p is bunched.

The proof is in Appendix A.3.

[6] The indirect utility function thus increases in skill α. In Appendix A.2, we discuss differentiability of the indirect utility function in more detail, and we also establish which incentive compatibility constraints are redundant.

[7] Alternatively, one could define a worker p being bunched when there exists another worker p' in its neighborhood such that $x(p) = x(p')$. Our definition of bunching is the closure of this set. While these definitions are economically equivalent, our definition facilitates the presentation of Proposition 3.

3.4 Taxation

In order to describe optimal distortions, we define tax wedges for each task. The tax wedge captures the difference between a worker's marginal rate of substitution between task x_s and consumption c, $v'\big(\frac{x_s(\alpha)}{\alpha_s}\big)\frac{1}{\alpha_s}/u'(c(\alpha))$, and the marginal rate of transformation, $zx_s(\alpha)$. We define the tax wedge as:

$$1-\tau_s := \frac{v'\big(\frac{x_s(\alpha)}{\alpha_s}\big)\frac{1}{\alpha_s}}{u'(c(\alpha))}\Bigg/\big(zx_s(\alpha)\big) = -\frac{p_s}{z}\frac{\mathcal{C}'(c(p))}{\mathcal{X}'(x_s(p))}, \tag{14}$$

where it follows from the inverse function theorem that $\mathcal{C}'(c(p)) = 1/u'(c(\alpha))$. A positive wedge plays a role of an implicit tax on marginal income on task s.[8]

4 Characterization

We next derive an optimality condition for the multidimensional tax problem that incorporates bunching.

4.1 Implementability Condition

The planner chooses consumption utility and labor disutility (c, x) to minimize the Lagrangian:

$$\mathcal{L}(c,x) = \int \Big(\mathcal{C}(c) + z\big(\mathcal{X}(x_c)+\mathcal{X}(x_m)\big) - \lambda\big(c - p_c x_c - p_m x_m - \mathcal{U}\big)\Big)\pi \mathrm{d}p. \tag{15}$$

subject to the incentive constraints (10), where λ denotes the multiplier on the promise keeping constraint (11).[9]

Proposition 1. *Implementability Condition.* Let (c, x) denote a solution to the planner problem, then the implementability condition:

$$\int \big(\mathcal{C}'(c)\hat{c} + z\big(\mathcal{X}'(x_c)\hat{x}_c + \mathcal{X}'(x_m)\hat{x}_m\big)\big)\pi \mathrm{d}p \geq \lambda \int \big(\hat{c} - p_c\hat{x}_c - p_m\hat{x}_m\big)\pi \mathrm{d}p \tag{16}$$

holds for any feasible allocation $(\hat{c}, \hat{x}) \in \mathcal{I}$. At a solution $(\hat{c}, \hat{x}) = (c, x)$, (16) holds with equality.

The proof is in Appendix A.4. Proposition 1 states that for any feasible allocation $(\hat{c}, \hat{x})$, the implementability condition is necessarily satisfied, where the marginal resource costs of providing consumption utility $\mathcal{C}'(c)$, as well as the marginal resource costs of providing disutility from work

[8] Using the definition for the tax wedge, we write $\frac{\tau_s}{1-\tau_s} = -\frac{z\mathcal{X}'(x_s(p))+p_s\mathcal{C}'(c(p))}{p_s\mathcal{C}'(c(p))}$.

[9] We suppress the dependence on p to streamline notation.

$(\mathcal{X}'(x_c), \mathcal{X}'(x_m))$, are evaluated at an optimum. Thus, the implementability condition places restrictions on the optimal (c, x) that need to satisfy (16) for any feasible allocation $(\hat{c}, \hat{x})$.

Proposition 1 combines two variational arguments. First, consider a small proportional change in consumption utility and labor disutility. This variation is feasible. Since this scaling is unrestricted, meaning that it can either increase or decrease the utility allocations, it implies that (16) holds with equality at the optimal allocation (c, x). Second, consider a convex combination of an optimal allocation and any other feasible allocation with a small weight. The convex combination is equivalent to scaling down the optimal allocation and adding a small positive perturbation. By the previous argument, rescaling does not change the Lagrangian at the optimum allocation. The positive perturbation should not decrease the Lagrangian. Since this perturbation is positive it gives an inequality condition.

Proposition 1 presents an implementability constraint for an incentive constrained economy. The implementability conditions are more common in the Ramsey taxation literature where they represent the distortions to allocations introduced by pre-specified taxes. In our model, we do not impose direct restrictions on the permissible taxes and, instead, an information friction endogenously restricts the set of allocations. Importantly, our implementability condition holds with inequality which, as we show, is essential for characterizing the bunching regions.

4.2 Optimal Tax Condition as Stochastic Dominance

We use Proposition 1 to derive an optimality condition for our multidimensional taxation problem in terms of a stochastic dominance condition.

We first use the indirect utility (12) for a feasible allocation $(\hat{c}, \hat{x})$ to write the implementability condition (16) as:

$$\int \big(\mathcal{C}'(c)\big(\hat{u} - \nabla\hat{u}\cdot p\big) - z\mathcal{X}'(x)\cdot\nabla\hat{u}\big)\pi \mathrm{d}p - \lambda\int \hat{u}\pi \mathrm{d}p \geq 0, \tag{17}$$

for any nonnegative, decreasing and convex indirect utility function $\hat{u}$. By Proposition 1 it follows that (17) holds with equality for an optimal indirect utility function. Integrating implementability condition (17) by parts we obtain:

$$\int \big(\partial_{p_c}\big(\pi(p_c\mathcal{C}'(c)+z\mathcal{X}'(x_c))\big)+\partial_{p_m}(\pi(p_m\mathcal{C}'(c)+z\mathcal{X}'(x_m)))\big)\hat{u}\mathrm{d}p \geq \int \pi(\lambda - C'(c))\hat{u}\mathrm{d}p + \Xi(\hat{u}), \tag{18}$$

for any nonnegative, decreasing and convex indirect utility function $\hat{u}$, where boundary conditions act on $\hat{u}$ as $\Xi(\hat{u}) = \int_{\underline{p}_m}^{\bar{p}_m} \pi(p_c\mathcal{C}'(c) + z\mathcal{X}'(x_c))\hat{u}\big|_{\underline{p}_c}^{\bar{p}_c}\mathrm{d}p_m + \int_{\underline{p}_c}^{\bar{p}_c} \pi(p_m\mathcal{C}'(c) + z\mathcal{X}'(x_m))\hat{u}\big|_{\underline{p}_m}^{\bar{p}_m}\mathrm{d}p_c$.

We now define second-order stochastic dominance (Shaked and Shanthikumar, 2007):

Definition. The measure μ second-order stochastically dominates the measure ν, or $\mu \succeq \nu$, if and only if for any nonnegative, decreasing and convex function $\hat{u}$:

$$\int \hat{u}(p)\mathrm{d}\mu \geq \int \hat{u}(p)\mathrm{d}\nu. \tag{19}$$

Second-order stochastic dominance states that equation (19) holds for any nonnegative, decreasing, and convex function $\hat{u}$. These conditions exactly correspond to the indirect utility being feasible (Lemma 1). Applying the definition for second-order stochastic dominance to equation (18) we thus obtain the following theorem.

Theorem. *Optimal Tax Condition as Stochastic Dominance.* Suppose the optimal allocation (c, x), density function, and assignment are all continuously differentiable. Then,

$$\partial_{p_c}\big(\pi(p_c\mathcal{C}'(c) + z\mathcal{X}'(x_c))\big) + \partial_{p_m}(\pi(p_m\mathcal{C}'(c) + z\mathcal{X}'(x_m))) \succeq \pi(\lambda - \mathcal{C}'(c)) + \Xi. \tag{20}$$

This theorem derives the optimality condition for the multidimensional taxation economy that incorporates bunching. This condition shows that, at the optimum, the measure over marginal tax revenues, $\pi\left(1/u'(\mathcal{C}(c)) - \lambda\right)$, second-order stochastically dominates the measure over marginal tax distortions,

$$\partial_{p_c}\Big(\frac{\pi}{u'(\mathcal{C}(c))}p_c\frac{\tau_c}{1-\tau_c}\Big) + \partial_{p_m}\Big(\frac{\pi}{u'(\mathcal{C}(c))}p_m\frac{\tau_m}{1-\tau_m}\Big) + \Xi, \tag{21}$$

where we use the definition of the skill wedge (14) and footnote 8.

Comparing the costs and the benefits of taxes is the key insight of the classic ABC formula and the analysis of Diamond (1998) and Saez (2001). In the classic unidimensional case, these costs and benefits are exactly equated for each of the skill levels. Our theorem shows that for the multidimensional tax case with bunching the logic of the ABC formula applies as the costs and the benefits of the taxes are compared. However, those are not necessarily equated at each skill level. Instead, our optimal tax condition (20) considers the benefits and the costs of the entire schedule of taxes and states that the entire schedule of benefits of taxes should second-order stochastically dominate the entire schedule of distortions − showing the nonlocal nature of the problem with multidimensional skills in the regions with bunching. Our formula applies both to the regions with and without bunching and, in the latter case reduces to equating the costs and the benefits of distortions at each skill level − thus making it a local problem for the regions without bunching.[10]

[10]Our optimal taxation condition as stochastic dominance is also related to the sweeping operator in Rochet and Choné (1998). More specifically, the existence of a version of the sweeping operator can be established by using a variation of the Strassen Theorem (see Shaked and Shanthikumar (2007), Theorem 4.A.5). The optimal tax formula goes beyond existence of such an operator by further relating the entire schedule of costs to the entire schedule of benefits of optimal taxes.

4.3 Global Optimal Tax Formula

We next provide an optimal tax formula for the multidimensional taxation problem as an equality. This representation also connects to the optimal tax formulas in unidimensional taxation problems which are derived as equality measuring the marginal costs and benefits of taxation (Mirrlees, 1971; Diamond, 1998; Saez, 2001).[11] The main difference with these results is that the optimal tax formula in our multidimensional taxation problem explicitly accounts for global incentive constraints.

Theorem. *Global Optimal Tax Formula.* Suppose the optimal allocation (c, x), density function, and assignment are all continuously differentiable. Then,

$$\partial_{p_c}\Big(\frac{\pi}{u'(\mathcal{C}(c))}p_c\frac{\tau_c}{1-\tau_c}\Big)+\partial_{p_m}\Big(\frac{\pi}{u'(\mathcal{C}(c))}p_m\frac{\tau_m}{1-\tau_m}\Big)=\pi\left(\frac{1}{u'(\mathcal{C}(c))}-\lambda\right)-\Delta M(p), \tag{22}$$

where $M(p)$ is a positive semidefinite matrix that enforces the convexity of the indirect utility function, and $\Delta M(p)=\sum_{i,j}\frac{\partial^2}{\partial p_i\partial p_j}M_{ij}(p)$.

The full proof is in Appendix A.6 and here we provide a sketch of the proof. First, we use the definition of the indirect utility function (12) to reformulate the planner problem as directly choosing an indirect utility function to minimize the resource cost of providing welfare. For the indirect utility function to be globally incentive compatible, the reformulated planning problem is constrained by the condition that the indirect utility function is convex and decreasing in worker type p following the characterization in Lemma 1.

Second, the indirect utility function $u(p)$ being convex is equivalent to its Hessian being positive semidefinite, $H(u)\succeq 0$ for all worker types p, which in turn is equivalent to $v^TH(u)v\geq 0$, for all vectors v. These inequalities are an infinite series of constraints parameterized by the vectors v for each worker type p. For each of these constraints, we introduce a multiplier $\lambda(v,p)\geq 0$ and include these constraints into the Lagrangian for the planning problem.

Third, we establish (Lemma 8) that one can represent the constraint that the indirect utility function has to be convex as a matrix condition by introducing a positive semidefinite Kuhn-Tucker matrix $M(p)$ for each worker $p\in P$. Instead of considering the infinite series of constraints for each worker p, a single positive semidefinite matrix $M(p)$ induces convexity of the indirect utility function. Upon integration by parts, this restriction appears as a modified social welfare weight

[11] In Appendix A.5, we develop the relationship between our general optimal tax conditions and the classic ABC formula (Diamond, 1998; Saez, 2001).

$\omega(p) = 1 + \frac{\Delta M(p)}{\lambda \pi}$. In summary, the main contribution of the convexity constraint to the planning problem is modifying the social welfare weights through a convexity correction. Finally, we derive in Appendix A.6.2 and Appendix A.6.3 how this modified social welfare weight translates into the optimal tax condition (22).

We now discuss in more detail how the optimal tax condition applies in regions without bunching. Specifically, we consider the domain where the indirect utility function is strongly convex and, therefore, there is no bunching.

The main difficulty in analyzing bunching in the multidimensional case is that the possible indirect utility perturbations $\hat{u}$ are required to be convex. The convexity of perturbations thus acts as an additional constraint on the entire tax schedule. Without bunching, the perturbation argument is straightforward to construct and leads to equating of cost and benefits of taxes at each skill level. Intuitively, if the underlying utility function is strongly convex, a small enough additive perturbation preserves convexity. As a result, the optimal tax condition (22) applies with the convexity correction $\Delta M = 0$ at the types where there is no bunching.

Corollary 2. *Multidimensional Optimal Tax Formula without Bunching.* If the indirect utility function is strongly convex for a worker p, then:

$$\pi \left(\frac{1}{u'(\mathcal{C}(c))} - \lambda \right) = \partial_{p_c} \left(\frac{\pi}{u'(\mathcal{C}(c))} p_c \frac{\tau_c}{1-\tau_c} \right) + \partial_{p_m} \left(\frac{\pi}{u'(\mathcal{C}(c))} p_m \frac{\tau_m}{1-\tau_m} \right). \tag{23}$$

The proof is in Appendix A.7. In order to provide intuition for Corollary 2, and to connect our expression to the existing literature, we also write this condition in the original worker type coordinates α:

$$\phi(\alpha) \left(\lambda - \frac{1}{u'(c(\alpha))} \right) = \frac{1}{\rho} \partial_{\alpha_c} \left(\frac{\phi(\alpha)}{u'(c(\alpha))} \alpha_c \frac{\tau_c}{1-\tau_c} \right) + \frac{1}{\rho} \partial_{\alpha_m} \left(\frac{\phi(\alpha)}{u'(c(\alpha))} \alpha_m \frac{\tau_m}{1-\tau_m} \right), \tag{24}$$

which is the same form as derived in Kleven, Kreiner, and Saez (2006, p. 23), Lehmann, Renes, Spiritus, and Zoutman (2021), and Golosov and Krasikov (2025). The left-hand side captures the marginal benefit of increasing taxes, lowering the resource cost by taxing worker α at the cost λ of tightening the promise keeping condition, and where ϕ denotes the density function over the types α. At an optimum, the marginal benefit of increasing taxes is equated to the marginal distortionary cost of increasing taxes, which is given by the right-hand side. The right-hand side captures the change in labor distortions inversely weighted by the marginal utility of consumption. Distortionary costs of taxation scale with the elasticity of labor supply, which is governed by ρ.

When the supply of tasks is elastic (low ρ), marginal distortionary costs are large. When the supply of tasks is inelastic (high ρ), marginal distortionary costs are small. All else equal, for high-skill workers with low marginal utility from consumption, $\lambda < 1/u'(c(\alpha))$, the skill distortion decreases with an increase in either cognitive or manual skills. When more workers ϕ are affected by a change in the skill distortions, or when the promise keeping constraint is tight, marginal labor distortions change more rapidly.

We next provide a converse to Corollary 2 that allows us to determine the regions of bunching.

Proposition 3. *Identifying Bunching.* If equation (23) does not hold for a worker type p, then this worker is bunched.

Proposition 3 thus provides a test to identify bunching. Whenever equation (23) is violated, the worker is bunched. We prove Proposition 3 in Appendix A.8. By the contrapositive to Corollary 2 it follows that when equation (23) does not hold, the indirect utility function is not strongly convex, meaning that the Hessian matrix is degenerate for worker p. We show that the Hessian matrix is also degenerate for all workers within the neighborhood of p, which we show is equivalent to worker p being bunched.

Finally, we show the relationship between our multidimensional optimal tax formula and the optimal tax condition for unidimensional taxation problems in Mirrlees (1971). The optimal tax formula in Mirrlees (1971), which is equation (33) of Theorem 2 in his paper, is:

$$\frac{1}{\rho}\frac{\phi(\alpha)}{u_c(\alpha)}\alpha\frac{\tau(\alpha)}{1-\tau(\alpha)} = \int_{\alpha}^{\infty}\left[\frac{1}{u_c(x)} - \lambda\right]\phi(x)\mathrm{d}x, \tag{25}$$

where we denote the unidimensional productivity type by α and the distribution function for types by ϕ for ease of comparison. Differentiating this expression with respect to type α, we obtain:

$$\frac{1}{\rho}\partial_{\alpha}\left(\frac{\phi(\alpha)}{u'(c(\alpha))}\alpha\frac{\tau}{1-\tau}\right) = \phi(\alpha)\left(\lambda - \frac{1}{u'(c(\alpha))}\right). \tag{26}$$

This is exactly our multidimensional optimal tax formula without bunching (24) specialized to the unidimensional environment. Without bunching, the multidimensional optimal tax formula is simply the multivariate analog of the unidimensional optimal tax formula.

4.4 Legendre Linearization

In this section, we discuss the main technique that enables the numerical solution of our problem. Specifically, we transform our problem into a linear problem using Legendre transformations for convex functions that translates convex functions into the upper envelopes of their tangent lines.

In order to explain the Legendre transform, and show its importance, we use the resource cost of providing consumption utility $\mathcal{C}$ as an example.

A convex function exceeds all tangent lines. For any consumption utility c, and for any point of tangency a:

$$\mathcal{C}(c) \geq \mathcal{C}(a) + (c-a)\mathcal{C}'(a) = \varphi c - \mathcal{C}^*(\varphi), \tag{27}$$

where the equality follows by parameterizing the tangent lines with their slope $\varphi := \mathcal{C}'(a)$ and by letting $\mathcal{C}^*(\varphi) := -\mathcal{C}(a) + a\mathcal{C}'(a)$ for $a = \mathcal{C}'^{-1}(\varphi)$. The function $\mathcal{C}^*$ is the Legendre transform for the resource cost of providing consumption utility $\mathcal{C}$. Since a convex function exceeds all its tangent lines, and since the function value equals the value of the tangent line at the point of tangency:

$$\mathcal{C}(c) = \max_{\varphi \geq 0} \ \varphi c - \mathcal{C}^*(\varphi). \tag{28}$$

The Legendre transformation converts the convex resource cost of providing consumption utility on the left side of (28) into a family of linear constraints on the right. The family of linear constraints is parameterized by the slopes of the tangent lines of the cost function. Since the resource cost increases with consumption utility, the slopes of the tangent lines are positive, or $\varphi \geq 0$.

The previous steps apply for any convex function, allowing us to use the same argument to transform the resource cost of providing work disutility into a family of linear constraints:

$$\mathcal{X}(x_s) = \max_{\psi_s \leq 0} \ \psi_s x_s - \mathcal{X}^*(\psi_s), \tag{29}$$

for each skill $s \in \mathcal{S}$. An increase in production disutility increases production and therefore lowers resource costs. The resource cost of production disutility is decreasing, implying negative slopes of the tangent lines, or $\psi_s \leq 0$.

To summarize, the transformed planning problem is to minimize the resource cost of providing utilitarian welfare $\mathcal{U}$:

$$\int \Big(\max_{\varphi(p) \geq 0} \big(\varphi(p)c(p) - \mathcal{C}^*(\varphi(p))\big) + z(p) \sum_s \max_{\psi_s(p) \leq 0} \big(\psi_s(p)x_s(p) - \mathcal{X}^*(\psi_s(p))\big)\Big)\pi(p)\mathrm{d}p \tag{30}$$

subject to incentive constraints (10) for all workers $(p, q) \in P \times P$, and the linear promise keeping condition (11).[12]

[12] In Appendix A.9, we show this problem is equivalent to maximizing utilitarian welfare subject to the resource constraint, and the incentive constraints. In Appendix A.10 we establish how to derive the stochastic dominance condition and the general optimal tax formula directly from the transformed problem.

Numerical Approach. The main insight of this analysis is that Legendre transform enables us to translate the planning problem into a linear problem (see Appendix A.11 for more detail). This is the reason why we are able to solve the model for a total of 40 thousand worker types, with 200 types in both the cognitive and the manual dimension, and a total of 1.6 billion incentive constraints. Importantly, a large number of types and numerical precision is not merely a technical and computational curiosity, it is essential to characterize the regions and nature of bunching. In addition, we use two other significant steps to reduce the number of effective incentive constraints.

First, we consider only a small set of incentive constraints by adding incentive constraints between two worker types only if the distance between them is small.[13] We then use an iterative procedure to update the set of incentive constraints. On each step, we add all violated incentive constraints to the problem.[14] With 40 thousand types, this procedure allows us to reduce the number of incentive constraints to about 4 million constraints instead of 1.6 billion. Second, an important step that helps us reduce the number of incentive constraints is that we do not need to consider reducible incentive constraints (see Appendix A.2). This observation additionally reduces the number of constraints by a factor of two. In Appendix A.11 we further prove the accuracy of the approximate planner problem and describe the algorithm that we use to characterize the numerical solution. We finally note that without introducing Legendre transforms the objective is nonlinear. Currently, even the state-of-the-art nonlinear solvers cannot handle the characterization of the solution even for small numbers of types.

5 Positive Economy

We describe and characterize an equilibrium in a positive model of workers with multidimensional skills sorting with heterogeneous firms.

Every firm z takes wage schedule w as given and chooses two workers to solve:

$$\Omega(z) = \max_{x_1, x_2} \; y(x_1, x_2, z) - w(x_1) - w(x_2). \tag{31}$$

We define the surplus S as output minus payments to the workers and the firm:

$$S(x_1, x_2, z) = y(x_1, x_2, z) - w(x_1) - w(x_2) - \Omega(z). \tag{32}$$

[13] Oberman (2013) shows that the solution to the problem with only local constraints is a reasonable initial guess.

[14] After the final step, the candidate solution satisfies all constraints to the strictly convex optimization problem and hence is the unique solution. In practice, we always obtain the same solution for different initial conditions.

Firm output cannot exceed payments to its workers and owner, that is, $S(x_1, x_2, z) \leq 0$ for any triplet (x_1, x_2, z).

Every worker takes the wage schedule w as given and chooses their cognitive and manual task inputs x to solve:

$$\max_{x_c, x_m} \; u(c) - v\Big(\frac{x_c}{\alpha_c}\Big) - v\Big(\frac{x_m}{\alpha_m}\Big) \tag{33}$$

subject to the budget constraint $c = (1-\tau)w(x)$, where $w(x) = w(x_c, x_m)$ is the wage as a function of cognitive and manual inputs, and the disutility from work is given by (2). The government taxes earnings at a rate τ to finance public expenditures G that are not valued by workers.

The resource constraint is given by:

$$\int y(x_1, x_2, z) \mathrm{d}\gamma(x_1, x_2, z) = \int c(\alpha) \mathrm{d}\Phi(\alpha) + \int \Omega(z) \mathrm{d}F_z(z) + G. \tag{34}$$

Total production, $\int y(x_1, x_2, z) \mathrm{d}\gamma(x_1, x_2, z)$, equals output distributed to workers, $\int c(\alpha) \mathrm{d}\Phi(\alpha)$, to firms $\int \Omega(z) \mathrm{d}F_z(z)$, and to public expenditures G.

Equilibrium. Given fiscal policy (τ, G), an equilibrium is a firm value function Ω, a wage schedule w, a worker input distribution F_x, a feasible assignment γ, and an allocation $\{(c(\alpha), x_c(\alpha), x_m(\alpha))\}$ such that firms solve their profit maximization problem (31), workers solve the worker's problem (33), the government budget constraint is satisfied $G = \tau \int w(x) \mathrm{d}\Phi(\alpha)$, and the resource constraint (34) is satisfied.

The equilibrium assignment is the assignment that maximizes aggregate output, that is, solves the primal problem, while the equilibrium wages w and firm value function Ω solve the corresponding dual problem. The characterization of the equilibrium assignment, wage schedule, and firm value function through primal and dual problems is discussed in Appendix A.12.

5.1 Characterizing Equilibrium

We note that solving for the equilibrium assignment in the positive economy follows the same steps as solving for the planner assignment (A.2) in Appendix A.1. It follows that the equilibrium features self-sorting between workers and coworkers, and positive sorting between team quality and firm project values.

In order to characterize wages and firm values, we solve the dual transport problem. Since the surplus is negative for any triplet in equilibrium, $S(x_1, x_2, z) \leq 0$, and since the aggregate

resource constraint (34), the government budget constraint and the household budget constraints hold in equilibrium, the surplus equals zero almost everywhere with respect to the equilibrium assignment, so $w(x_1) + w(x_2) + \Omega(z) = y(x_1, x_2, z)$. Output is distributed to the owner and to the workers. We use this condition to establish further properties of the firm value function and the wage schedule in Appendix A.14.

In Appendix A.14, we first note that wages are only a function of effective worker skills $X = x_c^2 + x_m^2$, and we define $h(X)$, the firm's total wage bill, as $h(X) = 2w(x)$. By applying standard arguments from optimal transport, wages are convex in effective worker skill X. In other words, small differences in effective worker skill translate into increasingly large differences in earnings.[15] Moreover, the firm value function is the Legendre transform of the wage bill, $\Omega = h^*$. As a result, $h(X) + h^*(z) = zX$.

In our quantitative analysis, we infer the distribution of project values F_z using earnings data. The key is to show that there exists a firm project z such that $h(X)+h^*(z) = zX$ for any pairing (z, X). When the wage bill h is continuously differentiable $h(X) + h^*(z) = zX$ implies $z = h'(X)$. That is, the derivative of the firm's wage bill equals its project value. Given an increasing and convex wage bill h, and effective skills X, this condition identifies increasing values for firm productivity z.

We apply this logic to the parametric continuously differentiable function $h(X) = X^\eta + 2\zeta$ where $\eta \geq 1$ governs the convexity of wages and ζ captures the lowest wage per worker. Using the derived fact that $z = h'(X)$, we can relate the distribution of firm projects z to the convexity parameter η of the wage bill. If $\eta = 1$, there is no dispersion in firm productivity. We formalize this in Lemma 3.

Lemma 3. For some firm distribution F_z there exists an equilibrium with (i) a self-sorted assignment, and (ii) a wage function:

$$w(x) = \frac{1}{2}\left(x_c^2 + x_m^2\right)^\eta + \zeta. \tag{35}$$

The proof is in Appendix A.15. The idea is to show there is a firm distribution F_z so that given wage schedule (35), workers and firms both optimize in a self-sorting equilibrium. Given the firm technology (4) and the wage equation (35), firm profits decrease in the difference between their

[15] The hedonic pricing condition $z = h'(X)$ delivers superstar effects in our model as well as a number of other assignment models (see, for example, Rosen (1981), Gabaix and Landier (2008), Tervio (2008), Scheuer and Werning (2017), and Boerma, Tsyvinski, and Zimin (2025)).

workers' skills. In order to minimize output losses, firms thus hire pairs of identical workers. Given wage equation (35), the worker problem (33) has a unique solution, so that the distribution of worker inputs F_x is uniquely determined by the worker problem. Finally, we infer the firm distribution that induces (35) as an equilibrium wage equation using $z = h'(X)$. We use these steps to pointwise identify the worker skill distribution as we show in Section 6.

6 Quantitative Analysis

In this section, we infer the distribution of cognitive and manual skills Φ. The inference of the underlying distributions of skills, a central input for the optimal tax formula, generalizes the approach with unidimensional skills in Saez (2001) to a labor market model with multidimensional skills, coworker and firm effects. We also calibrate the parameter ρ that governs the curvature of disutility with respect to effort.

6.1 Data Sources

We use data from the American Community Survey (ACS). We consider individuals between 25 and 60 years of age. The final sample from the ACS includes almost 16 million individuals between 2000 and 2019. For all our results, we use sample weights provided by the survey. Our measure of labor income is wage and salary income before taxes over the past 12 months.[16]

The ACS contains occupational information for every worker. We combine a worker with the task intensity for their occupation using O*NET task measures from Acemoglu and Autor (2011). Our cognitive measure is the average of their cognitive measures, and our manual measure is the average of their manual measures. Our resulting scores are approximately normally distributed across occupations.

We first construct a measure of relative task intensity by occupation. To obtain aggregated task production levels we use a Cobb-Douglas technology to map worker subtasks into final task production similar to Kremer (1993), Acemoglu and Autor (2011), and Deming (2017):

$$q_s = \exp\left(\frac{1}{|\mathcal{V}|}\sum_{\nu\in\mathcal{V}} \log q_{s\nu}\right). \tag{36}$$

Letting $\log q_{s\nu}$ be the Z-score by subtask ν, we obtain cognitive and manual task production levels. Since our aggregated cognitive and manual measure are approximately normally distributed, task

[16] This measure includes wages, salaries, commissions, cash bonuses, tips, and other money income received from an employer. We drop individuals with earnings below a threshold to focus on workers who are attached to the labor market. This minimum is one-half of the federal minimum wage times 13 weeks at 40 hours per week (as in Guvenen, Ozkan, and Song (2014)).

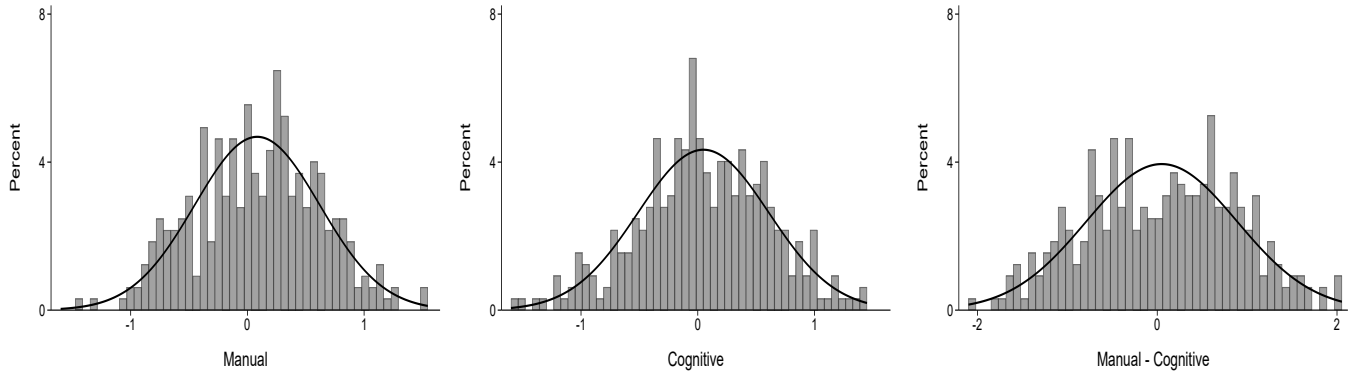


Figure 1: Task Intensity Across Occupations

Figure 1 shows the distribution of manual and cognitive task production levels across occupations in logs (left and center panel) together with the relative distribution of manual and cognitive task intensity (right panel). Each distribution is well-approximated by a lognormal distribution.

production levels are approximately lognormal. We now make an identification assumption that the relative task input is equal to the relative task production level, $x_m/x_c = q_m/q_c$, which is hence also approximately lognormally distributed across occupations.

Figure 1 shows the distribution of manual and cognitive task production levels across occupations in logs together with the relative distribution of manual and cognitive task intensity. The first two panels show that the distribution of manual task production levels and the distribution of cognitive task production levels can be described by a lognormal distribution. The right panel shows that the same holds for the relative manual task intensity.

Figure 2 displays the relationship between relative task intensity and average earnings across occupations. Earnings are low for occupations with high manual task intensity, such as gardeners and truck drivers, while earnings are high for occupations with high cognitive task intensity such as software developers and actuaries. Moving from the 25th percentile to the 75th percentile in relative manual task intensity decreases earnings from 62 to 35 thousand dollars.

6.2 Calibration

We now calibrate the positive model. We parameterize fiscal policy and preferences, and infer the underlying multidimensional skill distribution.

The government taxes labor earnings to finance expenditures G. If pre-tax earnings are w, then taxes are given by $T(w) = \tau w$. After-tax earnings are thus $(1 - \tau)w$, we set $\tau = 0.3$.

Firm heterogeneity governs the convexity of the wage schedule (see Lemma 3). We set the curvature parameter for the wage schedule η to align the added variation in log wages due to firm

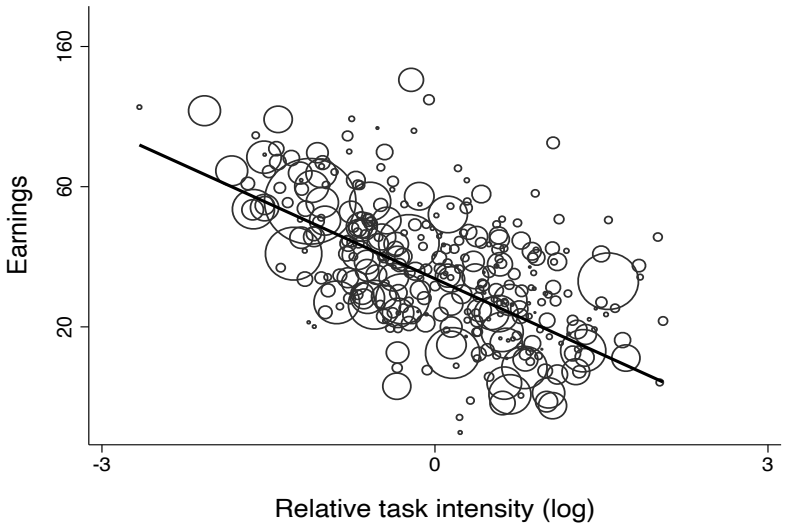


Figure 2: Earnings and Relative Task Intensity

Figure 2 shows the relation between average earnings (y-axis, logarithmic scale) and relative task intensity across occupations. Average earnings are decreasing in the relative manual task intensity. The size of each circle corresponds to the occupation's employment share.

heterogeneity with evidence from the literature on variation in log wages due to firm effects. Using the wage equation (35), the variation in firm projects multiplies the underlying variation across workers by η^2. We set $\eta = 1.1$ to attribute 17 percent of the added variation in wages to firm effects. Our target of 17 percent is in line with estimates from the literature.[17]

We next discuss the calibration of worker preferences. We use linear preferences with respect to consumption goods, $u(c) = c$, and estimate the parameter governing the curvature of the disutility function to efforts in each task ρ. We set ρ such that a regression of log market hours on hourly wages, holding constant the marginal value of wealth, yields a coefficient of 0.55. This target value comes from the meta-analysis of estimates of the intensive margin Frisch elasticity from Chetty, Guren, Manoli, and Weber (2012).

To use estimates for the Frisch elasticity for total hours with respect to hourly productivity to calibrate the curvature of the utility function with respect to effort, we derive this expression within our model. Given the specification for the disutility from work (2), the linear utility from consumption, and the worker technology (3), the worker's problem (33) is:

$$\max_{x_c, x_m} \frac{1}{2}(1-\tau)(x_c^2 + x_m^2)^\eta - \kappa\Big(\frac{x_c}{\alpha_c}\Big)^\rho - \kappa\Big(\frac{x_m}{\alpha_m}\Big)^\rho. \tag{37}$$

[17] For example, Abowd, Lengermann, and McKinney (2003) find that firm variation makes up 17 percent of the variance in wages while Song, Price, Guvenen, Bloom, and Von Wachter (2019) instead report that firm variation makes up between 8 percent and 12 percent.

The optimality condition to the worker's problem for each task $s \in \mathcal{S}$ is:

$$(1-\tau)\eta\big(2w(x)\big)^{\frac{\eta-1}{\eta}} = \kappa\rho\frac{x_s^{\rho-2}}{\alpha_s^{\rho}}, \tag{38}$$

where $w(x) = \frac{1}{2}(x_c^2 + x_m^2)^{\eta} + \zeta$ by wage equation (35) with ζ representing minimum earnings in our data. That is, the marginal consumption utility from supplying extra tasks equals the marginal cost of effort. Taking the ratio of these optimality conditions, this implies that the skill, effort and task intensity ratio are related by:

$$\frac{\alpha_m}{\alpha_c} = \left(\frac{x_m}{x_c}\right)^{\frac{\rho-2}{\rho}} = \left(\frac{\ell_m}{\ell_c}\right)^{\frac{\rho-2}{2}}, \tag{39}$$

where the second equality follows from the worker task technology (3). The marginal rate of substitution between activities, $\left(\frac{\ell_c}{\ell_m}\right)^{\rho-1}$, is equal to the ratio of marginal benefits between activities, $\left(\frac{\alpha_c}{\alpha_m}\right)^2 \frac{\ell_c}{\ell_m}$. Relative efforts are determined by relative skills $\frac{\alpha_c}{\alpha_m}$. Workers spend more effort on tasks in which they are more talented.

Using the first-order conditions for effort, and observing that the share of total efforts on each task is constant by (39), we can express the Frisch elasticity of total hours $\ell_c + \ell_m$ as:[18]

$$\varepsilon = \left.\frac{\partial \log(\ell_c + \ell_m)}{\partial \log \tilde{z}(x)}\right|_{\lambda} = \left.\frac{\partial \log(\ell_c + \ell_m)}{\partial \log(1-\tau)}\right|_{\lambda} = \frac{1}{\rho - 1}, \tag{40}$$

where λ is the marginal value of wealth, and $\tilde{z}(x) := w(x)/(\ell_c + \ell_m)$ is productivity per hour. We set $\rho = 2.8$ so that the Frisch elasticity ε is indeed 0.55. Finally, we normalize $\kappa = \frac{1}{2\rho}$.

Skill Distribution. We now identify the skill distribution pointwise. Using the solution to the worker's problem, together with data on both total earnings and occupational relative task intensity for each worker, we separately identify two sources of worker productivity (α_c, α_m) that rationalize the data as a model outcome. This identification argument is similar to Boerma and Karabarbounis (2020, 2021) who use explicit solutions for home production models to identify productivity at home and to identify permanent and transitory market productivity using data on consumption, home and market hours.

Using the O*NET task measures, we have information on the relative task intensity for each occupation $\frac{x_m}{x_c}$. In order to determine the level of tasks, we use the wage equation (35):

$$w(x) = \frac{1}{2}(x_c^2 + x_m^2)^{\eta} = \frac{x_c^{2\eta}}{2}\left(1 + \left(\frac{x_m}{x_c}\right)^2\right)^{\eta}. \tag{41}$$

Given the skill ratio for an individual's occupation, $\frac{x_m}{x_c}$, and an individual's earnings $w(x)$, this equation uniquely determines the level of cognitive tasks x_c, and hence the level of manual tasks

[18] See Appendix A.16.

Table 1: Example of Identification

		Relative Task	Wages	Task Intensity		Task Skills	
		x_m/x_c	$w(x)$	x_m	x_c	α_m^ρ	α_c^ρ
1	Baseline	1	1	1.00	1.00	0.50	0.50
2	Task intensity	3	1	1.35	0.45	0.63	0.26
3	Wages	1	4	2.00	2.00	0.87	0.87
4	Taxes $\tau = 0.3$	1	1	1.00	1.00	0.71	0.71
5	Firms $\eta = 1.1$	1	1	0.97	0.97	0.42	0.42

Table 1 illustrates the identification of workers' manual and cognitive skills through five examples. We infer higher levels of manual skills with higher manual task intensity (in Row 2), higher earnings (Row 3), higher taxes (Row 4), and with less dispersion in firms' project values (Row 5).

x_m. By the optimality condition (38), we identify both cognitive skills α_c and manual skills α_m for each worker.

Examples. To provide insight into the identification of worker skill heterogeneity, we consider a numerical example. We first consider an economy without taxes $\tau = 0$ and without heterogeneity in firm projects, $\eta = 1$.

Suppose a worker's occupational relative task intensity is equal to one, $\frac{q_m}{q_c} = \frac{x_m}{x_c} = 1$, and their earnings equal mean earnings, which we normalize to one. By equation (41), the worker's cognitive task intensity and the worker's manual task intensity are equal to 1. Using the optimality condition for task inputs (38), $\alpha_s^\rho = \frac{1}{2}$, implying the worker is equally skilled in both tasks. This worker is presented in the first row of Table 1.

Inferred manual skill increases with manual task intensity. Consider some worker with relative manual task intensity equal to three, $\frac{x_m}{x_c} = 3$, and average earnings. By equation (41), the cognitive task intensity is $x_c = \frac{1}{\sqrt{5}} < 1$ and hence the worker's manual task intensity is greater with $x_m = \frac{3}{\sqrt{5}} > 1$. Since $\alpha_s^\rho = \frac{1}{2}x_s^{\rho-2}$, it follows that the worker's inferred manual skill increases with relative manual task intensity, while the worker's cognitive skills decreases, as shown in the second row of Table 1.

Inferred skill levels increase with earnings. For a worker with a relative task intensity of one, but a high level of earnings, the relative skill intensity is one but the level of each task is greater. Consider a worker earning four times average earnings. By equation (41), we identify the worker's cognitive task intensity, and therefore the worker's manual task intensity, to be equal to 2. Using

Table 2: Illustration of Identification

Occupation	Relative $\log \frac{q_m}{q_c}$	Wages $\mathbb{E}w(x)$	Manual $\frac{\alpha_m - \mathbb{E}\alpha_m}{\sigma_m}$	Cognitive $\frac{\alpha_c - \mathbb{E}\alpha_c}{\sigma_c}$	Firm $\frac{\alpha_z - \mathbb{E}\alpha_z}{\sigma_z}$	SOC Code
Gardeners	1.7	23	0.93	-2.35	-1.28	37−3010
Cashiers	0.7	20	0.47	-1.16	-1.62	41−2010
Police officers	-0.1	64	0.82	0.48	0.82	33−3050
Physicians	-0.2	184	1.77	1.32	3.11	29−1060
Chief executives	-2.1	149	-2.39	1.71	2.63	11−1010
Actuaries	-2.7	136	-3.46	1.65	2.43	15−2010

Table 2 shows the identification of worker skills for a number of occupations. Holding constant the relative manual skill intensity, high earnings identify high skill levels as seen by comparing the manual and cognitive skills of police officers and physicians. Holding constant earnings, high manual task intensity identifies high manual skills as seen by comparing the skills of gardeners and cashiers.

the worker's optimality condition for task inputs (38), $\alpha_s^\rho = \frac{1}{2}2^{\rho-2}$, implying that the worker is equally skilled in both tasks, and almost 1.75 times as skilled as a worker in the same occupation earning average earnings. This worker is presented in the third row of Table 1.

The presence of taxes does not affect inferred task intensities x but does increase the inferred skill levels α. Since the identification of the task intensity is based on pretax earnings (41), inferred task intensities do not vary with taxes. For $\eta = 1$, since the task intensity does not change with taxes, we obtain $\alpha_s^\rho = \frac{1}{2(1-\tau)}x_s^{\rho-2}$. When workers are taxed, the marginal benefit from completing tasks is reduced. In order to rationalize the same levels of cognitive and manual task intensity supplied by a worker, it must be less costly for the worker to complete tasks due to increased levels of skills, as shown in the fourth row of Table 1.

Finally, increased dispersion in firm project values decreases wage dispersion that is attributed to dispersion in task intensity. Consider the dispersion in firm projects with $\eta > 1$. Reorganizing the wage equation (41), $x_c = (2w(x))^{\frac{1}{2\eta}} \Big/ \sqrt{1 + \left(\frac{x_m}{x_c}\right)^2}$, shows that higher values of η compress the dispersion in task intensity. Further, by combining the first-order condition (38) with wage equation (41), we obtain $\alpha_s^\rho \propto w(x)^{\frac{\rho}{2\eta}-1}$. An increase in η decreases the effective dispersion in skills. Dispersion in firm projects magnifies underlying differences in task intensity due to the positive sorting between workers and projects. Equivalently, small differences in effective worker skills generate large earnings differences.

Having illustrated the identification with examples, we turn to identification using earnings data. Table 2 illustrates the identification of underlying skills for representative workers in occupations listed in the first column. The second column shows the relative manual task intensity for these occupations from O*NET task measures. The third column shows average earnings of the workers by occupation in the ACS. Table 2 shows a negative relation between manual task intensity and average earnings by occupation, in line with Figure 2.

In order to identify manual and cognitive skills, we use equations (38), (39) and (41). First, we establish that higher earnings identify higher levels of skills, everything else equal. Consider an example of police officers and physicians. Since the relative task intensity for police officers and physicians is comparable, their relative skills are comparable by equation (39). Average earnings of physicians exceed the average earnings of police officers implying a higher level of both cognitive and manual skills for physicians. Indeed, the fourth and fifth column in Table 2 show that while both physicians and police officers' cognitive and manual skills exceed the population average, $\alpha_s > \mathbb{E}\alpha_s$, the skills of physicians exceed the skills of police officers in both dimensions.

Second, we consider two occupations with similar wages to show that high manual task intensity identifies high manual skill all else being equal. While the earnings of gardeners and cashiers are similar, gardening is more demanding in manual skills. By equation (41), the cognitive task requirements of gardeners are lower than the cognitive task requirements for cashiers. By equation (39) it follows that a gardener has more manual skills than a cashier, but less cognitive skills. The fourth and fifth column in Table 2 displays this pattern.

We apply the identification argument to all workers in the ACS to identify their skills. By identifying skills at the worker level, we allow for skill heterogeneity within occupations driven by earnings differences within occupation. As in the example, workers with high earnings have higher cognitive and manual skills than a worker with low earnings in the same occupation. Figure 3 shows the resulting distribution of cognitive and manual skills, after clipping the bottom 5 percent of the skill distribution in each dimension and after smoothing the pointwise identified distribution using a kernel density estimation.[19]

For illustrative purposes, we introduce representative occupations in Figure 3. Specifically, we provide nine representative occupations within the type space. For example, cashiers are workers with both low cognitive and low manual skills, executives have low manual skills but high cognitive skills, while physicians have both high cognitive and high manual skills.

[19]We correct our kernel density estimator at the boundaries of our rectangular type space by reflecting along all boundaries, see, e.g. Karunamuni and Alberts (2005).

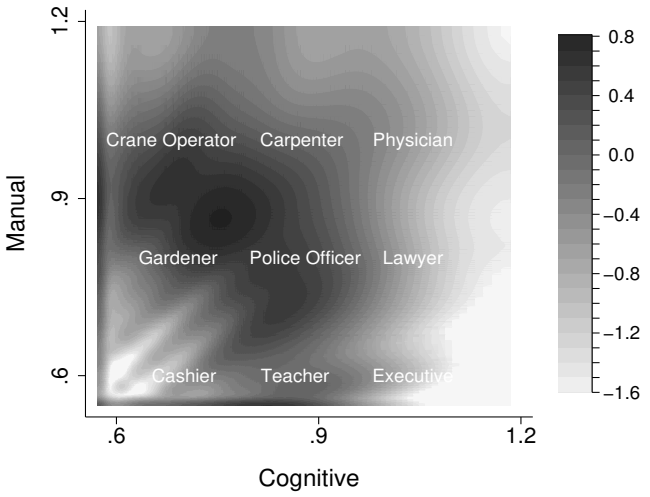


Figure 3: Inferred Skill Distribution

Figure 3 shows the inferred worker skill distribution, with bright colors indicating more mass. The panel shows the smoothed distribution of cognitive and manual skills that exactly rationalizes the data which is obtained using data on relative task intensity by occupation and worker earnings, through equations (38) to (41). The values are normalized such that one reflects a uniform distribution.

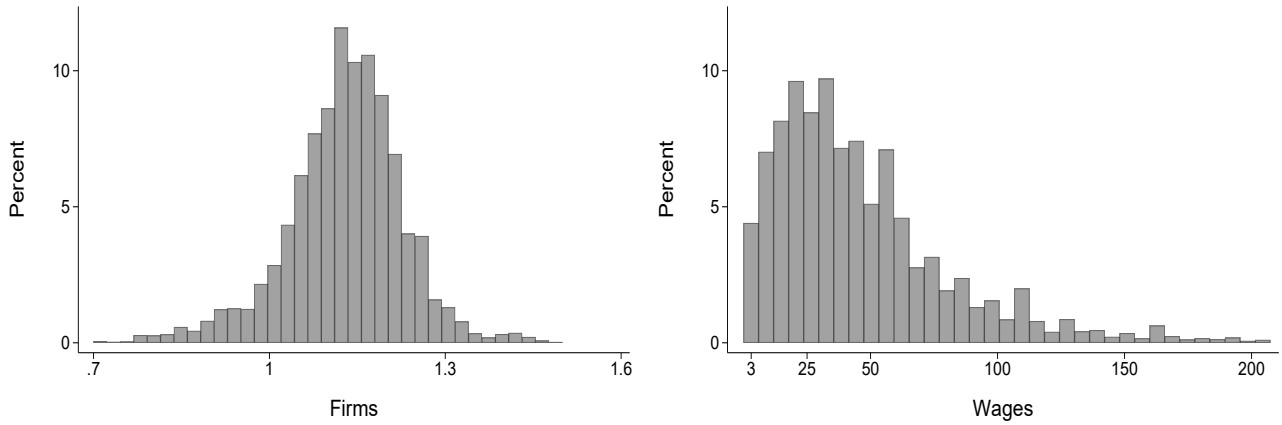


Figure 4: Firm and Wage Distribution

Figure 4 shows the histogram for the inferred firm distribution (left panel) and the model implied distribution of wages (right panel).

Finally, Figure 4 shows the inferred firm productivity distribution in the left panel and the implied wage distribution in the right panel. The left-hand distribution shows that the distribution of firm projects is relatively concentrated with project values ranging from 30 percent below the mean to 40 percent above the mean (1.1). By construction, the right panel replicates the empirical wage distribution.

7 Quantitative Results

In this section, we present the quantitative results to the planning problem using the empirically relevant model of Section 6.

7.1 Unconstrained Benchmark

In order to build intuition for the solution, we first present a benchmark without incentive constraints and firm heterogeneity. The planning problem then simplifies to minimizing resource costs (6) subject to the promise keeping condition (8). By using the functional form for preferences, the promise keeping condition simplifies to:

$$\int \Big(c(\alpha) - \kappa\big(x_c(\alpha)/\alpha_c\big)^{\rho} - \kappa\big(x_m(\alpha)/\alpha_m\big)^{\rho}\Big)\mathrm{d}\Phi \geq \mathcal{U}. \tag{42}$$

At the optimum, cognitive tasks are independent of workers' manual skills, and the elasticity of cognitive tasks with respect to cognitive skills is $\frac{\rho}{\rho-2}$. Furthermore, the solution does not feature bunching. In order to see this, note that the following condition has to be satisfied:

$$x_s \propto \alpha_s^{\frac{\rho}{\rho-2}}, \tag{43}$$

for each skill $s \in \{c, m\}$. Due to additive separability of tasks in preferences and technology, the efforts on task s depend only on the worker's skills in this task. Equivalently, there is no cross-dependence between tasks. Since (43) describes a one-to-one relation between the worker's skills and efforts in each task, there is no bunching at optimum. That is, in a neighborhood of worker α, every pair of distinct workers (α', α'') is assigned distinct allocations as $x(\alpha') \neq x(\alpha'')$.

Given the empirical description of the distribution for cognitive and manual skills in Figure 3, equation (43) gives the optimal allocation of both cognitive and manual tasks. Figure 5 visualizes the benchmark allocation of task intensity by worker's cognitive and manual skills. The left panel shows the allocation of manual tasks, the right panel shows the allocation of cognitive tasks. Since (43) rules out any cross-dependence between tasks, the optimal allocation is captured by parallel horizontal and vertical lines, respectively. Manual task intensity only varies with manual skill, while cognitive task intensity only varies with cognitive skill.

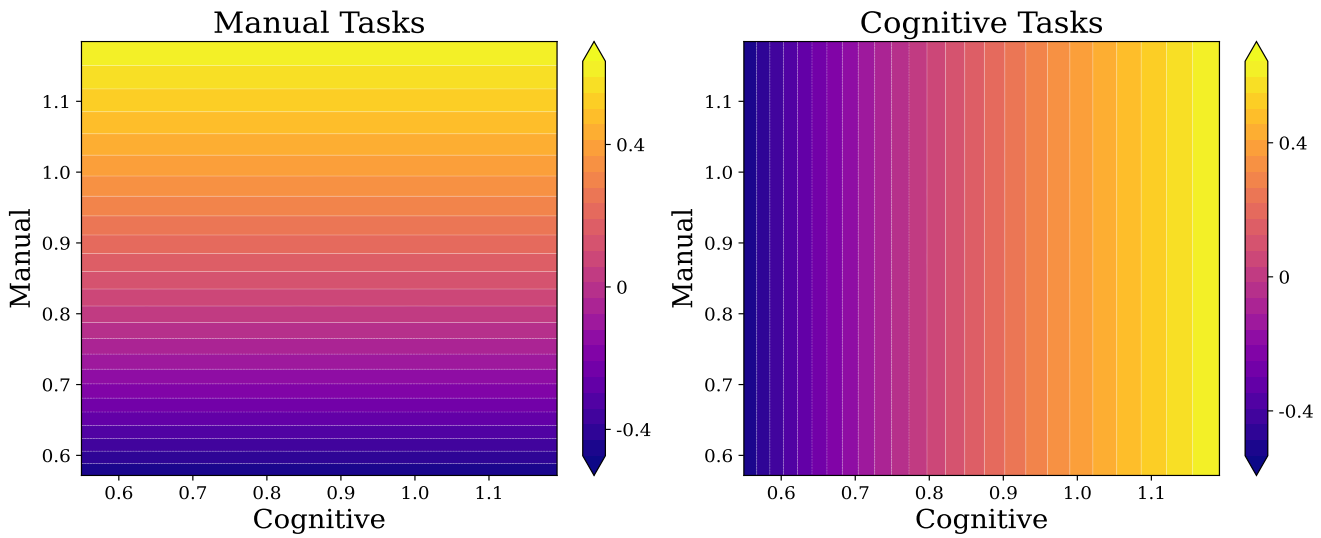


Figure 5: Benchmark Allocation

Figure 5 shows the benchmark allocation for task intensity by worker's cognitive and manual skills. The left panel shows the allocation of manual tasks, the right panel illustrates the allocation of cognitive tasks. The optimal allocation does not feature any cross-dependence between tasks: manual task intensity only varies with manual skill, while cognitive task intensity only varies with cognitive skill.

7.2 Optimal Solution

Figure 6 shows the solution to the planner problem.[20] The top row shows the allocation of manual and cognitive tasks, the bottom row shows the allocation of consumption and the assignment of workers to firms. In contrast to the benchmark, the optimal task intensity in one skill depends positively on a worker's other skills. Consider the manual task allocation in the top left panel. Similar to the benchmark, the manual task intensity increases with a worker's manual skills holding constant their cognitive skills. In contrast to the benchmark solution, the manual task intensity also increases with workers' cognitive skills. That is, workers with the same manual ability but with a higher cognitive ability conduct a higher level of manual tasks. Moreover, this codependence between cognitive skills and manual tasks intensifies at low levels of cognitive skill. This can be seen by the contour lines being almost negative 45 degree lines at low levels of manual ability, while being almost flat at high levels of manual ability. The same pattern holds for cognitive tasks.

In this economy, the binding incentive constraints are for high types to mimic low types, which is also generally the case with unidimensional skill. To prevent high types from pretending to be low types, the allocation for the low types is distorted. With multidimensional skills, the allocation for the low types is distorted both by reducing the level of task output as in the unidimensional

[20] In the numerical characterization, type space A includes a disability type, which has zero productivity and hence zero output in both tasks. Together with nonnegative consumption, incentive compatibility implies nonnegative utility for all worker types.

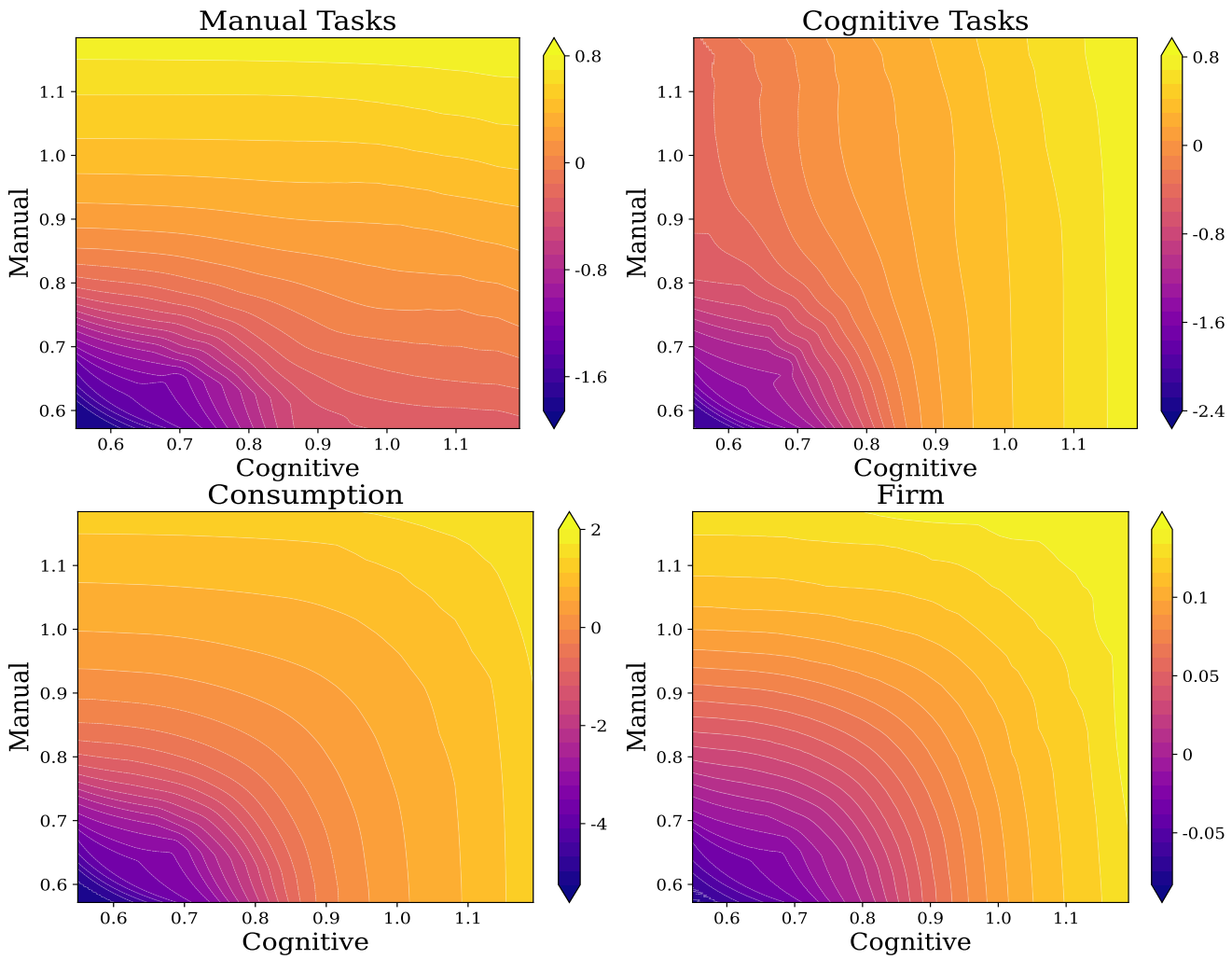


Figure 6: Planner Allocation

Figure 6 visualizes the solution by worker's cognitive and manual skills. The top row shows the manual and cognitive task allocation, the bottom row shows the consumption allocation and the assignment of workers to firms. The solution features positive dependence between tasks. For example, optimal cognitive task intensity increases with manual skills.

case, and by increasing the codependence between tasks. The latter is a new type of distortion that emerges in taxation problems with multidimensional skills. In short, with multidimensional skills, both the level and composition of tasks of low types are distorted to discourage high types from mimicking to be low types.

The bottom left panel shows the solution for consumption. Consumption increases with skills. Consumption of workers with top cognitive skills exceeds consumption of workers with top manual skills due to higher absolute levels of skill. The bottom right panel shows the assignment of workers to firms. Given the cognitive and manual tasks, the planner assigns workers with greatest effective skills $X = x_c^2 + x_m^2$ to projects of greater value following Proposition 4 in Appendix A.1. A physician thus works on a more valuable project than a cashier as in the positive economy. Since

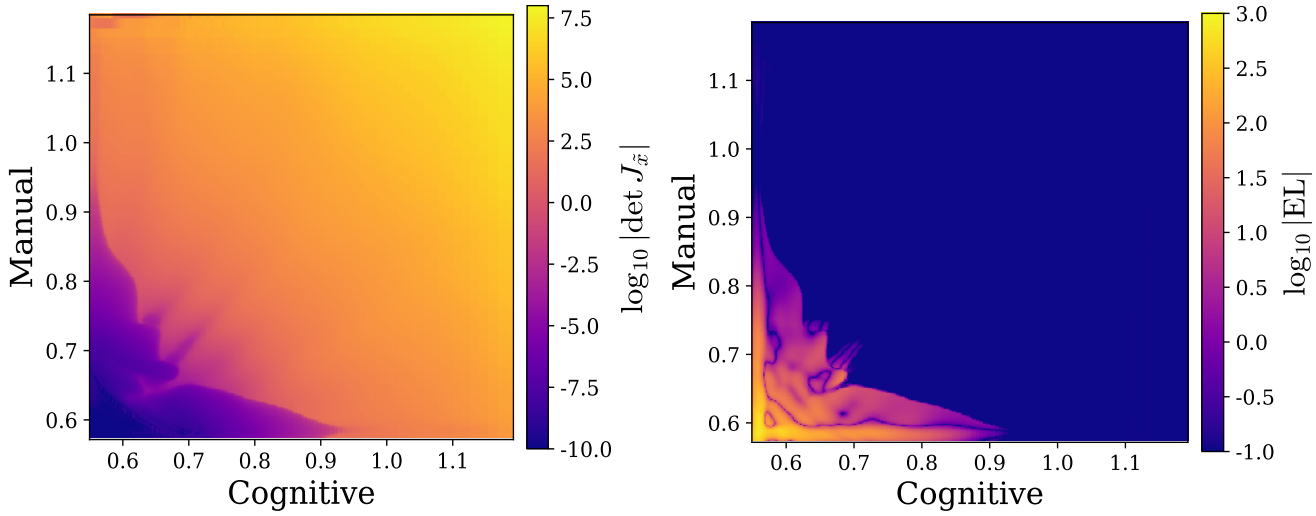


Figure 7: Bunching

Figure 7 illustrates bunching in the optimal allocation. The left panel identifies bunching by analyzing the determinant of the Hessian matrix of the indirect utility function, and shows the value of the determinant on a log base 10 scale. The right panel identifies bunching using Corollary 2 by analyzing deviations from the multidimensional optimal tax formula on a log base 10 scale. The variable on the horizontal axis is cognitive skill α_c; the variable on the vertical axis is manual skill α_m. Both panels identify that workers in the bottom left region of the type space are bunched under the optimal allocation.

the range of the cognitive skills is higher than the range of the manual skills, the high-value projects are assigned to workers with greater cognitive skills.

Bunching. We now describe bunching in the optimal solution. Bunching means that different workers are assigned identical labor supply allocations and, therefore, are also assigned identical consumption allocations (see Section 3.3). We use three distinct methods based on the theoretical analysis in Section 3 and Section 4 to comprehensively characterize the bunching patterns that emerge in the quantitative model.[21]

First, we use our theoretical results in Lemma 2 to identify bunching by analyzing the determinant of the Hessian matrix of the indirect utility function. By Lemma 2, if the indirect utility function is not strongly convex at all points in the neighborhood of type p, then worker p is bunched. If the Hessian matrix is not invertible, then the indirect utility function is not strongly convex. A matrix is invertible if and only if the determinant is not equal to zero. Therefore, if the determinant of the Hessian matrix equals zero, the matrix is not invertible, so the indirect utility function is not strongly convex and the worker is bunched. Thus bunching is present in regions where the determinant of the Hessian matrix of the indirect utility function is equal to zero. We

[21] We provide sensitivity analyses for the bunching patterns with respect to project heterogeneity, consumption utility, disutility from work, and the production technology in Numerical Appendix B.

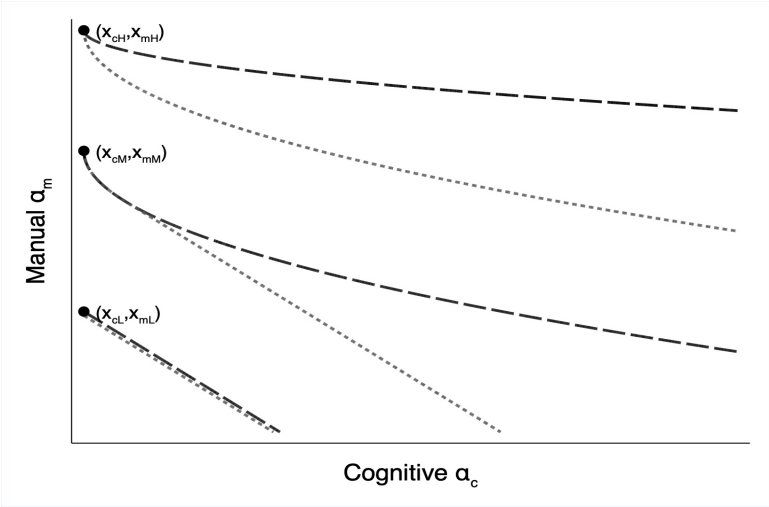


Figure 8: Illustration of Bunching

Figure 8 illustrates the procedure to classify bunching using isocurves. The long-dash lines represent isocurves for different cognitive task levels, while the short-dashed lines represent isocurves for different manual task levels. Worker α is bunched with worker α' if the isocurves for $x(\alpha)$ intersect α'.

apply this method in the left panel of Figure 7, which shows the value of the determinant on a log base 10 scale. This panel shows that workers in the bottom left (dark) region are bunched.

Second, we use our theoretical results in Corollary 2 to identify bunching. If the optimal tax formula without bunching does not hold, then the indirect utility function is not strongly convex for worker p. By Lemma 2, this implies that worker type p is bunched. Numerically, we analyze deviations from the multidimensional optimal tax formula without bunching to establish bunching. We apply this method in the right panel of Figure 7. The figure shows the deviations from the multidimensional optimal tax formula without bunching on a log base 10 scale. The right panel delivers the same bunching region as the left panel, in the bottom left (light) region.

Bunching reflects the distortion of the task composition to discourage high types from mimicking low types. This benefit is weighed against the output cost of distorting the task composition. This cost is low at the bottom of the productivity type distribution, where workers produce little output. As a result, bunching occurs primarily at the bottom of the type distribution in Figure 7.

The third approach to identify bunching is based directly on the definition of bunching (see Section 3.3). When different worker types bunch, they are assigned identical task levels. Worker α is bunched if there are other workers who are assigned the same task levels $x(\alpha)$. Visually, we draw the isocurves corresponding to both $x_c(\alpha)$ and $x_m(\alpha)$ displayed in Figure 8 in the worker space α and assess whether the isocurves intersect for any other worker α'. If there exists a worker $\alpha' \neq \alpha$ such that the isocurves intersect, then workers α and α' are bunched.

Figure 8 gives an example of the procedure to classify bunching using isocurves. The long-dash lines represent isocurves for different cognitive task levels, while the short-dash lines represent isocurves for different manual task levels. The dots indicate three hypothetical allocations. First, consider the isocurves corresponding to the allocation of crane operators in the top left corner. The long-dash isocurve represents workers with other combinations of skills (α_c, α_m) who produce the same cognitive tasks x_{cH} as the crane operator. The short-dash isocurve represents workers with other combinations of skills (α_c, α_m) who produce the same manual tasks x_{mH} as the crane operator. The lines intersect only at one point − the skills of the crane operator at the top left corner. That is, no other worker (α_c, α_m) receives the same task allocation (x_{cH}, x_{mH}) that is assigned to the crane operator. Next, consider the isocurves corresponding to the allocation of a gardener in the middle of Figure 8. The long-dash isocurve (workers producing the same cognitive tasks x_{cM} as the gardener) overlaps with the short-dash isocurve (workers producing the same manual tasks x_{mM} as the gardener) for high levels of manual skill α_m and for low levels of cognitive skill α_c. These workers (α_c, α_m) produce the same cognitive and manual tasks (x_{cM}, x_{mM}) as the gardener, indicating that gardeners bunch with workers whose comparative advantage also lies in manual work. We call this type of bunching *targeted bunching*. Finally, consider the bottom-left allocation corresponding to cashiers. In this case, the long-dash and short-dash isocurves for cognitive and manual tasks overlap throughout the type space. All workers with skills (α_c, α_m) on that line produce the same cognitive and manual tasks (x_{cL}, x_{mL}) as the cashier, despite their skill differences. We call this type of bunching *blunt bunching*, where the planner does not distinguish different worker types when allocating tasks.

Figure 9 shows bunching at the optimal solution and presents two complementary views of the issue. The left panel displays bunching through the allocation of tasks. We display combinations of cognitive and manual tasks (x_c, x_m) that are optimal. An allocation is marked green or blue if the allocation is assigned to more than one worker, while the allocation is marked orange if the allocation is assigned to one worker. There are two main regions of bunching in this picture. The first region is that of the blunt bunching and is given by the green line segment at the bottom of both cognitive and manual tasks. The second region is that of the targeted bunching and is represented by two blue line segments at the borders of the task trapezoid. That is, targeted bunching happens when the task input is low for only one task. The lower (flat) blue line segment represents low manual tasks, and the upper (vertical) line segment represents low cognitive tasks. Note that on these borders, when the task intensity becomes sufficiently high, the allocations are

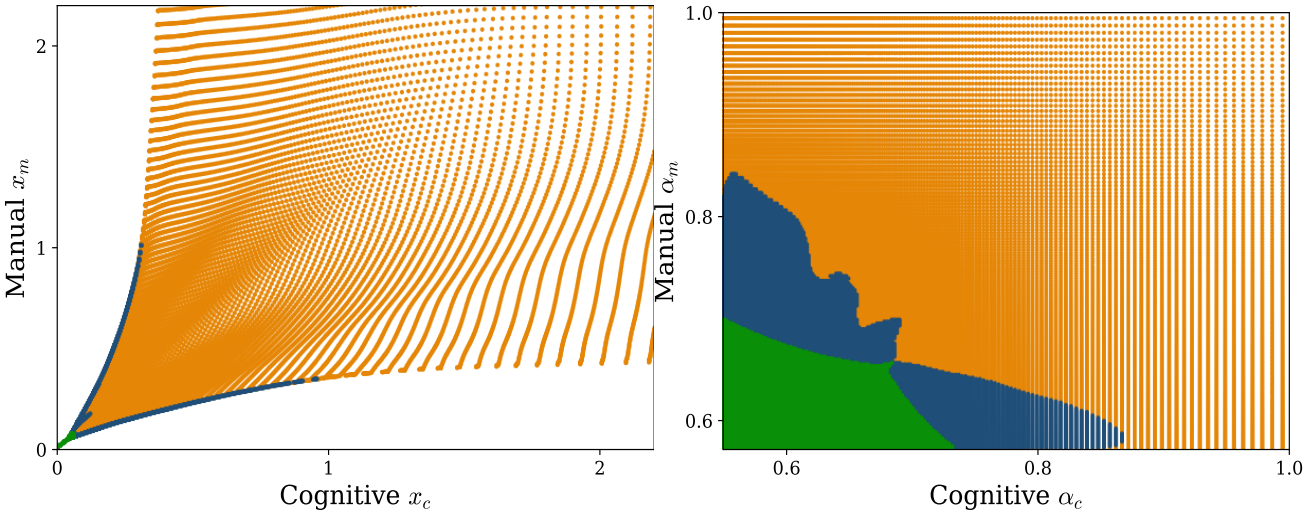


Figure 9: Optimal Targeted and Blunt Bunching

Figure 9 shows bunching at the solution. The left panel demonstrates bunching in the allocation space by displaying combinations of optimal cognitive and manual tasks (x_c, x_m). An allocation is marked in green or in blue if the allocation is assigned to more than one worker, while the allocation is marked orange if the allocation is assigned to one worker. The right panel displays bunching in the worker type space (α_c, α_m). In this figure, a worker type is marked in green or in blue if the worker's task allocation is also assigned to another worker. The green area indicates the blunt bunching region while the blue area indicates the targeted bunching region.

no longer bunched − that is, the blue line turns orange on the edges of the trapezoid.

Blunt and Targeted Bunching. In order to see which workers are bunched, the right panel shows bunched workers in the type space (α_c, α_m). In this figure, a worker type is marked in blue or green if the worker's task allocation is also assigned to another worker.[22] This panel shows that bunched workers have low cognitive or low manual skills (or both). Worker types are marked blue in the region of targeted bunching, and marked green in the region of blunt bunching. Workers with both low cognitive and manual skills are in the blunt bunching region. Workers with medium manual or medium cognitive skills are in the targeted bunching region when their skill set is asymmetric. Workers with medium cognitive skills bunch when their manual skills are low, and vice versa. To quantify the mass of bunched workers in the economy, we now combine the right panel with the worker type distribution of Figure 3. At the optimum, 10.4 percent of the workers are bunched. The blunt bunching region comprises 30 percent of bunched workers. The targeted bunching region comprises 70 percent of bunched workers.

Workers bunch with other workers both near and far. While Figure 9 shows at what allocations workers are bunched and which workers are bunched, it does not show with whom workers bunch.

[22] We construct the distance between two allocations $x(p)$ and $x(\hat{p})$ by considering the Euclidean distance between the allocations relative to the Euclidean distance between the respective types p and $\hat{p}$.

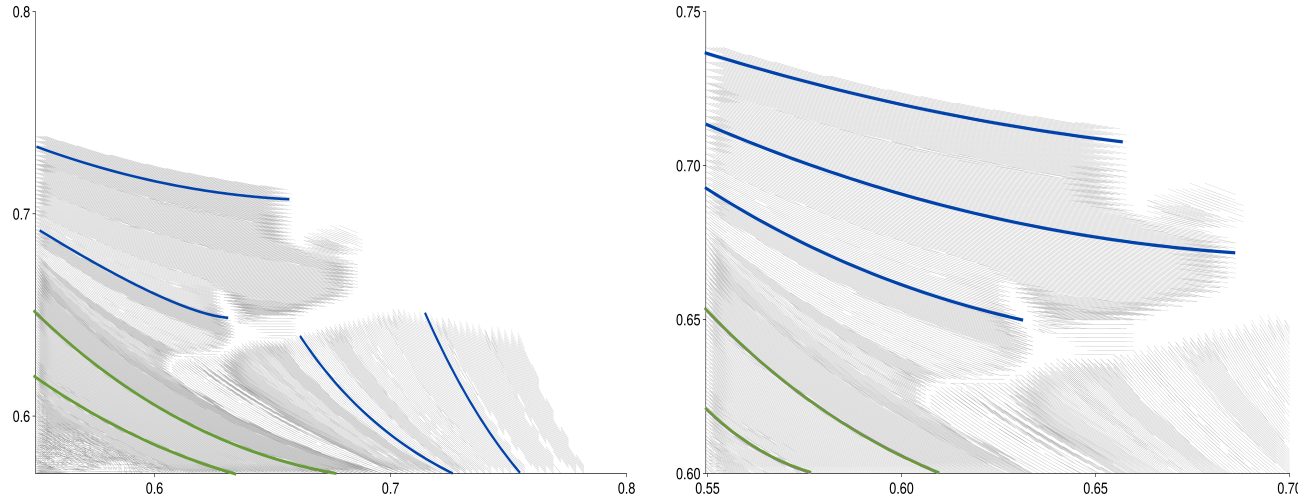


Figure 10: Bunching Patterns

Figure 10 shows with whom workers bunch by connecting workers with a line in the worker type space if their allocations are identical. Workers exclusively bunch with workers that are better in one skill dimension, but worse in another as represented by downward-sloping lines. Green lines indicate the blunt bunching region, blue lines indicate the targeted bunching region. Under blunt bunching, workers on the vertical boundary bunch with workers on the horizontal boundary, unlike under targeted bunching. The right-hand panel zooms in to distinguish the blunt bunching region from the targeted bunching region.

These bunching relations are shown in Figure 10, which connects two workers with a line in the type space if they are bunched.[23]

Workers do not bunch with workers over whom they have an absolute advantage or who have an absolute advantage over them. Workers exclusively bunch with workers that are better in one skill dimension but worse in another. In Figure 10 this is evident since all connections are represented by downward-sloping lines. Gardeners with somewhat better cognitive skills, but somewhat lower manual skills are bunched with gardeners with less cognitive skills and higher manual skills. Despite the slight difference in skills, the planner assigns both identical tasks.

Within our bunching regions we distinguish two distinct patterns. In the lower triangle, which we indicate by green lines, the planner bunches together cognitive and manual tasks for all workers, ranging from those with relatively high manual skills to those with relatively high cognitive skills. Workers with the same effective skill index produce the same cognitive and manual tasks. Work is not tailored towards workers' specific skills but to an overall level of skill. In this blunt region, the planner optimally satisfies the incentive constraints by assigning identical allocations to different workers. Deviations are deterred bluntly at the cost of efficiency.

The planner also bunches workers in the targeted bunching regions, but in a less rudimentary

[23]To facilitate the presentation, we display the bunching relations for one quarter of all workers. In addition, we display at most two relations for each worker.

way because the efficiency cost of bunching increases with worker skills. The planner separately bunches together cognitive and manual tasks for workers that are relatively skilled in manual tasks and similarly bunches together the cognitive and manual tasks for workers that are relatively skilled in cognitive tasks. Unlike in the lower triangle, however, workers that are relatively skilled in cognitive labor do not bunch with workers that are skilled in manual labor. In sum, the planner separates on worker's comparative advantage and bunches on worker's comparative disadvantage in the targeted bunching region.

Blunt bunching and targeted bunching can thus be understood in terms of the difference in output costs of distorting the worker task composition. At the low end of the skill distribution, the output cost of distorting the task composition is outweighed by the benefit of incentive provision for higher types, leading to blunt bunching. As manual skills increase, the cost of distorting the production of manual tasks increases. To reduce the distortionary output costs, the planner only bunches higher manual skill types with similarly skilled workers that have a comparative advantage in terms of producing manual tasks. This delivers the targeted bunching regions of workers who are separated on their dominant skill and bunched on their weaker skill.[24]

In the regions without bunching, the planner distorts allocations to deter workers from deviating to allocations they find desirable, as in the unidimensional case. This is incentive provision through distinct distorted allocations.

Tax Wedges. We next study the implications for optimal taxes by studying the labor wedge (14) for cognitive and manual tasks. Using linear preferences for consumption, the labor wedge is:

$$1-\tau_s = \frac{1}{2}\frac{x_s(\alpha)^{\rho-2}}{z(\alpha)\alpha_s^{\rho}} \qquad \Longrightarrow \qquad \log(1-\tau_s) \propto -\log z + (\rho-2)\log x_s - \rho \log \alpha_s. \tag{44}$$

The labor wedge is determined by the optimal assignment through $\log z$, by the allocation of tasks through $(\rho-2)\log x_s$, and by worker skills through $\rho \log \alpha_s$. Workers at better firms face a higher labor wedge τ_s. If the planner reduces task inputs, the labor wedge increases. Keeping allocations constant, the labor wedge increases with worker skills.

The optimal tax wedges are presented in Figure 11. First, the respective wedge is zero for the worker with the highest respective skill in either the cognitive or manual dimension. Consider the graph for the manual tax in the left panel. Workers with the highest manual skill are those represented by the top horizontal boundary of the graph. The manual tax on them is zero. Note

[24] In the robustness analysis of Numerical Appendix B, the targeted bunching region indeed decreases as manual skills and cognitive skills increase when the type distribution is uniform.

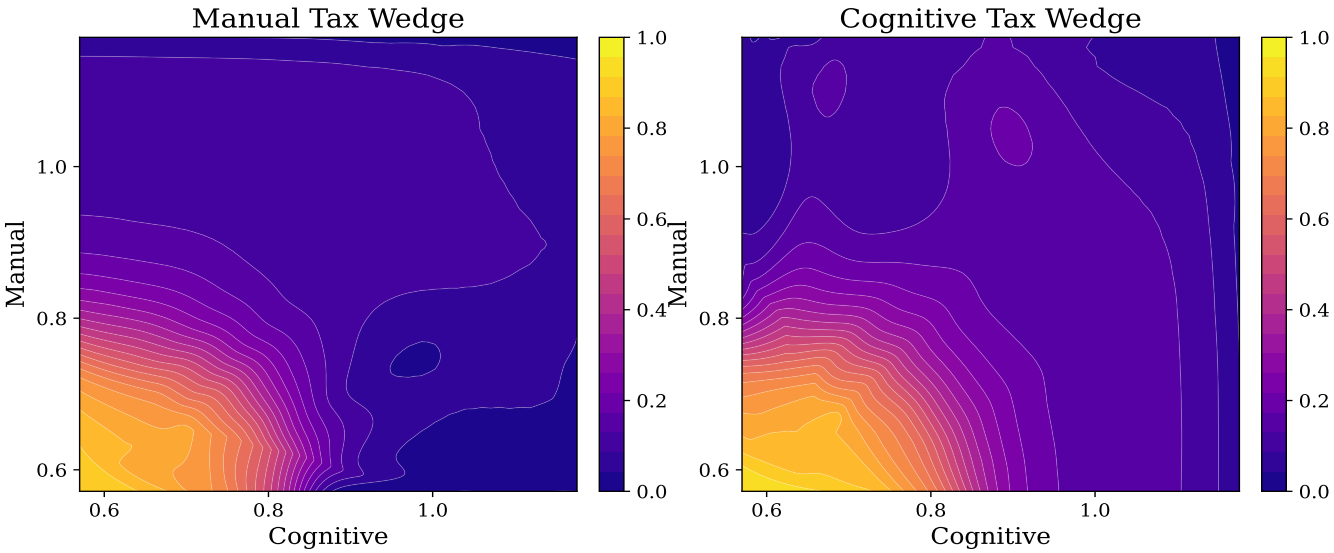


Figure 11: Tax Wedges

Figure 11 visualizes the tax wedges for the planner solution. The left panel displays the manual tax wedge, the right panel displays the cognitive tax wedge.

that for the workers with the top cognitive skill, who are represented as the right vertical boundary, the manual tax is not zero. The manual tax on the best crane operator (highest manual and low cognitive) and physician (highest manual and high cognitive) is zero while it is positive on the top executive (low manual and highest cognitive).

The cognitive tax is displayed in the right panel of Figure 11. The cognitive tax for workers with the highest cognitive skill is zero. For workers with the top manual skill the cognitive tax is not zero. The cognitive tax on top executives (low manual and highest cognitive) and the best physician (highest manual and high cognitive) is zero while it is positive on the best crane operator (highest manual and low cognitive).[25]

Second, we describe how taxes change with the level of skill. Figure 12 plots the manual (on the left) and cognitive taxes (on the right) for three groups of workers separated by their skills corresponding to the heatmap in Figure 11. Consider top manual workers. Those are physicians (high cognitive), carpenters (medium cognitive), and crane operators (low cognitive). These workers are located in the top horizontal strip of Figure 11 and represented by the dash-dotted line in the right panel of Figure 12. These workers are not in bunched regions, and the cognitive tax on them is low.

[25]In our setting with bounded worker skills and project types, the no-distortion-at-the-top result is directly derived from the variation in Appendix A.6, which generates the global optimal tax formula. With a unidimensional worker skill distribution and unbounded project distribution one can generate an unbounded earnings distribution and a top marginal tax rate that converges to a non-zero constant (Scheuer and Werning, 2017).

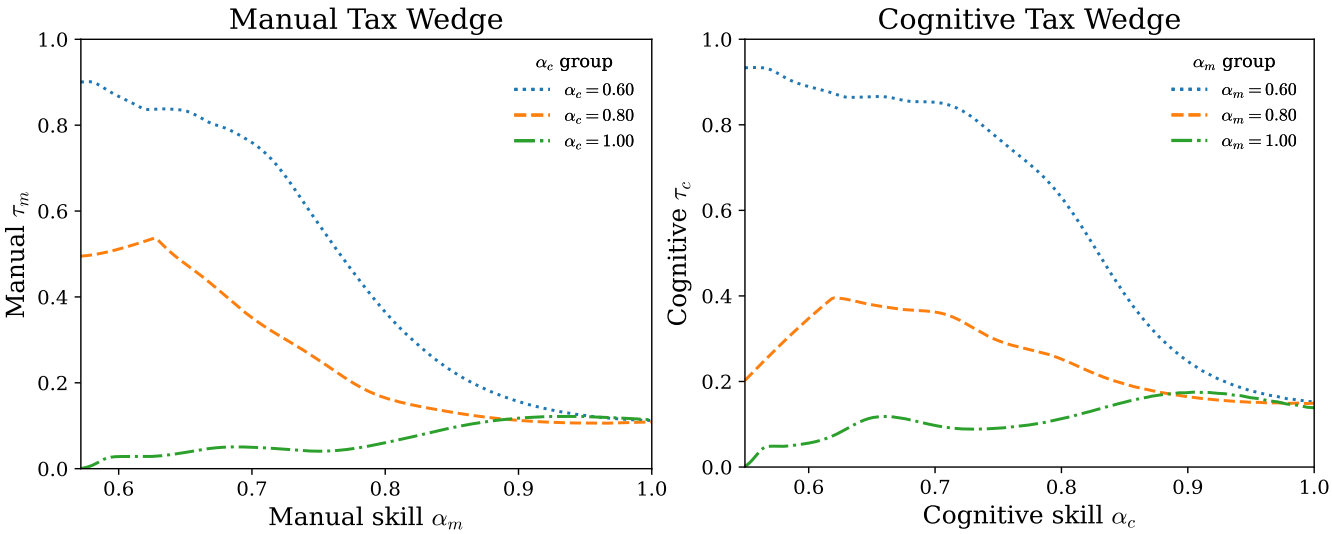


Figure 12: Manual and Cognitive Tax Wedge

Figure 12 describes how the tax wedges vary with the level of worker skill. Figure 12 displays the manual wedge in the left panel, and the cognitive tax wedge in the right panel, for three groups of workers separated by their cognitive and manual skills, respectively.

Consider medium manual workers. Those are lawyers (high cognitive), police officers (medium cognitive), and gardeners (low cognitive). These workers are located in the middle horizontal strip of Figure 11 and are represented by the long-dash line in the right panel of Figure 12. The cognitive tax on them is higher. Three forces give this shape from equation (44) − the assignment to heterogeneous firms, the task allocation, and the worker skill. The assignment is monotonic: lawyers are assigned to better firms than police officers and gardeners. This force gives a higher labor wedge on lawyers compared to the gardeners. Conversely, the task allocation increases with skills leading to a reduction in labor wedges. Finally, the labor wedge increases in worker skills. The low cognitive skill workers (gardeners) face a higher cognitive labor wedge than the highest cognitive skill workers (lawyers) which is hence driven by the low amounts of cognitive tasks they conduct. The high level of tax distortion on gardeners is also driven by the fact that they are in the targeted bunching region while lawyers and police officers are not bunched.

Consider low manual workers. Those are executives (high cognitive), teachers (medium cognitive), and cashiers (low cognitive). These workers are located in the bottom horizontal strip of Figure 11 and are represented by the dotted line in the right panel of Figure 12. The cognitive tax on them is higher than on medium manual workers. The cognitive tax is generally decreasing in skill with executives facing a lower marginal tax rate than cashiers. The high level of distortions on cashiers and teachers is also driven by the fact that they are bunched. The teachers are in the targeted bunching region − they have comparative advantage in their cognitive skills and are

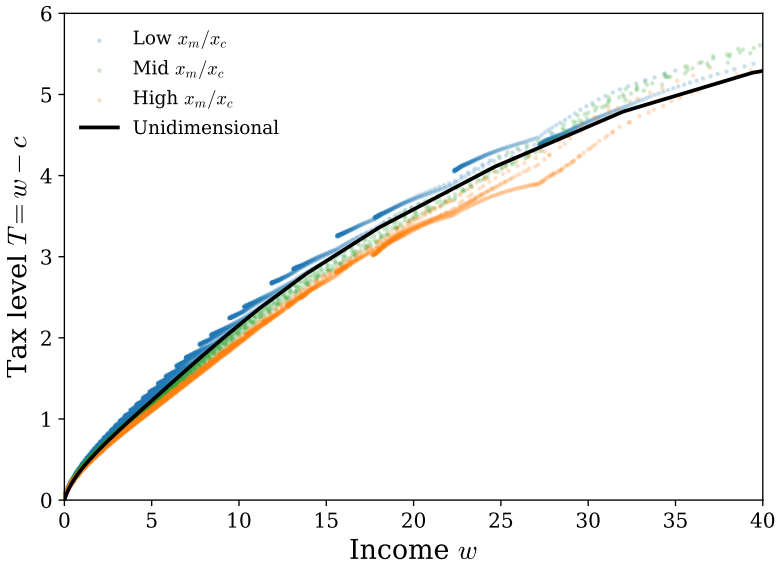


Figure 13: Taxation by Task Mix

Figure 13 shows the tax by income level for different groupings of task mixes, and thus displays how occupations with similar incomes, but different task mixes are taxed. The blue color indicates more cognitively intensive occupations, while the orange color indicates more manually intensive occupations. For a given level of income, occupations that are more manually intensive are taxed less. The solid black line shows the optimal tax system when taxes can only depend on earnings and not on task mixes.

separated along the cognitive dimension but bunched in the manual dimension. The cashiers are in the blunt bunching region. They are bunched in both the manual and the cognitive dimensions and this leads to their allocation being heavily distorted. In the left panel of Figure 12 we repeat the tax analysis for the manual tax wedge, using the left panel of Figure 11.

7.3 Implementation

The planning solution can be implemented with a tax system that only depends on labor income and the line of work x_m/x_c. To establish this result, we first prove in Appendix A.17 that the planner solution is implementable through a tax system over task inputs x. Since the effective worker skills $X = x_c^2 + x_m^2$ determine worker wages $w(X)$ where w is strictly increasing (Section 5), there is a one-to-one mapping between the skill index $x_c^2 + x_m^2$ and earnings w. Therefore, there is a one-to-one map between $(w, x_m/x_c)$ and (x_c, x_m). Since the optimum is implementable by a tax function $\overline{T}(x_m, x_c)$ it is also implementable by a tax function that is only a function of income and line of work $T(w, x_m/x_c)$.

We use the numerical solution to the planner problem to show how the optimal tax system differs by line of work. Figure 13 shows the tax by income level for different groupings of task mixes, and thus displays how occupations with similar incomes, but different task mixes are taxed.

The blue color indicates more cognitively intensive occupations (such as lawyers and executives), while the orange color indicates more manually intensive occupations (such as gardeners and crane operators). For a given level of income, occupations that are more manually intensive are taxed less. The solid black line shows the optimal tax system when taxes can only depend on earnings and not on task mixes. In this case, the optimal tax problem boils down to a unidimensional screening problem, as we show in Appendix A.18. In summary, Figure 13 shows that the optimal tax burden conditioning on line of work tracks the optimal tax burden which only depends on earnings, but lowers the tax burden on more manually intensive occupations.

8 Conclusion

We advance the understanding of optimal tax policy in multidimensional environments with bunching theoretically and quantitatively. Our optimal tax conditions generalize classic unidimensional optimal tax conditions of Mirrlees (1971), Diamond (1998), and Saez (2001) to multidimensional taxation problems and account for global incentive constraints and bunching. For an empirically relevant model, we show that bunching is both substantial and nuanced and, hence, importantly impacts optimal policy design.

References

Abowd, J. M., P. Lengermann, and K. L. McKinney (2003): "The Measurement of Human Capital in the U.S. Economy," U.S. Census Bureau Working Paper.

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.

Ambrosio, L., and N. Gigli (2013): "A User's Guide to Optimal Transport," in *Modelling and Optimisation of Flows on Networks*, pp. 1–155.

Armstrong, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64(1), 51–75.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81(4), 813–846.

Benabou, R. (2002): "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?," *Econometrica*, 70(2), 481–517.

Bertsekas, D. P., and H. Yu (2011): "A Unifying Polyhedral Approximation Framework for Convex Optimization," *SIAM Journal on Optimization*, 21(1), 333–360.

Blundell, R., and A. Shephard (2012): "Employment, Hours of Work and the Optimal Taxation of Low-Income Families," *Review of Economic Studies*, 79(2), 481–510.

Boerma, J., and L. Karabarbounis (2020): "Labor Market Trends and the Changing Value of Time," *Journal of Economic Dynamics and Control*, 115.

——— (2021): "Inferring Inequality with Home Production," *Econometrica*, 89(5), 2517–2556.

Boerma, J., A. Ottolini, and A. Tsyvinski (2025): "Multidimensional Sorting: Comparative Statics," NBER Working Paper No. 34591.

Boerma, J., A. Tsyvinski, and A. P. Zimin (2025): "Sorting with Teams," *Journal of Political Economy*, 133(2), 421–454.

Bogachev, V. I., and A. V. Kolesnikov (2012): "The Monge-Kantorovich Problem: Achievements, Connections, and Perspectives," *Russian Mathematical Surveys*, 67(5), 785–890.

Chen, S., A. Figalli, and Y. R.-Y. Zhang (2026): "Global Regularity and Free Boundary Geometry in the Planar Choné-Rochet Model," ETH Zürich Working Paper.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2012): "Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," *NBER Macroeconomics Annual*, 27, 1–56.

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.

Chiappori, P.-A., R. J. McCann, and B. Pass (2017): "Multi-to One-Dimensional Optimal Transport," *Communications on Pure and Applied Mathematics*, 70(12), 2405–2444.

Conesa, J. C., S. Kitao, and D. Krueger (2009): "Taxing Capital? Not a Bad Idea After All!," *American Economic Review*, 99(1), 25–48.

Deming, D. J. (2017): "The Growing Importance of Social Skills in the Labor Market," *Quarterly Journal of Economics*, 132(4), 1593–1640.

Diamond, P. A. (1998): "Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88(1), 83–95.

Dupuy, A., and A. Galichon (2014): "Personality Traits and the Marriage Market," *Journal of Political Economy*, 122(6), 1271–1319.

Duran, M. A., and I. E. Grossmann (1986): "An Outer-Approximation Algorithm for a Class of Mixed-Integer Nonlinear Programs," *Mathematical Programming*, 36, 307–339.

Ekeland, I., and S. Moreno-Bromberg (2010): "An Algorithm for Computing Solutions of Variational Problems with Global Convexity Constraints," *Numerische Mathematik*, 115(1), 45–69.

Gabaix, X., and A. Landier (2008): "Why has CEO Pay Increased so Much?," *Quarterly Journal of Economics*, 123(1), 49–100.

Galichon, A., and B. Salanié (2022): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," *Review of Economic Studies*, 89(5), 2600–2629.

Gayle, G.-L., and A. Shephard (2019): "Optimal Taxation, Marriage, Home Production, and Family Labor Supply," *Econometrica*, 87(1), 291–326.

Geoffrion, A. M. (1970): "Elements of Large Scale Mathematical Programming Part II: Synthesis of Algorithms and Bibliography," *Management Science*, 16(11), 676–691.

Golosov, M., and I. Krasikov (2025): "The Optimal Taxation of Couples," *Quarterly Journal of Economics*, 140(3), 2163–2211.

Griessler, C. (2018): "C-cyclical Monotonicity as a Sufficient Criterion for Optimality in the Multimarginal Monge-Kantorovich Problem," *Proceedings of the American Mathematical Society*, 146(11), 4735–4740.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2020): "Multidimensional Skill Mismatch," *American Economic Journal: Macroeconomics*, 12(1), 210–44.

Guvenen, F., S. Ozkan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122(3), 621–660.

Heathcote, J., K. Storesletten, and G. L. Violante (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132(4), 1693–1754.

Heathcote, J., and H. Tsujiyama (2021): "Optimal Income Taxation: Mirrlees meets Ramsey," *Journal of Political Economy*, 129(11), 3141–3184.

——— (2026): "Practical Optimal Income Taxation," FRB of Minneapolis Working Paper.

Judd, K., D. Ma, M. A. Saunders, and C.-L. Su (2017): "Optimal Income Taxation with Multidimensional Taxpayer Types," Stanford University Working Paper.

Karunamuni, R. J., and T. Alberts (2005): "A Generalized Reflection Method of Boundary Correction in Kernel Density Estimation," *Canadian Journal of Statistics*, 33(4), 497–509.

Kleven, H. J., C. T. Kreiner, and E. Saez (2006): "The Optimal Income Taxation of Couples," NBER Working Paper No. 12685.

——— (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77(2), 537–560.

Kocherlakota, N. R. (2010): *The New Dynamic Public Finance*. Princeton University Press.

Kremer, M. (1993): "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, 108(3), 551–575.

Lehmann, E., S. Renes, K. Spiritus, and F. Zoutman (2021): "Optimal Taxation with Multiple Incomes and Types," CEPR Discussion Paper No. 16571.

Lindenlaub, I. (2017): "Sorting Multidimensional Types: Theory and Application," *Review of Economic Studies*, 84(2), 718–789.

Lindenlaub, I., and F. Postel-Vinay (2023): "Multidimensional Sorting under Random Search," *Journal of Political Economy*, 131(12), 3497–3539.

Lions, P.-L. (1998): "Identification of the Dual Cone of Convex Functions and Applications," *Comptes Rendus de l'Academie des Sciences*, 12(326), 1385–1390.

Lise, J., and F. Postel-Vinay (2020): "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, 110(8), 2328–76.

McAfee, R. P., and J. McMillan (1988): "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory*, 46(2), 335–354.

McCann, R. J., C. Rankin, and K. S. Zhang (2026): "The Monopolist's Free Boundary Problem in the Plane," University of Toronto Working Paper.

McCann, R. J., and K. S. Zhang (2025): "Remark on Rochet and Choné's Square Screening Example," *Journal of Convex Analysis*, 32(2), 579–584.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38(2), 175–208.

Moser, C., and P. Olea de Souza e Silva (2019): "Optimal Paternalistic Savings Policies," Columbia Business School Working Paper.

Oberman, A. M. (2013): "A Numerical Method for Variational Problems with Convexity Constraints," *SIAM Journal on Scientific Computing*, 35(1), 378–396.

Rochet, J.-C., and P. Choné (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66(4), 783–826.

Rockafellar, R. T. (1999): *Network Flows and Monotropic Optimization*, vol. 9. Athena Scientific.

Rosen, S. (1981): "The Economics of Superstars," *American Economic Review*, 71(5), 845–858.

Rothschild, C., and F. Scheuer (2013): "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, 128(2), 623–668.

——— (2014): "A Theory of Income Taxation under Multidimensional Skill Heterogeneity," NBER Working Paper No. 19822.

——— (2016): "Optimal Taxation with Rent-Seeking," *Review of Economic Studies*, 83(3), 1225–1262.

Roys, N., and C. Taber (2022): "Skill Prices, Occupations, and Changes in the Wage Structure for Low Skilled Men," University of Wisconsin Working Paper.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–229.

Sanders, C., and C. Taber (2012): "Life-cycle Wage Growth and Heterogeneous Human Capital," *Annual Review of Economics*, 4(1), 399–425.

Scheuer, F. (2014): "Entrepreneurial Taxation with Endogenous Entry," *American Economic Journal: Economic Policy*, 6(2), 126–63.

Scheuer, F., and I. Werning (2017): "The Taxation of Superstars," *Quarterly Journal of Economics*, 132(1), 211–270.

Shaked, M., and J. G. Shanthikumar (2007): *Stochastic Orders*. Springer Series in Statistics.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019): "Firming Up Inequality," *Quarterly Journal of Economics*, 134(1), 1–50.

Tervio, M. (2008): "The Difference That CEOs Make: An Assignment Model Approach," *American Economic Review*, 98(3), 642–68.

Tinbergen, J. (1956): "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv*, pp. 155–175.

Yamaguchi, S. (2012): "Tasks and Heterogeneous Human Capital," *Journal of Labor Economics*, 30(1), 1–53.

Bunching and Taxing Multidimensional Skills

Supplemental Appendix

Job Boerma, Aleh Tsyvinski and Alexander Zimin

September 2026

A Proofs

In this appendix, we formally prove the results in the main text.

A.1 Assignment

The planner problem is to choose an allocation $\{(c(\alpha), x_c(\alpha), x_m(\alpha))\}_{\alpha \in A}$, where consumption and task inputs are both positive, and an assignment $\gamma \in \Gamma$ to minimize the resource cost of providing welfare $\mathcal{U}$:

$$\int c(\alpha) \mathrm{d}\Phi - \int y(x_1, x_2, z) \mathrm{d}\gamma, \tag{A.1}$$

subject to incentive constraints (7) and the promise keeping condition (8).

The planning problem contains an assignment problem. The planner pairs worker and coworker inputs with firm projects to maximize output given a distribution of worker inputs and firm projects. We show the planner optimally chooses a self-sorted assignment, meaning that workers are paired with identical coworkers, and also show that the planner pairs better teams with more valuable projects.[26]

The assignment problem embedded in the planning problem is to choose an assignment to maximize production given the distribution of worker tasks F_x and the project distribution F_z:

$$\max_{\gamma \in \Gamma} \int y(x_1, x_2, z) \mathrm{d}\gamma. \tag{A.2}$$

We construct an assignment γ that self-sorts workers and coworkers to obtain a unidimensional distribution for team quality, or effective worker skill, $X = x_c^2 + x_m^2$. The assignment γ combines

[26] This assignment problem falls into a class of multimarginal, multidimensional optimal transportation problems. Multidimensional skill and the dependence of worker output on coworkers are central to recent advances in sorting models that utilize optimal transport theory to characterize equilibrium (Chiappori, McCann, and Nesheim, 2010; Dupuy and Galichon, 2014; Lindenlaub, 2017; Chiappori, McCann, and Pass, 2017; Galichon and Salanié, 2022).

self-sorting of workers with positive sorting between the effective worker skill X and projects z.[27] This assignment γ solves the assignment problem (A.2).

Proposition 4. *Optimal Sorting.* The planner assignment γ_* is characterized by self-sorting of workers and by positive sorting between team quality and project values.

The proof is presented below.

We now develop the intuition for Proposition 4. Given a firm project, and since the technology for each task in equation (4) is supermodular, the planner optimally wants to positively sort both cognitive and manual inputs to project z. In our economy with multidimensional worker skills, positive sorting within each task is attained by self-sorting. An optimal assignment thus features self-sorting of workers with coworkers within projects z. Given that workers are optimally self-sorted, it remains for the planner to sort self-sorted workers with effective skill X to firms z. Since the effective production technology is supermodular in team quality X and project value z, the optimal assignment features positive sorting between teams and project values. Positive sorting of effective skill X with project values z follows the classical Beckerian analysis (Becker, 1973) as in the taxation framework of Scheuer and Werning (2017).

Given that the assignment features self-sorting, the output per worker produced by a team of two workers supplying task inputs (x_c, x_m) is $\frac{1}{2}z(x_c^2 + x_m^2)$. Aggregate output is $\int y(x_1, x_2, z)\mathrm{d}\gamma = \frac{1}{2}\int z(x_c^2 + x_m^2)\mathrm{d}\Phi$, and the resource cost (A.1) can be written as:

$$\int \Big(c(\alpha) - \frac{1}{2}z(\alpha)\big(x_c^2(\alpha) + x_m^2(\alpha)\big)\Big)\mathrm{d}\Phi. \tag{6}$$

Proof of Proposition 4

To understand the optimal assignment, we consider a discrete version of the problem with identical discrete worker distributions $\{x_{1s}\} = \{x_{2s}\}$, which we denote by F_x, and a discrete firm distribution $\{z_s\}$, which we denote by F_z, for $s = \{1, \ldots, n\}$. The discrete problem is to find an assignment γ to maximize output:

$$\max_{\gamma \in \underline{\Gamma}} \sum \gamma_{ijk}\, y(x_{1i}, x_{2j}, z_k), \tag{A.3}$$

where $\gamma \in \underline{\Gamma} := \{\gamma_{ijk} \geq 0 \,\big|\, \sum_{jk}\gamma_{ijk} = 1, \sum_{ik}\gamma_{ijk} = 1, \sum_{ij}\gamma_{ijk} = 1\}$. We next solve this problem in steps.

[27] Self-sorting is defined with respect to the distribution of effective task inputs that the workers supply, not necessarily with respect to the underlying worker skills α. In the presence of bunching, multiple workers α supply the same task levels x and hence self-sorting of effective tasks may imply matching different α.

First, we prove that without loss we can focus on assignments γ that are symmetric in worker inputs, so that $\gamma_{ijk} = \gamma_{jik}$. Suppose a solution γ is not symmetric, and use that the worker input samples are identical to define another feasible transport plan $\hat{\gamma}$ so that $\hat{\gamma}_{ijk} = \gamma_{jik}$ for all workers and projects. When γ solves the assignment problem, so does $\hat{\gamma}$ because $\sum \hat{\gamma}_{ijk} y(x_{1i}, x_{2j}, z_k) = \sum \gamma_{jik} y(x_{1i}, x_{2j}, z_k) = \sum \gamma_{jik} y(x_{2j}, x_{1i}, z_k) = \sum \gamma_{ijk} y(x_{1i}, x_{2j}, z_k)$, where the second equality follows as the production technology is symmetric in worker inputs, and the third equality follows by relabeling. This implies that the assignment $\frac{1}{2}(\gamma + \hat{\gamma})$ is also a solution, which is indeed a symmetric solution. In summary, for every optimal assignment γ there is a symmetric assignment $\frac{1}{2}(\gamma + \hat{\gamma})$ that also solves the assignment problem. Without loss of generality we can therefore focus on assignments γ that are symmetric in worker inputs.

Second, we prove it is optimal to self-sort workers so that $\gamma_{ijk} \neq 0$ implies that the workers are identical, or $i = j$. Consider an optimal symmetric assignment γ and some project z_k, and denote the joint distribution of workers assigned to this project by $\gamma^k_{ij} := \gamma_{ijk}$. We construct the marginal distributions of workers and coworkers assigned to this project as $\mu^k_{1i} := \sum_j \gamma^k_{ij}$ and $\mu^k_{2j} := \sum_i \gamma^k_{ij}$. Due to the symmetry of the assignment function γ, the worker and coworker distribution within the firm are identical, $\mu^k = \mu^k_1 = \mu^k_2$. Further, we let $\hat{\gamma}^k$ denote the optimal reassignment of workers and coworkers within the project:

$$\max_{\gamma^k \in \tilde{\Gamma}(\mu,\mu)} \sum \gamma_{ij}\, y\big(x_{1i}, x_{2j}, z_k\big), \tag{A.4}$$

that is, $\hat{\gamma}^k$ solves the assignment problem within a firm given worker and coworker distribution μ.

Within firms it is optimal to self-sort workers. Suppose some firm z is assigned some worker and coworker distribution μ, and consider two identical samples from this distribution. The within-firm assignment problem given these identical worker samples $\{x_{ij}\}$ for $i \in \{1, 2\}$ and $j \in \mathcal{J} := \{1, \ldots, J\}$ is to choose an assignment, or equivalently a permutation σ, to maximize output $z \sum_{j \in \mathcal{J}} \big(x^j_{1c} x^{\sigma(j)}_{2c} + x^j_{1m} x^{\sigma(j)}_{2m}\big)$. Using the rearrangement inequality, aggregate output is bounded by:

$$\max_{\sigma}\ z \sum_{j \in \mathcal{J}} \big(x^j_{1c} x^{\sigma(j)}_{2c} + x^j_{1m} x^{\sigma(j)}_{2m}\big) \leq z \sum_{j \in \mathcal{J}} \big(x^j_{1c} x^j_{2c} + x^j_{1m} x^j_{2m}\big) = z \sum_{j \in \mathcal{J}} \big((x^j_c)^2 + (x^j_m)^2\big). \tag{A.5}$$

The final equality follows as the worker and coworker distributions are identical. We conclude that self-sorting within every project attains maximum production. The rearrangement inequality implies optimality of positively sorting the skills of workers within each firm as the production technology for each unidimensional task is supermodular as in Becker (1973). In our environment with multidimensional skills positive sorting within each task is indeed attained by self-sorting, implying that $\hat{\gamma}^k$ is a diagonal matrix.

To formally establish that the optimal assignment function features self-sorting for each firm, also with continuous marginal distributions μ, we observe that (A.5) implies that the self-sorting set $M \subset X \times X$ is c-monotone (see, e.g., Bogachev and Kolesnikov (2012) or Ambrosio and Gigli (2013)).

Definition. The set M is c-monotone if for all pairings $(x_{11}, x_{21}), (x_{12}, x_{22}), \ldots, (x_{1n}, x_{2n}) \in M$:

$$\sum_{j \in \mathcal{J}} y(x_{1j}, x_{2j}) \geq \sum_{j \in \mathcal{J}} y(x_{1j}, x_{2\sigma(j)}) \tag{A.6}$$

for any permutation σ.

The c-monotonicity condition directly implies the weaker condition that the matching set M is c-cyclically monotone, or $\sum_{j \in \mathcal{J}} y(x_{1j}, x_{2j}) \geq \sum_{j \in \mathcal{J}} y(x_{1j}, x_{2j+1})$, where $x_{2n+1} = x_{21}$. The self-sorting assignment γ_z with support on matching set M is optimal as this statement is equivalent to the support of γ_z being c-cyclically monotone following Theorem 1.2.7 in Bogachev and Kolesnikov (2012) or Theorem 1.13 in Ambrosio and Gigli (2013).

Given optimal self-sorting of workers within each firm, we construct a diagonal assignment $\hat{\gamma}$ by replacing γ^k with the optimal self-sorted $\hat{\gamma}^k$ for every project k. Since $\hat{\gamma}^k$ solves the assignment problem within each firm, $\sum \hat{\gamma}_{ijk} y(x_{1i}, x_{2j}, z_k) \geq \sum \gamma_{ijk} y(x_{1i}, x_{2j}, z_k)$. By the construction of $\hat{\gamma}$, it holds that $\hat{\gamma}_{ijk} \neq 0$ implies that workers are identical $i = j$ for any k. Without loss of generality, an optimal assignment indeed features self-sorting of workers with coworkers within projects z. We define effective worker skills, or a team's quality, by $X := x_c^2 + x_m^2$.

Finally, the optimal assignment sorts the best teams with the most valuable firm projects. Since the optimal assignment of workers within firms is self-sorting, the Kantorovich problem (A.3) simplifies to finding assignment $\gamma_{ik} \in \underline{\Gamma} := \{\gamma_{ik} \geq 0 \,\big|\, \sum_k \gamma_{ik} = 1, \sum_i \gamma_{ik} = 1\}$ to solve:

$$\max_{\gamma \in \underline{\Gamma}} \sum \gamma_{ik}\, y(X_i, z_k), \tag{A.7}$$

that is, to assign teams to firms. Given that the reduced-form production technology is supermodular, the solution to this problem is a positive sorting between the team quality X_i and the project value z_k. The solution to the original multimarginal Kantorovich problem (A.3) is then constructed using $\gamma_{ijk} = \gamma_{ik}\delta_{ij}$, where δ is the Kronecker delta function.

While we constructed the solution to the Kantorovich formulation of the assignment problem in the main text, we observe that the optimal assignment to the discrete planning problem γ_{ijk} is a Monge solution, meaning $\gamma_{ijk} \in \{0, 1\}$. This means the optimal assignment is a solution to the

discrete planning problem of choosing permutations σ_i, to maximize output

$$\max_{\sigma_1,\sigma_2} \sum_s y\big(x_{1\sigma_1(s)}, x_{2\sigma_2(s)}, z_s\big). \tag{A.8}$$

given the identical worker samples $\{x_{is}\}$ drawn from the distribution F_x and a firm project sample $\{z_s\}$ drawn from the distribution F_z for $s = \{1, \ldots, n\}$.

Before proceeding, we observe that a transport problem with two identical worker distributions F_x with measure one for each role is equivalent to a transport problem with a single distribution of workers F_x with measure two. Owing to the symmetry of workers' skills in production (4), these are equivalent. Intuitively, any assignment for a problem with distinct worker distributions can be made symmetric. Consider assignment γ that solves the discrete assignment problem for distinct worker distributions F_{x_1} and F_{x_2}. The transpose of the assignment γ along the worker input dimensions, which we denote by γ', solves the discrete assignment problem with worker input distributions F_{x_2} and F_{x_1}. This implies that symmetric assignment $\hat{\gamma} := \frac{\gamma+\gamma'}{2}$ solves the discrete assignment problem with worker distributions $F_x = \frac{1}{2}(F_{x_1} + F_{x_2})$.

Continuous Distributions. To obtain the solution for continuous distributions of workers and coworkers, we extend our argument for the discrete distributions. We construct an assignment γ that self-sorts workers and coworkers to obtain a unidimensional distribution for team quality X. The assignment γ combines self-sorting of workers with positive sorting between worker skill index X and projects z.

This assignment γ solves the Kantorovich problem (A.2). To prove this claim, denote the support of the assignment by M, the matching set. Consider a collection of points within the matching set, $\{(x_{1s}, x_{2s}, z_s)\} \in M$, then for each of those points it holds $x_{1s} = x_{2s}$, and that $z_s \leq z_{s'}$ implies $X_s \leq X_{s'}$. Since the support is constructed by using a Monge solution for the discrete assignment problem, $\sum_s y(x_{1\sigma_1(s)}, x_{2\sigma_2(s)}, z_s) \leq \sum_s y(x_{1s}, x_{2s}, z_s)$ for all permutations σ_1, σ_2. Equivalently, the matching set M is c-monotone. By Theorem 1.2 in Griessler (2018), the assignment γ solves the Kantorovich problem. This concludes the proof of Proposition 4.

A.2 Incentive Constraints

We first discuss the differentiability of the indirect utility function u. Feasibility of an allocation is exactly equivalent to the indirect utility function u being feasible, which means that u is convex, nonnegative, and decreasing, together with the additional constraint $-x(p) \in \partial u(p)$, where ∂u denotes the subdifferential of u at a given point.

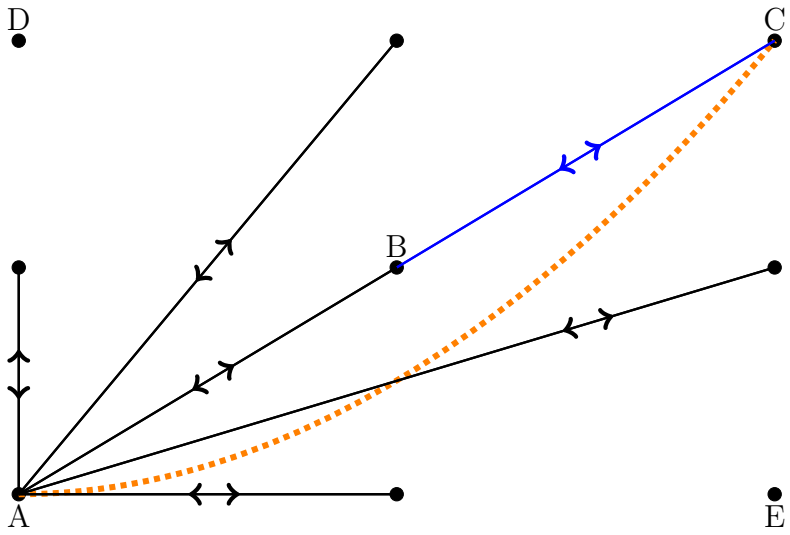


Figure A.1: Reducible and Irreducible Incentive Constraints

Figure A.1 shows reducible and irreducible incentive constraints for worker A. When the irreducible incentive constraints between workers A and B are satisfied (as indicated by the black solid line between workers A and B), and the irreducible incentive constraints between workers B and C are satisfied (as indicated by the blue solid line between workers B and C), then reducible incentive constraints between A and C are satisfied (as indicated by the orange dashed line). Every reducible incentive constraint is satisfied when the irreducible constraints are. The black solid lines denote all the irreducible incentive constraints for worker A.

Consider the incentive constraint (10):

$$c(p) - p_c x_c(p) - p_m x_m(p) \geq c(p') - p_c x_c(p') - p_m x_m(p')$$

Using the notation for the indirect utility function (12), that is, $u(p) = c(p) - p_c x_c(p) - p_m x_m(p)$, this can be equivalently rewritten as

$$u(p) - u(p') \geq \langle p - p', -x(p') \rangle.$$

This inequality for each pair of types (p, p') is equivalent to the convexity of u together with the constraint that $-x(p) \in \partial u(p)$, where ∂u denotes the subdifferential of u at a given point. In this case, we know by Alexandrov theorem that u, since it is a convex function, is differentiable almost everywhere, and hence $\partial u(p) = \nabla u(p)$ almost everywhere.

We next show which incentive compatibility constraints are redundant to the planner problem in the numerical analysis. We establish that every reducible incentive constraint is redundant in the presence of the irreducible constraints, which shrinks the set of incentive constraints that need to be taken into account by the planner. To show this result, we let $L \subseteq \mathbb{R}^2$ be a finite subset and we first define irreducible constraints.

Definition. A couple of points $(p, q) \subseteq L$ is irreducible if there is no point $m \in L$ on the interval between p and q.

Lemma 4. All reducible incentive constraints are implied by irreducible incentive constraints.

The proof is presented below.

Figure A.1 shows the reducible and the irreducible incentive constraints for worker A. When the irreducible incentive constraints between workers A and B are satisfied (as indicated by the black solid line between workers A and B), and the irreducible incentive constraints between workers B and C are satisfied (as indicated by the blue solid line between workers B and C), then reducible incentive constraints between A and C are satisfied (as indicated by the orange dashed line). Every reducible incentive constraint is satisfied when the irreducible constraints are. The black solid lines denote the irreducible incentive constraints for worker A. The incentive constraints between workers A and D as well as between workers A and E are also reducible.

We next establish that no other incentive constraints can be eliminated a priori. The set of feasible allocations strictly increases by removing any of the irreducible incentive constraints.

Lemma 5. Consider any irreducible incentive constraint where worker type p_0 does not want to report to be of type q_0. If we eliminate such an incentive constraint, then there exists an allocation that satisfies all other incentive constraints while worker type p_0 wants to report q_0. That is, for any irreducible pair $(p_0, q_0) \subseteq L$ there exist functions (u, x) such that:

$$u(p) - u(q) \geq \langle p - q, -x(q) \rangle$$

for all $(p, q) \in L$ where $(p, q) \neq (p_0, q_0)$ and

$$u(p_0) - u(q_0) < \langle p_0 - q_0, -x(q_0) \rangle.$$

The proof is presented below.

We denote the set of utility allocations that satisfy the set of irreducible linear incentive constraints by $\mathcal{I}$.

Proof of Lemma 4. Let v be a ray from a parameter point $p = (p_c, p_m)$, and let scalar parameters λ and β such that $0 < \lambda < \beta$. We consider points $p + \lambda v$ and $p + \beta v$.

We first show that if incentive constraints between points p and $p + \lambda v$ as well as $p + \lambda v$ and $p + \beta v$ are satisfied, then incentive constraints between p and $p + \beta v$ are implied. By (13), we

know the incentive constraint that p does not want to report $q = p + \lambda v$ implies:

$$u(p) - u(p + \lambda v) \geq -\lambda \langle v, \partial u(p + \lambda v) \rangle = \lambda \langle v, x(p + \lambda v) \rangle. \tag{A.9}$$

Similarly, the incentive constraint that $q = p + \lambda v$ does not want to report p implies:

$$u(p + \lambda v) - u(p) \geq \lambda \langle v, \partial u(p) \rangle = -\lambda \langle v, x(p) \rangle$$

Adding these two constraints, we obtain $\langle v, x(p) \rangle \geq \langle v, x(p + \lambda v) \rangle$.

Using the incentive constraint that $p + \beta v$ does not want to report $p + \lambda v$:

$$u(p + \beta v) - u(p + \lambda v) \geq -(\beta - \lambda) \langle v, x(p + \lambda v) \rangle, \tag{A.10}$$

we show that given these incentive constraints, the constraint between points $p + \beta v$ and p is implied. We evaluate

$$\begin{aligned} u(p + \beta v) - u(p) &= \big(u(p + \beta v) - u(p + \lambda v)\big) + \big(u(p + \lambda v) - u(p)\big) \\ &\geq -(\beta - \lambda) \langle v, x(p + \lambda v) \rangle - \lambda \langle v, x(p) \rangle \\ &\geq -(\beta - \lambda) \langle v, x(p) \rangle - \lambda \langle v, x(p) \rangle = -\beta \langle v, x(p) \rangle \end{aligned}$$

where the first inequality follows from the first and third incentive constraint, while the second inequality follows as $\langle v, x(p) \rangle \geq \langle v, x(p + \lambda v) \rangle$. The final equality indeed implies that $p + \beta v$ does not want to report p.

Similarly, we use the incentive constraint that $p + \lambda v$ does not want to report $p + \beta v$:

$$u(p + \lambda v) - u(p + \beta v) \geq -(\lambda - \beta) \langle v, x(p + \beta v) \rangle \tag{A.11}$$

in order to prove that p does not want to report $p + \beta v$:

$$\begin{aligned} u(p) - u(p + \beta v) &= \big(u(p) - u(p + \lambda v)\big) + \big(u(p + \lambda v) - u(p + \beta v)\big) \\ &\geq \lambda \langle v, x(p + \lambda v) \rangle - (\lambda - \beta) \langle v, x(p + \beta v) \rangle \\ &\geq \lambda \langle v, x(p + \beta v) \rangle - (\lambda - \beta) \langle v, x(p + \beta v) \rangle = \beta \langle v, x(p + \beta v) \rangle \end{aligned}$$

where the final inequality follows by adding (A.10) and (A.11), which implies $\langle v, x(p + \lambda v) \rangle \geq \langle v, x(p + \beta v) \rangle$. This shows we do not need to incorporate incentive constraints between p and $p + \beta v$ when we incorporate the incentive constraint between p and $p + \lambda v$, and between $p + \lambda v$ and $p + \beta v$.

The final step is that our result so far held for general scalar parameters λ and β such that $0 < \lambda < \beta$. We note that for $\underline{\lambda}$ so that $0 < \underline{\lambda} < \lambda < \beta$, we can show that the constraints between p and $p + \lambda v$ are implied by the constraints between p and $p + \underline{\lambda} v$ as well as $p + \underline{\lambda} v$ and $p + \lambda v$. Hence, for every point p we only need to consider the constraints for the lowest possible values for λ. These constraints are irreducible.

Proof of Lemma 5. By induction. The induction base is a set of points in L with $|L| = 2$. Let the points within the set be p_0 and q_0. We show there exist functions (u, x) so that:

$$u(q_0) - u(p_0) \geq \langle q_0 - p_0, -x(p_0) \rangle$$
$$u(p_0) - u(q_0) < \langle p_0 - q_0, -x(q_0) \rangle.$$

Construct the function $x(p_0) = x(q_0) = 0$, and $u(p_0) = 0$ and $u(q_0) = 1$.

Induction step for $|L| = n + 1$ points. Let z denote a vertex of the convex hull of set L which is neither p_0 nor q_0. Such a point indeed exists, else the convex hull is an interval between p_0 and q_0, implying that any other point of the set L would be a point between p_0 and q_0 contradicting that (p_0, q_0) is irreducible.

Remove the point z from the set L. By induction step at n, there exist functions (u, x) for the lattice $L \setminus \{z\}$ such that for the same irreducible pair $(p_0, q_0) \subseteq L \setminus \{z\}$:

$$u(p) - u(q) \geq \langle p - q, -x(q) \rangle$$

for all $(p, q) \in L \setminus \{z\}$ and $(p, q) \neq (p_0, q_0)$ and

$$u(p_0) - u(q_0) < \langle p_0 - q_0, -x(q_0) \rangle.$$

We need to extend the functions u and x onto the point z. Here, we will use that z is a vertex of the convex hull. We construct the value for the functions u and x at point z. At the point z, we require:

$$u(z) - u(p) \geq \langle z - p, -x(p) \rangle$$
$$u(p) - u(z) \geq \langle p - z, -x(z) \rangle$$

for all $p \in L \setminus \{z\}$. Reorganizing, the first inequality becomes

$$u(z) \geq \max_{p \neq z} \left\{ u(p) + \langle z - p, -x(p) \rangle \right\} =: \mathrm{C}$$

where we observe that the constant C is independent of both $u(z)$ and $x(z)$. We set $u(z) = \mathrm{C}$. As a result, the second inequality is written as:

$$u(p) - \mathrm{C} \geq \langle p - z, -x(z) \rangle.$$

To show that the inequality is satisfied, we use the following variation of Farkas' Lemma. For any convex polytope P and for any vertex v of this polytope, there exists a hyperplane such that v belongs to the hyperplane while all other points of the convex polytope P lie strictly on one side of it. Equivalently, there exists a vector h such that $\langle x - v, h \rangle < 0$ for all $x \in \mathrm{P} \setminus \{v\}$.

Since point z is a vertex of the convex hull of L there exists $\tilde{x}(z)$ so that $\langle p - z, -\tilde{x}(z) \rangle < 0$ for every $p \in L \setminus \{z\}$. Define the constant $\mathrm{C}_p = \langle p - z, -\tilde{x}(z) \rangle < 0$. Then there exists a positive value $t_p > 0$ so that:

$$\langle p - z, -t_p \tilde{x}(z) \rangle = t_p \mathrm{C}_p < u(p) - \mathrm{C}.$$

Further, let $t = \max_{p \in L \setminus \{z\}} t_p$, implying that $\langle p - z, -t\tilde{x}(z) \rangle = t\mathrm{C}_p \leq t_p \mathrm{C}_p < u(p) - \mathrm{C}$ for all $p \in L \setminus \{z\}$. Hence, we set $x(z) = t\tilde{x}(z)$ to conclude our claim.

A.3 Lemma 2

We first prove that if the indirect utility function (12) is strongly convex, then there is no bunching. By the inverse function theorem, using that the indirect utility function u is twice continuously differentiable, if the Jacobian matrix of the mapping from p to x is invertible then the labor task allocation is invertible. The Jacobian matrix of the mapping from p to x is the negative to the Hessian matrix of the indirect utility function $\begin{pmatrix} \partial x_c/\partial p_c & \partial x_c/\partial p_m \\ \partial x_m/\partial p_c & \partial x_m/\partial p_m \end{pmatrix}$ using $x(p) = -\nabla u(p)$. Since the utility function u is strongly convex for worker p, its Hessian matrix is invertible, and hence the Jacobian matrix is. Summarizing, if the utility function is strongly convex, then there is no bunching.

Now we prove that if the indirect utility function (12) is not strongly convex, i.e. the Hessian matrix is degenerate for all workers in the neighborhood of p, worker p is bunched, or $p \in \mathcal{B}$. To prove this statement, consider a mapping f from worker type p to the labor allocation x, and let $\mathcal{P}$ denote the neighborhood of workers around p such that the Hessian matrix $H(u)$ is degenerate for all workers $p \in \mathcal{P}$. Since the Jacobian matrix of the mapping f is the negative to the Hessian matrix of the indirect utility function, the Jacobian matrix is degenerate for all workers $p \in \mathcal{P}$. Equivalently, $\mathcal{P}$ is a critical set. By Sard's theorem it follows that the image $f(\mathcal{P})$ has Lebesgue measure zero.

To prove that worker p is bunched, suppose by contradiction they are not, $p \notin \mathcal{B}$. Equivalently, the mapping f is injective in a neighborhood $\hat{\mathcal{P}} \subseteq \mathcal{P}$. By the invariance of domain, the image $f(\hat{\mathcal{P}})$ is a non-empty open set. This implies that the Lebesgue measure is strictly positive for the image, contradicting the implication from Sard's theorem. Thus, workers are bunched when the optimality condition does not hold.

A.4 Proposition 1

To prove the proposition, we prove Lemma 6 and Lemma 7.

Lemma 6. Let (c, x) solve the planner problem. The following condition holds with equality at an optimum:

$$\int \big(\mathcal{C}'(c)c + z\big(\mathcal{X}'(x_c)x_c + \mathcal{X}'(x_m)x_m\big)\big)\pi \mathrm{d}p = \lambda \int \big(c - p_c x_c - p_m x_m\big)\pi \mathrm{d}p. \tag{A.12}$$

Proof. Consider an allocation (c, x) that satisfies the incentive compatibility constraints. Consider multiplying this allocation by a constant factor $\zeta > 0$, to obtain the scaled allocation $\zeta(c, x)$. The Lagrangian of the scaled allocation exceeds the Lagrangian of the optimal allocation (c, x), or $\mathcal{L}(c, x) \leq \mathcal{L}(\zeta(c, x))$. Therefore, we can consider a variation around the optimal allocation (c, x), where we scale the allocation by a small factor ε, so that the alternative allocation $(c + \varepsilon c, x + \varepsilon x)$ is feasible. Given such a variation, the implied change in the resource cost is:

$$\Delta = \varepsilon\Big(\int \big(\mathcal{C}'(c)c + z\big(\mathcal{X}'(x_c)x_c + \mathcal{X}'(x_m)x_m\big)\big)\pi \mathrm{d}p - \lambda \int \big(c - p_c x_c - p_m x_m\big)\pi \mathrm{d}p\Big) + o(\varepsilon). \tag{A.13}$$

At an optimum, neither a positive ($\varepsilon > 0$) nor a negative ($\varepsilon < 0$) small variation decreases the cost of resources, so $\Delta = o(\varepsilon)$, which establishes (A.12). □

Lemma 7. Let (c, x) solve the planning problem, then the implementability condition (16) holds for any feasible allocation $(\hat{c}, \hat{x}) \in \mathcal{I}$.

Proof. If two allocations (c, x) and $(\hat{c}, \hat{x})$ satisfy the incentive constraints, so does their convex combination $(\tilde{c}, \tilde{x}) = (1 - \varepsilon)(c, x) + \varepsilon(\hat{c}, \hat{x})$ for any $\varepsilon \in (0, 1)$. If (c, x) is a planner solution, it follows that for any $\varepsilon \in (0, 1)$ and for any feasible allocation $(\hat{c}, \hat{x})$, a convex combination of the alternative allocation and the solution increases the Lagrangian value relative to its optimum, $\mathcal{L}(\tilde{c}, \tilde{x}) - \mathcal{L}(c, x) \geq 0$. By construction of the convex combination, this is equivalent to $\mathcal{L}\big((c, x) + \varepsilon\big((\hat{c}, \hat{x}) - (c, x)\big)\big) - \mathcal{L}\big(c, x\big) \geq 0$.

To further develop this, note that $\mathcal{L}((c,x)+\varepsilon((\hat{c},\hat{x})-(c,x)))-\mathcal{L}(c,x)=\varepsilon(\int(\mathcal{C}'(c)(\hat{c}-c)+z\sum\mathcal{X}'(x_s)(\hat{x}_s-x_s))\pi\mathrm{d}p-\lambda\int((\hat{c}-c)-\sum p_s(\hat{x}_s-x_s))\pi\mathrm{d}p)+o(\varepsilon)=\varepsilon(\int(\mathcal{C}'(c)\hat{c}+z\sum\mathcal{X}'(x_s)\hat{x}_s)\pi\mathrm{d}p-\lambda\int(\hat{c}-\sum p_s\hat{x}_s)\pi\mathrm{d}p)+o(\varepsilon)$, where the final equality follows by the optimality condition in Lemma 6. Equation (16) follows because for any $\varepsilon\in(0,1)$ the previous condition is positive, that is:

$$\int\big(\mathcal{C}'(c)\hat{c}+z\big(\mathcal{X}'(x_c)\hat{x}_c+\mathcal{X}'(x_m)\hat{x}_m\big)\big)\pi\mathrm{d}p\geq\lambda\int\big(\hat{c}-p_c\hat{x}_c-p_m\hat{x}_m\big)\pi\mathrm{d}p, \tag{16}$$

for any allocation $(\hat{c},\hat{x})\in\mathcal{I}$. □

A.5 Global Optimal Tax Formula in One Dimension

In this appendix, we develop the connection of our general optimal tax condition with stochastic dominance to the classic ABC formula. First, we consider equation (18) under unidimensional skill heterogeneity. With a slight abuse of notation, we denote the unidimensional skill by p. In this case, equation (18) simplifies to:

$$\int\partial_p\big(\pi(p\mathcal{C}'(c)+z\mathcal{X}'(x))\big)\hat{u}\mathrm{d}p\geq\int\pi(\lambda-C'(c))\hat{u}\mathrm{d}p, \tag{A.14}$$

for any decreasing, nonnegative and convex indirect utility function $\hat{u}$ with $\hat{u}(\bar{p})=0$.[28] Moreover, with one dimension of worker heterogeneity, the measure f second-order stochastically dominates the measure g if and only if $\int_{\underline{p}}^{\hat{p}}F(p)\mathrm{d}p\geq\int_{\underline{p}}^{\hat{p}}G(p)\mathrm{d}p$, where F and G denote cumulative distribution functions (see the next paragraph). When the unidimensional measure $\partial_p\pi(p\mathcal{C}'(c)+z\mathcal{X}'(x))$ second-order stochastically dominates the unidimensional measure $\pi(\lambda-C'(c))$ it thus implies:

$$\int_{\underline{p}}^{p}\frac{\pi(s)}{u'(\mathcal{C}(c(s)))}s\frac{\tau}{1-\tau}\mathrm{d}s\leq\int_{\underline{p}}^{p}\int_{\underline{p}}^{t}\frac{\pi(s)}{u'(\mathcal{C}(c(s)))}\left(1-u'(\mathcal{C}(c(s)))\lambda\right)\mathrm{d}s\mathrm{d}t \tag{A.15}$$

for every worker p, where we use the definition of the skill wedge (14), which changes the inequality sign, and also use that $\mathcal{C}'(c)=1/u'(\mathcal{C}(c))$. At an optimum, the utility-weighted average benefit of increasing marginal tax rates for all workers below p, on the right, exceeds the corresponding costs. The benefit of an increase in a marginal tax rate is an increase in revenues collected from workers below p (high α) net of the cost of tightening the promise keeping constraint, $\int_{\underline{p}}^{t}\frac{\pi(s)}{u'(\mathcal{C}(c(s)))}\left(1-u'(\mathcal{C}(c(s)))\lambda\right)\mathrm{d}s$. The cost of increasing the marginal tax for all workers below p is captured by the marginal utility-weighted labor wedge. Our optimal tax formula as stochastic dominance (18) extends this logic to multidimensional skills.

[28] Asserting there is no bunching at the top of the unidimensional worker skill distribution, both the boundary conditions are zero under the additional condition that $\int\pi C'(c)\mathrm{d}p\geq\lambda$.

Second-Order Stochastic Dominance in One Dimension. Let $\Upsilon_a(p)$ denote a decreasing, nonnegative and convex function parameterized by a that is strictly positive for all $p < a$ and is equal to zero for all $p \geq a$. Specifically, we let $\Upsilon_a(p) := \max(a - p, 0)$. Given that $\Upsilon_a(p)$ is decreasing, nonnegative, and convex, measure f second-order stochastically dominating measure g implies that $\int \Upsilon_a f \mathrm{d}p \geq \int \Upsilon_a g \mathrm{d}p$ for all a following (19). Given the specification for $\Upsilon_a(p)$ this is equivalent to $\int_0^a (a-p) f \mathrm{d}p \geq \int_0^a (a-p) g \mathrm{d}p$ for all a, alternatively $\int F \mathrm{d}p \geq \int G \mathrm{d}p$. Since any unidimensional decreasing, nonnegative and convex indirect utility function $\hat{u}$ with $\hat{u}(\bar{p}) = 0$ can be considered as a positive combination of $\Upsilon_a(p)$, the claim holds.

Relationship to Mirrlees (1971), Diamond (1998), and Saez (2001). Second, we discuss in more detail how the optimal tax condition (22) relates to the ABC formulas in Diamond (1998) and Saez (2001), and to the optimal tax condition in Mirrlees (1971). To see the relationship, we first take the ABC formula in Saez (2001), which is equation (25) in his paper:

$$\frac{\tau_l(\theta)}{1-\tau_l(\theta)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{1 - F(\theta)}{\theta f(\theta)} \int_\theta^\infty \left[1 - \frac{u_c(x)}{p}\right] \frac{u_c(\theta)}{u_c(x)} \frac{f(x)}{1 - F(\theta)} \mathrm{d}x \tag{A.16}$$

Reorganizing this expression, letting $\rho = \left(1 + \frac{1}{\varepsilon}\right)$, we obtain:

$$\frac{1}{\rho} \frac{f(\theta)}{u_c(\theta)} \theta \frac{\tau_l(\theta)}{1-\tau_l(\theta)} = \int_\theta^\infty \left[\frac{1}{u_c(x)} - \frac{1}{p}\right] f(x) \mathrm{d}x, \tag{A.17}$$

which is, in fact, the representation of the optimal tax formula in equation (33) in Mirrlees (1971). Differentiating this expression with respect to type θ, we obtain:

$$\frac{1}{\rho} \partial_\theta \left(\frac{f(\theta)}{u_c(\theta)} \theta \frac{\tau_l(\theta)}{1-\tau_l(\theta)} \right) = \left[\lambda - \frac{1}{u_c(\theta)}\right] f(\theta). \tag{A.18}$$

This is the unidimensional analog of our characterization in equation (22), where we observe that the multiplier p on the resource constraint in Saez (2001) is the inverse of the multiplier on the promise keeping constraint $\frac{1}{\lambda}$, which follows directly from the Lagrangian (15).

A.6 Global Optimal Tax Formula

In this appendix, we derive the global optimal taxation formula (22). For ease of presentation, we assume that the convex indirect utility functions are smooth, meaning the second derivative is well-defined and continuous.

First, we reformulate the planner problem using the definition of the indirect utility function (12). Given the indirect utility function u, both consumption and effort allocations can be expressed in

terms of u and its gradient. The planner chooses an indirect utility function u to minimize the resource cost of providing welfare:

$$\min_{u\in C}\int\left(\mathcal{C}(u(p)-\nabla u(p)\cdot p)+z(p)\left(\mathcal{X}\left(-\frac{\partial u(p)}{p_c}\right)+\mathcal{X}\left(-\frac{\partial u(p)}{p_m}\right)\right)\right)\pi(p)\mathrm{d}p, \tag{A.19}$$

subject to the incentive constraint that requires the indirect utility to be convex and decreasing in worker type p (Lemma 1), or $u\in C$, and the promise keeping condition:

$$\int u(p)\pi(p)\mathrm{d}p\geq\mathcal{U}. \tag{A.20}$$

We introduce multiplier $\lambda\geq 0$ on the promise keeping condition (A.20) in order to formulate the Lagrangian:

$$\min_{u\in C}\int\left(\mathcal{C}(u(p)-\nabla u(p)\cdot p)+z(p)\left(\mathcal{X}\left(-\frac{\partial u(p)}{\partial p_c}\right)+\mathcal{X}\left(-\frac{\partial u(p)}{\partial p_m}\right)\right)\right)\pi\,\mathrm{d}p-\lambda\left(\int u(p)\pi\,\mathrm{d}p-\mathcal{U}\right)$$

In the remainder of this appendix, we refer to the problem of minimizing the Lagrangian shorthand as:

$$\min_{u\in C}J(u)=\min_{u\in C}\int_P L(p,u(p),\nabla u(p))\mathrm{d}p, \tag{A.21}$$

where $L(p,u(p),\nabla u(p))$ is the contribution of worker type p to the Lagrangian and where P denotes the type space, a bounded convex subset of $\mathbb{R}^2$.

A.6.1 Convex Indirect Utility

The restriction that the indirect utility function is convex is equivalent to the Hessian matrix of the indirect utility function being positive semidefinite, $H(u)\succeq 0$ for all $p\in P$, which in turn is equivalent to:

$$v^TH(u)v\geq 0, \tag{A.22}$$

for all $v\in\mathbb{R}^2$. We treat these inequalities as an infinite series of constraints parameterized by both v and p. For each of these constraints, we introduce a corresponding multiplier $\lambda(v,p)\geq 0$. The objective function is augmented to include these multipliers:

$$\min_{u\in C}\ J(u)-\int_P\int_{\mathbb{R}^2}\lambda(v,p)\big(v^TH(u)v\big)\mathrm{d}v\mathrm{d}p. \tag{A.23}$$

Next, fix a worker type p and consider the multipliers associated with this worker type. Using $\langle\cdot,\cdot\rangle$ to denote the inner products of matrices, we write:

$$\int_{\mathbb{R}^2}\lambda(v,p)\big(v^TH(u)v\big)\mathrm{d}v=\left\langle H(u),\int_{\mathbb{R}^2}\lambda(v,p)vv^T\mathrm{d}v\right\rangle. \tag{A.24}$$

This equation expresses the quadratic form as an inner product between the Hessian matrix $H(u)$ and a matrix defined by the integral $\int_{\mathbb{R}^2} \lambda(v,p) vv^T \mathrm{d}v$, which we denote by A.

We next characterize all matrices that can be represented as a convex combination of outer products vv^T over the directions $v \in \mathbb{R}^2$:

$$A = \int_{\mathbb{R}^2} \lambda(v,p) vv^T \mathrm{d}v \tag{A.25}$$

where $\lambda(v,p) \geq 0$ is a non-negative function.

Lemma 8. Matrix A is symmetric and positive semidefinite if and only if $A = \int_{\mathbb{R}^2} \lambda(v,p) vv^T \mathrm{d}v$ with respect to some function $\lambda(v,p) \geq 0$.

Proof. We first establish that $A = \int_{\mathbb{R}^2} \lambda(v,p) vv^T \mathrm{d}v$ with respect to some function $\lambda(v,p) \geq 0$ implies A is symmetric and positive semidefinite. Matrix A is symmetric as each outer product vv^T is a symmetric matrix, and any linear combination or integral of symmetric matrices is symmetric. Second, A is positive semidefinite because for any vector $w \in \mathbb{R}^2$, $w^T A w = \int_{\mathbb{R}^2} \lambda(v,p)(w^T v)^2 \mathrm{d}v \geq 0$, which shows A is positive semidefinite.

Next, we show the converse also holds. We start with a matrix that is symmetric and positive semidefinite and show that it can be written in the form of equation (A.25). Any symmetric matrix $A \in \mathbb{R}^{2\times 2}$ can be factorized using its eigenvalue decomposition. If A is a positive semidefinite matrix, it can be written as:

$$A = Q \Lambda Q^T \,,$$

where $Q \in \mathbb{R}^{2\times 2}$ is an orthogonal matrix whose columns are the eigenvectors of A. Since the matrix is positive semidefinite, it follows that Λ is a diagonal matrix with nonnegative eigenvalues $\lambda_i \geq 0$.

Since each eigenvalue λ_i corresponds to an eigenvector q_i, we rewrite A as a finite sum of outer products of the eigenvectors:

$$A = \sum_{i=1}^{2} \lambda_i q_i q_i^T.$$

The sum can be generalized to an integral, with the eigenvectors q_i replaced by the corresponding vectors $v \in \mathbb{R}^2$, and the eigenvalues λ_i replaced by the corresponding continuous nonnegative function $\lambda(v,p)$. In sum, a positive semidefinite matrix A can be expressed in the form:

$$A = \int_{\mathbb{R}^2} \lambda(v) vv^T \mathrm{d}v,$$

where $\lambda(v,p) \geq 0$. □

Instead of considering the continuous family of multipliers $\lambda(v,p)$, we represent the constraint that the indirect utility function has to be convex as a matrix condition by introducing the Kuhn-Tucker matrix $M(p)$ for each $p \in P$. By Lemma 8, the matrix $M(p)$ is required to be positive semidefinite, or $M(p) \succeq 0$, for all $p \in P$. The Kuhn-Tucker matrix substitutes the term $\int_{\mathbb{R}^2} \lambda(v,p) vv^T \mathrm{d}v$ in the objective function (A.23), where $\lambda(v,p) \geq 0$:

$$\min_u \; J(u) - \int_P \big\langle H(u), M(p) \big\rangle \mathrm{d}p, \tag{A.26}$$

where $M(p)$ is the positive semidefinite matrix that enforces the convexity of the indirect utility function.

We proceed by integrating by parts the term:

$$\int_P \big\langle H(u), M(p) \big\rangle \mathrm{d}p = \int_P \sum_{i,j} \frac{\partial^2 u(p)}{\partial p_i \partial p_j} M_{ij}(p) \mathrm{d}p. \tag{A.27}$$

Through integration by parts, we shift the derivatives from the indirect utility function to the Kuhn-Tucker matrix to obtain:[29]

$$\int_P \sum_{i,j} \frac{\partial^2 u(p)}{\partial p_i \partial p_j} M_{ij}(p) \mathrm{d}p = \int_P \sum_{i,j} \frac{\partial^2 M_{ij}(p)}{\partial p_i \partial p_j} u(p) \mathrm{d}p = \int_P u(p) \Delta M(p) \mathrm{d}p \,, \tag{A.28}$$

where $\Delta M(p) = \sum_{i,j} \frac{\partial^2 M_{ij}(p)}{\partial p_i \partial p_j}$. The resulting objective function for the problem is:

$$\min_u \int_P L(p, u(p), \nabla u(p)) \, \mathrm{d}p - \int_P u(p) \Delta M(p) \, \mathrm{d}p. \tag{A.29}$$

We can rewrite this by combining the promise keeping constraint with the convexity correction $u(p)\Delta M(p)$ as:

$$\begin{aligned} \min_{u \in C} \int \Bigg(\mathcal{C}(u(p) - \nabla u(p) \cdot p) + z(p) \left(\mathcal{X}\left(-\frac{\partial u(p)}{\partial p_c} \right) + \mathcal{X}\left(-\frac{\partial u(p)}{\partial p_m} \right) \right) \Bigg) \pi \, \mathrm{d}p \\ - \lambda \left(\int \pi u(p) \left(1 + \frac{\Delta M(p)}{\lambda \pi} \right) \mathrm{d}p - \mathcal{U} \right) \end{aligned} \tag{A.30}$$

This representation shows that the requirement that the indirect utility function is convex leads to the modified social welfare weight $1 + \frac{\Delta M(p)}{\lambda \pi} = 1 + \frac{\sum \frac{\partial^2}{\partial p_i \partial p_j} M_{ij}(p)}{\lambda \pi}$. In other words, the main difference that convexity adds to the planning problem is through modifying the welfare function by the convexity correction. We next show how this result carries over to the optimality conditions.

[29] Since boundary terms do not affect the derivation of the optimality condition and the general optimal taxation formula, we suppress them for ease of exposition. Similarly, we suppress non-negativity constraints on consumption as this constraint only binds for the least productive type.

A.6.2 Optimality Conditions

We derive the optimality conditions by using the objective function (A.29), and by considering a small variation in the indirect utility function. The first variation of the objective gives:

$$\frac{\partial L}{\partial u} - \sum_{k=1}^{2} \frac{\partial}{\partial p_k}\left(\frac{\partial L}{\partial u_k}\right) = \Delta M(p), \tag{A.31}$$

where $u_k = \frac{\partial u}{\partial p_k}$. The left-hand side is the standard optimality condition, $\frac{\partial L}{\partial u} - \sum \frac{\partial}{\partial p_k}\left(\frac{\partial L}{\partial u_k}\right)$. This yields the optimality condition without bunching, $\frac{\partial L}{\partial u} - \sum \frac{\partial}{\partial p_k}\left(\frac{\partial L}{\partial u_k}\right) = 0$, when the convexity constraint does not bind.

The right-hand side gives the additional term involving the second derivatives of the Kuhn-Tucker matrix. This term arises from the integration by parts of the Kuhn-Tucker matrix, and represents the effect of the convexity constraint, which is enforced through the positive semidefinite matrix M.

Our derivation shows that the minimizer of our variational problem over convex indirect utility functions satisfies an optimality condition with an additional term arising from the Kuhn-Tucker multipliers associated with the convexity constraint. Our results build on Lions (1998), which analyzes variational problems over convex functions through a duality approach. Lions (1998) shows that the optimality conditions can be understood in terms of the polar cone of convex functions, where elements of the dual space are represented by measures linked to second derivatives. Our result explicitly incorporates Kuhn-Tucker multipliers into the variational problem to enforce the convexity constraint and includes them into the optimality condition. This construction directly shows the effect of convexity by showing how additional measure terms arise in the optimality conditions.

A.6.3 Optimal Tax Formula

We next apply the optimality condition for the general Lagrangian (A.31) to the Lagrangian for the optimality multidimensional taxation problem (A.19). As a result, we write the optimal taxation formula as:

$$\partial_{p_c}\left(\pi(p_c\mathcal{C}'(c) + z\mathcal{X}'(x_c))\right) + \partial_{p_m}\left(\pi(p_m\mathcal{C}'(c) + z\mathcal{X}'(x_m))\right) = \pi(\lambda - C'(c)) + \Delta M(p). \tag{A.32}$$

Similar to our reformulation of the general optimal tax formula as stochastic dominance (20), we use the definition of the skill wedge (14) to rewrite the optimal taxation formula as:

$$\partial_{p_c}\left(\frac{\pi}{u'(\mathcal{C}(c))}p_c\frac{\tau_c}{1-\tau_c}\right) + \partial_{p_m}\left(\frac{\pi}{u'(\mathcal{C}(c))}p_m\frac{\tau_m}{1-\tau_m}\right) = \pi\left(\frac{1}{u'(\mathcal{C}(c))} - \lambda\right) - \Delta M(p), \tag{A.33}$$

which is the optimal taxation formula (22).

A.7 Corollary 2

We start with the region of strong convexity of the indirect utility function u and, hence, a region without bunching. To analyze properties of optimal tax distortions, we use a perturbation function. Specifically, we construct a variation of the indirect utility function for a specific worker p. Consider a worker p in the interior of the type space such that both the assignment function z and the distribution of worker types π are differentiable in a neighborhood around this worker. Moreover, suppose that the strongly convex utility function u is twice continuously differentiable within a neighborhood of the worker p.

Consider an arbitrary perturbation of the indirect utility u denoted $\hat{u} = u + \varepsilon V$, where V is a bump function that is concentrated in a small ball around p which lies within the neighborhood around p, and ε is small. The arbitrary perturbation function $u + \varepsilon V$ is convex for small enough values for ε within the support of the bump function, $|\varepsilon| < \bar{\varepsilon}$. Intuitively, if the underlying utility function is strongly convex, a small enough additive perturbation preserves convexity.[30]

The perturbation function is convex, positive and non-increasing, and therefore implementable (17). Since the implementability condition (17) is linearly separable and holds with equality for an optimal utility function by Proposition 1, the implementability also has to be satisfied for εV for all $|\varepsilon| \leq \bar{\varepsilon}$. Since ε can take either positive or negative values, the implementability condition holds with equality with respect to the bump function V:

$$\int \big(\mathcal{C}'(c)\big(V - \nabla V \cdot p\big) - z\mathcal{X}'(x) \cdot \nabla V\big)\pi \mathrm{d}p = \lambda \int V \pi \mathrm{d}p. \tag{A.34}$$

Integrating the left-hand side of this equation by parts and tending the bump function V to the Dirac delta function, we obtain the optimality condition equation in Corollary 2.

Convex Perturbation Function. We establish that the perturbation function is convex. We suppose that the indirect utility function u is strongly convex for interior worker type p and twice continuously differentiable within its neighborhood. Specifically, we suppose that $H(u) - \alpha_I I$ is positive semidefinite for worker p for some $\alpha_I > 0$, where H denotes the Hessian matrix and I denotes the identity matrix.

Since worker p is in the interior of the type space, the indirect utility function is strictly positive and strictly decreasing for worker p. By contradiction, suppose the indirect utility function equals

[30] The proof of this statement is presented below. See *Convex Perturbation Function*.

zero for worker p, $u(p) = 0$. Since the indirect utility function is non-increasing, $u(p + \varepsilon) = 0$ for small enough $\varepsilon \geq 0$, implying that the gradient of the indirect utility function for worker p is equal to zero, $\nabla u(p) = 0$. By implication, consider that the partial derivative of the indirect utility function with respect to cognitive type p_c equals zero, $\frac{\partial}{\partial p_c} u(p) = 0$. Since we consider a partial derivative for a convex function, the partial derivative increases with p_c so that $\frac{\partial}{\partial p_c} u(p_c + \varepsilon_c, p_m) = 0$ for all $\varepsilon_c \geq 0$, or $\frac{\partial^2}{\partial^2 p_c} u(p) = 0$. It hence follows that $H_{cc}(u) = 0$, and hence that $H_{cc}(u) - \alpha_I < 0$ for $\alpha_I > 0$ which contradicts that $H(u) - \alpha_I I$ is positive semidefinite by the Sylvester criterion. We conclude that the utility function is strictly positive and strictly decreasing for interior worker p.

Since the indirect utility function u is strongly convex for worker type p and twice continuously differentiable within its neighborhood, the utility function is strongly convex in this neighborhood. The restriction that $H(u) - \alpha_I I$ is positive semidefinite in a neighborhood around worker type p implies $H(u) - \frac{\alpha_I}{2} I$ is positive semidefinite in the neighborhood around p when the indirect utility function is twice continuously differentiable. Hence, the utility function u is indeed strongly convex in this neighborhood.

We consider a perturbation of the indirect utility u denoted by $u + \varepsilon V$, where V is a bump function that is concentrated in a small ball around p which lies within the neighborhood around p, and ε is small. The arbitrary perturbation function $u + \varepsilon V$ is convex for small enough values for ε within the support of the bump function, $|\varepsilon| < \bar{\varepsilon}$.

While intuitive, we prove that $u + \varepsilon V$ is convex for small enough values for ε within the support of the bump function in two steps. First, we observe that for some $\beta > 0$, it holds that $H(V) - \beta I$ is negative semidefinite and that $H(V) + \beta I$ is positive semidefinite. In the former case, negative semidefinite is equivalent to $x_c^2 V_{cc} + 2x_c x_m V_{cm} + x_m^2 V_{mm} \leq \beta(x_c^2 + x_m^2)$ for any (x_c, x_m). To see this, we first observe $x_c^2 V_{cc} + 2x_c x_m V_{cm} + x_m^2 V_{mm} \leq |x_c|^2 |V_{cc}| + 2|x_c||x_m||V_{cm}| + |x_m|^2 |V_{mm}|$. Furthermore, we use that $2|x_m||x_c| \leq |x_c|^2 + |x_m|^2$ to write $x_c^2 V_{cc} + 2x_c x_m V_{cm} + x_m^2 V_{mm} \leq x_c^2(|V_{cc}| + |V_{cm}|) + x_m^2(|V_{cm}| + |V_{mm}|)$. Therefore, there indeed exists $\beta = \max(|V_{cc}| + |V_{cm}|, |V_{cm}| + |V_{mm}|) > 0$ such that $H(V) - \beta I$ is negative semidefinite. Through a similar argument $H(V) + \beta I$ is positive semidefinite. Given $\beta > 0$, it holds that $\varepsilon H(V) + |\varepsilon| \beta I$ is positive semidefinite for positive ε, and that $\varepsilon(H(V) - \beta I) = \varepsilon H(V) + |\varepsilon| \beta I$ is positive semidefinite for negative ε.

Second, we note that the Hessian matrix for the perturbation function is additively separable, $H(u + \varepsilon V) = H(u) + \varepsilon H(V)$. Since the matrix $H(u) - \frac{\alpha_I}{2} I$ is positive definite, the matrix $H(u + \varepsilon V) - \frac{\alpha_I}{2} I - \varepsilon H(V)$ is positive definite. Finally, since the sum of positive semidefinite matrices

is itself positive semidefinite, it follows that $H(u+\varepsilon V)-\left(\frac{\alpha_I}{2}-|\varepsilon|\beta\right)I$ is positive semidefinite for ε small enough, which confirms that the perturbation function is indeed convex. Following analogous reasoning, the indirect utility function u is also decreasing and positive in a neighborhood around worker p.

Changing Coordinates. To connect our expression to the existing literature, we transform the optimal tax formula into the original type coordinates α. We illustrate this transformation by focusing on the partial derivative with respect to cognitive skill in (23),

$$\partial_{p_c}\left(\frac{\pi}{u'(\mathcal{C}(c))}p_c\frac{\tau_s}{1-\tau_c}\right), \tag{A.35}$$

where π is the probability distribution in the transformed worker space p. To convert this term into the original worker space, we recall the change of coordinates $p_s=\kappa\alpha_s^{-\rho}$, or equivalently $\alpha_s=(\kappa/p_s)^{\frac{1}{\rho}}$, implying that $\mathrm{d}\alpha_s=-\frac{1}{\rho}\frac{\alpha_s}{p_s}\mathrm{d}p_s=-\frac{1}{\kappa\rho}\alpha_s^{\rho+1}\mathrm{d}p_s$.

We first explicitly formulate the relationship between the distribution function in the original worker type space α given by ϕ, and the worker distribution function in transformed coordinates p given by π:

$$\phi(\alpha)\mathrm{d}\alpha_c\mathrm{d}\alpha_m=\phi(\alpha)\frac{\alpha_c^{\rho+1}\alpha_m^{\rho+1}}{(\kappa\rho)^2}\mathrm{d}p_c\mathrm{d}p_m=\pi(p)\mathrm{d}p_c\mathrm{d}p_m,$$

where the distribution function $\pi(p):=\phi(\alpha)\alpha_c^{\rho+1}\alpha_m^{\rho+1}/(\kappa\rho)^2$. As a result, we express (A.35) as:

$$\partial_{p_c}\left(\frac{\pi}{u'(c(\alpha))}\frac{\tau_c}{1-\tau_c}\kappa\alpha_c^{-\rho}\right)=\partial_{p_c}\left(\frac{\phi}{u'(c(\alpha))}\frac{\tau_c}{1-\tau_c}\frac{\alpha_c\alpha_m^{\rho+1}}{\kappa\rho^2}\right), \tag{A.36}$$

Next, by the chain rule we have that $\frac{\partial z}{\partial p_c}=\frac{\partial z}{\partial\alpha_c}\frac{\partial\alpha_c}{\partial p_c}$, which gives:

$$\partial_{p_c}\left(\frac{\pi}{u'(c(\alpha))}\frac{\tau_c}{1-\tau_c}\kappa\alpha_c^{-\rho}\right)=\partial_{\alpha_c}\left(\frac{\phi}{u'(c(\alpha))}\frac{\tau_c}{1-\tau_c}\frac{\alpha_c\alpha_m^{\rho+1}}{\kappa\rho^2}\right)\frac{\partial\alpha_c}{\partial p_c}=-\frac{\alpha_c^{\rho+1}\alpha_m^{\rho+1}}{\kappa^2\rho^3}\partial_{\alpha_c}\left(\frac{\phi}{u'(c(\alpha))}\frac{\tau_c}{1-\tau_c}\alpha_c\right)$$

The derivation for the manual skill term is symmetric, which allows us to summarize the previous two expressions for both tasks as:

$$\partial_{p_s}\left(\frac{\pi}{u'(c(\alpha))}\frac{\tau_s}{1-\tau_s}\kappa\alpha_c^{-\rho}\right)=-\frac{\alpha_c^{\rho+1}\alpha_m^{\rho+1}}{\kappa^2\rho^3}\partial_{\alpha_s}\left(\frac{\phi}{u'(c(\alpha))}\frac{\tau_s}{1-\tau_s}\alpha_s\right), \tag{A.37}$$

Finally, we rewrite the left side of equation (23), using the relation between density functions, as

$$\pi\left(\frac{1}{u'(\mathcal{C}(c))}-\lambda\right)=-\phi(\alpha)\frac{\alpha_c^{\rho+1}\alpha_m^{\rho+1}}{\kappa^2\rho^2}\left(\lambda-\frac{1}{u'(c(\alpha))}\right) \tag{A.38}$$

Combining equation (23) in the worker type space p, with (A.37) and (A.38), we obtain (24) in the worker space α.

A.8 Proposition 3

By Corollary 2 it follows that when the optimality condition does not hold, the Hessian matrix is degenerate for worker p. We next show that the Hessian matrix $H(u)$ is also degenerate for all workers within the neighborhood of p. By contradiction, suppose that in every neighborhood of point p we can find a worker $\hat{p}$ such that its Hessian is non-degenerate, or equivalently, has full rank. By Corollary 2, the optimality condition holds for worker $\hat{p}$. We can thus construct a sequence of points $\{\hat{p}_n\}$ that converges to p. Since the optimality equation is continuous in p, the sequence converges and the optimality equation holds for worker p, which is a contradiction.

A.9 Planner Duality

We prove duality between our cost minimization problem and a welfare maximization problem. The welfare maximization problem is to choose an allocation (c, x) to maximize utilitarian welfare:

$$\int \big(c - p_c x_c - p_m x_m\big)\pi \mathrm{d}p, \tag{A.39}$$

subject to the incentive constraints (10) and the linear resource constraint:

$$\int \big(\mathcal{C}(c) + z(p)\big(\mathcal{X}(x_c) + \mathcal{X}(x_m)\big)\big)\pi \mathrm{d}p \leq R, \tag{A.40}$$

for some exogenous level of federal resources R.

Proposition 5. Let (c, x) solve the cost minimization problem associated with maximum welfare level $\overline{\mathcal{U}}$ so that the minimum resource cost is less than government resources R. Then allocation (c, x) solves the welfare maximization problem given government resources R.

Conversely, if allocation (c, x) solves the welfare maximization problem for resources R and induces welfare $\overline{\mathcal{U}}$, then (c, x) solves the cost minimization problem for $\mathcal{U} = \overline{\mathcal{U}}$.

Proof. First, we establish that the welfare attained by the cost minimization problem and welfare maximization problem are identical. Consider the solution to the cost minimization problem with maximum welfare level $\overline{\mathcal{U}}$ such that the resource cost is below resource level R. Allocation (c, x) satisfies both the incentive constraints and the resource constraints of the welfare maximization problem and is thus a feasible solution to the welfare maximization problem. Welfare in the welfare maximization problem therefore exceeds $\overline{\mathcal{U}}$.

Conversely, take the solution to the welfare maximization and let $\overline{\overline{\mathcal{U}}}$ denote maximum welfare. Consider the allocation (c, x) that solves the welfare maximization problem. The allocation (c, x)

satisfies both the incentive constraints and the promise keeping constraint of the cost minimization problem. Further, the associated resource cost is below resource level R. Hence, $\overline{\mathcal{U}} \geq \overline{\overline{\mathcal{U}}}$, implying that welfare is identical for the two problems, $\overline{\mathcal{U}} = \overline{\overline{\mathcal{U}}}$.

Second, we show duality of allocations. Suppose allocation (c, x) solves the cost minimization problem with maximum welfare level $\mathcal{U}$ such that the cost is below resources R, but the allocation does not solve the welfare maximization problem. Then, there is an alternative allocation $(\hat{c}, \hat{x})$ that solves the welfare maximization problem, is feasible, and attains strictly greater welfare. This implies that there exists a welfare level $\hat{\mathcal{U}} > \mathcal{U}$ so that $(\hat{c}, \hat{x})$ has a cost below resources R, contradicting that $\mathcal{U}$ is the maximum welfare so that the minimum resource cost is below R.

Conversely, suppose allocation $(\hat{c}, \hat{x})$ solves the welfare maximization problem given resources R inducing welfare $\hat{\mathcal{U}}$, but does not solve the cost minimization problem. Then, there exists an alternative allocation (c, x) that solves the cost minimization problem for a welfare level $\mathcal{U} > \hat{\mathcal{U}}$ such that the minimum cost is below resources R. Allocation (c, x) is feasible and attains strictly greater welfare, contradicting that $(\hat{c}, \hat{x})$ solves the welfare maximization problem. □

A.10 Transformed Planner Problem

In this appendix, we analyze the planner problem of choosing an allocation (c, x) to minimize the resource cost of providing welfare as in Section 3.1. Using the Legendre transforms (28) and (29) to linearize the resource costs, the planning problem is equivalent to:

$$\min_{c,x} \max_{\varphi,\psi} \int \Big(\big(\varphi(p)c(p) - \mathcal{C}^*(\varphi(p))\big) + z(p) \sum_s \big(\psi_s(p)x_s(p) - \mathcal{X}^*(\psi_s(p))\big) \Big) \pi(p) \mathrm{d}p \tag{A.41}$$

subject to the set of linear irreducible incentive constraints (10) and the promise keeping constraint (11).

To develop properties of the solution we formulate a Lagrangian, where λ is the multiplier on the promise keeping constraint:

$$\mathcal{L}\big(c, x, \varphi, \psi\big) = \int \Big(\big(\varphi c - \mathcal{C}^*(\varphi)\big) + z \sum_s \big(\psi_s x_s - \mathcal{X}^*(\psi_s)\big) - \lambda \Big(\int \big(c - \sum p_s x_s\big) - \mathcal{U} \Big) \Big) \pi \mathrm{d}p. \tag{A.42}$$

The Lagrangian is a continuous function that is concave-convex. Since the Legendre transform of a convex function is itself convex, the Lagrangian is concave in the distortions (φ, ψ) holding constant the allocations (c, x), and convex in the allocations when holding constant the distortions. Further, since the set of allocations that satisfies the incentive constraints (10) is convex, we can apply the minimax theorem. We use the minimax relationship, $\min_{c,x \in \mathcal{I}} \max_{\varphi \geq 0, \psi \leq 0} \mathcal{L}\big(c, x, \varphi, \psi\big) = \max_{\varphi \geq 0, \psi \leq 0} \min_{c,x \in \mathcal{I}} \mathcal{L}\big(c, x, \varphi, \psi\big)$, to establish Lemma 9.

Lemma 9. For every incentive compatible allocation $(c, x) \in \mathcal{I}$, stochastic dominance has to be satisfied:

$$\int \Big(\varphi c + \sum \psi_s x_s\Big)\pi \mathrm{d}p \geq \lambda \int (c - \sum p_s x_s)\pi \mathrm{d}p. \tag{A.43}$$

The result follows by analyzing $\max_{\varphi \geq 0, \psi \leq 0} \min_{c,x \in \mathcal{I}} \mathcal{L}\big(c, x, \varphi, \psi\big)$. By contradiction, suppose instead that $\int(\varphi c + \sum \psi_s x_s)\pi \mathrm{d}p < \lambda \int (c - \sum p_s x_s)\pi \mathrm{d}p$. Consider an increase in the allocation (c, x) by a constant factor $\zeta > 1$. Since incentive compatible constraints are linear, the alternative allocation $\zeta(c, x)$ is feasible. By increasing the constant factor, $\zeta \to \infty$, optimization would lead to negative infinity, which is not optimal. At the solution to the planning problem, the stochastic dominance condition (A.43) will hold with equality.

Proposition 6. Let (c, x, φ, ψ) solve the planning problem, then the stochastic dominance condition holds with equality at optimum:

$$\int \Big(\varphi c + \sum \psi_s x_s\Big)\pi \mathrm{d}p = \lambda \int \big(c - \sum p_s x_s\big)\pi \mathrm{d}p. \tag{A.44}$$

Proof. To establish the result, we use two problems. First, define the maximization problem:

$$\max_{\varphi,\psi} \ \underline{\mathcal{L}}(\varphi, \psi, \lambda), \tag{A.45}$$

where $\underline{\mathcal{L}}(\varphi, \psi, \lambda) := \min_{c,x} \mathcal{L}(c, x, \varphi, \psi, \lambda)$. Let (φ^*, ψ^*) be a solution to this problem. Similarly, we define a minimization problem:

$$\min_{c,x} \ \bar{\mathcal{L}}(c, x), \tag{A.46}$$

where $\bar{\mathcal{L}}(c, x) := \max_{\varphi,\psi} \mathcal{L}(c, x, \varphi, \psi, \lambda)$, and let (c^*, x^*) be a minimizer to this problem.

Claim 1. We show that for the Lagrangian (A.42) evaluated at the optimum it holds that:

$$\mathcal{L}\big(c^*, x^*, \varphi^*, \psi^*, \lambda\big) = \min_{c,x} \ \max_{\varphi,\psi} \ \mathcal{L}\big(c, x, \varphi, \psi, \lambda\big) = \max_{\varphi,\psi} \ \min_{c,x} \ \mathcal{L}\big(c, x, \varphi, \psi, \lambda\big) \tag{A.47}$$

Proof. Necessarily it holds that $\mathcal{L}(c^*, x^*, \varphi^*, \psi^*, \lambda) \geq \min_{c,x} \mathcal{L}(c, x, \varphi^*, \psi^*, \lambda) = \underline{\mathcal{L}}(\varphi^*, \psi^*, \lambda)$. Since (φ^*, ψ^*) solves the optimization problem, $\underline{\mathcal{L}}(\varphi^*, \psi^*, \lambda) = \max_{\varphi,\psi} \underline{\mathcal{L}}(\varphi, \psi, \lambda) = \max_{\varphi,\psi} \min_{c,x} \mathcal{L}(c, x, \varphi, \psi, \lambda)$, and thus it follows $\mathcal{L}(c^*, x^*, \varphi^*, \psi^*, \lambda) \geq \max_{\varphi,\psi} \min_{c,x} \mathcal{L}(c, x, \varphi, \psi, \lambda)$.

Similarly, note that it necessarily holds that $\mathcal{L}(c^*, x^*, \varphi^*, \psi^*, \lambda) \leq \max_{\varphi,\psi} \mathcal{L}(c^*, x^*, \varphi, \psi, \lambda) = \bar{\mathcal{L}}(c^*, x^*)$. Since the utility allocation (c^*, x^*) is a solution to the minimization problem, $\bar{\mathcal{L}}(c^*, x^*) =$

$\min_{c,x} \bar{\mathcal{L}}(c,x) = \min_{c,x}\max_{\varphi,\psi} \mathcal{L}(c,x,\varphi,\psi,\lambda)$. Combining the previous two statements, we conclude $\mathcal{L}(c^*,x^*,\varphi^*,\psi^*,\lambda^*) \leq \min_{c,x}\max_{\varphi,\psi} \mathcal{L}(c,x,\varphi,\psi,\lambda)$, and hence:

$$\min_{c,x}\max_{\varphi,\psi}\ \mathcal{L}(c,x,\varphi,\psi,\lambda) \geq \mathcal{L}(c^*,x^*,\varphi^*,\psi^*,\lambda) \geq \max_{\varphi,\psi}\min_{c,x}\ \mathcal{L}(c,x,\varphi,\psi,\lambda). \tag{A.48}$$

By the minimax theorem it follows that (A.47) applies. □

Optimality Conditions. We obtain optimality conditions analyzing the planner problem using $\mathcal{L}(c^*,x^*,\varphi^*,\psi^*,\lambda^*) = \max\ \min\ \mathcal{L}(c,x,\varphi,\psi,\lambda)$ from Claim 1. By reorganizing terms:

$$\max_{\varphi,\psi}\min_{c,x\in\mathcal{I}} \int \Big(\varphi c + \sum \psi_s x_s - \lambda\Big(c - \sum p_s x_s\Big) - \mathcal{C}^*(\varphi) - \sum \mathcal{X}^*(\psi_s) - \lambda\mathcal{U}\Big)\pi \mathrm{d}p. \tag{A.49}$$

We observe that only the first four terms depend on the utility allocation, and observe further that these terms are necessarily jointly positive for some utility allocation to be incentive compatible following Lemma 9. Since allocation $(c,x)=0$ is incentive compatible and attains the minimum, the optimal utility allocation is chosen such that these terms jointly equal zero, implying:

$$\int (\varphi^* c^* + \sum \psi_s^* x_s^*)\pi \mathrm{d}p = \lambda \int (c^* - \sum p_s^* x_s^*)\pi \mathrm{d}p, \tag{A.44}$$

and thus concluding the proof. □

To obtain further optimality conditions for our problem, we analyze the planner problem using $\mathcal{L}(c^*,x^*,\varphi^*,\psi^*,\lambda) = \min\ \max\ \mathcal{L}(c,x,\varphi,\psi,\lambda)$ in Claim 1 to write:

$$\min_{c,x}\max_{\varphi,\psi} \int \Big(\varphi c - \mathcal{C}^*(\varphi) + \sum \psi_s x_s - \sum \mathcal{X}^*(\psi_s) - \lambda\big(c - \sum p_s x_s - \mathcal{U}\big)\Big)\pi \mathrm{d}p. \tag{A.50}$$

We observe that only the first four terms depend on the convex conjugates, and that only the final term depends on the multiplier, in terms of the inner maximization problem. Since the promise keeping condition requires $c - \sum p_s x_s \geq \mathcal{U}$, and $\lambda \geq 0$, it has to hold that $\lambda(c^* - \sum p_s^* x_s^* - \mathcal{U}) = 0$. Similarly, it has to hold that $\varphi^* = \mathcal{C}'(c^*)$ and $\psi_s^* = \mathcal{X}'(x_s^*)$.

A.11 Numerical Approach

Linearization of the Problem. We now discuss the linearization of the problem that is central to numerical tractability. The only nonlinear part of the optimization problem that remains to be linearized is the objective:

$$\min \int \big(\mathcal{C}(c(p)) + z(p)\big(\mathcal{X}(x_c(p)) + \mathcal{X}(x_m(p))\big)\big)\,\pi(p)\mathrm{d}p. \tag{A.51}$$

To illustrate our approach, we focus on the linearization of the convex resource cost function for consumption utility $\mathcal{C}$, and we suppose that bounds for the optimal solution are known a priori, or $\underline{c}(p) \leq c(p) \leq \bar{c}(p)$ and $\underline{x}_s(p) \leq x_s(p) \leq \bar{x}_s(p)$.

The idea is to approximate the convex cost for consumption utility $\mathcal{C}$ from below with the tangent lines on the bounded interval. For each worker type p, it follows from the definition of the Legendre transform (28) that $\mathcal{C}(c(p)) = \max_{\varphi} \varphi c(p) - \mathcal{C}^*(\varphi)$. We replace this continuous set of tangent slopes φ in (28) with a finite set of tangent lines. Specifically, we consider a list of slopes $\{\varphi_i(p)\}_{i=1}^n$ with corresponding tangent lines $l_{ip}^c(t) := \varphi_i(p)t - \mathcal{C}^*(\varphi_i(p))$ such that the inequality:

$$0 \leq \mathcal{C}(t) - \max_{1 \leq i \leq n} l_{ip}^c(t) \leq \varepsilon_c \tag{A.52}$$

holds for all t in the bounded interval $[\underline{c}(p), \bar{c}(p)]$. Analogously, to linearize the resource cost of labor disutility $\mathcal{X}$, we consider a list of slopes $\{\psi_i^s(p)\}_{i=1}^n$ with corresponding tangent lines $l_{ip}^s(t) := \psi_i^s(p)t - \mathcal{X}^*(\psi_i^s(p))$ such that the inequality:

$$0 \leq \mathcal{X}(t) - \max_{1 \leq i \leq n} l_{ip}^s(t) \leq \varepsilon_s \tag{A.53}$$

holds for each skill $s \in \mathcal{S}$ and for all t in the interval $[\underline{x}_s(p), \bar{x}_s(p)]$.

As a key step, we next introduce independent auxiliary variables $r(p)$ for each worker p satisfying the following set of linear inequalities for all i:

$$r(p) \geq \varphi_i(p)c(p) - \mathcal{C}^*(\varphi_i(p)). \tag{A.54}$$

It follows from the discussion above that $r(p) \gtrsim \mathcal{C}(c(p))$ for each worker p. For the resource cost of disutility from working, we similarly define independent auxiliary variables $r_s(p)$ satisfying the linear inequalities for all i:

$$r_s(p) \geq \psi_i^s(p)x_s(p) - \mathcal{X}^*(\psi_i^s(p)). \tag{A.55}$$

We substitute the auxiliary variables $r(p)$ and $r_s(p)$ for $\mathcal{C}(c(p))$ and $\mathcal{X}(x_s(p))$ into our nonlinear objective to define the approximate planner problem. The approximate planner problem chooses (c, x_s, r, r_s) to solve:

$$\min \int \big(r(p) + z(p)\big(r_c(p) + r_m(p)\big)\big)\, \mathrm{d}\pi, \tag{A.56}$$

subject to the incentive constraints (10), the promise keeping constraint (11), constraints on the auxiliary variables (A.54) and (A.55), and the approximation bounds for consumption utility $\underline{c}(p) \leq c(p) \leq \bar{c}(p)$ and task outputs $\underline{x}_s(p) \leq x_s(p) \leq \bar{x}_s(p)$.

Accuracy. We next describe the accuracy of the approximate planner problem and provide the algorithm that we use to characterize its solution. The precision of the solution to the approximate planner's problem naturally depends on the accuracy of the prior location of the solution. The criterion we evaluate to ensure that the location is accurate is the absence of binding boundary constraints at the optimal solution. In line with this criterion, we say that a solution is proper when no boundary constraints binds.

Definition. The solution to the approximate problem is proper if the solution is strictly interior, that is $\underline{c}(p) < c(p) < \bar{c}(p)$, $\underline{x}_s(p) < x_s(p) < \bar{x}_s(p)$ if $\underline{x}_s(p) \neq 0$ and $x_s(p) \geq \underline{x}_s(p)$ when $\underline{x}_s(p) = 0$.

Proposition 7 shows that one can readily verify that a proper solution approximates well the optimal solution to the initial planner problem.

Proposition 7. For the approximate problem, introduce the maximal approximation errors:

$$\varepsilon := \max_p \max_{\underline{c} \leq t \leq \bar{c}} \left[\mathcal{C}(t) - \max_i \; l_{ip}^c(t) \right] \qquad \text{and} \qquad \varepsilon_s := \max_p \max_{\underline{x}_s \leq t \leq \bar{x}_s} \left[z(p)\mathcal{X}(t) - z(p) \max_i l_{ip}^s(t) \right].$$

If the solution to the approximate planner problem is proper, then the overall approximation error is bounded from above by the sum of maximal approximation errors:

$$0 \leq \int \Big(\mathcal{C}(c(p)) + z(p) \big(\mathcal{X}(x_c(p)) + \mathcal{X}(x_m(p)) \big) \Big) \mathrm{d}\pi - \Omega \leq \varepsilon + \varepsilon_c + \varepsilon_m,$$

where Ω is the minimum value for the original problem.

Proof. We show that if the solution (c, x_s, r, r_s) to the approximation problem is proper, then the overall approximation error is bounded from above by the sum of maximal approximation errors. We next prove that both inequalities are satisfied.

The first inequality is satisfied since the approximate allocation (c, x) is feasible. Since the approximate planner's problem produces a feasible solution, we clearly have

$$\Omega \leq \int \big(\mathcal{C}(c(p)) + z(p) \big(\mathcal{X}(x_c(p)) + \mathcal{X}(x_m(p)) \big) \big) \, \mathrm{d}\pi. \tag{A.57}$$

To prove the second inequality, we use that the definition of the approximation error ε_c and the approximation constraints (A.54) implies $\mathcal{C}(c(p)) \leq r_c(p) + \varepsilon_c$. Denote by $(\hat{c}, \hat{x})$ the allocation that attains the minimum resource cost $\Omega := \int \big(\mathcal{C}(\hat{c}(p)) + z(p) \big(\mathcal{X}(\hat{x}_c(p)) + \mathcal{X}(\hat{x}_m(p)) \big) \big) \, \mathrm{d}\pi$. Since the solution to the approximate problem is proper, there exists a weight $\lambda \in (0, 1)$ such that the convex combination given by $\tilde{c}(p) = \lambda c(p) + (1 - \lambda)\hat{c}(p)$ and $\tilde{x}_s(p) = \lambda x_s(p) + (1 - \lambda)\hat{x}_s(p)$ is a

proper allocation. To construct an alternative allocation that is feasible under the approximate problem, we can set:

$$\tilde{r}_c(p) = \max_i \; l^c_{ip}(\tilde{c}(p)) \tag{A.58}$$

$$\tilde{r}_s(p) = \max_i \; l^s_{ip}(\tilde{x}_s(p)) \tag{A.59}$$

Since (c, x_s, r, r_s) solves the approximate problem, and since $(\tilde{c}, \tilde{x}_s, \tilde{r}, \tilde{r}_s)$ is feasible, we know that the cost under the alternative allocation exceeds the cost under the approximate solution. Since the pointwise maximum of convex functions is convex, we have that

$$\tilde{r}(p) \leq \lambda \max_i l^c_{ip}(c(p)) + (1-\lambda) \max_i l^c_{ip}(\hat{c}(p)) \leq \lambda r(p) + (1-\lambda)\mathcal{C}(\hat{c}(p)) \tag{A.60}$$

$$\tilde{r}_s(p) \leq \lambda \max_i l^s_{ip}(x_s(p)) + (1-\lambda) \max_i l^s_{ip}(\hat{x}_s(p)) \leq \lambda r_s(p) + (1-\lambda)\mathcal{X}(\hat{x}_s(p)) \tag{A.61}$$

where the final inequalities follow from the definition of the approximation constraints (A.54), and from the observation that approximations are from below. By combining the two previous claims we write that

$$\int \big(r(p) + z(p)\big(r_c(p) + r_m(p)\big)\big)\mathrm{d}\pi \leq \int \big(\tilde{r}(p) + z(p)\big(\tilde{r}_c(p) + \tilde{r}_m(p)\big)\big)\mathrm{d}\pi$$
$$\leq \lambda \int \big(r_c(p) + z(p)\big(r_c(p) + r_m(p)\big)\big)\mathrm{d}\pi + (1-\lambda)\int \big(\mathcal{C}(\hat{c}(p)) + z(p)\big(\mathcal{X}(\hat{x}_c(p)) + \mathcal{X}(\hat{x}_m(p))\big)\big)\mathrm{d}\pi,$$

which implies $\Omega \geq \int \big(r_c(p) + z(p)\sum r_s(p)\big)\mathrm{d}\pi$. Finally, we use the definition of the approximation errors to write:

$$\Omega \geq \int \big(r(p) + z(p)\big(r_c(p) + r_m(p)\big)\big)\mathrm{d}\pi \geq \int \big(\mathcal{C}(c(p)) + z(p)\big(\mathcal{X}(x_c(p)) + \mathcal{X}(x_m(p))\big)\big)\mathrm{d}\pi - \sum \varepsilon$$

which concludes the proof. □

Algorithm. We use an iterative algorithm to solve the approximate planner's problem for a given precision level.[31] We solve the planner problem for a worker type space with 200 types in both the cognitive and the manual dimension, equivalently, for a total of 40 thousand types. We display the structure of our numerical approach in Algorithm 1.

Having described how to characterize the planner problem given an arbitrary assignment, we next describe how to update the assignment to obtain a jointly optimal assignment and allocation.

[31] See Ekeland and Moreno-Bromberg (2010) for a discussion on numerically solving optimization problems subject to a convexity constraint on the function u, and Oberman (2013) for a practical approach of dealing with global incentive constraints.

Algorithm 1. Iterative Algorithm for Planner's Problem with Fixed Assignment.

Set initial location boundaries $\{\underline{c}, \overline{c}\}$ and $\{\underline{x}_s, \overline{x}_s\}$, define initial accuracy levels ε_c and ε_s

while $\varepsilon_c + \sum \varepsilon_s > \zeta$ **do**

for each p, construct piecewise linear approximations of $\mathcal{C}$ and $\mathcal{X}$ on bounded intervals $[\underline{c}, \overline{c}]$ and $[\underline{x}_s, \overline{x}_s]$ with precisions ε_c and ε_s

solve the approximate planner's problem

if *the approximate solution is proper* **then**

update precision levels $\varepsilon_c \to \alpha\varepsilon_c$ and $\varepsilon_s \to \alpha\varepsilon_s$ for some $\alpha < 1$

update location boundaries $[\underline{c}, \overline{c}]$ and $[\underline{x}_s, \overline{x}_s]$

else

relax location boundaries

end

return *solution* (c, x_s, r_c, r_s) *to the final approximate planner's problem.*

end

Given the optimality of positive sorting between workers and firms in Proposition 4, we update our assignment after each step by positively sorting the distribution of project values with the effective worker skill index $\mathcal{X}(x_c(p)) + \mathcal{X}(x_m(p))$. By doing so, we reassign projects across workers, which yields a new assignment. We then solve the planner's problem for the new assignment function using Algorithm 1. We proceed until the assignment converges. To the best of our knowledge, there is no proof of unique convergence for this iterative procedure. In practice, however, we find that our algorithm always converges to the same assignment function for distinct initial assignments.

Literature. Our numerical approach relates to outer linearization of a separable convex objective function. This approach is well-established, see for example, Bertsekas and Yu (2011).

Outer linearization of a separable convex objective is part of the outer linearization approach for general problems. For example, Geoffrion (1970) presents the idea of approximating a convex function with supporting hyperplanes, which is outer linearization. Duran and Grossmann (1986) applies outer linearization to general convex mixed-integer optimization problems. In our case, the objective is separable in variables, which leads to faster and more efficient algorithms for constructing the supporting hyperplanes. Bertsekas and Yu (2011) also has an objective function that is separable in variables and discuss that this problem has been explored under the framework

of extended monotropic programming, which builds on monotropic programming (Rockafellar, 1999). Our approach extends beyond outer linearization of a separable convex objective. Whereas the literature routinely focuses on problems with linear equality constraints, our approach also addresses inequality constraints, making it applicable to a broader class of problems.

A.12 Characterizing Equilibrium using Transport Problems

To characterize an equilibrium, we relate our positive economy to optimal transport problems.

Primal Problem. The primal problem is to choose an assignment to maximize production:

$$\max_{\gamma \in \Gamma} \int y(x_1, x_2, z) \mathrm{d}\gamma. \tag{A.62}$$

The choice of an assignment is restricted by the feasibility constraint, $\gamma \in \Gamma(F_x, F_x, F_z)$, where Γ denotes the set of probability measures on the product space $\mathbf{X} \times \mathbf{X} \times Z$ such that the marginal distributions of γ onto $\mathbf{X}$ and Z are F_x and F_z respectively.

Dual Problem. The dual transport problem is to choose functions w and Ω that solve:

$$\min_{w,\Omega} \int w(x_1) \mathrm{d}F_x + \int w(x_2) \mathrm{d}F_x + \int \Omega(z) \mathrm{d}F_z, \tag{A.63}$$

subject to the constraint that the surplus is weakly negative for any triplet (x_1, x_2, z), that is, $S(x_1, x_2, z) \leq 0$.

We connect the primal problem and the dual problem to equilibrium in Lemma 10.

Lemma 10. The equilibrium assignment γ solves the primal problem (A.62), equilibrium wages w and firm values Ω solve the dual problem (A.63).

We use Lemma 10 to characterize the equilibrium.[32] We solve the primal problem (A.62) to characterize the equilibrium assignment and the dual problem (A.63) to characterize wages w and firm values Ω. To prove Lemma 10, we use Lemma 11.

Lemma 11. Suppose the objectives of the primal problem (A.2) and the dual problem (A.63) coincide $\int y(x_1, x_2, z) \mathrm{d}\gamma = \int w(x_1) \mathrm{d}F_x + \int w(x_2) \mathrm{d}F_x + \int \Omega(z) \mathrm{d}F_z$. Then γ solves the primal problem, and the functions w and Ω solve the dual transport problem.

[32] We note that a transport problem with two identical worker distributions F_x with unit mass for each role is equivalent to a transport problem with a single worker distribution Φ_x with mass two (Appendix A.13).

The proof of Lemma 11 only uses of a notion of weak duality.

Weak Duality. Let $\gamma \in \Gamma(F_{x_1}, F_{x_2}, F_z)$ be a joint probability measure, and (f, g, h) be functions such that $y(x_1, x_2, z) \leq f(x_1) + g(x_2) + h(z)$ for all (x_1, x_2, z). Then

$$\min_{f,g,h} \int f(x)\mathrm{d}F_{x_1} + \int g(x)\mathrm{d}F_{x_2} + \int h(z)\mathrm{d}F_z \;\geq\; \max_{\gamma\in\Gamma} \int y(x_1, x_2, z)\mathrm{d}\gamma. \tag{A.64}$$

Proof. For any functions (f, g, h) so that $y(x_1, x_2, z) \leq f(x_1) + g(x_2) + h(z)$ we have:

$$\max_{\gamma\in\Gamma} \int y(x_1, x_2, z)\mathrm{d}\gamma \leq \int \big(f(x_1) + g(x_2) + h(z)\big)\mathrm{d}\gamma = \int f(x)\mathrm{d}F_{x_1} + \int g(x)\mathrm{d}F_{x_2} + \int h(z)\mathrm{d}F_z,$$

where the equality follows as $\gamma \in \Gamma(F_{x_1}, F_{x_2}, F_z)$. Since the above inequality holds for any (f, g, h) it holds for (f, g, h) that minimize the right-hand side.

We use weak duality to establish Lemma 11 by contradiction.

Proof of Lemma 11. Suppose by contradiction that γ does not solve the planning problem, then

$$\begin{aligned}\max_{\pi\in\Gamma} \int y(x_1, x_2, z)\mathrm{d}\pi > \int y(x_1, x_2, z)\mathrm{d}\gamma &= \int w(x)\mathrm{d}F_x + \int w(x)\mathrm{d}F_x + \int \Omega(z)\mathrm{d}F_z \\ &\geq \min_{f,g,h} \int f(x)\mathrm{d}F_x + \int g(x)\mathrm{d}F_x + \int h(z)\mathrm{d}F_z,\end{aligned} \tag{A.65}$$

where the equality follows by assumption. This contradicts weak duality (A.64).

Suppose by contradiction that the functions $\hat{f}, \hat{g}$, and $\hat{h}$ do not solve the dual problem. Then there exist functions f, g, and h such that

$$\begin{aligned}\min_{f,g,h} \int f(x)\mathrm{d}F_x + \int g(x)\mathrm{d}F_x + \int h(z)\mathrm{d}F_z &< \int w(x)\mathrm{d}F_x + \int w(x)\mathrm{d}F_x + \int \Omega(z)\mathrm{d}F_z \\ &= \int y(x_1, x_2, z)\mathrm{d}\gamma \leq \max_{\pi\in\Gamma} \int y(x_1, x_2, z)\mathrm{d}\pi,\end{aligned} \tag{A.66}$$

where the equality follows by the assumption. This inequality contradicts weak duality (A.64).

We now use Lemma 11 to show that the equilibrium assignment γ solves the primal transport problem, and that the wage and firm value function solve the dual transport problem.

In equilibrium, the surplus is negative for any triplet (x_1, x_2, z), which implies that $y(x_1, x_2, z) \leq w(x_1) + w(x_2) + \Omega(z)$. By substituting the household budget constraints $c = (1-\tau)w(x)$, and the government budget constraint $G = \tau \int w(x)\mathrm{d}\Phi(\alpha)$, into the aggregate resource constraint (34), we write:

$$\int y(x_1, x_2, z)\mathrm{d}\mu = \int w(x_1)\mathrm{d}F_x(x_1) + \int w(x_2)\mathrm{d}F_x(x_2) + \int \Omega(z)\mathrm{d}F_z(z). \tag{A.67}$$

By Lemma 11 it thus follows that μ solves the primal problem and w and Ω solve the dual problem.

A.13 Symmetric Equilibrium

We prove that we can restrict our attention to symmetric equilibria without loss of generality.

Lemma 12. For any equilibrium with wages w and assignment function $\gamma \in \Gamma(F_{x_1}, F_{x_2}, F_z)$, there exists an equilibrium with wages w and a symmetric assignment $\hat{\gamma} = \frac{\gamma+\gamma'}{2}$.

Proof. Lemma 12 states that for any competitive equilibrium with wages w, firm value Ω, and assignment $\gamma \in \Gamma(F_{x_1}, F_{x_2}, F_z)$, there is an equilibrium with identical wages w, firm value function Ω with a symmetric assignment function $\hat{\gamma} := \frac{\gamma+\gamma'}{2} \in \Gamma(F_x, F_x, F_z)$, where $F_x := \frac{1}{2}(F_{x_1}, F_{x_2})$.

To prove this result, we first define γ' as a pushforward measure of the assignment function γ. We then show that the symmetric assignment function $\hat{\gamma}$ indeed solves the primal transport problem. Using Lemma 10, this establishes the result.

Definition. Given spaces M_1 and M_2, a measure γ concentrated on M_1, and a map $T : M_1 \to M_2$, the pushforward measure of γ through T, which we denote by $T_{\#}\gamma$, is defined so that:

$$\int f(y)\mathrm{d}T_{\#}\gamma = \int f(T(x))\mathrm{d}\gamma. \tag{A.68}$$

We define γ' as the pushforward measure of γ through a mapping T, or $\gamma' := T_{\#}\gamma$. Our mapping T maps from the matching set onto itself interchanging the position of the worker and the coworker, that is, $T : M \to M$ so that $(x_1, x_2, z) \to (x_2, x_1, z)$. If γ is a feasible assignment, γ' is a feasible assignment, that is, for $\gamma \in \Gamma(F_{x_1}, F_{x_2}, F_z)$ we have $\gamma' \in \Gamma(F_{x_2}, F_{x_1}, F_z)$.

Using the definition of γ', we construct the symmetric assignment function $\hat{\gamma} := \frac{\gamma+\gamma'}{2}$, and observe that the symmetric assignment function is feasible given F_x, that is, $\hat{\gamma} \in \Gamma(F_x, F_x, F_z)$. Moreover, we observe that:

$$\int y(x_1, x_2, z)\mathrm{d}\hat{\gamma}(x_1, x_2, z) = \int w(x_1)\mathrm{d}F_x(x_1) + \int w(x_2)\mathrm{d}F_x(x_2) + \int \Omega(z)\mathrm{d}F_z(z). \tag{A.69}$$

The left-hand side is unchanged as the production of equilibrium pairings does not change, while the right-hand side is unchanged as the skill distribution is unchanged. By Lemma 11, $\hat{\gamma}$ solves the primal transport problem, and functions w and Ω solve the dual transport problem. By Lemma

10, this shows that the symmetric assignment $\hat{\gamma}$ is an equilibrium assignment, w are equilibrium wages, and Ω are equilibrium firm values.

Finally, we remark that the equilibrium assignment need not be $\hat{\gamma}$. Specifically, we can replace $\hat{\gamma}$ with any other optimal primal solution. Suppose that there exists another solution to the primal problem $\tilde{\gamma}$, then

$$\int y(x_1, x_2, z)\mathrm{d}\tilde{\gamma}(x_1, x_2, z) = \int w(x_1)\mathrm{d}F_x(x_1) + \int w(x_2)\mathrm{d}F_x(x_2) + \int \Omega(z)\mathrm{d}F_z(z), \tag{A.70}$$

and hence $S(x_1, x_2, z) = 0$ for $\tilde{\gamma}$ almost everywhere. □

A.14 Wages and Firm Values

To see why only effective skill matters, consider two workers (x_c, x_m) and $(\hat{x}_c, \hat{x}_m)$ with identical effective skill $X = \hat{X}$. Since the surplus is zero almost everywhere under the equilibrium assignment γ, and using production technology (4), $2w(x)+\Omega(z) = zX$ and $2w(\hat{x})+\Omega(\hat{z}) = \hat{z}\hat{X}$. By the constraints to the dual problem (A.63), $2w(x)+\Omega(\hat{z}) \geq \hat{z}X = \hat{z}\hat{X}$ and $2w(\hat{x})+\Omega(z) \geq z\hat{X} = zX$, where the equalities follow since the workers' effective skills are identical. Combining these expressions, $w(\hat{x}) \geq w(x)$ and $w(x) \geq w(\hat{x})$, so that $w(\hat{x}) = w(x)$. It is useful to define $h(X)$, the firm's total wage bill, as $h(X) := 2w(x)$.

Wages are convex in effective skill X, so small differences in effective worker skill X translate into increasingly large differences in worker earnings. The dual constraints imply $h(X) \geq zX - \Omega(z)$ for any z. Since the surplus is zero almost everywhere with respect to the equilibrium assignment, it follows that $h(X) := \sup_z(zX - \Omega(z))$ implying that $h = \Omega^*$, the firm's wage bill is the Legendre transform of the firm value function. Since $h(X)$ is the supremum of linear functions in X, the wage function is convex.

The firm value function is the Legendre transform of the wage bill. The dual constraints also imply that for any x it holds that $\Omega(z) \geq zX - h(X)$ and therefore $\Omega(z) := \sup_X(zX - h(X))$. This implies that the firm value function is convex and indeed the Legendre transform of the wage bill $\Omega = h^*$. As a result, $h(X) + h^*(z) = zX$.

A.15 Lemma 3

We show there exists a firm distribution F_z such that given wage schedule w, workers and firms both optimize in a self-sorting equilibrium, where the distribution of worker skills F_x is determined

by the worker problems given a talent distribution Φ. We verify this claim by studying the firm and worker problem given the postulated wage schedule (35).

Firm. Taking the wage schedule w as given, the firm problem of choosing two workers to employ can be written as:

$$\max_{x_1,x_2} y(x_1,x_2,z) - w(x_1) - w(x_2) = \max_{x_1,x_2} z\left(x_{1c}x_{2c} + x_{1m}x_{2m}\right) - \frac{1}{2}(x_{1c}^2 + x_{1m}^2)^\eta - \frac{1}{2}(x_{2c}^2 + x_{2m}^2)^\eta$$

where the equality follows from substituting in the production technology (4) and wage schedule (35). The solution to this problem is that firm z wants to hire two identical workers.

To establish that each firm wants to hire two identical workers, we show that a firm that hires different workers (x_1, x_2) such that $x_1 \neq x_2$ can increase its profits by hiring two identical workers $\left(\frac{1}{2}(x_1+x_2), \frac{1}{2}(x_1+x_2)\right)$. By hiring two identical workers, the firm increases its production and decreases its wage bill. Production increases since the worker production technology is concave, that is, $\left(\frac{x_{1c}+x_{2c}}{2}\right)^2 + \left(\frac{x_{1m}+x_{2m}}{2}\right)^2 \geq x_{1c}x_{2c} + x_{1m}x_{2m}$ is implied by $(x_{1c}-x_{2c})^2 + (x_{1m}-x_{2m})^2 \geq 0$. To see that the firm decreases its wage bill by hiring two identical workers, observe that $(x_{1c}-x_{2c})^2 + (x_{1m}-x_{2m})^2 \geq 0$ implies $\left(\frac{x_{1c}+x_{2c}}{2}\right)^2 + \left(\frac{x_{1m}+x_{2m}}{2}\right)^2 \leq \frac{1}{2}\left(x_{1c}^2 + x_{1m}^2\right) + \frac{1}{2}\left(x_{2c}^2 + x_{2m}^2\right)$, which implies that for $\eta \geq 1$ we have $\left(\left(\frac{x_{1c}+x_{2c}}{2}\right)^2 + \left(\frac{x_{1m}+x_{2m}}{2}\right)^2\right)^\eta \leq \left(\frac{1}{2}\left(x_{1c}^2 + x_{1m}^2\right) + \frac{1}{2}\left(x_{2c}^2 + x_{2m}^2\right)\right)^\eta$. Since the function ς^η is convex $\left(\frac{1}{2}\varsigma + \frac{1}{2}\hat{\varsigma}\right)^\eta \leq \frac{1}{2}\varsigma^\eta + \frac{1}{2}\hat{\varsigma}^\eta$, we obtain $\left(\left(\frac{x_{1c}+x_{2c}}{2}\right)^2 + \left(\frac{x_{1m}+x_{2m}}{2}\right)^2\right)^\eta \leq \frac{1}{2}\left(x_{1c}^2 + x_{1m}^2\right)^\eta + \frac{1}{2}\left(x_{2c}^2 + x_{2m}^2\right)^\eta$ by applying this inequality to the right-hand side. Equivalently, by hiring two identical workers $\left(\frac{1}{2}(x_1+x_2), \frac{1}{2}(x_1+x_2)\right)$ a firm lowers their wage bill relative to hiring two different workers (x_1, x_2), $w(\frac{1}{2}(x_1+x_2)) + w(\frac{1}{2}(x_1+x_2)) \leq w(x_1) + w(x_2)$. Finally, since firms hire identical workers, the firm's optimality condition hiring effective worker skill X gives $z = h'(X)$. Since the wage schedule is convex, equilibrium sorting is positive.

Worker. The distribution of worker skills is uniquely induced by the worker problems. Given the wage schedule, a worker's problem is:

$$\max_{x_c,x_m} u\big((1-\tau)w(x)\big) - \kappa\Big(\frac{x_c}{\alpha_c}\Big)^\rho - \kappa\Big(\frac{x_m}{\alpha_m}\Big)^\rho. \tag{A.71}$$

Using a transformation $\tilde{x}_s := x_s^\rho$, the problem is:

$$\max_{\tilde{x}_c,\tilde{x}_m} u\big((1-\tau)\tilde{w}(\tilde{x})\big) - \frac{\tilde{x}_c}{p_c} - \frac{\tilde{x}_m}{p_m}, \tag{A.72}$$

where $\tilde{w}(\tilde{x}) := \left(\frac{1}{2}\sum \tilde{x}_s^{\frac{2}{\rho}}\right)^\eta$.

We prove strict concavity of the objective by examining each of the terms in the objective. The second and third terms are linear and thus concave. It remains to verify that the first term,

$u((1-\tau)\tilde{w}(\tilde{x}))$, is strictly concave. First, since the consumption utility u is strictly concave:

$$\lambda u((1-\tau)\tilde{w}(x)) + (1-\lambda)u((1-\tau)\tilde{w}(\tilde{x})) < u((1-\tau)(\lambda\tilde{w}(x) + (1-\lambda)\tilde{w}(\tilde{x}))$$

Since u is an increasing and concave function, the first term is strictly concave if $\tilde{w}(\tilde{x})$ is strictly concave. To establish this, we note that the transformed wage equation is a composite function of a concave CES aggregate with an increasing concave function as long as $\eta \leq \frac{\rho}{2}$. The worker problem is strictly concave and thus has a unique solution, which implies the distribution of worker skills is uniquely induced by the worker's problem.

To complete the proof we show there exists a distribution of firm projects such that the wage equation (35) is an equilibrium wage function. Since the wage bill is continuously differentiable, $z = h'(X)$. We can use this expression to uniquely pin down a distribution of firm project values that rationalizes the wage equation. Since the wage bill is convex, the inferred distribution indeed implies positive sorting between effective worker skills and firm project values.[33]

A.16 Frisch Elasticity

We next show how to derive the expression for the Frisch elasticity of labor supply within our model. Adding the optimality conditions across tasks gives $\ell_c^\rho + \ell_m^\rho = \mathcal{C}w(x)\lambda$, with constant $\mathcal{C} := (1-\tau)\frac{\eta}{\kappa\rho}$. Using the constant effort shares implied by (39), and multiplying and dividing by $\ell^\rho := (\ell_c + \ell_m)^\rho$ we write $\ell^\rho = \mathbb{C}w(x)\lambda$, where $\mathbb{C} := \mathcal{C}/\left(\left(\frac{\ell_c}{\ell}\right)^\rho + \left(\frac{\ell_m}{\ell}\right)^\rho\right)$ is constant across workers. To obtain the Frisch elasticity implied by our model, we relate a worker's total efforts to earnings per hour $\tilde{z}(x) = w(x)/(\ell_c + \ell_m)$ as $\ell^{\rho-1} = \mathbb{C}\tilde{z}(x)\lambda$ to obtain (40):

$$\varepsilon = \left.\frac{\partial \log(\ell_c + \ell_m)}{\partial \log \tilde{z}(x)}\right|_\lambda = \left.\frac{\partial \log(\ell_c + \ell_m)}{\partial \log(1-\tau)}\right|_\lambda = \frac{1}{\rho - 1}. \tag{40}$$

A.17 Implementation

In this appendix, we describe an implementation of the optimum through an income tax system. This argument follows the analysis in Kocherlakota (2010) for an economy with multidimensional private information. In order to illustrate this argument, let $\{c(\alpha), x(\alpha)\}$ denote the solution to the planning problem for a finite type space A.

[33] For the parametric specification, $\Omega(z) := \sup(zX - h(X))$ can be characterized as $\Omega(z) = \mathcal{C}_z z^{\frac{\eta}{\eta-1}} - 2\zeta$, where $\mathcal{C}_z$ is a multiplicative constant independent of z. Together with the dual constraint, this closed-form expression for the firm value allows us to characterize project value z without relying on the derivative of the firm's wage bill in the quantitative section.

We first note that the consumption allocation depends on type α only through the allocation of tasks $x(\alpha)$. If $x(\alpha) = x(\hat{\alpha})$ while $c(\alpha) > c(\hat{\alpha})$, worker $\hat{\alpha}$ would pretend to be type α to attain more consumption for identical task inputs and the planner allocation is not incentive compatible. The consumption allocation can thus be written as a function of the task inputs $c(\alpha) = \tilde{c}(x(\alpha))$, where the domain of the consumption function is finite and given by $\mathbf{X} := \cup_\alpha \{x(\alpha)\}$.

The consumption function is increasing in both cognitive and manual task disutility as the tax system has to reward workers for providing higher levels of task inputs. Suppose $x_c(\alpha) = x_c(\hat{\alpha})$, $x_m(\alpha) > x_m(\hat{\alpha})$ while $c(\alpha) \leq c(\hat{\alpha})$. In this case, worker α reports $\hat{\alpha}$, so the planner allocation is not incentive compatible.

While the consumption allocation function $\tilde{c}$ has a finite domain, we next extend to all x such that $x_c \geq \min_{\alpha \in A} x_c(\alpha)$ and $x_m \geq \min_{\alpha \in A} x_m(\alpha)$. For all such task inputs, define

$$\hat{c}(x) := \max_{x' \leq x} \tilde{c}(x') \tag{A.73}$$

subject to $x' \in \mathbf{X}$ if the maximizer exists and $\hat{c}(x) = 0$ otherwise. We also set $\hat{c}(x) = -\infty$ for all x outside the domain. We define the tax system $T(x)$ over the same values of x as:

$$T(x) = w(x) - \hat{c}(x). \tag{A.74}$$

Proposition 8. Tax system T implements the planner allocation.

Proof. In order to establish the result, we show that workers of skill α indeed choose the allocation for worker α under the planner allocation.

The problem of worker α given the tax system T is to choose the level of cognitive and manual task inputs x to solve:

$$\max_x \; u(c) - v\left(\frac{x_c}{\alpha_c}\right) - v\left(\frac{x_m}{\alpha_m}\right) \tag{A.75}$$

subject to the constraint $c \leq w(x) - T(x) = \hat{c}(x)$. The problem can thus be written as:

$$\max_x \; u(\hat{c}(x)) - v\left(\frac{x_c}{\alpha_c}\right) - v\left(\frac{x_m}{\alpha_m}\right). \tag{A.76}$$

Choosing cognitive and manual task inputs below their minimum would be suboptimal as the worker pays infinite taxes. We thus restrict attention to task inputs $x \geq \big(\min_\alpha x_c(\alpha), \min_\alpha x_m(\alpha)\big)$.

We next show that workers choose a task allocation $x \in \mathbf{X}$. By contradiction, suppose that the worker instead chooses an allocation $\hat{x} \notin \mathbf{X}$. Hence, the worker attains consumption $\hat{c}(\hat{x})$, which by the definition of the consumption allocation function is given by:

$$\hat{c}(\hat{x}) = \max_{x' \leq \hat{x}} \tilde{c}(x') \tag{A.77}$$

subject to $x' \in \mathbf{X}$. Since $\hat{x} \notin \mathbf{X}$, the worker can do better by reducing their work effort choosing $x \in \mathbf{X}$ that delivers the same consumption level. The worker consumes the same, exerting less effort. As a result, we restrict the choice to $x \in \mathbf{X}$ without loss of generality.

Worker α chooses the optimal bundle of task inputs $x \in \mathbf{X}$, which boils down to choosing x such that for all $x' \in \mathbf{X}$

$$u\big(\tilde{c}(x)\big) - v\left(\frac{x_c}{\alpha_c}\right) - v\left(\frac{x_m}{\alpha_m}\right) \geq u\big(\tilde{c}(x')\big) - v\left(\frac{x'_c}{\alpha_c}\right) - v\left(\frac{x'_m}{\alpha_m}\right) \tag{A.78}$$

By definition of the consumption allocation function over the domain $\mathbf{X}$ it follows that this is equivalent to:

$$u\big(c(\alpha)\big) - v\left(\frac{x_c(\alpha)}{\alpha_c}\right) - v\left(\frac{x_m(\alpha)}{\alpha_m}\right) \geq u\big(c(\alpha')\big) - v\left(\frac{x_c(\alpha')}{\alpha_c}\right) - v\left(\frac{x_m(\alpha')}{\alpha_m}\right) \tag{A.79}$$

for all $\alpha' \in A$. That worker α optimally chooses allocation $(c(\alpha), x(\alpha))$, implementing the planner allocation, then follows from the incentive constraints. □

A.18 Optimal Taxation of Labor Earnings

In this appendix, we analyze the planner problem when only worker earnings are observed. We formulate the planner problem when the planner observes individual output but cannot condition taxes on task mixes. We show that this problem simplifies to a unidimensional problem. Since this argument does not depend on the presence of firm heterogeneity, we omit firm heterogeneity for ease of exposition.

Planner Problem. The planner chooses $\{(c(\alpha), y(\alpha))\}$ to maximize:

$$\int V(c(\alpha), y(\alpha), \alpha)\pi(\alpha)\mathrm{d}\alpha \tag{A.80}$$

where the indirect utility is:

$$V(c(\alpha), y(\alpha), \alpha) = \max_{\{(x_c, x_m)\}} u\big(c(\alpha), x_c(\alpha)/\alpha_c, x_m(\alpha)/\alpha_m\big), \tag{A.81}$$

where maximization is subject to $y(\alpha) = \frac{1}{2}\big(x_c(\alpha)^2 + x_m(\alpha)^2\big)$. The planner thus prescribes a total level of output $y(\alpha)$ to be produced by type α, while the individual chooses the task inputs to generate this output level. The planner problem is subject to incentive constraints:

$$V(c(\alpha), y(\alpha), \alpha) \geq V(c(\alpha'), y(\alpha'), \alpha), \tag{A.82}$$

for all workers (α, α'), and an aggregate resource constraint $\int \left(c(\alpha) - y(\alpha)\right)\pi(\alpha)\mathrm{d}\alpha \leq 0$.

We now show this problem can be recast as a unidimensional optimal taxation problem by characterizing the indirect utility. The indirect utility of the agent is:

$$V(c(\alpha), y(\alpha), \alpha) = \max_{x_c, x_m} \; c - \kappa \left(\frac{x_c}{\alpha_c}\right)^{\rho} - \kappa \left(\frac{x_m}{\alpha_m}\right)^{\rho} \tag{A.83}$$

subject to $y(\alpha) = \frac{1}{2}\big(x_c(\alpha)^2 + x_m(\alpha)^2\big)$. Letting λ denote the multiplier on the individual production constraint, the optimality conditions give $\lambda = \frac{1}{2}\frac{x_s^{\rho-2}}{\alpha_s^{\rho}}$ for each task $s \in \{c, m\}$, and hence $x_s = (2\lambda\alpha_s^{\rho})^{\frac{1}{\rho-2}}$. Substituting into the production technology we obtain $2y = (2\lambda)^{\frac{2}{\rho-2}}\left(\alpha_c^{\frac{2\rho}{\rho-2}} + \alpha_m^{\frac{2\rho}{\rho-2}}\right)$. We substitute the optimal choices for x_c and x_m using the characterization of the multiplier λ into the expression for the disutility of work of the indirect utility function as:

$$V(c(\alpha), y(\alpha), \alpha) = c(\alpha) - \kappa\Big(\alpha_c^{\frac{2\rho}{\rho-2}} + \alpha_m^{\frac{2\rho}{\rho-2}}\Big)^{1-\frac{\rho}{2}}(2y(\alpha))^{\frac{\rho}{2}} \tag{A.84}$$

This transforms the planning problem to a unidimensional screening problem where the effective heterogeneity is $\big(\alpha_c^{\frac{2\rho}{\rho-2}} + \alpha_m^{\frac{2\rho}{\rho-2}}\big)^{1-\frac{\rho}{2}} 2^{\frac{\rho}{2}} = \theta$. We solve this problem using the tools for unidimensional problems. What matters is the skill index θ, not the higher-dimensional underlying type.

Bunching and Taxing Multidimensional Skills

Numerical Appendix

Job Boerma, Aleh Tsyvinski and Alexander Zimin

September 2026

B Additional Numerical Results

In this appendix, we analyze the sensitivity of our results with respect to changes in the problem inputs for a simple benchmark economy and the empirically relevant model. The simple benchmark model is identical to the empirically relevant model except (a) there is no firm heterogeneity, and (b) worker types p are uniformly distributed on $[0.5, 1.0]^2$.

Project Distribution. First, Figure B.1 displays sensitivity analysis with respect to the project distribution. In the simple benchmark economy, projects are uniformly distributed between $1-\Delta$ and $1+\Delta$. We consider $\Delta \in \{0.00, 0.05, 0.10, 0.25\}$. The results for different values of Δ are presented in Figure B.1. As Δ increases, the allocation features higher levels of allocations for higher skill types. The benefit of higher task allocations for high skill individuals is greater due to the complementarity with the high project values. This is seen by the upper bounds on both task allocations increasing. The middle and right hand panels display that the pattern of bunching is largely invariant to changes in the range of the productivity distribution in terms of rank coordinates. We observe that correspondingly the levels at which there is bunching vary.

Figure B.2 displays the sensitivity of our numerical results to changes in the distribution of firms for the empirically relevant economy. In this case, the firm distribution is given by a mixture of the empirical project distribution and a uniform distribution over the worker type space. The mixture probability on the uniform distribution is parameterized by Δ, and we consider $\Delta \in \{0.00, 0.1, 0.3, 0.5\}$. In this case, we see that both the allocations and the bunching regions are largely invariant to the changes in the project distributions. In this case, the range of the project values is unaltered, and the variation Δ merely captures reversion to the mean, compressing the distribution of project values.

Concave Utility from Consumption. We next show that bunching patterns are robust to including curvature in the utility from consumption. Under the baseline calibration, we adopt a power specification for consumption utility with the coefficient of relative risk aversion $\gamma = 0$. We next show the results for $\gamma \in \{0, 0.25, 0.5, 0.75\}$.

The results are presented in Figure B.3. As the curvature in the utility from consumption increases, the task allocation is shifted towards more asymmetric allocations, meaning that agents are asked to supply either great effort in cognitive tasks, or great effort in manual tasks, but not in both. Moreover, the effort asked to be supplied by bluntly bunched types increases. The middle and right panels display the changes in the pattern of bunching. Relative to the prior cases, the types that are bluntly bunched are unchanged, while the targeted bunching region slightly increases in size. More asymmetric types tend to be bunched less as the curvature of the edge of the targeted bunching region increases.

The same forces are present in the empirically relevant model, for which we show the sensitivity analysis in Figure B.4. We find that the task allocation is also shifted towards more asymmetric allocations and that the effective task inputs supplied by bluntly bunched types increase. In this case, we see no change in the bunching regions in the middle and right panel of Figure B.4. In sum, the bunching pattern is robust with respect to the curvature in the utility function.

Power of Disutility and Technology. We next analyze sensitivity of our numerical results with respect to changes in the curvature with respect to both the disutility for labor supply and the production technology.

We consider the planning problem in Section 3.1, with a slight generalization of the resource cost of providing disutility to $\mathcal{X}(x_s(p)) = -\frac{1}{2}x_s(p)^{\frac{\phi}{\rho}}$. The baseline calibration features $\zeta = \frac{\phi}{\rho} = 2/2.8$. We consider variations where $\zeta \in \{2/2.5, 2/2.8, 2/3, 2/5\}$. These variations simultaneously correspond to changes in the labor supply elasticity ρ and the power of the production technology ϕ as $\zeta = \frac{\phi}{\rho}$.

Figure B.5 shows our numerical findings for the simple baseline economy. As labor becomes more inelastic the allocation of cognitive and manual labor inputs compresses, that is, puts the allocations closer towards the origin. The middle and right panels, however, show that the bunching patterns are largely invariant to the choice of the labor supply elasticity.

Figure B.6 analyzes the sensitivity of our numerical results to changes in the empirically relevant economy. For the empirically relevant economy, the allocation patterns are remarkably similar, where the allocation of cognitive and manual labor inputs compresses as labor becomes

more inelastic. The middle and right panels show that the bunching region slightly increases as labor becomes more inelastic and it becomes more costly to separate types.

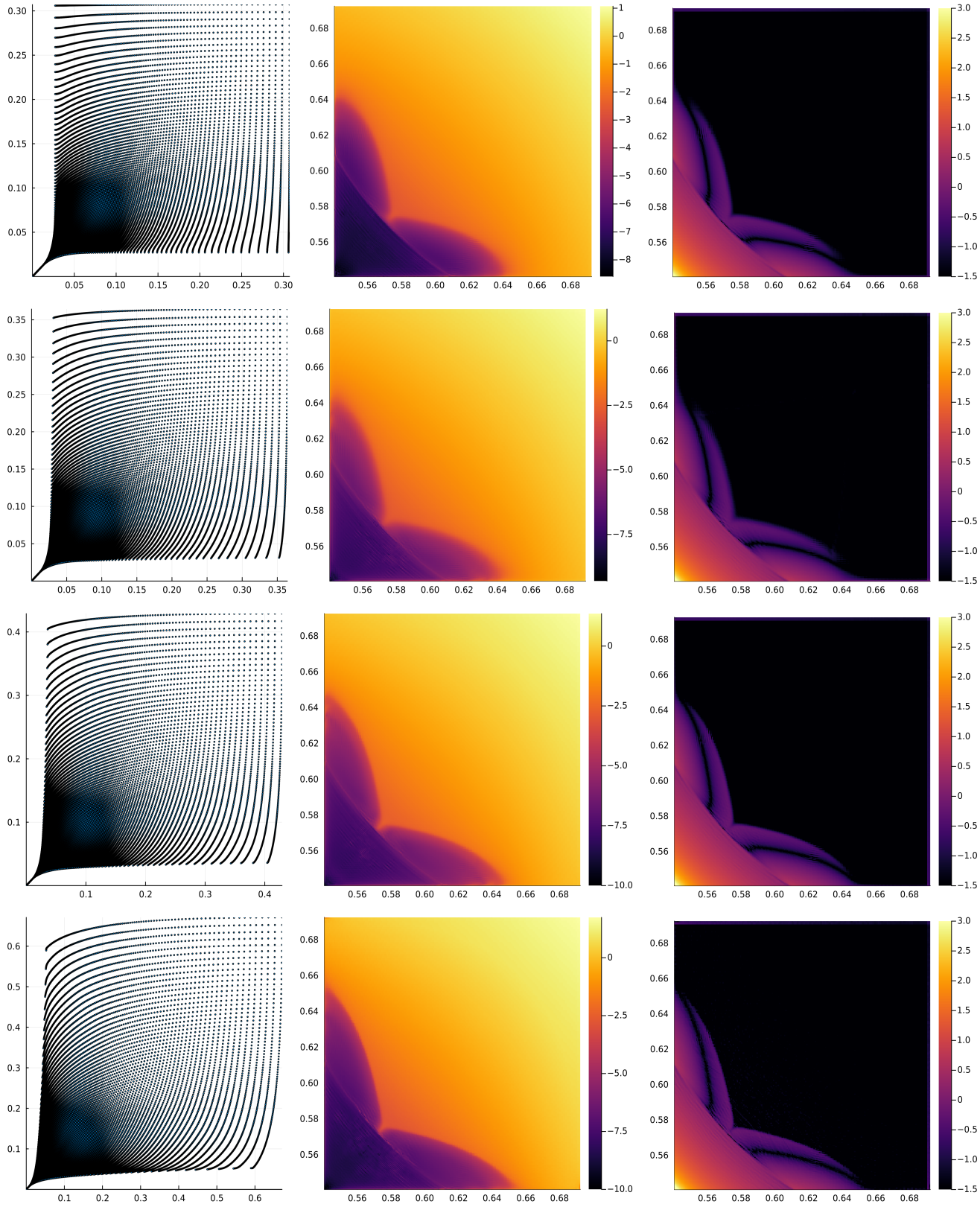


Figure B.1: Bunching in the Baseline Model with Variation in the Project Distribution

Figure B.1 illustrates bunching in the baseline model and variations with respect to $\Delta \in \{0.00, 0.05, 0.10, 0.25\}$. The left panel displays the allocation. The middle panel identifies bunching by analyzing the determinant of the Hessian matrix of the indirect utility function. The right panel identifies bunching using Corollary 2 by analyzing deviations from the multidimensional optimal tax formula. The variable on the horizontal axis is cognitive skill α_c; the variable on the vertical axis is manual skill α_m.

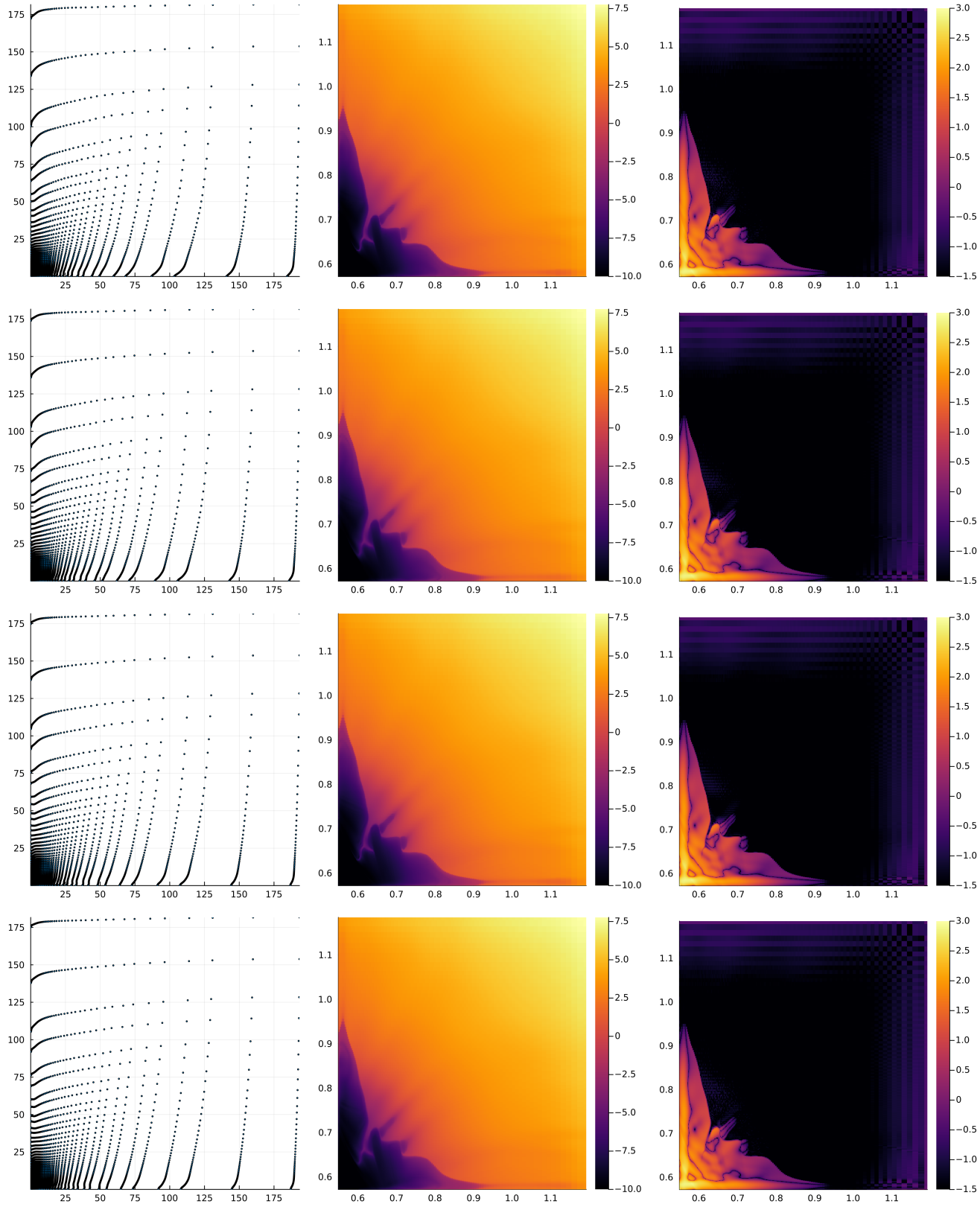


Figure B.2: Bunching in the Empirical Model with Variation in the Project Distribution

Figure B.2 illustrates bunching in the empirically relevant economy with respect to variations in $\Delta \in \{0.00, 0.1, 0.3, 0.5\}$. The left panel displays the allocation. The middle panel identifies bunching by analyzing the determinant of the Hessian matrix of the indirect utility function. The right panel identifies bunching using Corollary 2 by analyzing deviations from the multidimensional optimal tax formula. The variable on the horizontal axis is cognitive skill α_c; the variable on the vertical axis is manual skill α_m.

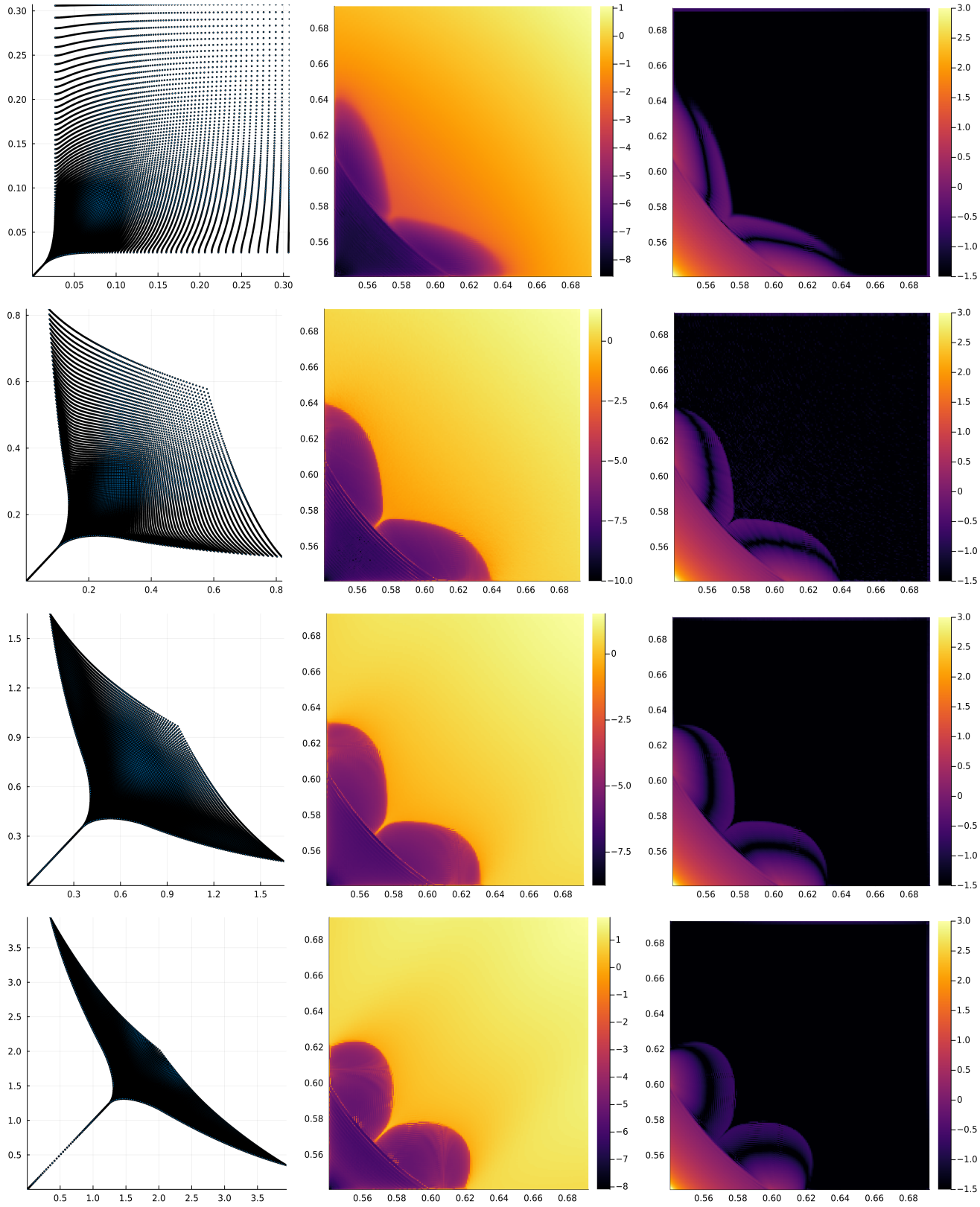


Figure B.3: Bunching in the Baseline Model with Variation in Consumption Utility

Figure B.3 illustrates bunching in the baseline model and variations with respect to $\gamma \in \{0, 0.25, 0.5, 0.75\}$. The left panel displays the allocation. The middle panel identifies bunching by analyzing the determinant of the Hessian matrix of the indirect utility function. The right panel identifies bunching using Corollary 2 by analyzing deviations from the multidimensional optimal tax formula. The variable on the horizontal axis is cognitive skill α_c; the variable on the vertical axis is manual skill α_m.

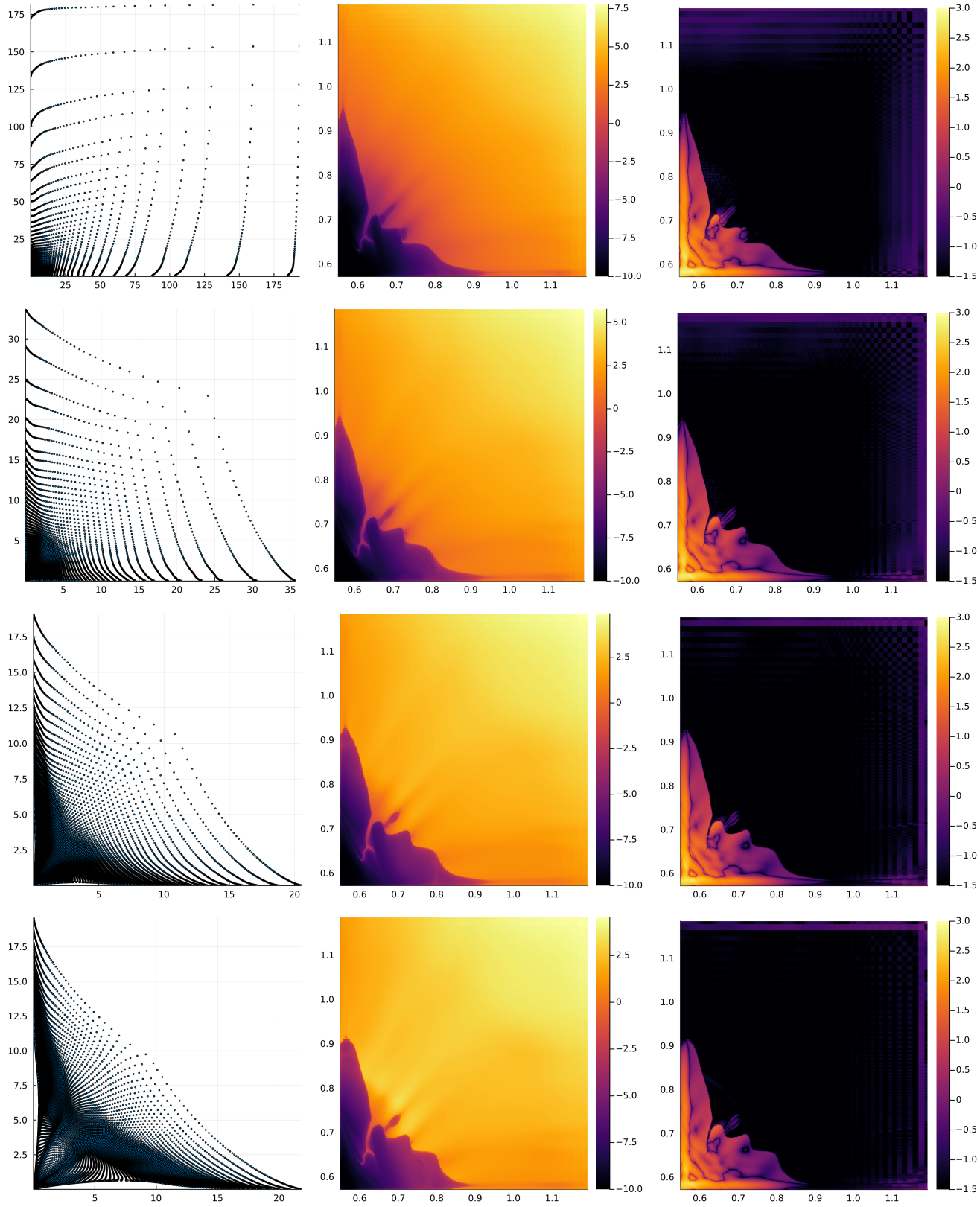


Figure B.4: Bunching in the Empirical Model with Variation in Consumption Utility

Figure B.4 illustrates bunching in the empirically relevant model with respect to variations in $\gamma \in \{0.00, 0.25, 0.50, 0.75\}$. The left panel displays the allocation. The middle panel identifies bunching by analyzing the determinant of the Hessian matrix of the indirect utility function. The right panel identifies bunching using Corollary 2 by analyzing deviations from the multidimensional optimal tax formula. The variable on the horizontal axis is cognitive skill α_c; the variable on the vertical axis is manual skill α_m.

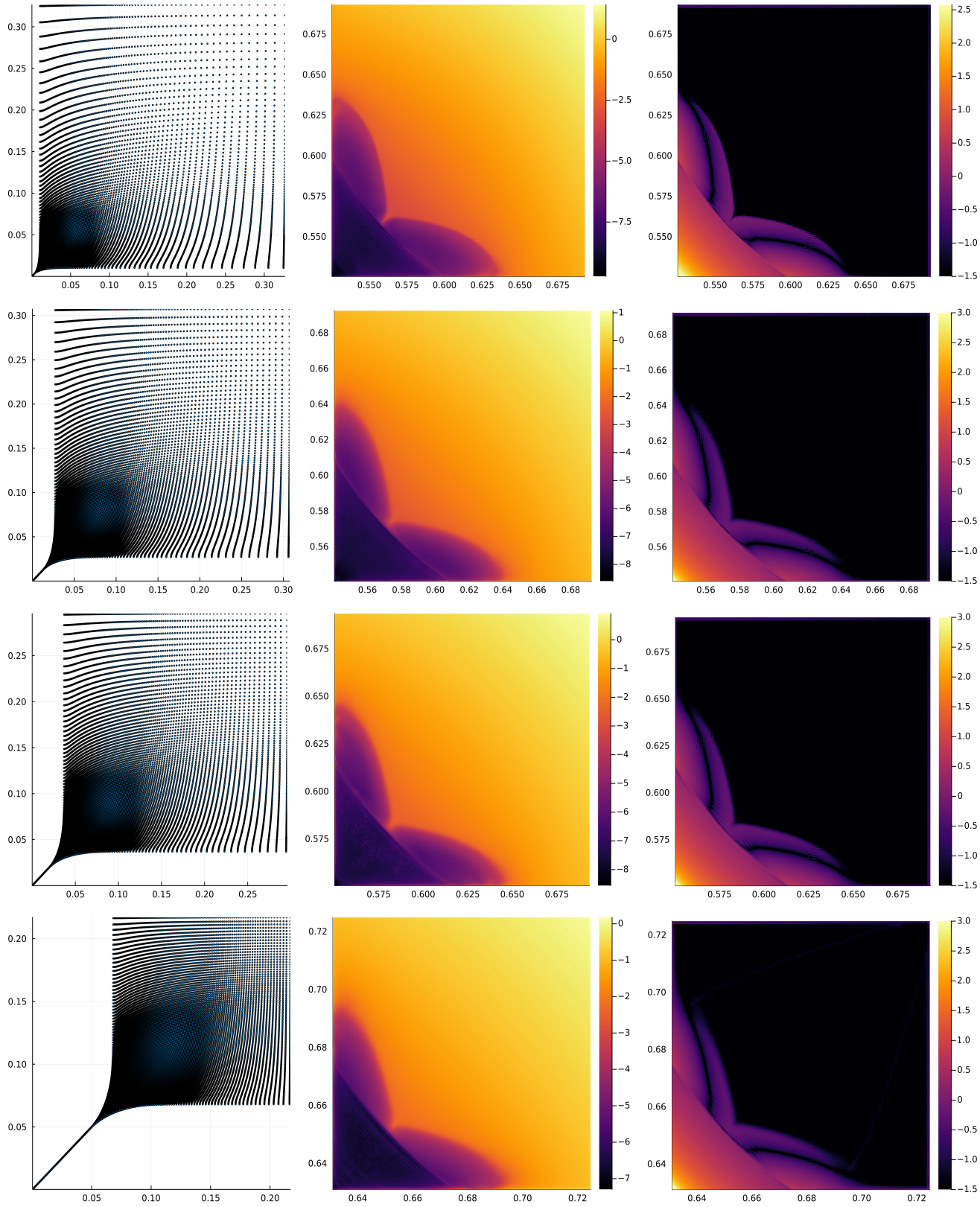


Figure B.5: Bunching in the Baseline Model with Variation in Labor Utility and Technology

Figure B.5 illustrates bunching in the baseline model with respect to variations in $\zeta \in \{2/2.5, 2/2.8, 2/3, 2/5\}$. The left panel displays the allocation. The middle panel identifies bunching by analyzing the determinant of the Hessian matrix of the indirect utility function. The right panel identifies bunching using Corollary 2 by analyzing deviations from the multidimensional optimal tax formula. The variable on the horizontal axis is cognitive skill α_c; the variable on the vertical axis is manual skill α_m.

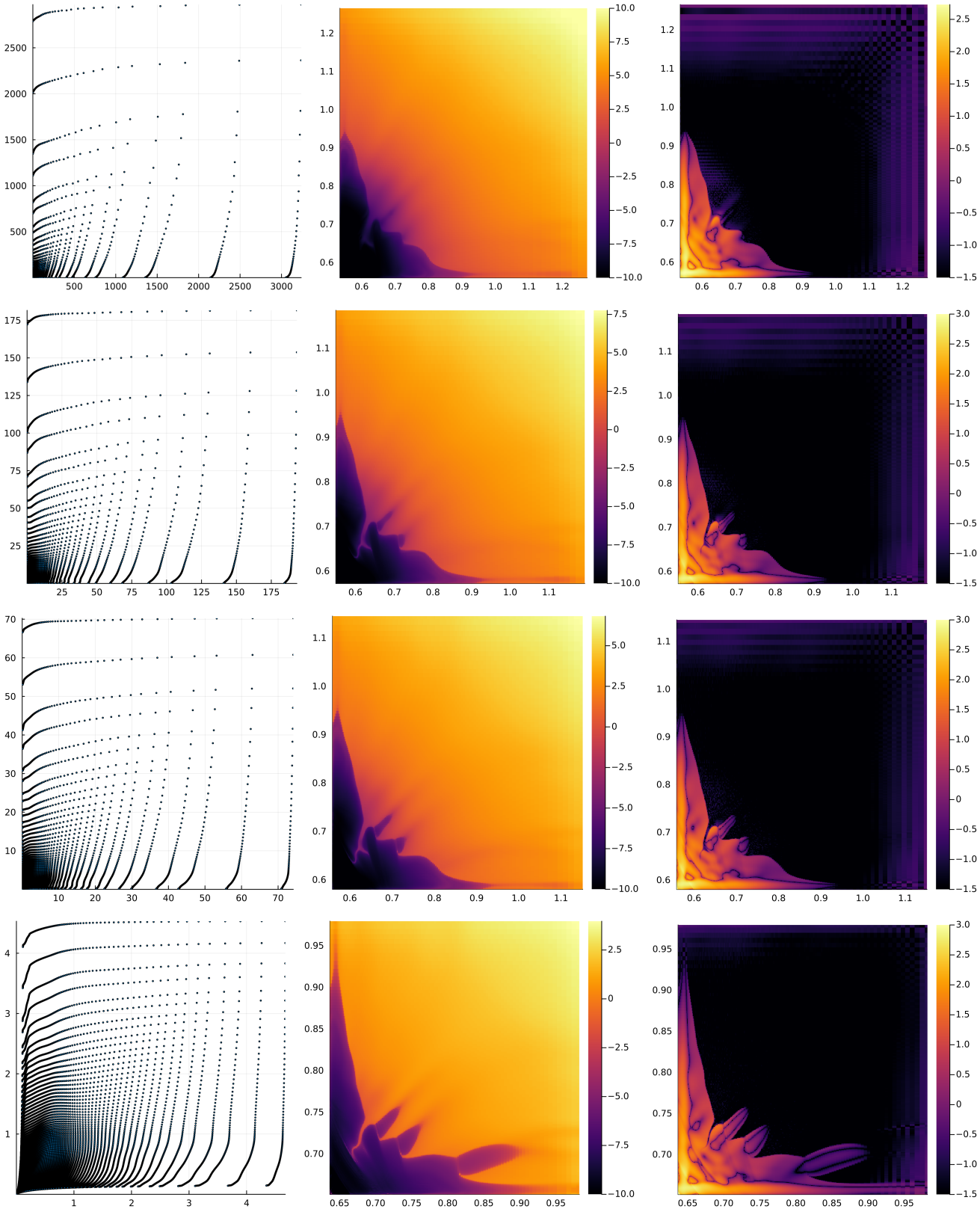


Figure B.6: Bunching in the Empirical Model with Variation in Labor Utility and Technology

Figure B.6 illustrates bunching in the empirically relevant model with respect to variations in $\zeta \in \{2/2.5, 2/2.8, 2/3, 2/5\}$. The left panel displays the allocation. The middle panel identifies bunching by analyzing the determinant of the Hessian matrix of the indirect utility function. The right panel identifies bunching using Corollary 2 by analyzing deviations from the multidimensional optimal tax formula. The variable on the horizontal axis is cognitive skill α_c; the variable on the vertical axis is manual skill α_m.