

NBER WORKING PAPER SERIES

THE OPTIMAL TAXATION OF COUPLES

Mikhail Golosov
Ilya Krasikov

Working Paper 31140
<http://www.nber.org/papers/w31140>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2023

Golosov thanks the NSF for support (grant #36354.00.00.00). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Mikhail Golosov and Ilya Krasikov. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Optimal Taxation of Couples
Mikhail Golosov and Ilia Krasikov
NBER Working Paper No. 31140
April 2023
JEL No. D82,H21

ABSTRACT

We consider optimal joint nonlinear earnings taxation of couples. We use multi-dimensional mechanism design techniques to study this problem and show that the first-order approach – that restricts attention to only local incentive constraints – is valid for a broad set of primitives. Optimal taxes are characterized by the solution to a certain second-order partial differential equation. Using the Coarea Formula, we solve this equation for various conditional averages of optimal tax rates and identify key forces that determine the optimal tax rates; show how these rates depend on earnings of each spouse, correlation in spousal earnings, and redistributive objectives of the planner; compare optimal rates for primary and secondary earners; identify both the conditions under which simple tax systems are optimal and the sources of welfare gains from more sophisticated taxes when those conditions are not satisfied. Under realistic assumptions, optimal tax rates for married individuals are increasing in correlation of spousal earnings. However, they remain lower than the tax rates for single individuals, and the marginal rates for one spouse increase (decrease) in the earnings of the other if both spouses have low (high) earnings.

Mikhail Golosov
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
and NBER
golosov@uchicago.edu

Ilia Krasikov
Department of Theoretical Economics
11 Pokrovsky Blvd, S1037
Moscow 109028
Russia
krasikovis.main@gmail.com

1 Introduction

A significant part of income taxes in developed countries is paid by households that consist of several adult members. For example, in the U.S., over 70 percent of federal individual income taxes are collected from married couples. Yet, the theory of optimal taxation of family income is poorly understood. What economic forces determine the shape and magnitude of the optimal joint earnings tax schedule? How should one member's taxes depend on the earnings of the other member? Is it ever optimal to tax each individual in a couple separately or to use total family income as a base for earnings taxation?

In this paper, we aim to make progress in answering these questions. To this end, we consider a canonical unitary household model. Each household comprises two spouses who choose their labor supplies and earnings to maximize joint utility. Productivities of spouses are heterogeneous and are drawn from some joint distribution. We consider the problem of designing a joint tax of earnings of each spouse that maximizes the sum of couples' utilities weighted with Pareto weights that are some decreasing function of couples' productivities.

We use the mechanism design approach in our study. The optimal tax problem can be equivalently cast as a problem of a fictitious mechanism designer who chooses allocations for all couples based on reports about their productivities subject to the incentive compatibility constraints. Since the vector of productivities is two-dimensional, our model falls into the class of multi-dimensional screening problems. The properties of solutions to such problems are not well understood.

We make progress in studying such problems in two directions. Firstly, we explore conditions under which our problem can be simplified by restricting attention to only local incentive constraints, the so-called first-order approach (FOA). The FOA has been the standard technique to analyze uni-dimensional tax models since the seminal work of Mirrlees (1971), but it is not clear whether it is valid in higher dimensions.¹ In the special case of our economy – in which productivity draws are independent between spouses and Pareto weights are separable in spousal productivities – we can explicitly characterize the necessary and sufficient condi-

¹For example, in their classic study of the optimal taxation of couples, Kleven et al. (2009) write (p. 538) “very few studies in the optimal tax literature have attempted to deal with multidimensional screening problems. The nonlinear pricing literature in industrial organization has analyzed such problems extensively. A central complication of multidimensional screening problems is that first order conditions are often not sufficient to characterize the optimal solution. The reason is that solutions usually display “bunching” at the bottom (Armstrong (1996), Rochet and Chone (1998)), whereby agents with different types are making the same choices.” To sidestep this perceived difficulty, Kleven et al. (2009) further restrict agents' choices by allowing one of the spouses to make only binary labor supply decisions. They explain (p. 538) “Our framework with a binary labor supply outcome for the secondary earner along with continuous earnings for the primary earner avoids the bunching complexities and offers a simple understanding of the shape of optimal taxes based on graphical exposition.”

tions that the primitives of our economy need to satisfy for the FOA to be valid. We compare these conditions to the analogous conditions in the uni-dimensional version of our model and show that the former set is strictly larger, so the FOA is *more likely* to hold in bi-dimensional settings.² For our general economy, in which productivity draws may not be independent and social weights may not be separable, we show that a sufficient condition for the FOA to hold is that Pareto weights are not, in a certain sense, too redistributive.

Secondly, we explore the properties of the optimal joint taxes under the assumption that the FOA is valid. Optimal taxes are characterized by the solution to a certain second-order non-linear partial differential equation. While there are no known techniques to solve this equation explicitly, we show that this roadblock can be partially side-stepped using a mathematical result known as the Coarea formula. It allows us to derive closed-form expressions for various conditional moments that the optimal taxes must satisfy. These moments isolate main economic forces determining optimal taxes and yield sharp comparative statics results. We show that, in a very broad sense, all these conditional moments are determined by a trade-off between benefits from redistribution, captured by Pareto weights, and deadweight costs of taxation, captured by two sets of elasticities: the pair of elasticities of labor supplies of the spouses, and various elasticities of the joint distribution of spousal productivities. Different distribution elasticities capture different aspects of the optimal joint taxes.

We use this general formula to obtain insights about various properties of the optimal tax schedule. First, we study its implication for the optimal average distortions³ that an individual of any given productivity faces. We show that these distortions are generally higher when couples are matched more assortatively in their productivities, with the strength of assortativeness captured by the standard positive quadrant dependence (PQD) order. This also implies that in the economy in which single and married households co-exists, the optimal taxes on married individuals should be lower than on single individuals, assuming same labor supply elasticities, redistributory objectives, and distribution of productivities for all, single or married, individuals.

These results are driven by the equity-efficiency trade-off. Our formulas show that average deadweight costs of tax distortions for any spouse depend only on spouse-specific characteristics but not on how they are matched with their partners. In contrast, average redistributive

²This result also allows us to explain that the failure of the FOA in multi-dimensional settings that was observed by Armstrong (1996) and Rochet and Chone (1998) is driven not by incentive constraints per se but by their interaction with participation constraints, which are typically absent in public finance applications.

³As it is standard in the optimal tax literature since the work of Mirrlees (1971), we characterize properties of the monotone transformation of the marginal tax rates, $\frac{\partial}{\partial y_i} T(y_1, y_2) / \left(1 - \frac{\partial}{\partial y_i} T(y_1, y_2)\right)$, where T is the tax function and y_i is earnings of spouse i . We refer to this object as the optimal tax distortion.

benefits of taxation depend on matching patterns. When individuals form couples, they pool their resources, implicitly redistributing resources within families. It is costly to crowd out this intra-family redistribution through distortionary taxation. The more random the matching is, the more redistribution occurs within families and the smaller tax rates the planner sets.

Second, we use our general formula to study jointness, i.e., how marginal tax rates of one spouse should depend on earnings of the other. We show that various average measures of optimal jointness depend on two moments of the joint distribution of spousal productivities, which can be written as elasticities of copula and the survival copula of this distribution. These moments have specific economic interpretation as they capture additional distortions that jointness introduces relative to any separable tax schedule (i.e., a tax schedule under which taxes of earnings of one spouse are independent of earnings of the other spouse).

Optimal jointness is determined by several competing forces. Positive jointness facilitates redistribution by targeting taxes to the richest couples at the costs of exacerbating tax distortions from separable taxes, which are summarized by the respective survival copula elasticity. On the other hand, negative jointness facilitates redistribution by targeting transfers to the poorest couples at the costs of exacerbating distortions from phase-outs of transfers from separable taxes, which are captured by the copula elasticity. Despite complex interaction of various forces, optimal average jointness in the tails, for the richest and poorest couples, is determined essentially by one property of the joint distribution of productivities – its tail (in)dependence. When the joint distribution is tail dependent (so that both copula and survival copula elasticities are small), very productive or unproductive individuals are likely to be matched with other very productive or unproductive individuals.⁴ In this case, the redistributive motives dominate and optimal jointness is positive at the top and negative at the bottom. If the joint distribution is tail independent, so that such extreme pairings are unlikely, then the distortion-reducing motive dominates. With the exception of several special cases that we discuss in the text, the optimal jointness for tail independent distributions is negative at the top and positive at the bottom.

Third, we also derive expressions that characterize optimal average family distortions and that compare distortions between two spouses in the same couple. We use these expressions to describe conditions under which taxation based on total family earnings is optimal and conditions under which secondary earners face on average higher or lower distortions than

⁴A symmetric joint distributions of two variables w_1, w_2 is right- (left-) tail independent if $\lim_{t \rightarrow \infty} \mathbb{P}(w_2 \geq t | w_1 \geq t) = 0$ (if $\lim_{t \rightarrow 0} \mathbb{P}(w_2 \leq t | w_1 \leq t) = 0$), it is right- and/or left- tail dependent if these limits are strictly positive. To give an example, a joint log-normal distribution is always tail-independent, a distribution with the perfectly assortative matching in the tails is tail dependent. In Section 4.3 we define tail dependence and independent formally for arbitrary joint distributions and discuss some of their implications.

primary earners.

Finally, we characterize optimal taxes if the planner is exogenously restricted to imposing taxes that are simpler than the fully unrestricted joint tax schedule. We consider three kinds of such taxes: anonymous taxes, taxes that are separable in spouses' earnings, and taxes that depend only on total family earnings. We show that these optimal restricted taxes can be analyzed by imposing additional measurability restrictions on our unrestricted mechanism design problem. Moreover, we prove that the optimal distortions under restricted taxes are always equal to the average distortions under optimal unrestricted taxes. This result helps us clarify sources of additional welfare gains that more sophisticated taxes attain.

In the last part of the paper, we characterize optimal taxes numerically. We use data on the earnings of married households and the U.S. tax schedule to obtain the joint distribution of productivities. We show that a Gaussian copula with Pareto-lognormal marginal distributions can well approximate this distribution. We then consider optimal taxes under a rich family of Pareto weights. We find that our analytical formulas provide excellent guidance about numerical properties of the optimal tax schedule. In the U.S. data, spousal productivities are positively but not perfectly dependent, so optimal taxes on married individuals are higher than in the economy with random matching but lower than in the uni-dimensional models such as Diamond (1998). The Gaussian copula is tail independent, so consistent with our analytical results, optimal jointness is positive for low earners and negative for high earners. The quantitative magnitude of this optimal jointness is small, so individual earnings-based taxes provide a good approximation to the optimal unrestricted tax schedule. In contrast, taxation based only on total family earnings is generally quite far from the unrestricted optimum, even when Pareto weights are chosen to explicitly favor family earnings-based taxation.

Our paper is related to several strands of literature. Small literature in public finance uses the multi-dimensional mechanism design approach to study optimal taxation. Mirrlees (1986) was perhaps the first to derive the partial differential equation that characterizes such taxes under the assumption that the FOA is valid but noted that it is much more difficult to solve than its uni-dimensional analog. Several authors imposed additional assumptions to simplify the multi-dimensional tax environment. For example, Kleven et al. (2009) studied taxation of couples but restricted one spouse to make only binary labor supply choices. Frankel (2014) considered the case in which a binary distribution describes spouses' productivities. Ales and Sleet (2022) studied couples taxation in a discrete choice environment. Moser and de Souza e Silva (2019) analyzed paternalistic savings policies in a model with two-dimensional discrete heterogeneity. Alves et al. (2021) considered the optimal tax problem of couples but imposed

enough structure to collapse it into a uni-dimensional problem. Golosov et al. (2013) and Lockwood and Weinzierl (2015) pursued a similar approach in labor and commodity taxation with preference heterogeneity. Hellwig and Werquin (2022) discussed the generalization of their ideas of redistributive arbitrage to multi-dimensional type spaces. In a series of papers, Rothschild and Scheuer (2013; 2014; 2016) developed a mechanism design approach to study optimal taxation in models with multi-dimensional private information but with uni-dimensional tax instruments. In contrast to these papers, we develop an approach that allows us to analytically characterize properties of optimal taxes in a fairly unrestricted multi-dimensional environment and shed light on economic forces that are hard to see in more specialized settings.⁵

The most closely related to our study is the unpublished Section 3 of the working paper by Kleven et al. (2007), henceforth KKS. In that section, KKS considered an economy similar to ours and made several very insightful observations. They noted that one should expect the FOA to hold in optimal tax settings when the planner is close to utilitarian, derived a formula that is analogous to our formula for the optimal average distortion, and characterized the sign of optimal jointness under the assumption that spousal productivity draws are independent. Our work systematically builds on these insights and allows us to obtain many novel results, such as a comparison of conditions for the validity of FOA in uni- and bi-dimensional settings, comparative statics results, implications of productivity dependence on optimal taxation, comparison of optimal tax rates for single and married individuals or spouses in the same couple, identification of the key conditions for optimality of separable taxation and taxation of family income.

Several authors, such as Golosov et al. (2014), Spiritus et al. (2022), Ferey et al. (2022) study optimal multidimensional taxation using an alternative, variational approach. They consider perturbations of tax schedules and derive expressions for optimal rates in terms of sufficient statistics. While their approach has many appealing features, its key limitation for our purposes is that the optimal tax rates are expressed in terms of endogenous objects that are themselves functions of the optimal tax schedule. This makes it difficult to use those expressions to understand how the model's structural parameters affect optimal taxes. In contrast, our formulas are in terms of exogenous primitives, which allows us to prove sharp

⁵In addition to these papers, our work is also related to the New Dynamic Public Finance literature (see., e.g., Golosov et al. (2003), Albanesi and Sleet (2006), Farhi and Werning (2013), Golosov et al. (2016), Stantcheva (2017), Ndiaye (2018)) that studies optimal nonlinear taxes in dynamic environments in which information is revealed over time. In those models, optimal taxes in a given period are a nonlinear function of earnings in previous periods, but the dynamic nature of information revelation allows collapsing the mechanism design problem to a sequence of problems with uni-dimensional incentive constraints. Also related is the recent work by Kushnir and Shourideh (2022) who explore alternative ways to relax multidimensional mechanism design problems.

theoretical results. That being said, we show in the paper how our formulas can be obtained using variational techniques by constructing perturbations that allows one to express optimal taxes in terms of model's primitives. Those perturbations differ from ones typically considered in the literature and should be of independent interest.

Gayle and Shephard (2019) and Spiritus et al. (2022) use numerical methods to study the optimal joint taxation of couples. Boerma et al. (2022) developed techniques to tackle multi-dimensional mechanism design problems when the FOA fails. Our work is complementary to theirs. Analytical results we derive provide insights about forces that determine optimal taxes that are often hard to see with numerical work.

The rest of the paper is organized as follows. In Section 2, we present our environment. In Section 3, we describe the mechanism design approach and conditions for the validity of the FOA. Section 4 characterizes optimal taxes. Section 6 provides calibration and quantitative analysis. Section 7 concludes.

2 Environment

We consider an economy with a continuum of couples. Each couple consists of two persons, whom we refer to as spouse 1 and spouse 2. Each couple is characterized by vector $\mathbf{w} = (w_1, w_2)$, which we call *type*, where w_i is the productivity of spouse i . We denote the distribution of types by F and its domain by W . We assume that W is a closed, simply-connected subset of \mathbb{R}_+^2 with a Lipschitz boundary ∂W and a non-empty interior W^o , and that it satisfies any of the following conditions: W is either (i) compact, or (ii) $W = \mathbb{R}_+^2$, or (iii) there exists some $\underline{\mathbf{w}} > \mathbf{0}$ such that $\mathbf{w} \geq \underline{\mathbf{w}}$ for all $\mathbf{w} \in W$. We assume that F has density f that is strictly positive on W^o , continuous on W , and continuously differentiable on W^o . We use G_1, G_2 and g_1, g_2 to denote marginals of F and f . If the domain is unbounded from above, then we require $\lim_{w_i \rightarrow \infty} -d \ln w_i / d \ln (1 - G_i(w_i)) < \infty$; and, if the domain includes 0, then we assume $\lim_{w_i \rightarrow 0} d \ln w_i / d \ln G_i(w_i) < \infty$. Finally, we use $f_i(\cdot|w_{-i})$ and $F_i(\cdot|w_{-i})$ to denote the density and the distribution of productivities of spouse i conditional on the productivity of spouse $-i$ being w_{-i} . These definitions imply that $f(\mathbf{w}) = f_i(w_i|w_{-i})g_{-i}(w_{-i})$ for all \mathbf{w} .

Spouse $i = 1, 2$ supplies earnings $y_i \geq 0$, and both spouses share consumption c . Let $\mathbf{y} = (y_1, y_2)$ be the vector of earnings of a couple. The utility of couple \mathbf{w} from bundle (c, \mathbf{y}) is given by

$$v = c - \gamma_1 \left(\frac{y_1}{w_1} \right)^{1/\gamma_1} - \gamma_2 \left(\frac{y_2}{w_2} \right)^{1/\gamma_2}, \quad (1)$$

where elasticity parameters⁶ $\gamma = (\gamma_1, \gamma_2)$ satisfy $\gamma_i \in (0, 1)$ for each i . We assume that F and γ satisfy

$$\mathbb{E} \left[w_1^{1/(1-\gamma_1)} + w_2^{1/(1-\gamma_2)} \right] < \infty, \quad (2)$$

which ensures that the surplus is finite in our economy.

Couples choose jointly (c, \mathbf{y}) to maximize preferences (1) subject to the budget constraint

$$c \leq y_1 + y_2 - T(y_1, y_2), \quad (3)$$

where $T : \mathbb{R}_+^2 \rightarrow \mathbb{R} \cup \{\infty\}$ is a tax function. We are interested in characterizing properties of tax functions that maximize social welfare. Social welfare is defined as $\int_W v(\mathbf{w}) \alpha(\mathbf{w}) f(\mathbf{w}) d\mathbf{w}$, where $\alpha : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is a decreasing, continuous and bounded function satisfying $\mathbb{E}[\alpha] = 1$. We refer to α as Pareto weights. A tax function is said to be *budget-feasible* if the total tax revenue collected from couples is non-negative⁷ and *optimal* if it is budget-feasible and there is no other budget-feasible tax function under which social welfare is strictly higher.

We postulate that T is a function of labor earnings \mathbf{y} but not productivities \mathbf{w} . This restriction can be micro-founded if we assume that \mathbf{w} is unobservable to the government. However, the analysis is the same if the government is exogenously restricted to use taxes that are a function only of earnings \mathbf{y} , irrespective of whether couples' productivities are observable. Thus, we keep this slightly more general formulation to be agnostic about observability of \mathbf{w} .

Our paper's primary focus is understanding properties of the optimal tax schedule T^* without imposing any further ad-hoc restrictions on this joint tax function. The economy is purposefully designed to focus on the simplest bi-dimensional model of taxation. The joint distribution of types F is given exogenously so that neither formation nor dissolution of couples is affected by taxes. Moreover, our model abstracts from income effects, bargaining within couples, and home production. While all these extensions are interesting and important, we focus in this study on the specification that allows us to highlight key mechanisms and trade-offs in designing optimal joint taxation most transparently. As a part of our analysis, we also discuss conditions under which this unrestricted tax function satisfies additional properties. We say that a tax function is *anonymous* if $T(y_1, y_2) = T(y_2, y_1)$, *individual earnings-based* if $T(y_1, y_2) = \tilde{T}_1(y_1) + \tilde{T}_2(y_2)$ for some functions \tilde{T}_1 and \tilde{T}_2 , and *family earnings-based* if $T(y_1, y_2) = T(y_1 + y_2)$ for all \mathbf{y} . In Section 5, we ask the opposite question of how to design optimal taxes that are exogenously restricted also to be anonymous, individual earnings-based,

⁶Note that the elasticity of earnings of spouse i is $\gamma_i/(\gamma_i - 1)$, so it is monotonically increasing in γ_i .

⁷Formally, T is budget-feasible if one can construct (Borel) measurable functions $(c, \mathbf{y}) : W \rightarrow \mathbb{R} \times \mathbb{R}_+^2$ such that for each \mathbf{w} , the pair $(c(\mathbf{w}), \mathbf{y}(\mathbf{w}))$ is an optimal choice of couple \mathbf{w} and $\int_W T(\mathbf{y}(\mathbf{w})) f(\mathbf{w}) d\mathbf{w} \geq 0$.

or family earnings-based, and we uncover a tight relationship between optimal restricted and unrestricted taxation.

Our general economy is characterized by a triple (F, α, γ) . Several special cases of this economy will be particularly useful. We call Pareto weights α *separable* if they can be written as

$$\alpha(\mathbf{w}) = \beta_1 \tilde{\alpha}_1(w_1) + \beta_2 \tilde{\alpha}_2(w_2), \quad (4)$$

where $(\tilde{\alpha}_1, \tilde{\alpha}_2)$ are non-negative, non-increasing, continuous, bounded functions that, without loss of generality, satisfy $\mathbb{E}[\tilde{\alpha}_i] = 1$, $\beta_i \geq 0$ for all i and $\beta_1 + \beta_2 = 1$. We call an economy (F, α, γ) *symmetric* if F and α are symmetric and $\gamma_1 = \gamma_2$, and *benchmark* if it is symmetric and Pareto weights are separable. Any (bi-dimensional) benchmark economy (F, α, γ) has a corresponding uni-dimensional economy, parameterized by $(G, \tilde{\alpha}, \gamma)$, that is essentially the one studied by Diamond (1998).⁸

The benchmark economy is a useful case to consider for at least two reasons. The first reason is that non-separable, in particular super- and sub-modular Pareto weights, sometimes have opposing implications for the value of redistribution in different parts of productivity space. The benchmark economy, with its separable weights and no intrinsic distinction between spouses, presents a natural starting point for the analysis. The second reason is that the benchmark economy allows for a simple and natural comparison of our results to the existing results about optimal taxation in uni-dimensional settings. Consider an economy in which all individuals have the same elasticity of labor supply and draw productivities from the same distribution G , and in which some individuals remain single while others form couples with joint distribution F . If the relationship between Pareto weights for single-person households $\tilde{\alpha}$ and married-persons households α is given by $\alpha(w_1, w_2) = \frac{1}{2}\tilde{\alpha}(w_1) + \frac{1}{2}\tilde{\alpha}(w_2)$ then the optimal taxes on couples coincide with optimal taxes in our benchmark economy (F, α, γ) while optimal taxes on singles coincides with the optimal taxes in the corresponding uni-dimensional economy $(G, \tilde{\alpha}, \gamma)$. See the appendix for the formal description of this environment and proofs.

2.1 Dependence concepts and copulas

As we shall see, optimal taxes depend crucially on how individuals are matched into couples, which is captured by properties of the joint distribution F . Recalling several standard dependence concepts for bi-variant random variables will be useful.⁹ Types are *independent* if

⁸Like us, Diamond focuses on the economy without income effects. Unlike our model, he allows for the general disutility of earnings $\nu(\frac{y}{w})$ rather than the constant elasticity specification $\gamma(\frac{y}{w})^{1/\gamma}$. He describes welfare by social welfare functions rather than Pareto weights. Our specification with constant labor supply elasticity and Pareto weights allows us to obtain a sharper characterization of optimal taxes.

⁹See Nelsen (2006), Shaked and Shanthikumar (2007), or Balakrishnan and Lai (2009) for textbook treatment.

$F(\mathbf{w}) = G_1(w_1)G_2(w_2)$ for all $\mathbf{w} \in W$ (or $F = G_1G_2$, for short) and positively dependent (also known as *positively quadrant dependent* or PQD) if $F \geq G_1G_2$. Positive dependence is equivalent to the condition that $\text{Cov}(\phi_1(w_1), \phi_2(w_2)) \geq 0$ for any two increasing functions ϕ_1 and ϕ_2 . A distribution F^b is *more dependent* than F^a if F^a and F^b have the same marginals and $F^b \geq F^a$. We denote this by $F^b \geq_{PQD} F^a$. We write $F^b >_{PQD} F^a$ when $F^b \geq_{PQD} F^a$ but not $F^a \geq_{PQD} F^b$. Any F satisfies bounds $F^* \geq_{PQD} F \geq_{PQD} F_*$, where F^* and F_* are distributions under perfect positive and negative assortative matching given marginal G_1, G_2 .

Any distribution F has corresponding functions C and \bar{C} , called *copula* and *survival copulas*, that satisfy $C(G_1(t_1), G_2(t_2)) = \mathbb{P}(\mathbf{w} \leq \mathbf{t})$ and $\bar{C}(1 - G_1(t_1), 1 - G_2(t_2)) = \mathbb{P}(\mathbf{w} \geq \mathbf{t})$ for all $\mathbf{t} \in W$, respectively. Conceptually, $C(u_1, u_2)$ is the joint probability that the productivity of spouse 1 is in the u_1^{th} quantile of her marginal distribution and the productivity of spouse 2 is in the u_2^{th} quantile. Copulas allow one to isolate dependence properties of F from properties of marginal distributions G_1, G_2 in general settings. There exists many families of copulas that are used to model dependence between two variables with arbitrary marginals, e.g., FGM, Gaussian, or t-copulas.¹⁰

3 The mechanism design problem

We study our optimal tax problem using the mechanism design approach. Consider a direct mechanism, in which a fictitious mechanism designer asks each couple their type \mathbf{w} and allocates bundles (c, \mathbf{y}) directly as a function of these reports, subject to feasibility and incentive compatibility of these allocations. A direct mechanism that maximizes social welfare can be written as a choice of functions (v, c, \mathbf{y}) that solve the following problem:

$$\max_{(v, c, \mathbf{y})} \int_W v(\mathbf{w}) \alpha(\mathbf{w}) f(\mathbf{w}) d\mathbf{w} \quad (5)$$

subject to $\mathbf{y} \geq \mathbf{0}$,

$$v(\mathbf{w}) = c(\mathbf{w}) - \gamma_1 \left(\frac{y_1(\mathbf{w})}{w_1} \right)^{1/\gamma_1} - \gamma_2 \left(\frac{y_2(\mathbf{w})}{w_2} \right)^{1/\gamma_2} \quad \forall \mathbf{w} \in W, \quad (6)$$

¹⁰The FGM copula is defined by $C(u_1, u_2) = u_1 u_2 [1 + \rho(1 - u_1)(1 - u_2)]$. The Gaussian and t-copulas generalize notions of bi-variate normal and Student t-distribution to arbitrary marginals. They are defined as $C(u_1, u_2) \propto \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \exp\left[-\frac{(s^2 - 2\rho st + t^2)}{2(1 - \rho^2)}\right] ds dt$ and $C(u_1, u_2) \propto \int_{-\infty}^{\Gamma^{-1}(u_1; \nu)} \int_{-\infty}^{\Gamma^{-1}(u_2; \nu)} \left[1 + \frac{(s^2 - 2\rho st + t^2)}{\nu(1 - \rho^2)}\right]^{-\frac{\nu+2}{2}} ds dt$ respectively, where $\Phi(\cdot)$ is the standard normal distribution and $\Gamma(\cdot; \nu)$ is the Student t-distribution with ν degrees of freedom. Note that bi-variate log-normal distributions are a special case of the Gaussian copula. For all three copulas, the parameter $\rho \in (-1, 1)$ captures dependence, with a higher value of ρ corresponding to stronger dependence in the PQD sense.

$$\int_W (y_1(\mathbf{w}) + y_2(\mathbf{w}) - c(\mathbf{w})) f(\mathbf{w}) d\mathbf{w} \geq 0, \quad (7)$$

and

$$c(\mathbf{w}) - \sum_{i=1}^2 \gamma_i \left(\frac{y_i(\mathbf{w})}{w_i} \right)^{1/\gamma_i} \geq c(\widehat{\mathbf{w}}) - \sum_{i=1}^2 \gamma_i \left(\frac{y_i(\widehat{\mathbf{w}})}{w_i} \right)^{1/\gamma_i} \quad \forall \mathbf{w}, \widehat{\mathbf{w}} \in W, \quad (8)$$

where maximization is over the space of functions measurable on W . The last two equations are feasibility and incentive compatibility, respectively.¹¹ The close connection between the optimal tax problem and the mechanism design problem is described by the following lemma.

Lemma 1. *For any budget-feasible tax function T , an optimal choice (c, \mathbf{y}) must satisfy (6) and (8). Conversely, for any (c, \mathbf{y}) that satisfy (7) and (8), there exists a budget-feasible T under which $(c(\mathbf{w}), \mathbf{y}(\mathbf{w}))$ is an optimal choice for every couple \mathbf{w} .*

The proof of this lemma is standard and given in the appendix. It shows that we can characterize optimal taxes in two steps: first solve the mechanism design problem (5) to find optimal allocations, then use these allocations to find a tax function that supports them as an equilibrium. Due to Lemma 1, this tax function is budget-feasible, and there cannot be any other budget-feasible tax function that yields higher welfare. Therefore, this characterizes the optimal tax function. We say that such T decentralizes (v, c, \mathbf{y}) . It is easy to show that in a symmetric economy such T can be constructed to be anonymous.

Let $\nabla v := (\nabla_1 v, \nabla_2 v)$ be the gradient of v . Constraints (6), (7) and (8) imply the following result.

Lemma 2. *Suppose (v, c, \mathbf{y}) satisfy (6), (7) and (8). Then, v is (i) increasing and bounded below, (ii) locally Lipschitz continuous on W° , (iii) both weakly and a.e. differentiable on W° , satisfies (iv) $\int_W |v(\mathbf{w})| f(\mathbf{w}) d\mathbf{w} < \infty$, (v) $\int_W \sum_{i=1}^2 w_i^{1+\gamma_i} (\nabla_i v(\mathbf{w}))^{\gamma_i} f(\mathbf{w}) d\mathbf{w} < \infty$ and (vi) $\int_W \sum_{i=1}^2 \gamma_i w_i \nabla_i v(\mathbf{w}) f(\mathbf{w}) d\mathbf{w} < \infty$. The (weak) gradient of v is given by*

$$\nabla v(\mathbf{w}) = \left(\frac{y_1^{1/\gamma_1}(\mathbf{w})}{w_1^{1+1/\gamma_1}}, \frac{y_2^{1/\gamma_2}(\mathbf{w})}{w_2^{1+1/\gamma_2}} \right). \quad (9)$$

Conditions (i)-(vi) are smoothness, monotonicity, and boundedness conditions that will be useful in various parts of our analysis. We use \mathcal{V} to denote the space of functions that satisfy these conditions. While v may not be differentiable everywhere in the classical sense, it must be weakly differentiable so that its gradient can be used in integration similar to the classical derivatives.¹² Equation (9) gives the explicit expression for this gradient. This gradient can

¹¹Slightly abusing notations, we set $y_i(\mathbf{w})/w_i = 0$ for all \mathbf{w} with $w_i = 0$.

¹²In Section B of the appendix, we overview the main properties of weak derivatives and Sobolev spaces that are used in our analysis.

be interpreted as local incentive constraints as it captures the marginal effect of misreporting their type by either of the spouses. It is a bi-dimensional version of the condition familiar from the uni-dimensional analysis.

We now can use Lemma 2 to simplify the mechanism design problem (5). Observe that at the optimum the feasibility constraint (7) must hold with equality. We can use equations (6) and (9) to substitute for c and \mathbf{y} , respectively. Bringing the feasibility constraint (7) into the objective function, problem (5) can be re-written as¹³

$$\max_{v \in \mathcal{V}} \int_W v(\mathbf{w}) (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} + \int_W \sum_{i=1}^2 \left(w_i^{1+\gamma_i} (\nabla_i v(\mathbf{w}))^{\gamma_i} - \gamma_i w_i \nabla_i v(\mathbf{w}) \right) f(\mathbf{w}) d\mathbf{w} \quad (10)$$

subject to

$$v(\mathbf{w}) \geq v(\widehat{\mathbf{w}}) + \sum_{i=1}^2 \gamma_i \widehat{w}_i \nabla_i v(\widehat{\mathbf{w}}) \left(\left(\frac{\widehat{w}_i}{w_i} \right)^{1/\gamma_i} - 1 \right) \quad \forall \mathbf{w}, \widehat{\mathbf{w}} \in W. \quad (11)$$

The objective function in this reformulated problem depends only on function v , and we can refer to this functional as $\Upsilon(v)$. Thus, the mechanism design problem can be written succinctly as

$$\max_{v \in \mathcal{V}} \Upsilon(v) \quad \text{s.t. (11)}. \quad (12)$$

Observe that all the local incentive constraints are already inside the functional Υ , so constraints (11) can affect the solution to this maximization problem only if some of the non-local constraints bind. The *relaxed problem* is the mechanism design problem in which these constraints are dropped, i.e., the problem defined as

$$\max_{v \in \mathcal{V}} \Upsilon(v). \quad (13)$$

Let v^* be a solution to the relaxed problem (13). We say that *the first-order approach (FOA) is valid* if v^* is also a solution to (12).

3.1 Validity of the FOA

The analysis of the mechanism design problem substantially simplifies when the FOA is valid. In uni-dimensional settings, essentially all analytical characterization of optimal taxation assumes that the FOA holds (seminal papers by Mirrlees (1971), Diamond (1998), and Saez (2001) all fall into this category). While it is known that the FOA may fail in uni-dimensional

¹³The solution to this problem is defined up to a constant, i.e., if v is a solution, then for any $b \in \mathbb{R}$, $v + b$ is also a solution. Once this problem is solved, the specific value of b can be found from the resource constraint (7).

settings for some parameter values, those cases appear to be rare in realistic applications that verify FOA validity numerically (see, e.g., Farhi and Werning (2013), Golosov et al. (2016), or Heathcote and Tsujiyama (2021)). One common concern, exemplified by the quote from Kleven et al. (2009) given in the introduction, is whether the first-order approach is ever valid in multi-dimensional settings. In this section, we examine conditions for validity of the FOA in our bi-dimensional settings and compare them to condition for validity of the FOA in a uni-dimensional version of our model.

One way to check the validity of the FOA in our model is to consider a transformation of space W into space X defined as $X := \left\{ \left(w_1^{-1/\gamma_1}, w_2^{-1/\gamma_2} \right) \mid \mathbf{w} \in W \right\}$. The relationship between our function v on space W and its transformation v^x on space X is given by $v^x(\mathbf{x}) := v\left(x_1^{-\gamma_1}, x_2^{-\gamma_2}\right)$. The advantage of this transformation is that utility of couples is linear in types in the X -space,

$$v^x(\mathbf{x}) = c^x(\mathbf{x}) - \gamma_1 x_1 (y_1^x(\mathbf{x}))^{1/\gamma_1} - \gamma_2 x_2 (y_2^x(\mathbf{x}))^{1/\gamma_2},$$

and the incentive constraint (11) becomes

$$v^x(\mathbf{x}) \geq v^x(\hat{\mathbf{x}}) + \sum_{i=1}^2 \nabla_i v^x(\mathbf{x})(x_i - \hat{x}_i) \quad \forall \mathbf{x}, \hat{\mathbf{x}} \in X. \quad (14)$$

A well-known result by Rochet (1987) gives a simple geometric interpretation of this constraint.

Lemma 3. (*Proposition 2 in Rochet (1987)*) *Suppose that W is such that X is a convex set. Let $v^{x,*}(\mathbf{x}) := v^*\left(x_1^{-\gamma_1}, x_2^{-\gamma_2}\right)$ for all $\mathbf{x} \in X$. The FOA is valid if and only if $v^{x,*}$ is convex.*

This lemma shows that validity of the FOA is equivalent to convexity of $v^{x,*}$, and the rest of this section focuses on conditions under which this convexity can be verified.¹⁴

We first examine validity of the FOA in one special case of our environment: the benchmark economy with independent types. In this case, we can solve (13) analytically and directly describe necessary and sufficient conditions that the primitives must satisfy for $v^{x,*}$ to be convex. These conditions can then be compared to their analogues in the uni-dimensional economy.

Proposition 1. *Consider the benchmark economy with independent types and $W = [\underline{w}, \bar{w}]^2$, where $0 \leq \underline{w} < \bar{w} \leq \infty$. Suppose $\tilde{\lambda}(t) := \frac{\int_{\underline{w}}^{\bar{w}} (1 - \tilde{\alpha}(w)) g(w) dw}{\gamma t g(t)}$ is continuously differentiable with bounded derivatives.*

¹⁴In uni-dimensional settings, Mirrlees (1971) showed that the FOA is valid if earning function $y^*(w)$, implied by the solution to the relaxed problem, is increasing in w . It can be shown that Mirrlees's result is equivalent to convexity of $v^{x,*}$.

(2D). The FOA is valid if and only if

$$x \cdot \left(1 + \frac{1}{2} \tilde{\lambda}(x^{-\gamma})\right) \text{ is increasing in } x. \quad (15)$$

(1D). In the corresponding uni-dimensional economy, the FOA is valid if and only if

$$x \cdot \left(1 + \tilde{\lambda}(x^{-\gamma})\right) \text{ is increasing in } x. \quad (16)$$

In particular, (15) holds whenever (16) holds.

Proposition 1 fully describes the set of (G, α, γ) under which the FOA holds and does not hold in both uni-dimensional and bi-dimensional settings with independent types. Examination of these conditions reveals that the set of parameters under which the FOA is valid is strictly larger in the bi-dimensional economy. Therefore, the FOA is *more likely* to hold in bi- than in uni-dimensional settings.

It is useful to consider the economic interpretation of these conditions. We start with the uni-dimensional case. Condition (16) can equivalently be written as

$$\left[1 + \tilde{\lambda}(w) \left(1 + \gamma \frac{\partial \ln(wg(w))}{\partial \ln w}\right)\right] + [1 - \tilde{\alpha}(w)] \geq 0 \text{ for all } w. \quad (17)$$

The term in the first square brackets is typically positive; the term in the second square brackets is negative for low w and positive for high w . Thus, inequality (17) is violated if the second term is sufficiently negative relative to the first, which occurs if the planner puts sufficiently high Pareto weights on some low types. In other words, the FOA holds in the uni-dimensional settings if the planner is not “too redistributive” in the precise sense given by equation (17). In the bi-dimensional economy, the analog of condition (17) is similar but allows for a larger set of weights $\tilde{\alpha}$. We discuss the intuition for this result once we characterize optimal taxes in Section 4.

The conclusion that the set of parameters for which FOA holds is strictly larger in a bi-dimensional economy depends on the assumption that $\tilde{\alpha}$ is a continuous function. This result does not extend to the cases when $\tilde{\alpha}$ is discontinuous on the boundary of the domain of G . One example of such weights is Rawlsian, $\tilde{\alpha}^R$, represented by the Dirac delta function that puts all the mass on the lowest type. With Rawlsian weights, the FOA typically fails in the bi-dimensional settings but holds for all (G, γ) that satisfy (16) on the *interior* of the domain of G in the uni-dimensional settings. However, the Rawlsian case in uni-dimensional settings is quite singular since the FOA fails for all continuous weights $\tilde{\alpha}$ that are close to $\tilde{\alpha}^R$ (such

weights must necessarily violate equation (17)). We provide formal statements and proofs of these results in the appendix.¹⁵

It is harder to check validity of the FOA for arbitrary (F, α) . As we show in the next section, v^* (or its transformation $v^{x,*}$) is characterized by a non-linear second-order elliptic PDE with Neumann boundary conditions, and convexity of solutions to even linear problems of this type is an open question in the theory of partial differential equations. Instead, we build on the ideas presented in the working paper by Kleven et al. (2007). KKS pointed out that the FOA must always hold for the utilitarian weights, $\alpha^0 \equiv 1$, since, in this case, the planner does not value redistribution and the *laissez-faire* allocation, which trivially satisfies the FOA, is optimal. KKS suggested that the Implicit Function Theorem can be used to establish the validity of the FOA for Pareto weights that are close to utilitarian.

To make these ideas operational, we need to overcome two technical challenges. First, we must verify the invertibility conditions required by the Implicit Function Theorem at the utilitarian solution. Second, we must ensure that convergence implied by the Implicit Function Theorem preserves convexity.

Proposition 2. *Suppose that $W \subset \mathbb{R}_{++}^2$ is compact, and the corresponding space X is convex. Suppose that F and ∂X are in $\mathcal{C}^{2,a}$ for some $a \in (0, 1]$. Take any sequence of weights $(\alpha^\varepsilon)_\varepsilon$ parameterized by $\varepsilon \geq 0$ such that α^ε is in $\mathcal{C}^{0,a}$, α^0 is identically equal to one, and $\alpha^\varepsilon \rightarrow \alpha^0$ in the $\mathcal{C}^{0,a}$ -norm. Then, the FOA is valid in the $(F, \alpha^\varepsilon, \gamma)$ economy for all ε small enough.*

In this proposition, we overcome the two challenges mentioned above in the following way. First of all, we formulate our problem in Hölder space¹⁶ $\mathcal{C}^{2,a}$. This is the “largest” Banach space that includes twice continuously differentiable functions and for which general results concerning existence of solutions to elliptic PDEs (Schauder Theory) are available.¹⁷ The

¹⁵This discussion is also pertinent for understanding the relationship between our results in Proposition 1 and the results of Rochet and Chone (1998) on failure of the FOA in multi-dimensional non-linear pricing models. An important difference between our model (as well as most models of optimal taxation) and non-linear pricing models is the existence of participation constraints. In our model, the participation constraint is absent. The government imposes taxes to maximize agents’ welfare, and agents cannot walk away from their tax obligations. In non-linear pricing models, the participation constraint plays a central role. Monopolist aims to extract maximum surplus from consumers, and only consumers’ ability to walk away from the contract restricts the amount of resources the monopolist can extract. The Lagrange multipliers on participation constraints act like Pareto weights in our settings. Since private information implies that more productive types can always attain the utility of less productive types, only the Lagrange multiplier on the lowest type can bind. So, Lagrangians in non-linear pricing models with participation constraints resemble our model with Rawlsian weights.

¹⁶In Section B of the appendix, we overview Hölder and other functional spaces that will be used in the sequel. Space $\mathcal{C}^{k,a}$ is a space of k -times continuously differentiable functions with sufficiently well-behaved (i.e., a -Hölder continuous) derivatives. Every $(k + 1)$ -times continuously differentiable function with bounded derivatives belongs to Hölder space $\mathcal{C}^{k,a}$.

¹⁷It is well-known that even the simplest PDE $\nabla_{11}v + \nabla_{22}v = \phi$ might fail to admit a twice continuously differentiable solution if ϕ is merely continuous.

utilitarian solution $v^{x,*,0}$ can be easily found explicitly. It is smooth and strongly convex, which ensures both that $v^{x,*,0} \in \mathcal{C}^{2,a}$ and that all other functions in the neighborhood of $v^{x,*,0}$ (in $\mathcal{C}^{2,a}$ space and its norm topology) are strongly convex as well. The Implicit Function Theorem for Banach spaces then establishes the conclusion of Proposition 2. The hardest part is checking the invertibility conditions necessary for applying the Implicit Function Theorem. This condition is equivalent to showing that a solution to a certain linear PDE exists. We show it by using results in Nardi (2015), and smoothness and compactness assumptions given in Proposition 2 are sufficient conditions for his results to hold.

The general message that emerges from Propositions 1 and 2 is that, similarly to the unidimensional models, the FOA is valid in our settings as long as the social planner is not too redistributive. Motivated by this finding, for the rest of the paper, we assume that the FOA is valid and describes properties of the optimal taxes implied by the solution to the relaxed problem (13).

3.2 Optimality conditions to the relaxed problem and optimal taxes

We now describe the necessary and sufficient conditions that solutions to the relaxed problem (13) must satisfy. To streamline the exposition, we do not provide formal statements of the results in this section but present all conditions heuristically, implicitly assuming that all functions are sufficiently smooth and well-behaved. The appendix provides all formal statements and proofs.

It is more convenient to derive optimality conditions not for v itself but for its transformation $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, which we define by

$$\lambda_i(\mathbf{w}) := (\nabla_i v(\mathbf{w}))^{\gamma_i - 1} w_i^{\gamma_i} - 1. \quad (18)$$

Transformation $\boldsymbol{\lambda}$ has a natural economic interpretation. If v is incentive compatible then the tax function T that decentralizes this allocation satisfies the following condition:

$$\frac{\nabla_i T(\mathbf{y}(\mathbf{w}))}{1 - \nabla_i T(\mathbf{y}(\mathbf{w}))} = \lambda_i(\mathbf{w}). \quad (19)$$

Thus, $\lambda_i(\mathbf{w})$ is a monotone transformation of the marginal taxes on earnings of spouse i in the couple of type \mathbf{w} . Moreover, this equation shows that dependence of the marginal tax rate of spouse i on earnings of spouse $-i$ is captured by the dependence of λ_i on w_{-i} . In particular, equation (19) implies that $\text{sign}(\nabla_{12} T(\mathbf{y}(\mathbf{w}))) = \text{sign}(\nabla_{-i} \lambda_i(\mathbf{w}))$. We say that taxes are positively (negatively) jointed at \mathbf{w} if this sign is positive (negative). We also refer to separable taxes, for which this sign is also zero, as disjointed.

We use v^* to denote the solution to the relaxed problem (13), and $\boldsymbol{\lambda}^*$ is the transformation of v^* . We refer to the $\boldsymbol{\lambda}^*$ as to the *optimal distortions*. The optimal marginal tax rates can be backed out from $\boldsymbol{\lambda}^*$ using (19).

Standard variational arguments can be used to show that if a solution v^* to the relaxed problem (13) is sufficiently smooth, then it must satisfy

$$\sum_{i=1}^2 \nabla_i \left(\lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) \right) = (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) \text{ on } W^o, \quad (20)$$

with a corresponding boundary condition

$$\sum_{i=1}^2 \lambda_i^*(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) = 0 \text{ on } \partial W, \quad (21)$$

where $\mathbf{n}(\mathbf{w})$ is the outward unit normal to ∂W at \mathbf{w} . Moreover, cross-partials of v^* must agree, in the sense that $\nabla_1 \nabla_2 v^* = \nabla_2 \nabla_1 v^*$, which, using $\boldsymbol{\lambda}^*$ transformation, can be written as

$$\nabla_2 \left((1 + \lambda_1^*(\mathbf{w}))^{1/(\gamma_1-1)} w_1^{\gamma_1/(\gamma_1-1)} \right) = \nabla_1 \left((1 + \lambda_2^*(\mathbf{w}))^{1/(\gamma_2-1)} w_2^{\gamma_2/(\gamma_2-1)} \right) \text{ on } W^o. \quad (22)$$

These conditions are not only necessary but also sufficient, in the sense that if there exists any smooth function $v^* \in \mathcal{V}$ that satisfies (20), (21), and (22), then there cannot be any other $v \in \mathcal{V}$, smooth or not, that gives higher welfare than $\Upsilon(v^*)$.

The optimality conditions show that the solution to the relaxed can be equivalently characterized either as a second-order non-linear partial differential equation in v^* or a system of first-order non-linear partial differential equations in $\boldsymbol{\lambda}^*$ subject to Neumann boundary condition. Equations (20) and (21) are generalizations of optimality and boundary conditions of uni-dimensional problems; equation (22) does not have a uni-dimensional analog.

Various versions of these conditions appeared in previous literature (e.g., Mirrlees (1986), Kleven et al. (2007), Renes and Zoutman (2017), Spiritus et al. (2022)). Unfortunately, solving such equations explicitly is hard, except in a few special cases, since mathematical literature about such equations is not well-developed.¹⁸ Our approach is to sidestep the difficult task of characterizing $\boldsymbol{\lambda}^*$ analytically at every point \mathbf{w} . Instead, we exploit the fact that equation (20) is a relatively tractable linear differential equation in $\boldsymbol{\lambda}^*$. As such, conditions (20) can be integrated over various subsets of W . The Divergence Theorem and the Coarea Formula allow then us to find various conditional averages of distortions $\boldsymbol{\lambda}^*$ on the boundaries of these

¹⁸Renes and Zoutman (2017) show that if $\boldsymbol{\lambda}^*$ is a conservative vector field (i.e., it is a gradient of some function so that $\nabla_2 \lambda_1^* = \nabla_1 \lambda_2^*$) then it can be characterized in some cases using so-called Green functions. Unfortunately, there is no reason in general for $\boldsymbol{\lambda}^*$ to be such a field; instead, the relevant auxiliary condition is (22) which equivalent to requiring that ∇v is a conservative vector field. This equation is highly non-linear which makes it difficult to find $\boldsymbol{\lambda}^*$ or to use Green functions.

subsets. As we shall see, these conditional averages contain much information about forces that determine the optimal taxation of couples; provide insights about how taxes on married individuals depend on earnings of their spouses and how those taxes compare to optimal taxes on single persons; describe conditions under which separable taxation of each spouse’s income or taxation of the joint family income are optimal; characterize dependence of taxes on the correlation structure in productivities of spouses and on social weight. We collaborate on these insights in Section 6, in which we numerically solve for the optimal taxes in a calibrated economy.

4 Optimal taxes under the first order approach

In this section, we study implications of equations (20), (21), (22) for optimal taxation of couples. To streamline our exposition, throughout this section we assume that the first order approach is valid, $W = \mathbb{R}_+^2$, and the optimal distortions are continuous and bounded on W with derivatives that are continuous and bounded on W° .

The uni-dimensional version of our model provides a natural point of comparison, and we first review existing results about optimal taxation in such a economy. We use $\lambda^{sng,*}$ to denote the optimal distortions in the uni-dimensional model $(G, \tilde{\alpha}, \gamma)$.

Proposition 3. *$\lambda^{sng,*}$ is non-negative and satisfies*

$$\lambda^{sng,*}(t) = \frac{\int_0^t (\tilde{\alpha}(w) - 1) g(w) dw}{\gamma t g(t)} = \frac{\int_t^\infty (1 - \tilde{\alpha}(w)) g(w) dw}{\gamma t g(t)}. \quad (23)$$

Proof. In the uni-dimensional case, equation (20) is an ordinary differential equation. Integrate it on $[0, t]$ using the boundary condition $\lambda^{sng,*}(0) = 0$ to obtain the first equality in (23). The rest follows from the fact that $\tilde{\alpha}$ is decreasing and integrates to one. \square

Equation (23) is a version of Diamond’s “ABC” formula. To understand its intuition, consider the following thought experiment. Suppose individuals face some, not necessarily optimal, tax schedule $\tilde{T}(y)$ and let $y(w)$ be earnings of type w under that schedule. Let’s perturb T^{sng} by increasing the marginal tax rate for all types on some interval $[t, t + dt]$ by $d\tau$. We assume that $d\tau$ and dt are small and allow $d\tau$ to take any sign. Equivalently, this perturbation can be described as increasing tax rates on earnings in the $[y(t), y(t) + \nabla y(t) dt]$ interval by $d\tau$. Panels (a) and (b) of Figure 1 show it graphically in both w - and y -spaces.

This perturbation has two effects: (i) it reallocates resources from agents with productivities $w > t$ to agents $w < t$, and (ii) it reduces earnings of the agents around $w = t$ threshold. We refer to these effects as benefits of redistribution and costs of distortions. To

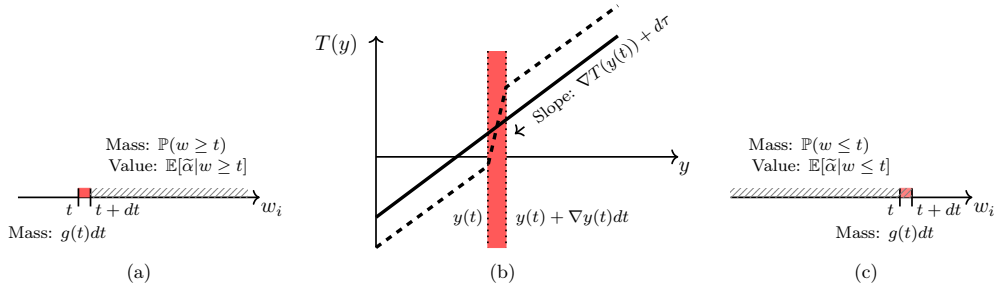


Figure 1: Perturbations describing intuition for equation (26).

compute the benefits of redistribution, observe that we increased the tax level of all individuals $w > t$ by $d\tau \nabla y(t) dt$ without changing their tax rates. In the absence of income effects, earnings of such individuals remain unchanged and, therefore, the government collects additional $d\tau \nabla y(t) dt \int_t^\infty g(w) dw$ in tax revenues. These revenues are returned in an equal amount to all households. The social value of \$1 in the hands of an average household is normalized 1 and the social value of \$1 in the hands of type w is $\alpha(w)$. Therefore, the net benefit of redistribution is

$$\mathfrak{B} = d\tau \nabla y(t) dt \left[\int_t^\infty g(w) dw - \int_t^\infty \tilde{\alpha}(w) g(w) dw \right].$$

To calculate the cost of distortion, first observe that each type in the $[t, t + dt]$ interval reduces their earnings¹⁹ by $\delta d\tau$, where

$$\delta = \frac{\gamma t \nabla y(t)}{1 - \nabla \tilde{T}(y(t))}. \quad (24)$$

This decreases the amount of taxes that each individual pays by $\nabla \tilde{T}(y(t)) \delta d\tau$. There are $g(t) dt$ of such individuals, so the total reduction in tax revenues collected from them is

$$\mathfrak{C} = d\tau \nabla y(t) dt \gamma t g(t) \frac{\nabla \tilde{T}(y(t))}{1 - \nabla \tilde{T}(y(t))}.$$

The net welfare effect of this perturbation is given by the difference between costs and benefits,

$$\mathfrak{B} - \mathfrak{C} = d\tau \nabla y(t) t \left[\int_t^\infty (1 - \tilde{\alpha}(w)) g(w) dw - \gamma t g(t) \frac{\nabla \tilde{T}(y(t))}{1 - \nabla \tilde{T}(y(t))} \right]. \quad (25)$$

If the tax system is optimal then there can be no perturbation that improves welfare and, therefore, the expression in square brackets must be zero. This gives the first equality in equation (23).

¹⁹To show this result, differentiate the optimality condition $1 - \nabla \tilde{T}(y(w)) = (y(w))^{1/\gamma-1} / w^{1/\gamma}$.

We could have considered an alternative perturbation of \tilde{T} whereby we increase transfers to all agents $w < t$, phased them out linearly on the interval $[t, t + dt]$, and adjust lump-tax taxes to satisfy the government budget constraint. This perturbation, depicted in Figure 1(c) in the w -space, directly leads to the second equality in equation (23). It yields no additional insights in uni-dimensional settings as it is equivalent in the y -space to the perturbation shown in panel (b).

Let $\bar{\theta}(t) := -\frac{\partial \ln(1-G)}{\partial \ln t}$ and $\theta(t) := \frac{\partial \ln G}{\partial \ln t}$ and observe that (23) can be written as

$$\lambda^{sg,*}(t) = \frac{1 - \mathbb{E}[\tilde{\alpha}|w \geq t]}{\gamma \bar{\theta}(t)} = \frac{\mathbb{E}[\tilde{\alpha}|w \leq t] - 1}{\gamma \theta(t)}. \quad (26)$$

Thus, optimal distortions in the uni-dimensional case are given by the ratio of benefits of redistribution, captured by $1 - \mathbb{E}[\tilde{\alpha}|w \geq t]$ (or, equivalently, $\mathbb{E}[\tilde{\alpha}|w \leq t] - 1$) to the costs of distortions, captured by the product two elasticities: the labor supply elasticity parameter γ and the elasticity of marginal distribution of productivities $\bar{\theta}(t)$ (or, equivalently, $\theta(t)$).

The key difficulty in transitioning from uni- to bi-dimensional analysis lies in the fact that the optimality conditions (20), (21), and (22) no longer yield closed-form expressions for λ^* , except in some special cases discussed below. We sidestep this difficulty by focusing on equations (20) and (21). Systematically working with them, one can obtain a variety of insightful conditional averages that describe λ^* . The unified treatment of all these conditional averages can be obtained by using the Coarea Formula. This formula states that for any function $Q : \mathbb{R}_{++}^2 \rightarrow \mathbb{R}_{++}$ that satisfies mild technical restriction (see appendix for the details and proofs) equations (20) and (21) imply

$$\mathbb{E} \left[\sum_{i=1}^2 \lambda_i^* \gamma_i \frac{\partial \ln Q}{\partial \ln w_i} \Big| Q = t \right] = \frac{1 - \mathbb{E}[\alpha|Q \geq t]}{-\partial \ln \mathbb{P}(Q \geq t) / \partial \ln t} = \frac{\mathbb{E}[\alpha|Q \leq t] - 1}{\partial \ln \mathbb{P}(Q \leq t) / \partial \ln t}. \quad (27)$$

Formula (27) shows that optimal distortions in the bi-dimensional case can be represented in the very general sense by the same trade-off between the benefits from redistribution and the costs of distortions. The benefits of redistribution are captured by $1 - \mathbb{E}[\alpha|Q \geq t]$ or $\mathbb{E}[\alpha|Q \leq t] - 1$. The costs of distortions are captured by a pair of labor elasticity parameters γ and various elasticities of the joint distribution F , summarized by $\partial \ln \mathbb{P}(Q \leq t) / \partial \ln t$ or $-\partial \ln \mathbb{P}(Q \geq t) / \partial \ln t$. Equation (27) should hold for *any* function Q , and by considering various such functions, one can obtain a rich characterization of optimal distortions.

The intuition for (27) can be obtained by studying perturbations of the joint tax function T whereby we increase tax levels by $\$ \varepsilon$ for all couples \mathbf{w} that satisfy $Q(\mathbf{w}) > q$ (or, alterna-

tively, $Q(\mathbf{w}) < q$) and adjust the lump sum tax component to satisfy the government budget constraint. We illustrate examples of such perturbations later in this section.

4.1 Optimality of individual earnings-based taxation

Before considering the general implications of (27), it will be useful to discuss one special case when optimal taxes can be found explicitly: the benchmark economy with independent types. This is the same economy that we considered in Proposition 1. It is easy to verify that $\tilde{\lambda}$ given in that proposition satisfies equations (20), (21), and (22) and, therefore, it is the optimal distortion. Examining it yields the following result.

Proposition 4. *In the benchmark economy with independent types, λ_i^* is independent of w_{-i} and satisfies*

$$\lambda_i^*(t, w_{-i}) = \frac{1}{2} \lambda^{sg,*}(t),$$

where $\lambda^{sg,*}$ is the optimal distortion in the corresponding uni-dimensional economy. Optimal taxes are separable, i.e., individual earnings-based.

This proposition shows two insights useful to understand intuition for our discussion below. First, optimal distortions on married individuals are precisely one-half of those on single individuals. The intuition for this result can be understood from our discussion of the costs and benefits of taxation. The costs of tax distortion for any individual with productivity t depends only on the characteristics of that individual: her labor supply elasticity, productivity, and the total mass of individuals with such productivity; it does not depend on how individuals match. In contrast, the benefits of redistribution depend crucially on matching patterns because married individuals share earnings with their spouses. When matching is random, every person t marries, in expectation, an average person. This reduces the benefits of redistribution exactly in half, leading to the result in Proposition 4.

The second insight of Proposition 4 is that optimal taxes in the benchmark economy with independent types are disjointed. Before discussing the intuition for this result, it will be useful to extend this insight first to non-benchmark economies. To state this result, we let $\bar{\theta}_i$ and θ_i denote the elasticities of marginals G_i defined by analogy with $\bar{\theta}$ and θ in equation (26).

Corollary 1. *In any economy with independent types and separable Pareto weights, λ_i^* is independent of w_{-i} and satisfies*

$$\lambda_i^*(t, w_{-i}) = \beta_i \frac{1 - \mathbb{E}[\tilde{\alpha}_i | w_i \geq t]}{\gamma_i \bar{\theta}_i(t)}.$$

The optimal taxes are separable, i.e., individual earnings-based.

This corollary shows that individual earnings-based taxation is optimal if Pareto weights are separable and types are independent. Separable social weights treat each individual the same irrespective of their partner’s identity, embedding implicit social preferences for individual earnings-based taxation. Despite this preference, individual earnings-based taxation is generally not optimal unless matching is random. If there is some assortativeness, the planner would like to sacrifice pure individual earnings-based taxation in favor of some jointness in the tax system, as this jointness can facilitate redistribution across individuals. This will be the focus of our analysis in Section 4.3.

4.2 Average distortions

In this section, we study the optimal average distortions that we define as $\mathbb{E} \left[\lambda_i^* \middle| w_i = t \right] = \int_0^\infty \lambda_i^*(t, w_{-i}) F(dw_{-i}|t)$. Economically, this is the mean optimal level of distortion of spouse i with productivity t . Graphically, these are the distortions of individuals along the line $w_i = t$ in the w -space, as shown in Figure 2. When W is compact, the expressions for the optimal distortions can be easily obtained by integrating (20) over a half-plane $\{\mathbf{w} : w_i \leq t\}$ or $\{\mathbf{w} : w_i \geq t\}$, solving for the value of the integral on the boundary using the Divergence Theorem, and simplifying using the boundary condition (21).²⁰ The Divergence Theorem does not directly apply when W is unbounded, but we can still obtain expressions for the average distortions by setting $Q(\mathbf{w}) = w_i$ in formula (27):

$$\mathbb{E} \left[\lambda_i^* \middle| w_i = t \right] = \frac{1 - \mathbb{E}[\alpha | w_i \geq t]}{\gamma_i \bar{\theta}_i(t)} = \frac{\mathbb{E}[\alpha | w_i \leq t] - 1}{\gamma_i \theta_i(t)}. \quad (28)$$

Equation (28) is remarkably similar to the formula (26) in the uni-dimension case. The intuition for it can be understood either by considering either a perturbation that raises $\$ \varepsilon$ of tax revenues from couples in the half-plane $\{\mathbf{w} : w_i > t\}$ or by a perturbation that gives $\$ \varepsilon$ of transfers to couples in the half-plane $\{\mathbf{w} : w_i < t\}$, with appropriate adjustment of the rest of the tax system,²¹ see Figure 2. The benefits from redistribution of the first perturbation is $1 - \mathbb{E}[\alpha | w_i \geq t]$. This is the social value of $\$1$ in the hands of an average couple minus the value of $\$1$ in the hands of couples in the half-plane $\{\mathbf{w} : w_i > t\}$. The cost of distortion per dollar

²⁰Kleven et al. (2007) used this approach to derive similar expression in their model.

²¹In the appendix, we formally describe this perturbation and use this perturbation to provide an alternative derivation of (28). The crucial step is to perturb joint tax T in such a way that we change marginal tax rates only on spouses in the $[t_i, t_i + dt]$ interval in the w -space. This interval in the w -space maps into a non-trivial band in the y -space that we characterize in the appendix. This is needed in order to express optimal tax distortions in terms of the exogenous parameters of the model. This contrasts with perturbations considered by Golosov et al. (2014) or Spiritus et al. (2022), which can be described as perturbations of taxes on some $[y_i, y_i + dy]$ interval in the y -space. Effects of such perturbations are easier to derive but lead to formulas that express optimal distortions in terms of endogenous conditions, such as moments of earnings distribution that depend on the optimal tax code. This makes them less suitable for conducting the comparative statics analysis.

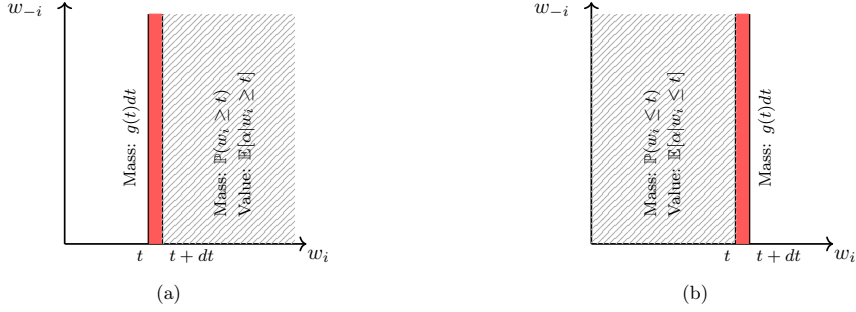


Figure 2: Perturbations describing intuition for equation (28)

of revenue is $\gamma_i \bar{\theta}_i(t) = \gamma_i t g_i(t) / \mathbb{P}(w_i \geq t)$. This cost depends on the labor supply elasticity, the productivity of spouse i , and the ratio of the density of spouses i with productivity t to the total mass of couples whose spouse i has productivity $w_i > t$.

We first use equation (28) to study how optimal average distortions depend on the assortativeness of matching. We start with the benchmark economy, where analysis is particularly transparent, before extending our results to the general economy. In the benchmark economy, Pareto weights are separable, and we can explicitly unpack integral $1 - \mathbb{E}[\alpha | w_i \geq t]$ as

$$1 - \mathbb{E}[\alpha | w_i \geq t] = \frac{1}{2} \int_t^\infty (1 - \tilde{\alpha}(w_i)) \frac{dG(w_i)}{1 - G(t)} + \frac{1}{2} \int_0^\infty (1 - \tilde{\alpha}(w_{-i})) dF(w_{-i} | w_i \geq t), \quad (29)$$

where $F(w_{-i} | w_i \geq t) := \int_t^\infty F(w_{-i} | w_i) \frac{dG(w_i)}{1 - G(t)}$. This equation decomposes the redistributive benefit from a marginal tax on individual w_i into two components: the benefit of extracting tax revenues from individuals $w_i > t$ (the first term on the right-hand side of (29)) and the benefit of extracting revenues from their spouse (the second term on the right-hand side of (29)). This second term is zero when matching as random, as it was in the economy considered in Proposition 4. More generally, this term suggests that the redistributive benefits and optimal distortions should be higher in economies with more assortative matching between spouses. The following proposition formalizes this intuition.

Proposition 5. *Consider benchmark economies (F^b, α, γ) and (F^a, α, γ) with $F^a \leq_{PQD} F^b$. Then,*

$$\mathbb{E}^a[\lambda_i^{a,*} | w_i = t] \leq \mathbb{E}^b[\lambda_i^{b,*} | w_i = t] \leq \lambda^{snq,*}(t) \text{ for all } t,$$

where $\lambda^{snq,*}$ are optimal distortions in the corresponding uni-dimensional economy.

Proposition 5 shows that optimal average distortions are ranked by dependence, so stronger dependence implies higher average distortions. The average distortions are lowest and highest in economies with perfect negative and positive assortative matching; in the latter case,

they also coincide with distortions in the corresponding uni-dimensional economy. As long as matching is not perfectly positive assortative, optimal distortions for a married person are, on average, lower than for a single person of the same productivity. Some redistribution occurs within couples even in the absence of taxation, and it is costly to crowd it out via distortionary taxation.

Proposition 5 extends directly to any economy with separable Pareto weights:

Corollary 2. *Consider economies (F^b, α, γ) and (F^a, α, γ) with $F^a \leq_{PQD} F^b$ and separable α . Then,*

$$\mathbb{E}^a[\lambda_i^{a,*} | w_i = t] \leq \mathbb{E}^b[\lambda_i^{b,*} | w_i = t] \text{ for all } t.$$

We now focus on understanding the role of non-separability of social weights. We focus on implications of super- and sub-modularity, which restricts the cross-partial derivatives of α to have the same sign. Mathematically, α is (strictly) supermodular if the difference $\alpha(w'', \cdot) - \alpha(w', \cdot)$ is (strictly) decreasing for all w'', w' with $w'' \geq w'$; submodularity is defined analogously through increasing difference. Economically, supermodularity means that the social planner puts similarly low weights on couples with both very productive spouses; submodularity means that the social planner puts similarly high weights on couples with both very unproductive spouses. Thus, supermodularity favors raising revenues from couples in the upper right orthants; submodularity implicitly favors targeting transfers to couples in the lower left orthants.

We now characterize implications of dependence for super- and sub-modular planners. Note that F and α are related through the normalization $\mathbb{E}[\alpha] = 1$. Therefore, in order to compare two distributions with different degree of dependence we need to re-normalize Pareto weights to ensure that they integrate to one. We say that $\alpha^b \sim \alpha^a$ if α^a, α^b are the same up to the normalization constant, that is if the ratio $\alpha^b(\mathbf{w})/\alpha^a(\mathbf{w})$ is independent of \mathbf{w} . Note that this re-normalization is not needed when α is separable, since in that case $\mathbb{E}^a[\alpha] = \mathbb{E}^b[\alpha]$ for any F^a, F^b with the same marginals.

Corollary 3. *Consider economies (F^b, α^b, γ) and (F^a, α^a, γ) with $F^a \leq_{PQD} F^b$ and $\alpha^b \sim \alpha^a$. Then,*

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}^b[\lambda_i^{b,*} | w_i = t]}{\mathbb{E}^a[\lambda_i^{a,*} | w_i = t]} \geq 1 \text{ if } \alpha^a \text{ is supermodular,}$$

$$\lim_{t \rightarrow 0} \frac{\mathbb{E}^b[\lambda_i^{b,*} | w_i = t]}{\mathbb{E}^a[\lambda_i^{a,*} | w_i = t]} \geq 1 \text{ if } \alpha^a \text{ is submodular.}$$

Moreover, both inequalities are strict if $F^a <_{PQD} F^b$ and super- and sub-modularities are strict.

When social weights are non-separable, the results of Proposition 5 apply in the tails. A supermodular planner has a strong desire to extract resources from couples in orthants $\{\mathbf{w} : \mathbf{w} > \mathbf{t}\}$ with large \mathbf{t} . When productivities are more dependent, there are more couples in those orthants and hence larger redistributive benefits of raising taxes from spouses with high t_i . On the other hand, a submodular planner has a strong desire to transfer resources to couples in orthants $\{\mathbf{w} : \mathbf{w} < \mathbf{t}\}$ with small \mathbf{t} . Stronger dependence means that there are more couples in those orthants and hence larger benefits of transfers to spouses with low t_i . Higher transfers, in turn, lead to higher optimal distortions when those transfers are phased-out.

We finish this section by considering the effect of redistributiveness of Pareto weights on optimal average distortions. We say that weights α^b are *more redistributory* than α^a if $\alpha^a(\mathbf{w}')\alpha^b(\mathbf{w}'') \geq \alpha^b(\mathbf{w}' \vee \mathbf{w}'')\alpha^a(\mathbf{w}' \wedge \mathbf{w}'')$ for all $\mathbf{w}', \mathbf{w}''$. A reader might recognize that this definition is the multivariate likelihood order.²² It is a multi-dimensional generalization of a more familiar uni-dimensional version that states that $\tilde{\alpha}^b$ dominates in likelihood order $\tilde{\alpha}^a$ if the ratio $\tilde{\alpha}^b(w)/\tilde{\alpha}^a(w)$ is decreasing in w .

Proposition 6. *Consider economies (F, α^b, γ) and (F, α^a, γ) and suppose that f is log-supermodular. If either α^b are more redistributory than α^a , or the economies are benchmark and $\tilde{\alpha}^b$ are more redistributory than $\tilde{\alpha}^a$, then*

$$\mathbb{E}[\lambda_i^{a,*}|w_i = t] \leq \mathbb{E}[\lambda_i^{b,*}|w_i = t] \text{ for all } t.$$

This proposition shows that more redistributive planners generally choose higher optimal average distortions. Log-supermodularity of f is satisfied by many common distributions, e.g., by the bivariate log-normal distribution with a non-negative correlation parameter (see Karlin and Rinott (1980) for other examples and proofs).

4.3 Average jointness

In this section, we discuss jointness of the optimal tax schedule. We focus on the following measure of jointness,

$$J_i(\mathbf{t}) = \frac{\mathbb{E}[\lambda_i^*|w_i = t_i, w_{-i} \geq t_{-i}]}{\mathbb{E}[\lambda_i^*|w_i = t_i, w_{-i} \leq t_{-i}]} - 1, \quad (30)$$

which we refer to as *average jointness*. This measure is positive (negative) if distortions of spouse i increase (decrease) in the productivity of their partner. If separable taxes are optimal, then $J_i \equiv 0$ for both spouses. A useful device to analyze jointness are functions H_i and \bar{H}_i defined as

$$H_i(\mathbf{t}) := \frac{\mathbb{E}[\lambda_i^*|w_i = t_i, w_{-i} \leq t_{-i}]}{\mathbb{E}[\lambda_i^*|w_i = t_i]}, \quad \bar{H}_i(\mathbf{t}) := \frac{\mathbb{E}[\lambda_i^*|w_i = t_i, w_{-i} \geq t_{-i}]}{\mathbb{E}[\lambda_i^*|w_i = t_i]}.$$

²²See Chapter 6.E of Shaked and Shanthikumar (2007) for a discussion of this order and its properties.

These functions compare conditional average distortions (e.g., the average distortion of spouse i with productivity t_i conditional on their partner's productivity being either less or more than t_{-i}) and unconditional average distortions. Clearly, $J_i = \bar{H}_i/H_i - 1$ and $\text{sign}(J_i(\mathbf{t})) = \text{sign}(\bar{H}_i(\mathbf{t}) - 1) = \text{sign}(1 - H_i(\mathbf{t}))$.

As the first step to our characterization, use $Q(w_1, w_2) = \max\{k_1 w_1, k_2 w_2\}$ for some $\mathbf{k} \in \mathbb{R}_{++}^2$ in formula (27). By setting $t_i = t/k_i$, it implies that

$$\sum_i \gamma_i \bar{\theta}_i(t_i) \bar{\eta}_i(\mathbf{t}) \mathbb{E}[\lambda_i^* | w_i = t_i, w_{-i} \geq t_{-i}] = 1 - \mathbb{E}[\alpha | \mathbf{w} \geq \mathbf{t}], \quad (31)$$

$$\sum_i \gamma_i \theta_i(t_i) \eta_i(\mathbf{t}) \mathbb{E}[\lambda_i^* | w_i = t_i, w_{-i} \leq t_{-i}] = \mathbb{E}[\alpha | \mathbf{w} \leq \mathbf{t}] - 1, \quad (32)$$

where $\eta_i(\mathbf{t}) := \frac{\mathbb{P}(w_{-i} \geq t_{-i} | w_i = t_i)}{\mathbb{P}(w_{-i} \geq t_{-i} | w_i \geq t_i)}$ and $\bar{\eta}_i(\mathbf{t}) := \frac{\mathbb{P}(w_{-i} \leq t_{-i} | w_i = t_i)}{\mathbb{P}(w_{-i} \leq t_{-i} | w_i \leq t_i)}$.

Two statistics $\boldsymbol{\eta}$ and $\bar{\boldsymbol{\eta}}$ play an important role in determining optimal jointness. We refer to these statistics as *copula* and *survival copula elasticities* since they can be represented as

$$\begin{aligned} \eta_i(\mathbf{t}) &= \frac{\partial \ln C(\mathbf{u})}{\partial \ln u_i} = 1 + \frac{\partial \ln [C(\mathbf{u})/u_i]}{\partial \ln u_i}, \\ \bar{\eta}_i(\mathbf{t}) &= \frac{\partial \ln \bar{C}(\mathbf{u})}{\partial \ln u_i} = 1 + \frac{\partial \ln [\bar{C}(\mathbf{u})/u_i]}{\partial \ln u_i}, \end{aligned} \quad (33)$$

where derivatives of copulas are evaluated at quantiles of \mathbf{t} , i.e., $\{u_i = G(t_i)\}$ and $\{u_i = 1 - G(t_i)\}$, respectively. The copula representation is helpful as it highlights the economic intuition behind these elasticities: η_i is the percentage change of the conditional quantile of the spouse $-i$'s productivity, C/u_i , when the quantile of spouse i 's productivity, u_i , increases by 1%; similarly, $\bar{\eta}_i$ is the percentage change of the conditional survival quantile, \bar{C}/u_i . These elasticities are equal to one when types are independent and are decreasing in the strength of orthant dependence.²³

4.3.1 Average jointness in the benchmark economy

Before analyzing the general case, it will be helpful to consider optimal jointness in the benchmark economy along the diagonal $t_1 = t_2 = t$. This case illustrates forces determining optimal

²³Recall (see, e.g., Chapter 9.B in Shaked and Shanthikumar (2007)) that F^b dominates F^a in the *lower orthant decreasing ratio order* (LORD), $F^a \leq_{LORD} F^b$, if F^a and F^b have the same marginals and the ratio of corresponding copulas $C^b(\mathbf{u})/C^a(\mathbf{u})$ is decreasing in \mathbf{u} . Similarly, F^b dominates F^a in the *upper orthant decreasing ratio order* (UORD), $F^a \leq_{UORD} F^b$, if the same relationship holds for their survival copulas. LORD or UORD order implies PQD order, so these notions of dependence are somewhat stronger than PQD. It is easy to verify the following relationship between orthant dependence and copula elasticities for all \mathbf{w} :

$$\begin{aligned} F^a \leq_{LORD} F^b &\implies \boldsymbol{\eta}^a(\mathbf{w}) \geq \boldsymbol{\eta}^b(\mathbf{w}), \\ F^a \leq_{UORD} F^b &\implies \bar{\boldsymbol{\eta}}^a(\mathbf{w}) \geq \bar{\boldsymbol{\eta}}^b(\mathbf{w}). \end{aligned}$$

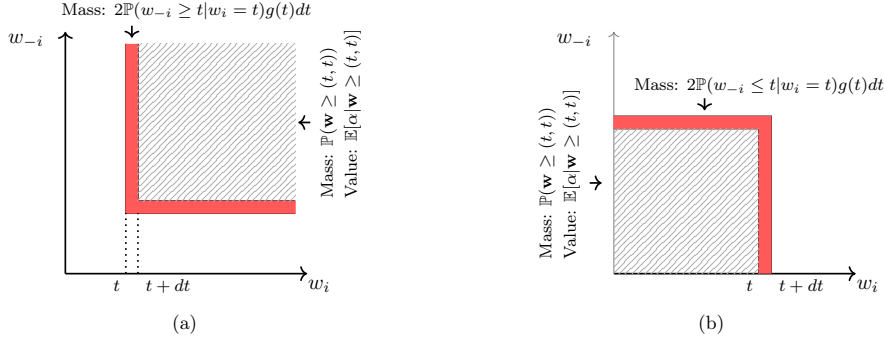


Figure 3: Perturbations describing effects of introducing jointness in upper right (panel (a)) and lower left (panel (b)) orthants.

jointness in a transparent way. In the benchmark economy, we can explicitly solve for $\bar{H}(t, t)$. Indeed, invert (31) to obtain

$$\mathbb{E}[\lambda_i^* | w_i = t, w_{-i} \geq t] = \frac{1 - \mathbb{E}[\alpha | \mathbf{w} \geq (t, t)]}{2\bar{\eta}(t, t)\gamma\bar{\theta}(t)}. \quad (34)$$

Then, combine equations (34) and (28) to get the following expression for $\bar{H}(t, t)$:

$$\bar{H}(t, t) = \frac{1}{2\bar{\eta}(t, t)} \frac{1 - \mathbb{E}[\alpha | \mathbf{w} \geq (t, t)]}{1 - \mathbb{E}[\alpha | w_i \geq t_i]}. \quad (35)$$

Equation (35) highlights the main trade-off determining optimal jointness. To develop the intuition for this expression, consider the following thought experiment. Take a disjointed tax schedule $T(y_1, y_2) = \tilde{T}(y_1) + \tilde{T}(y_2)$ and slightly perturb it introducing jointness. Specifically, increase the level of taxes by some amount ε for all couples whose productivities satisfy $\min\{w_1, w_2\} > t$ and adjust the lump sum component of the tax schedule to satisfy the government budget constraint. Such perturbation increases marginal taxes for all individuals on the boundary of orthant $\{\mathbf{w} : \mathbf{w} > (t, t)\}$ by $d\tau$ (see Figure 3(a)). Jointness is positive if $d\tau > 0$ and negative if $d\tau < 0$.

This perturbation has redistributory benefits and distortionary costs. The redistributory benefits are calculated by the analogy with the example at the beginning of Section 4 and can be written as

$$\mathfrak{B} = \nabla y(t) dt d\tau \int_t^\infty \int_t^\infty (1 - \alpha(\mathbf{w})) f(\mathbf{w}) d\mathbf{w}.$$

The distortionary costs arise because of behavioral responses. To calculate these costs, observe that, since we started with a disjointed tax schedule, the reduction in earnings of every distorted individual equals to $\delta d\tau$, where δ is the same as in the uni-dimensional case (see equation (24)). There are $2g(t)(1 - F(t|t))dt$ of distorted individuals. Therefore, the total cost of distortions

is

$$\mathfrak{C} = \nabla y(t) dt d\tau \gamma t 2g(t)(1 - F(t|t)) \frac{\nabla \tilde{T}(y(t))}{1 - \nabla \tilde{T}(y(t))}.$$

The net welfare effect of introducing jointness is the difference between benefits and costs:

$$\begin{aligned} \mathfrak{B} - \mathfrak{C} &= \nabla y(t) dt d\tau \mathbb{P}(\mathbf{w} \geq (t, t)) \\ &\times \left[(1 - \mathbb{E}[\alpha | \mathbf{w} \geq (t, t)]) - \underbrace{\frac{2\mathbb{P}(w_i \geq t)\mathbb{P}(w_{-i} \geq t | w_i = t)}{\mathbb{P}(w_{-i} \geq t, w_i \geq t)}}_{=2\bar{\eta}(t,t)} \underbrace{\frac{tg(t)}{1 - G(t)}}_{=\bar{\theta}(t)} \gamma \frac{\nabla \tilde{T}(y(t))}{1 - \nabla \tilde{T}(y(t))} \right]. \end{aligned} \quad (36)$$

In this formula, $\bar{\theta}\gamma$ captures distortions from the separable tax schedule \tilde{T} and is familiar from Section 4.2. The new term, $2\bar{\eta}$, measures additional distortions introduced by jointness.²⁴ The numerator of $2\bar{\eta}$ captures the number of couples distorted by jointness, while the denominator is the number of couples paying higher average taxes due to jointness. Thus, $2\bar{\eta}$ reflects distortions from jointness in the similar way $g/(1 - G)$ captures distortions from taxing individual. As formula (36) shows, positive jointness can improve welfare by targeting taxes to the richest couples, negative jointness can improve welfare by reducing distortions associated with separable taxes \tilde{T} .

Formula (36) applies to an arbitrary separable tax schedule. To see the connection to equation (35), evaluate (36) at the optimal separable tax, \tilde{T}^* . While we study such taxes formally in Section 5, it is easy to use perturbational arguments to show that \tilde{T}^* must satisfy

$$\frac{\nabla \tilde{T}^*(y(t))}{1 - \nabla \tilde{T}^*(y(t))} = \frac{1 - \mathbb{E}[\alpha | w_i \geq t]}{\gamma \bar{\theta}(t)}. \quad (37)$$

Substitute equation (37) into (36) and re-arrange to obtain that $\mathfrak{B} - \mathfrak{C}$ is proportional to

$$\frac{1}{2\bar{\eta}(t, t)} \frac{1 - \mathbb{E}[\alpha | \mathbf{w} \geq (t, t)]}{1 - \mathbb{E}[\alpha | w_i \geq t_i]} - 1,$$

which is exactly the same as $\bar{H}(t, t) - 1$ due to equation (35).

This thought experiment highlights that $\bar{H}(t, t)$ in equation (35) is a product of two terms: the costs of additional distortions due to positive jointness, captured by $1/2\bar{\eta}(t, t)$, and additional benefits of redistribution that such jointness provides. When types are independent, these additional benefits and costs exactly offset each other, so that $\bar{H}(t, t) = 1$, and average jointness is zero.²⁵ Positive dependence increases the redistributory benefits of jointness

²⁴Elasticity $\bar{\eta}$ is multiplied by 2 because our perturbation introduced additional distortions on two individuals simultaneously, spouse 1 with productivity t and spouse 2 with productivity t .

²⁵In fact, it is zero everywhere as we showed in Proposition 4.

$1 - \mathbb{E}[\alpha | \mathbf{w} \geq (t, t)]$, but at the same time it decreases its costs $2\bar{\eta}(t, t)$. For a *fixed* tax schedule \tilde{T} , equation (36) shows that the net welfare effect of positive jointness is higher with stronger dependence. However, as equation (37) and our analysis in Proposition 5 showed, higher dependence also increases the *optimal separable* tax \tilde{T}^* and hence the distortions associated with it. Thus, whether positive dependence favors positive or negative jointness depends on which of these two forces dominates. The answer to this question turns out to be closely related to concepts of right-tail dependence and independence.

Recall that a joint distribution is right-tail independent (dependent) if $\lim_{u \rightarrow 0} \bar{C}(u, u)/u$ is equal to zero (strictly positive).²⁶ Intuitively, if the probability that partners of extremely productive spouses are also extremely productive goes to zero, then the distribution is right-tail independent; if this probability is bounded away from zero then it is right-tail dependent. To see the distinction, suppose that $\ln w_1$ and $\ln w_2$ are drawn from a standard bivariate normal distribution with correlation parameter $\rho > 0$. Positive correlation implies that the joint distribution of productivities is positively dependent, so that high value of w_i implies that w_{-i} is likely to be high as well. However, the joint distribution is right-tail independent for any ρ since $\lim_{w_i \rightarrow \infty} \mathbb{P}(w_{-i} \geq t | w_i \geq t) = 0$. Contrast this with the case of perfect assortative matching, which can be informally thought of as a limit $\rho = 1$ in this example. Perfectly dependent distributions are both positively dependent and right-tail dependent since for them $\mathbb{P}(w_{-i} \geq t | w_i \geq t) = 1$ for all w_i .

Statistics $\bar{\kappa} := \lim_{u \rightarrow 0} \ln u / \ln \bar{C}(u, u)$ is used to measure the speed of convergence to the right-tail independence. $\bar{\kappa} = 1$ for right-tail dependent distributions and $\bar{\kappa} \in [\frac{1}{2}, 1)$ for positively dependent but right-tail independent distributions, with higher values of $\bar{\kappa}$ corresponding to the slower speed of convergence.²⁷

These definitions allows us to characterize jointness in the right tail. Observe, using

²⁶See Nelsen (2006) or Shaked and Shanthikumar (2007) for overview of these concepts. To simplify our illustrations, we implicitly focus on the case of symmetric distributions in this paragraph, which are the only distributions that we consider in Section 4.3.1. This allows us to drop quantile representation and work with conditional probabilities of productivities directly. Analogous discussions, just translated into quantile spaces, apply to asymmetric distributions.

²⁷Ledford and Tawn (1996, 1998) showed that under weak conditions, the right-tail behavior of a symmetric distribution could be written as

$$\mathbb{P}(w_1 \geq t, w_2 \geq t) \sim L(t) \cdot [\mathbb{P}(w_i \geq t)]^{1/\bar{\kappa}} \text{ as } t \rightarrow \infty,$$

where L is some function such that $L(G^{-1}(e^{-1/t}))$ is “slowly varying”, which is a certain generalization of functions that converge as $t \rightarrow \infty$. As can be seen from the above expression, larger values of $\bar{\kappa} \in [1/2, 1)$ mean slower convergence to independence. Coles et al. (1999); Heffernan (2000); Hua and Joe (2014) discuss the theoretical properties and empirical estimation of these statistics.

L'Hospital's rule, that $\lim_{t \rightarrow \infty} 1/2\bar{\eta}(t, t) = \bar{\kappa}$ and, therefore, equation (35) implies²⁸

$$\lim_{t \rightarrow \infty} \bar{H}(t, t) = \frac{\bar{\kappa}(1 - \tilde{\alpha}(\infty))}{1 - \frac{1}{2}\tilde{\alpha}(\infty) - \frac{1}{2}\mathbb{E}[\tilde{\alpha}(w_{-i})|w_i = \infty]}. \quad (38)$$

Since $\mathbb{E}[\tilde{\alpha}(w_{-i})|w_i = \infty] \in [\tilde{\alpha}(\infty), 1]$ for positively dependent distributions, we immediately obtain the following result.

Lemma 4. *Suppose F is positively dependent and $\tilde{\alpha}$ is strictly decreasing.*

If F is right-tail dependent, then $\lim_{t \rightarrow \infty} \bar{H}(t, t) \geq 1$. This inequality is strict if $F(\cdot|\infty)$ is non-degenerate (as, e.g., in the case when F is given by a t -copula).

If F is right-tail independent and either $\bar{\kappa} = \frac{1}{2}$ or $F(\cdot|\infty)$ is degenerate, then $\lim_{t \rightarrow \infty} \bar{H}(t, t) \leq 1$. This inequality is strict if either $\bar{\kappa} < 1$ and $F(\cdot|\infty)$ is degenerate (as, e.g., in the case when F is given by a Gaussian copula) or if $\bar{\kappa} = \frac{1}{2}$ and F is strictly first-order dominated by G (as, e.g., in the case when F is given by an FGM copula).

This lemma shows that the sign of optimal jointness is essentially determined by tail-(in)dependence. For right-tail dependent distributions, very productive spouses are married to very productive partners with positive probability. In this case, the benefits of additional redistribution outweigh the cost of distortions and optimal average jointness is positive for very productive couples. For right-tail independent distributions, generally the opposite is the case.²⁹

Our discussion so far has focused on how jointness helps the planner to target taxes to the richest couples. We can alternatively consider how jointness helps to target transfers to the poorest couples (see panel (b) of Figure 3). As can be expected, our previous discussions apply directly with minor modifications. For example, we would consider $H(t, t)$ rather than $\bar{H}(t, t)$, left-tail (in)dependence captured by the value of $\lim_{u \rightarrow 0} C(u, u)/u$, and the speed of convergence $\kappa := \lim_{u \rightarrow 0} \ln u / \ln C(u, u)$. Despite many parallels, there is one critical difference. While targeting of taxes to the richest couples is facilitated by positive jointness, targeting of transfers to the poorest couples is facilitated by negative jointness. This implies that the pros and cons of jointness flip: the redistributive benefits of transfer-targeting favor negative jointness while the reduction in the costs of distortions from separable transfers favors positive jointness. We summarize the results for both tails in the following proposition.

²⁸We use $\tilde{\alpha}(\infty)$ and $\mathbb{E}[\alpha|w_i = \infty]$ to denote limits of $\tilde{\alpha}(t)$ and $\mathbb{E}[\alpha|w_i \geq t]$ as $t \rightarrow \infty$. Such limits exist since α (and $\tilde{\alpha}$) are continuous, bounded, and decreasing. Similarly, $F(\cdot|\infty)$ denotes the limit of $F(\cdot|t)$, in the weak convergence sense. We say that $F(\cdot|\infty)$ is *degenerate* if $F(t|\infty) = 0$ for all finite t . Analogous definitions and considerations apply also to the left limit, $t \rightarrow 0$.

²⁹Lemma 4 does not include cases of right-tail independent distributions for which $\bar{\kappa} \in (\frac{1}{2}, 1)$ and $F(\cdot|\infty)$ is non-degenerate. For such distributions, the average jointness may take either sign. While such distributions exist, they are scarce in the sense that most commonly used copulas satisfy one of the conditions listed in Proposition 7 (see Hua and Joe (2011, 2014) for details).

Proposition 7. *Consider a benchmark economy with $F \geq_{PQD} G \cdot G$.*

If F is right-tail dependent (left-tail dependent), then the redistributive benefit of jointness dominates, and optimal average jointness is positive at the top, $\lim_{t \rightarrow \infty} J(t, t) \geq 0$ (negative at the bottom, $\lim_{t \rightarrow 0} J(t, t) \leq 0$).

If F is right-tail independent (left-tail dependent) and either the speed of convergence is $\frac{1}{2}$ or the limiting distribution is degenerate, then the distortion costs of jointness dominates, and optimal average jointness is negative at the top, $\lim_{t \rightarrow \infty} J(t, t) \leq 0$ (positive at the bottom, $\lim_{t \rightarrow 0} J(t, t) \geq 0$).

While Proposition 7 describes jointness in the tails, in some cases we can characterize it more generally. The next proposition shows that if F is given by the FGM copula then there is a unique point on the diagonal at which average jointness switches its sign from positive to negative:

Corollary 4. *Consider a benchmark economy in which $\tilde{\alpha}$ is strictly decreasing and F is given by the FGM copula with $\rho > 0$. Then, there exists some $\hat{t} > 0$ such that $J(t, t) > 0$ for all $t < \hat{t}$ and $J(t, t) < 0$ for all $t > \hat{t}$.*

4.3.2 Average jointness in general

We now turn to our general economy and show that most of the insights from Section 4.3.1 continue to hold. We show this results in two steps: Corollary 5 considers implications for average jointness along the diagonal $\mathbf{t} = (t, t)$ for symmetric economies with super- and sub-modular Pareto weights, while Corollary 6 extends those results to the general case.

Corollary 5. *Consider a symmetric economy with $F \geq_{PQD} G \cdot G$.*

If α is supermodular, then all conclusions of Proposition 7 hold except that the sign of $\lim_{t \rightarrow 0} J(t, t)$ is ambiguous if F is left-tail independent and $\kappa = \frac{1}{2}$.

If α is submodular, then all conclusions of Proposition 7 hold except that the sign of $\lim_{t \rightarrow \infty} J(t, t)$ is ambiguous if F is right-tail independent and $\bar{\kappa} = \frac{1}{2}$.

The main takeaway from this corollary is that all but one result from Proposition 7 continue to apply to sub- and supermodular social weights. The intuition for this parallels that of Corollary 3: supermodularity strengthens the redistributive benefits from targeting taxes to couples in upper right quadrants but has an ambiguous effect on these benefits in bottom left quadrants; submodularity strengthens the redistributive benefits from targeting transfers to couples in lower left quadrants. Keeping these observations in mind, the analysis of different cases considered in Proposition 7 goes through with minimal changes.

Conclusions of Corollary 5 extend to the general economy for at least one spouse, provided that appropriate limits are well-behaved along any ray.

Corollary 6. *Consider a general economy and fix $\mathbf{s} \in \mathbb{R}_{++}^2$.*

(a). *If $(\overline{H}_i(\mathbf{t}^u), H_i(\mathbf{t}^u))_{i=1,2}$, where $\mathbf{t}^u = (G_i^{-1}(1 - s_i u))_{i=1,2}$, converges as $u \rightarrow 0$ and each $\nabla_i \overline{C}(\mathbf{u})$ is non-decreasing in u_i in a neighborhood of $\mathbf{0}$, then conclusions of Proposition 7 and Corollary 5 hold for at least one spouse.*

(b). *If $(\overline{H}_i(\mathbf{t}^u), H_i(\mathbf{t}^u))_{i=1,2}$, where $\mathbf{t}^u = (G_i^{-1}(s_i u))_{i=1,2}$, converges as $u \rightarrow 0$ and each $\nabla_i C(\mathbf{u})$ is non-decreasing in u_i in a neighborhood of $\mathbf{0}$, then conclusions of Proposition 7 and Corollary 5 hold for at least one spouse.*

The monotonicity conditions for $\nabla_i C$ and $\nabla_i \overline{C}$ are sufficient to ensure that the copula and the survival copula of F are well-behaved in the limits along rays. These conditions are often encountered in the extreme value theory (see, e.g., Hua and Joe (2011)) and hold for FGM, Gaussian, and t-copulas when $\rho > 0$.

Relationship to KKS

The results in Corollary 5 are related to the analysis of jointness in Section 3 of the working paper by Kleven et al. (2007). Those authors focus on independent distributions and welfare captured by social welfare functions (SWFs). They show that the optimal jointness takes the opposite sign from the third derivative of the SWF. In our setting, it can be re-stated in terms of modularity of Pareto weights (see the appendix for details): in the symmetric economy with independent types, $J_i(\mathbf{t}) \leq 0$ and $\nabla_{-i} \lambda_i^*(\mathbf{t}) \leq 0$ for all \mathbf{t} ($J(\mathbf{t}) \geq 0$ and $\nabla_{-i} \lambda_i^*(\mathbf{t}) \geq 0$ for all \mathbf{t}) if α is supermodular (submodular).

This result hinges critically on the assumption of independence and can break down even with small positive dependence. To see that, consider the Gaussian copula. First, suppose that $\rho = 0$, so that types are independent. In this case, the limiting distribution $F(\cdot|0)$ satisfies $F(\cdot|0) = G(\cdot)$ and, therefore, $\lim_{t \rightarrow 0} H(t, t) = \frac{\alpha(0,0)-1}{2(\mathbb{E}[\alpha(\cdot,0)]-1)}$. By independence and symmetry, we have

$$2(\mathbb{E}[\alpha(\cdot, 0)] - 1) - (\alpha(0, 0) - 1) = \mathbb{E}[\alpha(w_1, 0) + \alpha(0, w_2) - \alpha(0, 0) - \alpha(w_1, w_2)] \leq 0,$$

where the last inequality follows from the definition of supermodularity. As a result, $\lim_{t \rightarrow 0} J(t, t) \leq 0$ and average jointness at the bottom is negative when types are independent. Contrast that with the case when there is *any* positive dependence, $\rho > 0$. In this case, since the conditional distribution $F(\cdot|0)$ is degenerate, $\lim_{t \rightarrow 0} H(t, t) = \frac{1+\rho}{2} \frac{\alpha(0,0)-1}{\alpha(0,0)-1} < 1$. This implies that $\lim_{t \rightarrow 0} J(t, t) > 0$, so that strictly positive jointness at the bottom is optimal.

4.4 Distortions within and across families

In Sections 4.2 and 4.3, we focused on comparing distortions across married individuals. We now turn our attention to comparing distortions within couples. To make our analysis more transparent, we focus here on the case when $\gamma_1 = \gamma_2$, leaving the discussion of the general case to the appendix.

It will be convenient to change coordinates that describe the type-space W . Define new coordinates (r, ι) using a bijection

$$(r, \iota) \leftrightarrow \left(\left(w_1^{1/(1-\gamma)} + w_2^{1/(1-\gamma)} \right)^{1-\gamma}, \frac{w_2}{w_1} \right) := (R(\mathbf{w}), I(\mathbf{w})).$$

In these new coordinates, r is a measure of total family productivity, and ι is the relative productivity of the two spouses. We use $L(r, \iota)$ to denote the joint distribution of productivities in the (r, ι) -space, and $L_r(r)$ and $L_\iota(\iota)$ to denote the two marginals of L . We use $\bar{\theta}_r, \theta_r$ and $\bar{\theta}_\iota, \theta_\iota$ to denote elasticities of $L_r(r)$ and $L_\iota(\iota)$. We say that α is *measurable only w.r.t. r* if it takes the form $\alpha(\mathbf{w}) = \tilde{\alpha}(R(\mathbf{w}))$ for some function $\tilde{\alpha} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$.

In what follows, we use our general formula (27) to derive implications for average and relative distortions for couples. By setting $Q(\mathbf{w}) = R(\mathbf{w})$ and $Q(\mathbf{w}) = I(\mathbf{w})$ respectively, we obtain

$$\mathbb{E} \left[\sum_{i=1}^2 \frac{w_i^{1/(1-\gamma)}}{w_1^{1/(1-\gamma)} + w_2^{1/(1-\gamma)}} \lambda_i^* \middle| R = r \right] = \frac{1 - \mathbb{E}[\alpha | R \geq r]}{\gamma \bar{\theta}_r(r)} \quad (39)$$

and

$$\mathbb{E}[\lambda_2^* - \lambda_1^* | I = \iota] = \frac{1 - \mathbb{E}[\alpha | I \geq \iota]}{\gamma \bar{\theta}_\iota(\iota)}. \quad (40)$$

Formula (39) derives the expression for the average distortion of couples that have total productivity r , where the distortion of each spouse is weighted with a measure of their relative productivity, or earnings, $w_i^{1/(1-\gamma)} / (w_1^{1/(1-\gamma)} + w_2^{1/(1-\gamma)})$. The expectation is taken over all couples with the total productivity of r . Formula (40) describes how these average distortions are allocated optimally between the two spouses. Both formulas resemble the expression for the optimal average distortions of individuals, equation (28), but each time a different elasticity of the joint distribution of productivities summarizes the relevant measure of deadweight losses. The optimal average distortions of individuals are determined by the elasticity of marginal distribution of individual productivity, $\bar{\theta}$; whereas the optimal average distortions of couples are determined by the elasticity of the marginal distribution of total couple productivity, $\bar{\theta}_r$, and the optimal relative distortions of the spouses are determined by the elasticity of the marginal of their relative productivities, $\bar{\theta}_\iota$.

Equation (40) allows us to compare distortions between the two spouses and study conditions under which the primary and the secondary earner (i.e., the spouse with higher and lower productivity, respectively) face identical marginal taxes.

Proposition 8. *Consider an economy with $\gamma_1 = \gamma_2$ and in which $\lim_{t \rightarrow 0} \theta_\iota(t), \theta_r(t)$ and $\lim_{t \rightarrow \infty} \bar{\theta}_\iota(t), \bar{\theta}_r(t)$ are well-defined and finite.*

(a). *If $\alpha(w, 0) \geq 1$ for all w , then $\lim_{\iota \rightarrow 0} \mathbb{E}[\lambda_2^* - \lambda_1^* | I = \iota] \geq 0$;*

(b). *Suppose that α is measurable only w.r.t. r , that r and ι are independent, and that $\tilde{\lambda}(r) := \frac{\int_r^\infty (1 - \tilde{\alpha}(r)) l_r(s) ds}{\gamma r l_r(r)}$ is continuously differentiable with a bounded derivative. Then, the optimal tax system is family earnings-based, and it satisfies $\lambda_2^*(\mathbf{w}) = \lambda_1^*(\mathbf{w}) = \tilde{\lambda}(R(\mathbf{w}))$ for all \mathbf{w} .*

(c). *Consider any two economies, (L^b, α, γ) and (L^a, α, γ) with α measurable only w.r.t. r . If $L^a \leq_{PQD} L^b$, then $\mathbb{E}^b[\lambda_2^{b,*} - \lambda_1^{b,*} | I = \iota] \geq \mathbb{E}^a[\lambda_2^{a,*} - \lambda_1^{a,*} | I = \iota]$ for all ι .*

Proposition 8 provides several sets of conditions that shed light on how distortions are allocated within families. Part (a) considers the case in which the social planner values some redistribution to couples with a very unproductive spouse, captured by the condition $\alpha(\cdot, 0) \geq 1$. In this case, the optimal distortion on the secondary earner is on average higher than on the primary earner, provided that the gap in their productivities $1/\iota$ is sufficiently large. The secondary earner faces higher distortions because of the phasing out of transfers. Under conditions of part (a), the planner would like to target some transfers to couples with a very low-earning spouse. As the productivity of that spouse increases, it is optimal to phase such transfers out, leading to high distortions for the secondary earner.

Parts (b) and (c) explore different social weights, the ones that value each couple only based on their total productivity r . Part (b) shows that in this case the optimal tax system is family earnings-based if the total productivity r and the relative productivity ι are independent. When these productivities are not independent, family earnings-based taxation is typically suboptimal, and part (c) of Proposition 8 shows how the strength of dependence affects the magnitude of the relative distortions, $\lambda_2^* - \lambda_1^*$.

The results in parts (b) and (c) of Proposition 8 bear a strong resemblance to the results of Propositions 4 and 5. In Propositions 4 and 5, we considered separable Pareto weights. Such social weights favor individual earning-based taxation. Despite preference for such taxation, the planner chooses individual earnings-based taxes only if individual productivities are independent. Similarly, the planner in parts (b) and (c) of Proposition 8 inherently favors family-based taxation but finds it optimal to depart of it if total family productivity r is not independent from the relative productivity ι of the spouses.

5 Optimal restricted taxation

In Sections 3 and 4, we considered properties of the optimal taxes $T^*(\mathbf{y})$ where tax function T^* have no further ad-hoc restrictions. We showed conditions under which this unrestricted T^* is anonymous, individual, or family earnings-based. One may want to consider the opposite question: how should the planner set taxes optimally that are exogenously restricted to satisfy any of these conditions? Such taxes are frequently used in practice, and knowing how they can be optimally designed is valuable.

As in Section 4, we assume that the first-order approach is valid, $W = \mathbb{R}_+^2$, and the optimal distortions are continuous and bounded on W with derivatives that are continuous and bounded on W^o . The following lemma shows that adapting our techniques to study the optimal form of such restricted taxes is remarkably easy.

Lemma 5. *Consider an economy with $\gamma_1 = \gamma_2$.*

(a) *Optimal anonymous taxes are characterized by (13) with an additional restriction that v is symmetric;*

(b) *Optimal individual earnings-based taxes are characterized by (13) with an additional restriction that v is separable;*

(c) *Optimal family earnings-based taxes are characterized by (13) with an additional restriction that v is measurable only w.r.t. r .*

To understand the economic implications of these measurability restrictions, consider case (a). In the unrestricted case, the tax system allows the planner to condition taxes on the identity of spouse i . Thus, if the planner observes a vector of earnings (y_1, y_2) , she knows that it is supported by the marginal distribution $F(w_1, w_2)$ and applies Pareto weights $\alpha(w_1, w_2)$. However, when taxes are anonymous, the planner cannot distinguish between vectors (y_1, y_2) and (y_2, y_1) . Thus, conditional on observing (y_1, y_2) , the planner only knows that productivities were drawn from the unconditional probability $F^{anon}(w_1, w_2) = \frac{1}{2}F(w_1, w_2) + \frac{1}{2}F(w_2, w_1)$ and can only use the unconditional weights

$$\alpha^{anon}(w_1, w_2) = \frac{\alpha(w_1, w_2) f(w_1, w_2) + \alpha(w_2, w_1) f(w_2, w_1)}{f(w_1, w_2) + f(w_2, w_1)}.$$

As a result, the mechanism design problem that characterizes optimal anonymous taxes and its necessary and sufficient conditions are the same as (13) and (20), (21), (22) respectively, except F and α are replaced with F^{anon} and α^{anon} . Similar logic applies to (b) and (c).

Let $\lambda^{anon,*}$, $\lambda^{ind,*}$, and $\lambda^{fam,*}$ be optimal distortions under anonymous, individual earnings- and family earnings-based taxation. As before, $\bar{\theta}^{anon}$ denotes the right tail elasticity of the marginal of F^{anon} .

Proposition 9. Consider an economy with $\gamma_1 = \gamma_2$ and suppose that both expressions in (42) and (43) are bounded from below by some $\underline{\lambda} > -1$.

(a) Optimal distortions of anonymous taxes satisfy

$$\mathbb{E} \left[\lambda_i^{anon,*} \mid w_i = t \right] = \frac{1 - \mathbb{E}^{anon} [\alpha^{anon} \mid w_i \geq t]}{\gamma \bar{\theta}^{anon}(t)}. \quad (41)$$

(b) Optimal distortions of individual earnings-based taxes satisfy

$$\lambda_i^{ind,*}(t) = \frac{1 - \mathbb{E}[\alpha \mid w_i \geq t]}{\gamma \bar{\theta}_i(t)}. \quad (42)$$

(c) Optimal distortions of family earnings-based taxes satisfy

$$\lambda^{fam,*}(r) = \frac{1 - \mathbb{E}[\alpha \mid R \geq r]}{\gamma \bar{\theta}_r(r)}. \quad (43)$$

This proposition derives expressions for the optimal distortions under restricted taxation. One of the key observations is that the right-hand sides of equations (41), (42) and (43) bear a strong resemblance to optimality conditions for unrestricted taxes, equations (28) and (39). Comparing these conditions, we conclude that optimal distortions under restricted taxation are the same on average as under unrestricted taxation.

Corollary 7. Under the condition of Proposition 9, the relationship between optimal restricted and unrestricted taxes is

$$\begin{aligned} \mathbb{E} \left[\lambda_i^{anon,*} \mid w_i = t \right] &= \sum_{i=1}^2 \frac{g_i(t)}{g_1(t) + g_2(t)} \mathbb{E} \left[\lambda_i^* \mid w_i = t \right], \\ \lambda_i^{ind,*}(t) &= \mathbb{E} \left[\lambda_i^* \mid w_i = t \right], \\ \lambda^{fam,*}(r) &= \mathbb{E} \left[\sum_{i=1}^2 \frac{w_i^{1/(1-\gamma)}}{w_1^{1/(1-\gamma)} + w_2^{1/(1-\gamma)}} \lambda_i^* \mid R = r \right]. \end{aligned}$$

This corollary shows that the analysis of unrestricted taxes also provides insights into the design of restricted taxes. The unrestricted and restricted optimal marginal taxes are chosen to equalize benefits from redistribution to costs of distortions due to changes in labor supply. The main difference is that the unrestricted planner can perturb taxes flexibly; as a result, the optimal marginal taxes can be tagged to (w_1, w_2) , equalizing benefits and costs point-wise. In contrast, the restricted planner cannot freely perturb taxes because certain couples cannot be distinguished and must be bunched; thus the optimal taxes equalize benefits and costs on average.

6 Quantitative analysis

In this section, we illustrate theoretical implications of our analysis using a quantitative model.

6.1 Calibration

To calibrate our model, we use data on earnings of couples from 2020 CPS survey. We restrict attention to couples in which both individuals are between 25 and 65 years old and worked at least 20 weeks in 2019. We invert the joint distribution of earnings to obtain the distribution of productivities. In order to do so, we assume that the environment is symmetric and set $\gamma = 1/4$, so that the implied labor supply elasticity of $1/3$ is the mid-range of values considered by Diamond (1998). Following Guner et al. (2014) and Heathcote et al. (2017), who argue that the U.S. tax schedule is such that family post-tax earnings are approximately a log-linear function of family pre-tax earnings, we assume that households in the data face taxes of the form $T(y_1, y_2) = (y_1 + y_2) - \nu(y_1 + y_2)^{1-\tau}$, where τ and ν are parameters. We refer to this functional form as the HSV tax schedule. With such taxes, the relationship between an observed vector of earnings \mathbf{y} and an unobserved vector of productivities \mathbf{w} is given by

$$w_i^{1/\gamma} = \frac{1}{(1-\tau)\nu} y_i^{1/\gamma-1} (y_1 + y_2)^\tau. \quad (44)$$

To invert this mapping, we use the value of (τ, ν) that Guner et al. (2014) estimate for the U.S. married couples.

We choose a parsimonious representation of the marginal and joint distribution of \mathbf{w} . Consistent with earlier literature (e.g., Badel et al. (2020) or Golosov et al. (2016)), we find that the marginal distribution of productivities G can be well approximated by a Pareto lognormal (PLN) distribution.³⁰ We choose its three parameters (μ, σ, a) to match the mean level of productivity, its Gini coefficient, and the tail parameter. These three moments can be expressed analytically in terms of parameters (μ, σ, a) so that these parameters can be obtained by a simple inversion of those equations (see appendix for the details). Panel (a) of Figure 4 shows the empirical and calibrated distribution G .

Our theory emphasizes several key statistics of the joint distribution F that determine the optimal shape of taxes: the degree of dependence in productivities, right- and left-tail (in)dependence, and the speeds of convergence $\kappa, \bar{\kappa}$. In the data, both observed earnings and backed-out productivities are positively dependent, with Kendall's tau measure of dependence

³⁰The PLN family was introduced in Colombi (1990) as a model of the income distribution, and since then, it has been used extensively in various studies. It is defined as $G(t) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right) - t^{-a} \exp(a\mu + a^2\sigma^2/2) \Phi\left(\frac{\ln t - \mu - a\sigma^2}{\sigma}\right)$, where Φ is the standard normal distribution.

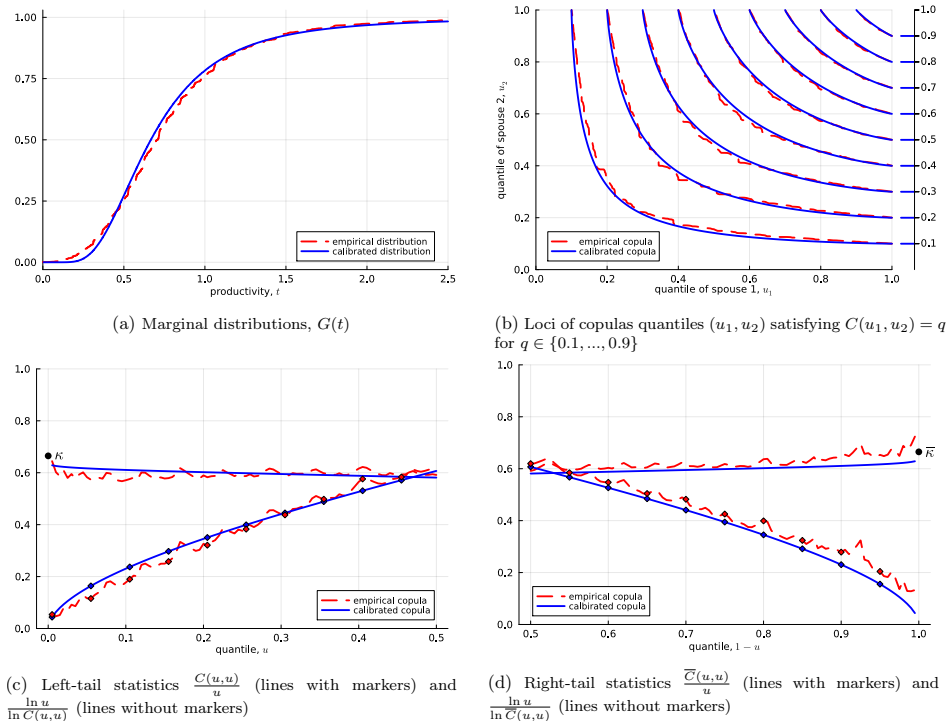


Figure 4: Empirical and calibrated joint distributions of productivities

for productivities equaling 0.21.³¹ The joint distribution appears to be both left- and right-tail independent, but the rate of convergence to independence is fairly low. Panels (c) and (d) illustrate this by plotting in dashed lines the value of the empirical copula $C(u, u)/u$ for difference percentiles of productivities (red line), and the value of $\ln u / \ln C(u, u)$ (blue lines). Consistent with left tail-independence, $C(u, u)/u$ approaches zero for low u ; consistent with the slow speed of convergence $\ln u / \ln C(u, u)$ remains much above $1/2$. Panel (d) plots similar statistics summary statistics for the right tails, using moments of the empirical survival copula.

We experimented with different families of copulas to capture these patterns and found that the Gaussian copula fits the data very well. Its parameter ρ , when chosen to match the Kendall's tau dependence coefficient,³² also fits well the measures of left- and right-tail dependence and speeds of convergence. This can be seen from panels (c) and (d) of Figure 4, where in solid lines we plot the counterparts of calibrated Gaussian copula of the empirical objected plotted in dashed lines. Black dots show the speeds of convergence for the Gaussian

³¹Kendell's tau is the standard measure of strength dependence of two variables (see Chapter 5 in Nelsen (2006)). Kendall's tau measure has an advantage over Pearson's correlation coefficient because it is independent of the marginal distributions. In our data, Pearson's correlation coefficients of earnings and productivities are 0.21 and 0.25, respectively.

³²The relationship between Kendall's tau and parameter ρ of the Gaussian copula is given by Kendall's tau = $\frac{2 \arcsin \rho}{\pi}$.

| Parameter | Value | Definition | Target |
|-----------|-------|------------------------------------------|-------------------------------------------------------------|
| γ | 0.25 | Measure of labor supply elasticity | Elasticity of labor supply, 0.33 |
| a | 2.95 | Pareto tail of PLN cdf | Pareto statistics at 99% of individual productivities, 2.95 |
| μ | -0.71 | Location parameter of PLN cdf | Mean individual productivity, 0.81 |
| σ | 0.40 | Shape parameter of PLN cdf | Gini of individual productivities, 0.31 |
| ρ | 0.33 | Correlation parameter of Gaussian copula | Kendell's tau of spousal productivities, 0.21 |

Table 1: Calibrated parameters

copula, $\kappa = \bar{\kappa} = \frac{1+\rho}{2}$. Panel (b) of Figure 4 shows the “isoquants” of both empirical and calibrated joint distribution, where each line plots, for a given q , all pairs (u_1, u_2) that satisfy $C(u_1, u_2) = q$. Table 1 summarizes all our parameters and their empirical counterparts.

We want to make several remarks. If one were to assume that the HSV tax schedule applies to individual rather than family earnings, then calibration of the unobservable distribution of productivities using the observed distribution of earnings is particularly simple and transparent. If the joint distribution of earnings is PLN-Gaussian and individuals face HSV taxes, then the joint distribution of productivities is also PLN-Gaussian. Moreover, the parameters of (μ, σ, a, ρ) of this distribution of productivities can be expressed in closed form as functions of mean, Gini, Pareto tail, and Kendell's tau measures of the joint distribution of earnings, as well as parameters (τ, ν) of the HSV tax function. Thus, the entire calibration can be done using the moments of the raw data directly, sidestepping the need for the inversion described in equation (44). The calibration approach we present in the text is more general and uses a more realistic specification of the tax function. In any case, we tried both approaches and obtained very similar results.

Secondly, in the appendix we show how the optimal taxes would change if the joint distribution were given by the FGM rather than the Gaussian copula, calibrated to match the same Kendell's tau coefficient. The calibrated FGM copula, like the Gaussian copula we use, fits empirical isoquants reasonably well and is tail independent. However, it converges to tail independence much faster, with $\kappa, \bar{\kappa}$ equal to $\frac{1}{2}$ that are clearly rejected by our data. Thus, comparing the results reported in the text for the Gaussian copula with those reported in the appendix for the FGM copula highlights the role of the speed of convergence of tail dependence for optimal taxation.

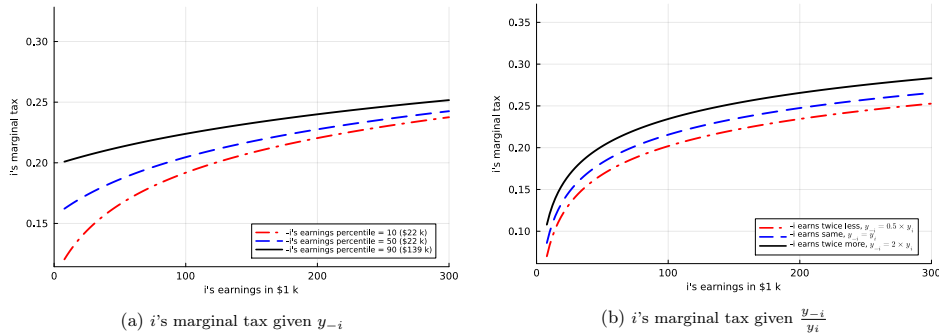


Figure 5: U.S. taxes implied by the estimated HSV schedule

6.2 Optimal taxes in the calibrated economy

To study optimal taxes, we consider a family of Pareto weights given by $\tilde{\alpha}(w) = e^{-mw^{1/(1-\gamma)}}$ and $\alpha(\mathbf{w}) \propto \left[\frac{1}{2}\tilde{\alpha}(w_1)^k + \frac{1}{2}\tilde{\alpha}(w_2)^k \right]^{1/k}$, where k, m are parameters. This rich family allows us to illustrate how optimal taxes are affected by various properties of Pareto weights. Parameter m measures the strength of the redistributive motive, with higher values of m corresponding to more redistributive planners. Parameter k measures the strength of modularity. Weights α are submodular when $k \geq 1$ and supermodular when $k \leq 1$. Weights are separable when $k = 1$, corresponding to a benchmark economy. The case $k = 0$ corresponds to weights α being measurable only w.r.t. to r , i.e., the case when the planner has an inherent preference for family earnings-based taxation.

To compute the optimal taxes, we first solve the relaxed problem (13) and then verify that the solution satisfies global incentive constraints. In all cases, which we report here and in the appendix, we found that the FOA was valid. We provide additional computational details in the appendix. We summarize properties of the optimal marginal taxes in two sets of figures. The first set reports the marginal taxes $\nabla_i T^*(y_i, y_{-i})$ as a function of y_i , holding y_{-i} fixed at different levels. The second set reports $\nabla_i T^*(y_i, by_i)$ as a function of y_i , holding ratio of earnings y_{-i}/y_i fixed at different values of b . For ease of comparison, Figure 5 reports these statistics of the U.S. tax schedule implied by the estimated HSV functional form.

We now discuss the optimal taxes. We start with a benchmark economy, which restricts $k = 1$. We set the parameter m to 0.35, which in our calibration implies that the average optimal marginal tax rate coincides with the average marginal tax rate in the data. This way, the total amount of redistribution is similar in our model and the data. Figure 6 reports the optimal taxes in the same format as Figure 5. It also plots, in gray lines, optimal tax rates for two alternative assumptions about dependence of productivities: perfect dependence (dashed-dotted) and independence (solid). As discussed in Section 4, the optimal taxes under perfect

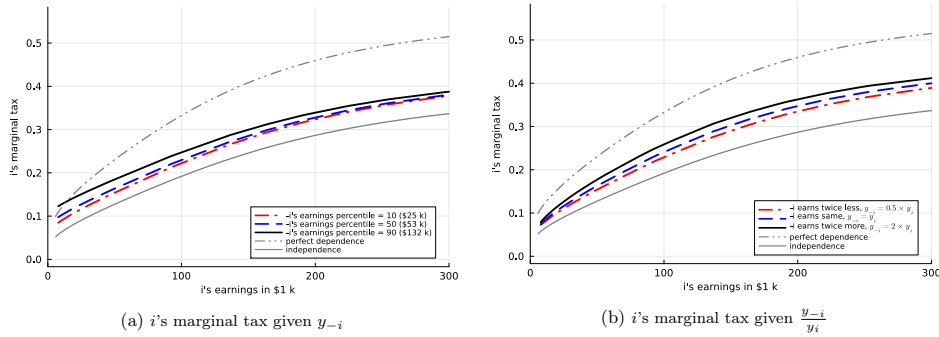


Figure 6: Optimal taxes, $m = 0.35$ and $k = 1$

dependence can also be interpreted as the optimal tax on single individuals. The optimal taxes with independent types are separable and thus do not depend on the other spouse's earnings.

Optimal taxes in the calibrated economy lie between the two gray lines, consistent with the comparative statics results we established in Proposition 5. One can also see from Figure 5 that the optimal marginal taxes are positively jointed for low earners since the marginal tax $\nabla_i T^*(y_i, y_{-i})$ is increasing in y_{-i} for low values of y_i . This result follows from Proposition 7. The same proposition also established that the optimal taxes must be negatively-jointed for high-earners.³³ This occurs at much higher earnings levels ($> \$8.5$ mln) than the scale of the x-axis we use. That being said, optimal jointness is very modest for all earnings levels, with marginal taxes for one spouse changing by, at most, several percentage points as a function earnings of the other spouse. This feature is driven by the properties of the Gaussian copula. In the appendix, we plot the optimal taxes for the FGM copula and show that optimal jointness is much more pronounced in that case. This is consistent with our discussion in Section 4.3, where we showed that a slower pace of convergence to tail independence implies a smaller force for positive jointness at the bottom (negative jointness at the top), that eventually switched to negative (positive) jointness for tail-dependence distributions. The Gaussian copula, with its lower speed of convergence, implies smaller jointness than the FGM copula.

The optimal tax schedule shares properties implied by our analytical formulas, both qualitatively and quantitatively. In particular, in the appendix, we plot the optimal distortions $\lambda^*(\cdot, w_{-i})$ where wages of the spouse w_{-i} are held at the 50th percentile of the productivity distribution. These distortions align very closely with the analytical expression for the average distortions $\mathbb{E} \left[\lambda_i^* \mid w_i = t \right]$ that we derived in equation (28).

Formula (28) together with the bounds established in Propositions 4 and 7 allow us to

³³Spiritus et al. (2022) solved numerically a related optimal joint taxation problem and also found that optimal jointness may be positive or negative.

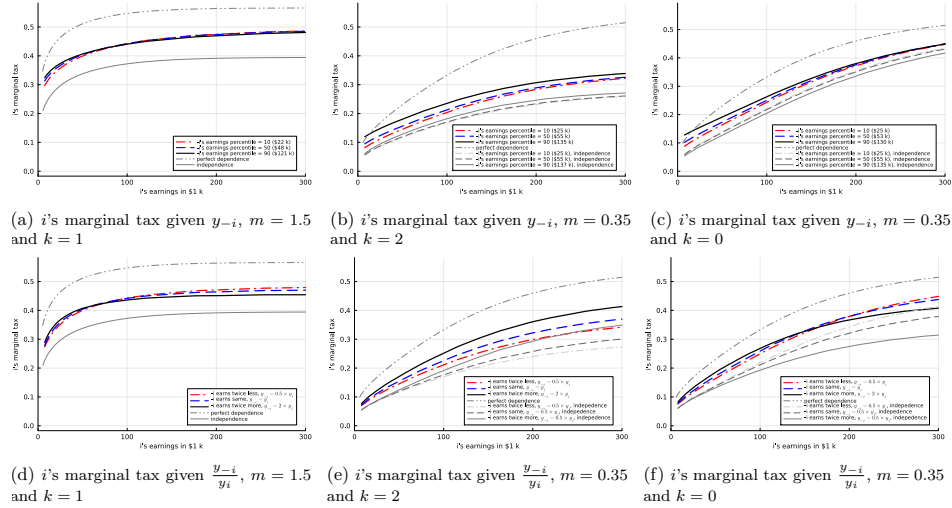


Figure 7: Optimal taxes, robustness to m and k

understand some quantitative properties of the optimal tax schedule shown in Figure 6. On the left tail, the bounds established in Propositions 4 and 7 imply that distortions from the optimal taxes should be between $\frac{1}{2}\gamma\theta(0)(\tilde{\alpha}(0) - 1)$ and $\gamma\theta(0)(\tilde{\alpha}(0) - 1)$, where $\theta(0) := \lim_{t \rightarrow 0} \theta(t)$. The PLN distribution has a thin left tail, which implies that $\theta(0) = 0$ and, therefore, optimal marginal taxes must go to zero as $y_i \rightarrow 0$, at least when averaged over y_{-i} . This explains why the optimal marginal tax rates go to zero as $y_i \rightarrow 0$ in panel (b) of Figure 6. The right tail of the PLN distribution is thick, with $\lim_{t \rightarrow \infty} \theta(t) = a > 0$. This implies that the limiting average distortions in our calibrated economy, using formula (28), are

$$\lim_{t \rightarrow \infty} \mathbb{E} \left[\lambda_i^* \mid w_i = t \right] = \frac{1}{\gamma a} \approx 1.35,$$

corresponding to the average tax rate on high-earner to be around 0.55. However, this limit is reached very slowly as the convergence rate of $\theta(t)$ to a is low. The slow speed of convergence explains why the optimal marginal tax rates in Figure 6 are substantially lower than this limit, even for individuals with earnings of \$300K.

In Figure 7, we explore sensitivity of these results to parameters governing Pareto weights. In the top row we plot the optimal taxes for the benchmark economy ($k = 1$) but with a much more redistributive planner ($m = 1.5$). The optimal taxes are higher, reflecting a stronger desire for redistribution. Negative jointness now occurs at much lower earnings levels than in Figure 6 and can be easily seen on the graphs. Nonetheless, the level of optimal jointness remains low.

The second and third columns in Figure 7 plot optimal taxes for $(m, k) = (0.35, 2)$ and for $(m, k) = (0.35, 0)$ illustrating effects of sub- and super-modularity. The level of the optimal

marginal tax rates in these two cases is broadly similar to our starting point, $(m, k) = (0.35, 1)$. Consistent with Corollary 5, submodularity amplifies the benefits of positive jointness at the bottom, so the positive jointness is now much more visible and substantial, especially in panel (b). On the other hand, supermodularity decreases these benefits and amplifies gains from the negative jointness at the top. As a result, both negative jointness at the bottom and positive jointness at the top are now clearly visible. Nonetheless, the optimal magnitude of this jointness is very low, and the optimal taxation is close to the individual earnings-based tax schedule. This remains the case even when $k = 0$, i.e., social weights explicitly favor family earnings-based taxation.

Figure 7 also plots the optimal tax rate under perfect dependence and independence. The optimal tax schedule is no longer separable with independent types when $k \neq 1$. Therefore, in the last two columns of Figure 7, we use several gray lines to plot optimal tax rates under independent types. One can easily see from these graphs that under independence optimal taxes are positively (negatively) jointed if α is submodular (supermodular), consistent with the results in Kleven et al. (2007) and our discussion in Section 4.3.

The U.S. tax system is family earnings-based, and the departure of the optimal taxation from the family earnings-based one is difficult to see from Figures 6 and 7. To make this comparison easier, we represent the optimal tax schedule in the form $T^*(\mathbf{y}) = T^{fam,*}(Y(\mathbf{y}), \tau(\mathbf{y}))$, where $Y(\mathbf{y}) = y_1 + y_2$ are the total family earnings and $\varrho(\mathbf{y}) = \min\{y_1, y_2\}/Y(\mathbf{y})$ is the share of a secondary earner in total earnings. In Figure 8 we plot $\frac{\partial}{\partial Y} T^{fam,*}(\cdot, \varrho)$ for different values of ϱ . Panel (a) uses this representation to show the U.S. tax code implied by the estimated HSV function. Since U.S. tax schedule is family earnings-based, $\frac{\partial}{\partial Y} T^{fam,US}(\cdot, \varrho)$ is the same for all ϱ . In panels (b)-(f) we plot the optimal tax on family earnings for the same specifications that we used in Figures 6 and 7. The marginal tax rates vary substantially with the share of earnings of the secondary earner, with a higher share corresponding to a lower marginal family tax. In all cases, pure family earnings-based taxation is a poor approximation of the optimal tax code.

7 Conclusion

Multidimensional screening problems are ubiquitous in public finance applications. In this paper, we consider one of the simplest versions of such problems - the optimal taxation of joint earnings of couples. We show that despite superficial similarity to multidimensional screening problems in industrial organization, our problem is much easier to analyze and can often be studied using the first-order approach. We identify the lack of participation constraints in our

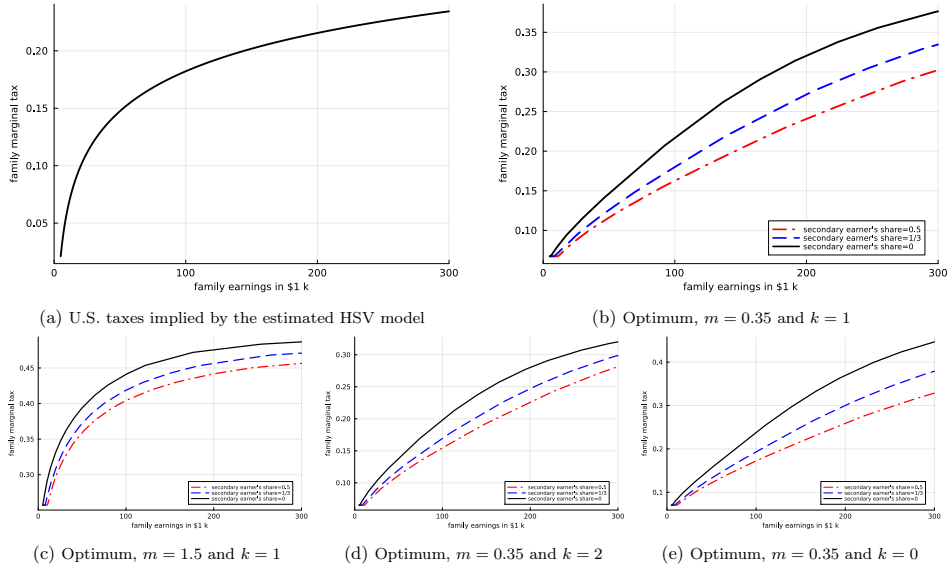


Figure 8: Marginal taxes on family earnings

application as the key reason for this simplification.

We also characterize the optimal taxes in these settings. Such taxes are a solution to a second-order partial differential equation, which is very complex and does not generally have an analytical solution. We show that this problem can be overcome by focusing on various conditional average moments of taxes. These conditional moments are very illuminating about the economic mechanisms that drive the shape of the optimal tax schedule, both qualitatively and quantitatively.

In the calibrated economy, we find that the optimal taxes are negatively jointed at the bottom and positively at the top. However, this jointness is small, and the optimal taxes can be well approximated by separable, individual earnings-based taxation. In contrast, family earnings-based taxes provide a poor approximation to the optimal tax code, even when Pareto weights explicitly favor this type of taxation.

References

- Albanesi, Stefania, and Christopher Sleet.** 2006. “Dynamic optimal taxation with private information.” *Review of Economic Studies*, 73(1): 1–30.
- Ales, Laurence, and Christopher Sleet.** 2022. “Optimal Taxation of Income-Generating Choice.” *Econometrica*, 90(5): 2397–2436.
- Alves, Cassiano Breno Machado, Carlos Eugenio da Costa, and Humberto Ataide Moreira.** 2021. “Intrahousehold Inequality and the Joint Taxation of Household Earnings.” FGV EPGE Economics Working Papers 825.
- Armstrong, Mark.** 1996. “Multiproduct Nonlinear Pricing.” *Econometrica*, 64(1): 51–75.
- Badel, Alejandro, Mark Huggett, and Wenlan Luo.** 2020. “Taxing Top Earners: a Human Capital Perspective.” *Economic Journal*, 130(629): 1200–1225.
- Balakrishnan, Narayanaswamy, and Chin Diew Lai.** 2009. *Continuous Bivariate Distributions.*: Springer Berlin.
- Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin.** 2022. “Bunching and Taxing Multidimensional Skills.” Working Paper 30015, National Bureau of Economic Research.
- Coles, Stuart, Janet Heffernan, and Jonathan Tawn.** 1999. “Dependence Measures for Extreme Value ABCnalsyses.” *Extremes*, 2(4): 339–365.
- Colombi, Roberto.** 1990. “A New Model of Income Distribution: The pareto-lognormal Distribution.” In *Income and wealth distribution, inequality and poverty.*: Springer, 18–32.
- Diamond, Peter.** 1998. “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates.” *American Economic Review*, 88(1): 83–95.
- Evans, Lawrence C.** 2010. *Partial differential equations.* 19: American Mathematical Soc..
- Evans, Lawrence C, and Ronald F Garzepy.** 2018. *Measure theory and fine properties of functions.*: Routledge.
- Farhi, Emmanuel, and Ivan Werning.** 2013. “Insurance and Taxation over the Life Cycle.” *Review of Economic Studies*, 80(2): 596–635.

- Ferey, Antoine, Benjamin B. Lockwood, and Dmitry Taubinsky.** 2022. “Sufficient Statistics for Nonlinear Tax Systems with General Across-Income Heterogeneity.” CESifo Working Paper Series 9958, CESifo.
- Frankel, Alexander.** 2014. “Taxation of Couples under Assortative Mating.” *American Economic Journal: Economic Policy*, 6(3): 155–77.
- Gayle, George-Levi, and Andrew Shephard.** 2019. “Optimal Taxation, Marriage, Home Production, and Family Labor Supply.” *Econometrica*, 87(1): 291–326.
- Gilbarg, David, Neil S Trudinger, David Gilbarg, and NS Trudinger.** 1977. *Elliptic partial differential equations of second order*. 224: Springer.
- Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski.** 2003. “Optimal Indirect and Capital Taxation.” *Review of Economic Studies*, 70(3): 569–587.
- Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski.** 2016. “Redistribution and Social Insurance.” *American Economic Review*, 106(2): 359–86.
- Golosov, Mikhail, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl.** 2013. “Preference heterogeneity and optimal capital income taxation.” *Journal of Public Economics*, 97 160–175.
- Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin.** 2014. “A Variational Approach to the Analysis of Tax Systems.” Working Paper 20780, National Bureau of Economic Research.
- Grisvard, P.** 2011. *Elliptic Problems in Nonsmooth Domains*. Classics in Applied Mathematics: Society for Industrial and Applied Mathematics.
- Guner, Nezh, Remzi Kaygusuz, and Gustavo Ventura.** 2014. “Income Taxation of U.S. Households: Facts and Parametric Estimates.” *Review of Economic Dynamics*, 17(4): 559–581.
- Hajłasz, Piotr.** 1999. *Sobolev mappings, co-area formula and related topics*.: Institut Mittag-Leffler.
- Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante.** 2017. “Optimal Tax Progressivity: An Analytical Framework.” *Quarterly Journal of Economics*, 132(4): 1693–1754.

- Heathcote, Jonathan, and Hitoshi Tsujiyama.** 2021. “Optimal Income Taxation: Mirreles Meets Ramsey.” *Journal of Political Economy*, 129(11): 3141–3184.
- Heffernan, Janet E.** 2000. “A Directory of Coefficients of Tail Dependence.” *Extremes*, 3(3): 279–290.
- Hellwig, Christian, and Nicolas Werquin.** 2022. “A Fair Day’s Pay for a Fair Day’s Work: Optimal Tax Design as Redistributive Arbitrage.” TSE Working Papers 22-1284, Toulouse School of Economics (TSE).
- Hua, Lei, and Harry Joe.** 2011. “Tail order and intermediate tail dependence of multivariate copulas.” *Journal of Multivariate Analysis*, 102(10): 1454–1471.
- Hua, Lei, and Harry Joe.** 2014. “Strength of tail dependence based on conditional tail expectation.” *Journal of Multivariate Analysis*, 123 143–159.
- Karlin, Samuel, and Yosef Rinott.** 1980. “Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions.” *Journal of Multivariate Analysis*, 10(4): 467–498.
- Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez.** 2007. “The Optimal Income Taxation of Couples as a Multi-Dimensional Screening Problem.” CESifo Working Paper Series 2092, CESifo.
- Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez.** 2009. “The Optimal Income Taxation of Couples.” *Econometrica*, 77(2): 537–560.
- Kushnir, Alexey, and Ali Shourideh.** 2022. “Optimal Multi-Dimensional Mechanisms: A Heuristic Approach.” working paper, Carnegie Mellon University.
- Ledford, Anthony W, and Jonathan A Tawn.** 1996. “Statistics for Near Independence in Multivariate Extreme Values.” *Biometrika*, 83(1): 169–187.
- Ledford, Anthony W, and Jonathan A Tawn.** 1998. “Concomitant Tail Behaviour for Extremes.” *Advances in applied Probability*, 30(1): 197–215.
- Lockwood, Benjamin B., and Matthew Weinzierl.** 2015. “De Gustibus non est Taxandum: Heterogeneity in preferences and optimal redistribution.” *Journal of Public Economics*, 124 74–80.

- Mirrlees, James.** 1971. “An Exploration in the Theory of Optimum Income Taxation.” *Review of Economic Studies*, 38(2): 175–208.
- Mirrlees, James.** 1986. “The Theory of Optimal Taxation.” *Handbook of Mathematical Economics*, Volume 3 1197–1249.
- Moser, Christian, and Pedro Olea de Souza e Silva.** 2019. “Optimal Paternalistic Savings Policies.” Opportunity and Inclusive Growth Institute Working Papers 17, Federal Reserve Bank of Minneapolis.
- Nardi, Giacomo.** 2015. “Schauder Estimate for Solutions of Poisson’s Equation with Neumann Boundary Condition.” *L’enseignement Mathématique*, 60(3): 421–435.
- Ndiaye, Abdoulaye.** 2018. “Flexible Retirement and Optimal Taxation.” working paper, NYU.
- Nelsen, Roger B.** 2006. *An Introduction to Copulas*. New York, NY, USA: Springer, , 2nd edition.
- Renes, Sander, and Floris Zoutman.** 2017. “As Easy as ABC? Multidimensional Screening in Public Finance.” CESifo Working Paper Series 6301, CESifo.
- Rindler, Filip.** 2018. *Calculus of Variations*.: Springer Cham.
- Rochet, Jean.** 1987. “A Necessary and Sufficient Condition for Rationalizability in a Quasi-linear Context.” *Journal of Mathematical Economics*, 16(2): 191–200.
- Rochet, Jean-Charles, and Philippe Chone.** 1998. “Ironing, Sweeping, and Multidimensional Screening.” *Econometrica*, 66(4): 783–826.
- Rockafellar, Ralph Tyrell.** 2015. *Convex Analysis*.: Princeton university press.
- Rothschild, Casey, and Florian Scheuer.** 2013. “Redistributive Taxation in the Roy Model.” *The Quarterly Journal of Economics*, 128(2): 623–668.
- Rothschild, Casey, and Florian Scheuer.** 2014. “A Theory of Income Taxation under Multidimensional Skill Heterogeneity.” Working Paper 19822, National Bureau of Economic Research.
- Rothschild, Casey, and Florian Scheuer.** 2016. “Optimal Taxation with Rent-Seeking.” *Review of Economic Studies*, 83(3): 1225–1262.

- Saez, Emmanuel.** 2001. “Using Elasticities to Derive Optimal Income Tax Rates.” *Review of Economic Studies*, 68(1): 205–229.
- Shaked, M., and J.G. Shanthikumar.** 2007. *Stochastic Orders*. Springer Series in Statistics: Springer New York.
- Spiritus, Kevin, Etienne Lehmann, Sander Renes, and Floris T. Zoutman.** 2022. “Optimal Taxation with Multiple Incomes and Types.” IZA Discussion Papers 15036.
- Stantcheva, Stefanie.** 2017. “Optimal Taxation and Human Capital Policies over the Life Cycle.” *Journal of Political Economy*, 125(6): 1931–1990.
- Zeidler, Eberhard.** 2012. *Applied functional analysis: applications to mathematical physics*. 108: Springer Science & Business Media.
- Zeidler, Eberhard.** 2013. *Nonlinear functional analysis and its applications: I: nonlinear monotone operators.*: Springer Science & Business Media.

Appendix

A Motivation for the benchmark economy

In this section, we formalize the results discussed in Section 2 about the relation between optimal taxes on single and married individuals. Formally, consider the following economy, which we refer to as *extended*. There is a measure one of individuals who all draw their productivities from the same marginal distribution G and have the same labor supply elasticity γ . Each individual, independently from his or her productivity, enters the marriage market with probability $\mu \in (0, 1)$ and with probability $1 - \mu$ stays single. All individuals who entered the marriage market are matched so that F gives the joint distribution of productivities of married individuals. By construction, G is the marginal of F .

The utility \tilde{v} of a single individual with productivity w is $\tilde{v} = c - \gamma \left(\frac{y}{w}\right)^{1/\gamma}$, and utility v of two married individuals with productivities $\mathbf{w} = (w_1, w_2)$ is given by (1). The social planner chooses taxes on single individuals $\tilde{T}(y)$ and married individuals $T(\mathbf{y})$ to maximize a weighted average of their utilities

$$(1 - \mu) \int_{\underline{w}}^{\bar{w}} \tilde{v}(w) \tilde{\alpha}(w) g(w) dw + \frac{\mu}{2} \int_{\mathbf{W}} v(\mathbf{w}) \alpha(\mathbf{w}) f(\mathbf{w}) d\mathbf{w},$$

where $\tilde{\alpha}(w)$ and $\alpha(\mathbf{w}) = \frac{1}{2}\tilde{\alpha}(w_1) + \frac{1}{2}\tilde{\alpha}(w_2)$ are Pareto weights on singles and couples respectively, i.e., $\tilde{\alpha}$ is decreasing, bounded, continuous and satisfies $\int \tilde{\alpha}(w) dG(w) = 1$.

Lemma 6. *For any $\mu \in (0, 1)$, the optimal tax on married individuals coincides (up to a constant) with the optimal tax in the benchmark economy $(F, \alpha, (\gamma, \gamma))$ and the optimal tax on single individuals coincides (up to a constant) with the optimal tax in the corresponding uni-dimensional economy $(G, \tilde{\alpha}, \gamma)$.*

Proof. Let \tilde{T} and T be tax functions that are budget-feasible in the extended economy. Denote the corresponding earnings chosen by single and married individuals by y and $\mathbf{y} = (y_1, y_2)$. Since welfare in the first-best allocation is finite due to (2), the total expected tax revenue is finite. Define two new tax functions \tilde{T}^{new} and T^{new} as follows:

$$\begin{aligned} \tilde{T}^{new} &\equiv \tilde{T} - \mu \mathbb{E} \left[\tilde{T}(y(w)) \right] - \frac{\mu}{2} \mathbb{E} [T(\mathbf{y}(\mathbf{w}))], \\ T^{new} &\equiv T + (1 - \mu) \mathbb{E} \left[\tilde{T}(y(w)) \right] - \left(1 - \frac{\mu}{2}\right) \mathbb{E} [T(\mathbf{y}(\mathbf{w}))]. \end{aligned}$$

It is easy to see that welfare is the same under these new taxes due to the relationship between $\tilde{\alpha}$ and α . By construction, T^{new} is budget-feasible in the benchmark economy, and \tilde{T}^{new} is budget-feasible in its uni-dimensional counterpart. As a result, the optimal taxes for married households are the same (up to a constant) as in the benchmark economy, and the optimal taxes for single individuals are the same (up to a constant) as in the uni-dimensional economy. \square

Lemma 6 establishes that in the extended economy with Pareto weights that satisfy $\alpha(\mathbf{w}) = \frac{1}{2}\tilde{\alpha}(w_1) + \frac{1}{2}\tilde{\alpha}(w_2)$ the optimal marginal taxes are the same irrespective of how individuals are matched into couples, i.e., μ and F . It can be shown that this condition is also necessary in the following sense. Consider the extended economy but follow for general Pareto weights α on married individuals. Then, the optimal marginal taxes are the same for all μ and F only if $\alpha(\mathbf{w}) = \frac{1}{2}\tilde{\alpha}(w_1) + \frac{1}{2}\tilde{\alpha}(w_2)$ for all \mathbf{w} .

B Mathematical preliminaries

This section lists some basic mathematical concepts necessary to characterize a solution to the taxation problem. We refer the reader to Evans (2010) and Rindler (2018) for additional background reading.

Let $U \subseteq \mathbb{R}^2$ be open. A function $v : U \rightarrow \mathbb{R}$ is compactly supported in U if it is zero outside some compact $C \subset U$. A measurable $v : U \rightarrow \mathbb{R}$ is *locally integrable* on U if $\int_C |v(\mathbf{w})| d\mathbf{w} < \infty$ for every compact $C \subset U$, and it is called *integrable* on U if $\int_U |v(\mathbf{w})| d\mathbf{w} < \infty$. A measurable $v : U \rightarrow \mathbb{R}$ is said to be *essentially bounded* on U if there exists m such that $|v(\mathbf{w})| \leq m$ a.e. on U .

A locally integrable function $v : U \rightarrow \mathbb{R}$ is *weakly differentiable* on U if there exists a locally integrable vector field $\nabla v : U \rightarrow \mathbb{R}^2$ such that for all infinitely differentiable ϕ with a compact support in U ,

$$\int_U \nabla_i \phi(\mathbf{w}) v(\mathbf{w}) d\mathbf{w} = - \int_U \phi(\mathbf{w}) \nabla_i v(\mathbf{w}) d\mathbf{w}.$$

The vector field ∇v is called a *weak gradient*, it is unique up to a set of zero measure. If v is differentiable, it is weakly differentiable, and its weak gradient coincides with the classical one.

In the sequel, we will need the following Sobolev spaces:

$$\begin{aligned} \mathcal{W}^{1,1}(U) &:= \{v : U \rightarrow \mathbb{R} | v \text{ weakly differentiable s.t. } v \text{ and } \nabla v \text{ are integrable}\}, \\ \mathcal{W}^{1,\infty}(U) &:= \{v : U \rightarrow \mathbb{R} | v \text{ weakly differentiable s.t. } v \text{ and } \nabla v \text{ are essentially bounded}\}. \end{aligned}$$

Clearly, for bounded U , $\mathscr{W}^{1,1}(U) \supset \mathscr{W}^{1,\infty}(U) \supset \mathscr{C}^1(\overline{U})$, where $\mathscr{C}^1(\overline{U})$ is the space of continuously differentiable functions with derivatives that are uniformly continuous on bounded subsets. In general, we will use $\mathscr{C}^k(U)$ to denote the space of k -times continuously differentiable functions. If a function admits continuous derivatives of all orders, it will be called smooth. Similarly, $\mathscr{C}^k(\overline{U})$ will be used for the space of k -times continuously differentiable functions with derivatives that are uniformly continuous on bounded subsets. In addition, for every $a \in (0, 1]$, $\mathscr{C}^{k,a}(\overline{U})$ stays for the space of k -times continuously differentiable functions with derivatives that are a -Hölder continuous on U , i.e., v is in $\mathscr{C}^{0,a}(\overline{U})$ if $\sup_{\mathbf{w}, \widehat{\mathbf{w}} \in U, \mathbf{w} \neq \widehat{\mathbf{w}}} \frac{|v(\mathbf{w}) - v(\widehat{\mathbf{w}})|}{\|\mathbf{w} - \widehat{\mathbf{w}}\|^a} < \infty$. To simplify notations, when referring to various functional spaces we may omit a domain if it is clear from the context, no confusion should arise.

We will oft manipulate integrals by integrating by parts and changing variables of integration. The Divergence Theorem is a multidimensional analogue of integration by parts, and the Coarea Formula is a generalization of change of variables in an integral.

If $v \in \mathscr{C}^1(\overline{U})$, then $v|_{\partial U}$ is well-defined by definition. In general, for $v \in \mathscr{W}^{1,1}(U)$, $v|_{\partial U}$ stays for a *trace* of v on ∂U , which is defined uniquely for bounded U with a Lipschitz boundary, i.e., the boundary that can be thought of as locally being the graph of $\mathscr{C}^{0,1}$ function. More generally, a boundary is said to be $\mathscr{C}^{k,a}$ with $a \in (0, 1]$ if it can be locally represented as a graph of $\mathscr{C}^{k,a}$ function (Definition 1.2.1.2 in Grisvard (2011)). For sets with a Lipschitz boundary, Theorem 1.5.1.3 in this book asserts that there exists a unique continuous operator \mathcal{T} from $\mathscr{W}^{1,1}(U)$ onto the space of integrable functions on ∂U ; thus, we set $v|_{\partial U} := \mathcal{T}(v)$.

It is well known that for $\phi \in \mathscr{W}^{1,\infty}(U)$ and $v \in \mathscr{W}^{1,1}(U)$, the product ϕv in an element of $\mathscr{W}^{1,1}(U)$ with $\nabla(\phi v) = \nabla\phi v + \phi \nabla v$. Then, the following identity, known as the Divergence Theorem (Theorem 1.5.3.1 in Grisvard (2011)), is satisfied for any bounded U with a Lipschitz boundary:

$$\int_{\overline{U}} \nabla_i(\phi(\mathbf{w})v(\mathbf{w})) d\mathbf{w} = \int_{\partial U} \phi(\mathbf{w})v(\mathbf{w})n_i(\mathbf{w}) d\sigma(\mathbf{w}),$$

where $n_i(\mathbf{w})$ is the i -th component of the outward unit vector to ∂U at \mathbf{w} and σ is the Lebesgue measure on ∂U .

The Coarea Formula for Sobolev spaces is Theorem 11 in Hajlasz (1999). Let $Q : \mathbb{R}_{++}^2 \rightarrow \mathbb{R}^n$, where n is either one or two, be an element of $\mathscr{W}^{1,1}(\mathbb{R}_{++}^2 \cap B_r^o)$, where B_r is a closed ball of radius r , for each $r > 0$. Then, for every (Borel) measurable $v : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ that is either non-negative or is such that $v \cdot \mathcal{J}Q$ is integrable, where $\mathcal{J}Q(\mathbf{w})$ is the Jacobian of Q at \mathbf{w} ,

$$\int v(\mathbf{w}) \cdot \mathcal{J}Q(\mathbf{w}) d\mathbf{w} = \int \left(\int_{Q^{-1}(q)} v(\mathbf{w}) d\mathcal{H}(\mathbf{w}) \right) dq,$$

where \mathcal{H} is the so called Hausdorff measure. As explained in Chapter 2 of Evans and Garzepy (2018), the Hausdorff measure coincides with the standard Lebesgue measure for "nice" sets, i.e., $n = 1$ and $Q^{-1}(q)$ is a plane curve. On the other hand, if $n = 2$ and Q is injective, then the Hausdorff measure is a counting measure; thus, the inner integral on the right-hand side is $v(Q^{-1}(q))$, and we recover the standard change of variables formula.

C Proofs for Section 2

C.1 Proof of Lemma 1

Proof. First of all, $c(\mathbf{w}) = y_1(\mathbf{w}) + y_2(\mathbf{w}) - T(\mathbf{y}(\mathbf{w}))$ for all $\mathbf{w} \in W$ due to optimality of (c, \mathbf{y}) , and (7) directly follows from non-negativity of $\mathbb{E}[T(\mathbf{y}(\mathbf{w}))]$. Since for every couple $\mathbf{w} \in W$, each pair $(c(\widehat{\mathbf{w}}), \mathbf{y}(\widehat{\mathbf{w}}))$ with $\widehat{\mathbf{w}} \neq \mathbf{w}$ is feasible but not chosen, equation (8) ensues.

Conversely, take any (c, \mathbf{y}) that satisfies (7) and (8) and choose the tax function T as follows. For any \mathbf{y} for which there exists $\mathbf{w} \in W$ such that $\mathbf{y} = \mathbf{y}(\mathbf{w})$, set $T(\mathbf{y}) := y_1 + y_2 - c(\mathbf{w})$; otherwise, set $T(\mathbf{y}) := \infty$. This assignment uniquely defines T because $c(\mathbf{w}) = c(\widehat{\mathbf{w}})$ whenever $\mathbf{y}(\mathbf{w}) = \mathbf{y}(\widehat{\mathbf{w}})$ due to (8). By (8), $(c(\mathbf{w}), \mathbf{y}(\mathbf{w}))$ is an optimal choice of every couple \mathbf{w} , and the non-negativity of taxes follows from (7). \square

C.2 Proof of Lemma 2

Proof. Use (6) to substitute for $c(\mathbf{w})$ in (7) and (8) to obtain

$$\int_W \left(\sum_{i=1}^2 \left(y_i(\mathbf{w}) - \gamma_i \left(\frac{y_i(\mathbf{w})}{w_i} \right)^{1/\gamma_i} \right) - v(\mathbf{w}) \right) f(\mathbf{w}) d\mathbf{w} \geq 0, \quad (45)$$

$$v(\mathbf{w}) \geq v(\widehat{\mathbf{w}}) - \sum_{i=1}^2 \gamma_i (y_i(\widehat{\mathbf{w}}))^{1/\gamma_i} \left(w_i^{-1/\gamma_i} - \widehat{w}_i^{-1/\gamma_i} \right) \quad \forall \mathbf{w}, \widehat{\mathbf{w}} \in W. \quad (46)$$

(i). Monotonicity of v follows directly from (46) and $\mathbf{y} \geq \mathbf{0}$. As for boundedness, if $W = \mathbb{R}_+^2$, then $v(\mathbf{w}) \geq v(\mathbf{0})$ for all $\mathbf{w} \in W$ due to monotonicity of v .

The other cases are more subtle, and we shall prove boundedness of v by contradiction. Suppose v is unbounded from below, i.e., there exists some sequence $(\mathbf{w}^n) \subset W$ such that $v(\mathbf{w}^n) \rightarrow -\infty$ as n tends to ∞ . If every limit point of this sequence is bounded away from $\partial\mathbb{R}_+^2$, which is trivially true when there exists $\underline{\mathbf{w}} > \mathbf{0}$ such that every point in W lies above $\underline{\mathbf{w}}$, then (46) gives

$$-\infty = \lim_{n \rightarrow \infty} v(\mathbf{w}^n) \geq v(\widehat{\mathbf{w}}) - \underbrace{\limsup_{n \rightarrow \infty} \sum_{i=1}^2 \gamma_i (y_i(\widehat{\mathbf{w}}))^{1/\gamma_i} \left((w_i^n)^{-1/\gamma_i} - \widehat{w}_i^{-1/\gamma_i} \right)}_{< \infty} \forall \widehat{\mathbf{w}}.$$

This is a clear contradiction.

The only case that hasn't been covered is one in which W is compact. Given such W , let $\widehat{\mathbf{w}}$ be a limit point of (\mathbf{w}^n) that lies on $\partial\mathbb{R}_+^2$. By compactness, $\widehat{\mathbf{w}}$ is an element of W . Recall our convention that $y_i(\widehat{\mathbf{w}})/\widehat{w}_i = 0$ whenever $\widehat{w}_i = 0$, thus taking a limit along the subsequence (\mathbf{w}^{n_k}) that gives $\widehat{\mathbf{w}}$, the following contradiction ensues:

$$-\infty = \lim_{n \rightarrow \infty} v(\mathbf{w}^n) = v(\widehat{\mathbf{w}}) - \underbrace{\lim_{k \rightarrow \infty} \sum_{i=1}^2 \gamma_i (y_i(\widehat{\mathbf{w}}))^{1/\gamma_i} \left((w_i^{n_k})^{-1/\gamma_i} - \widehat{w}_i^{-1/\gamma_i} \right)}_{< \infty}.$$

(ii). Using the transformation onto space X , which is discussed in Section 3.1, for every closed ball $B \subset X$,

$$v^x(\mathbf{x}) \geq v^x(\widehat{\mathbf{x}}) - \sum_{i=1}^2 \gamma_i \left(y_i \left(\widehat{x}_1^{-\gamma_1}, \widehat{x}_2^{-\gamma_2} \right) \right)^{1/\gamma_i} (x_i - \widehat{x}_i) \quad \forall \mathbf{x}, \widehat{\mathbf{x}} \in B. \quad (47)$$

The right-hand side of this equation is a collection of functions affine in \mathbf{x} , therefore v^x restricted to B is a pointwise maximum of affine functions, thus convex. By Theorem 10.4 in Rockafellar (2015), convexity implies that v^x is Lipschitz on B . Since the change of coordinates $\mathbf{x} \in B \mapsto \mathbf{w} \in \left\{ \left(x_1^{-\gamma_1}, x_2^{-\gamma_2} \right) \mid \mathbf{x} \in X \right\}$ is a continuous diffeomorphism, v is Lipschitz on the pre-image of B . Finally, since v is Lipschitz continuous for every closed ball that is contained in X , it is locally Lipschitz on the interior of W .

(iii). By Theorem 6 in on p. 296 (Rademacher Theorem) in Evans (2010), v is differentiable a.e. on W° due to local Lipschitz continuity. Then, Theorem 5 and Remark on p. 295 in Evans (2010) imply that v is weakly differentiable on W° . Since v is differentiable a.e., the envelope theorem applied to (47) together with the fact that the maximum on the right-hand side is attained at $\widehat{\mathbf{x}} = \mathbf{x}$ establishes that (9) holds at every point of differentiability.

(iv). As shown above, v is bounded from below by some \underline{v} . On the other hand, the surplus is lower than what can be obtained by pointwise optimization, i.e.,

$$\sum_{i=1}^2 \left(y_i(\mathbf{w}) - \gamma_i \left(\frac{y_i(\mathbf{w})}{w_i} \right)^{1/\gamma_i} \right) \leq \sum_{i=1}^2 (1 - \gamma_i) w_i^{1/(1-\gamma_i)} =: \overline{S}(\mathbf{w}),$$

which is integrable (with respect to f) due to (2). Combining both bounds, equation (45) gives

$$\begin{aligned} \int_W \left[\sum_{i=1}^2 \left(y_i(\mathbf{w}) - \gamma_i \left(\frac{y_i(\mathbf{w})}{w_i} \right)^{1/\gamma_i} \right) - \overline{S}(\mathbf{w}) \right] f(\mathbf{w}) d\mathbf{w} + \int_W [\underline{v} - v(\mathbf{w})] f(\mathbf{w}) d\mathbf{w} + \\ + \int_W (\overline{S}(\mathbf{w}) - \underline{v}) f(\mathbf{w}) d\mathbf{w} \geq 0, \end{aligned}$$

where we used integrability of \bar{S} and the fact that each term in the square brackets is non-positive to interchange integration and summation. It follows that v is integrable with respect to f .

(v). The argument in (iv) also implies that the surplus is integrable (with respect to f), thus (7) is equivalent to the following:

$$\int_W \sum_{i=1}^2 \left(y_i(\mathbf{w}) - \gamma_i \left(\frac{y_i(\mathbf{w})}{w_i} \right)^{1/\gamma_i} \right) f(\mathbf{w}) d\mathbf{w} \geq \int_W v(\mathbf{w}) f(\mathbf{w}) d\mathbf{w} > -\infty. \quad (48)$$

Consider the following auxiliary problem parameterized by $b \geq 0$:

$$S_1(b) := \max_{\tilde{\mathbf{y}} \geq \mathbf{0}} \int_W \sum_{i=1}^2 \tilde{y}_i(\mathbf{w}) f(\mathbf{w}) d\mathbf{w} - b \quad \text{s.t.} \quad \int_W \sum_{i=1}^2 \gamma_i \left(\frac{y_i(\mathbf{w})}{w_i} \right)^{1/\gamma_i} f(\mathbf{w}) d\mathbf{w} = b.$$

It is easy to see that $\lim_{b \rightarrow \infty} S_1(b) = -\infty$, thus (48) implies $\int_W \sum_{i=1}^2 \gamma_i \left(\frac{y_i(\mathbf{w})}{w_i} \right)^{1/\gamma_i} f(\mathbf{w}) d\mathbf{w} < \infty$. To get (v), use (9) to substitute $\nabla_i v(\mathbf{w})$ for $y_i(\mathbf{w})$.

(vi). It can be shown similarly to (v) by considering the following auxiliary problem:

$$S_2(b) := \max_{\tilde{\mathbf{y}} \geq \mathbf{0}} b - \int_W \sum_{i=1}^2 \gamma_i \left(\frac{y_i(\mathbf{w})}{w_i} \right)^{1/\gamma_i} f(\mathbf{w}) d\mathbf{w} \quad \text{s.t.} \quad \int_W \sum_{i=1}^2 \tilde{y}_i(\mathbf{w}) f(\mathbf{w}) d\mathbf{w} = b$$

and noting that $\lim_{b \rightarrow \infty} S_2(b) = -\infty$. □

C.3 Optimality conditions

In this section, we formally derive conditions that are necessary and sufficient for optimality in the relaxed problem.

Proposition 10. *(Necessity). Suppose $v^* \in \mathcal{C}^1(W^o)$, the corresponding $\boldsymbol{\lambda}^*$ is essentially bounded on $W^o \cap B_r^o$ and $(\lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w}))_{i=1,2}$ is an element of $\mathcal{W}^{1,1}(W^o \cap B_r^o)$, where B_r is a closed ball of radius $r > 0$ centered at $\mathbf{0}$, for all r . Then,*

$$\sum_{i=1}^2 \nabla_i (\lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) = (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) \quad \text{on } W^o, \quad (49)$$

$$\sum_{i=1}^2 \lambda_i^*(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) = 0 \quad \text{on } \partial W, \quad (50)$$

$$\nabla_2 \left((1 + \lambda_1^*(\mathbf{w}))^{1/(\gamma_1-1)} w_1^{\gamma_1/(\gamma_1-1)} \right) = \nabla_1 \left((1 + \lambda_2^*(\mathbf{w}))^{1/(\gamma_2-1)} w_2^{\gamma_2/(\gamma_2-1)} \right) \quad \text{on } W^o, \quad (51)$$

where all derivatives are weak, $\mathbf{n}(\mathbf{w})$ is the outward unit normal to ∂W at \mathbf{w} on the boundary and the term in the brackets in (50) is understood in the trace sense. In addition, if $\boldsymbol{\lambda}^*$ is in $\mathcal{C}^0(W) \cap \mathcal{C}^1(W^o)$, then the above partial differential equations hold in the classical sense.

(Sufficiency). Suppose $v \in \mathcal{C}^0(W) \cap \mathcal{C}^1(W^o)$, the corresponding λ is an essentially bounded element of $\mathcal{W}^{1,\infty}(W^o \cap B_r^o)$ for all $r > 0$, satisfies $\lambda \geq (\underline{\lambda}, \underline{\lambda})$ for some $\underline{\lambda} > -1$ and (49), (50), (51) hold. Then,

$$\Upsilon(v) \geq \Upsilon(\hat{v}) \quad \forall v \in \mathcal{V},$$

thus $v^* = v$.

The necessity part of this proposition shows that if v^* is sufficiently well-behaved, it is characterized by equations (49), (50), and (51). The sufficiency part looks at the converse of this statement demanding in addition λ^* and its weak derivatives to be essentially bounded (boundedness of $\nabla \lambda^*$ on compact subsets is enough). The additional requirement of $\lambda \geq (\underline{\lambda}, \underline{\lambda})$ is a sufficient condition for the function v to be an element of \mathcal{V} , and it can be substantially weakened. This result shows that if one can find sufficiently well-behaved v such that the corresponding λ solves these equations, then that v must be a solution to the relaxed problem.

Proof. Let $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ be continuously differentiable such that $\nabla v^* + \varepsilon \nabla \phi \geq 0$ on W^o for all sufficiently small $\varepsilon > 0$. It is easy to see that $v^* + \varepsilon \phi$ is in \mathcal{V} for all sufficiently small $\varepsilon > 0$, and so we will refer to such ϕ as *admissible*.

Since v^* is a solution to the relaxed problem, $\Upsilon(v^* + \varepsilon \phi) - \Upsilon(v^*) \leq 0$, where

$$\begin{aligned} \frac{\Upsilon(v^* + \varepsilon \phi) - \Upsilon(v^*)}{\varepsilon} &= \int_W \phi(\mathbf{w}) (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} - \int_W \sum_{i=1}^2 \nabla_i \phi(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) d\mathbf{w} + \\ &+ \int_W \sum_{i=1}^2 \frac{(\nabla_i v^*(\mathbf{w}) + \varepsilon \nabla_i \phi(\mathbf{w}))^{\gamma_i} - (\nabla_i v^*(\mathbf{w}))^{\gamma_i}}{\varepsilon} w_i^{1+\gamma_i} f(\mathbf{w}) d\mathbf{w}. \end{aligned}$$

Since $v^* \in \mathcal{V}$, it is increasing; therefore, its weak gradient is non-negative a.e. Monotonicity and concavity of $w \mapsto w^{\gamma_i}$ implies that if $\nabla_i \phi(\mathbf{w}) \geq 0$, then the integrand in the last line is non-negative (non-positive) for all $\varepsilon > 0$, and it is increasing (decreasing) in $\varepsilon > 0$. By the Monotone Convergence Theorem, the limit can be taken under the integral sign, thus

$$0 \geq \int_W \phi(\mathbf{w}) (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} + \int_W \sum_{i=1}^2 \nabla_i \phi(\mathbf{w}) \lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) d\mathbf{w},$$

In particular, if ϕ is compactly supported on some closed ball B and $-\phi$ is also admissible, then we must have

$$0 = \int_{W \cap B} \phi(\mathbf{w}) (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} + \int_{W \cap B} \sum_{i=1}^2 \nabla_i \phi(\mathbf{w}) \lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) d\mathbf{w}. \quad (52)$$

Due to our assumptions on λ^* , the function $\phi(\mathbf{w}) \lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w})$ belongs to $\mathscr{W}^{1,1}(W^\circ \cap B^\circ)$. As a result, the second term in (52) can be transformed by the Divergence Theorem as follows:

$$\begin{aligned} \int_{W \cap B} \nabla_i \phi(\mathbf{w}) \lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) d\mathbf{w} &= \int_{\partial W \cap B} \phi(\mathbf{w}) \lambda_i^*(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) d\sigma(\mathbf{w}) + \\ + \underbrace{\int_{W \cap \partial B} \phi(\mathbf{w}) \lambda_i^*(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) d\sigma(\mathbf{w})}_{=0} &- \int_{W \cap B} \phi(\mathbf{w}) \nabla_i (\lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) d\mathbf{w}. \end{aligned} \quad (53)$$

Finally, note that the first term in the second line in (53) is zero due to ϕ vanishing on ∂B .

Necessity, equation (49). Take a closed ball $B \subset W^\circ$. Since $\partial W \cap B = \emptyset$, the second term in (53) is zero. Combine (52) and (53) to obtain

$$0 = \int_B \phi(\mathbf{w}) \left((\alpha(\mathbf{w}) - 1) f(\mathbf{w}) - \sum_{i=1}^2 \nabla_i (\lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) \right) d\mathbf{w}, \quad (54)$$

which has to hold for all admissible $\phi \in \mathcal{C}^1(\mathbb{R}^2)$ that are compactly supported on B such that $-\phi$ is admissible.

We claim that every continuously differentiable function ϕ that vanishes on outside of B is admissible. Indeed, invert (18) to obtain

$$\nabla_i v^*(\mathbf{w}) = (1 + \lambda_i^*(\mathbf{w}))^{1/(\gamma_i - 1)} w_i^{\gamma_i/(\gamma_i - 1)}. \quad (55)$$

Since λ_i^* is essentially bounded on B , and v is continuously differentiable, $\nabla_i v$ is bounded away from 0 on B due to (55).

Taking all pieces together, since equation (54) holds for all continuously differentiable (hence for all smooth) ϕ that vanish on ∂B , the Fundamental Lemma of Calculus of Variations (Lemma 3.10 in Rindler (2018)) gives that the term in the brackets in (54) is zero a.e. on B° . Since this is true for every closed ball contained in W° , (49) follows.

Necessity, equation (50). Let $\widehat{\mathbf{w}}$ be a point on ∂W and take some closed ball B centered at $\widehat{\mathbf{w}}$. Combine (52), (53) and (54) to obtain

$$0 = \int_{\partial W \cap B} \phi(\mathbf{w}) \left(\sum_{i=1}^2 \lambda_i^*(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) \right) d\sigma(\mathbf{w}), \quad (56)$$

which has to hold for all admissible $\phi \in \mathcal{C}^1(\mathbb{R}^2)$ that are compactly supported on B such that $-\phi$ is admissible.

If $\widehat{\mathbf{w}} \in \mathbb{R}_{++}^2$, then by shrinking B (if necessary), we can ensure $B \subset \mathbb{R}_{++}^2$. By exactly the same argument as above, $\nabla_i v^*$ is bounded away from 0 on B due to (55). As a result, every

continuously differentiable ϕ supported on B is admissible, and the Fundamental Lemma of Calculus of Variations gives that the term in the brackets in (56) is zero a.e. on B^o . This proves equation (50) for the portion of ∂W that doesn't intersect $\partial\mathbb{R}_+^2$.

If $\widehat{\mathbf{w}} \in \partial\mathbb{R}_+^2$, then the argument is more subtle. Consider $\widehat{\mathbf{w}} = (0, \widehat{w}_{-i}) \neq 0$ that admits some closed ball B of radius $r > 0$ such that $B \setminus \{\widehat{\mathbf{w}}\}$ has a non-empty intersection with exactly one of the axis, i.e., the set of the form $\{\mathbf{w} \in \mathbb{R}_+^2 | w_i = 0\}$. This restriction on \mathbf{w} is without loss of generality because the set of \mathbf{w} 's that belong to $\partial\mathbb{R}_+^2 \cap \partial W$ but don't satisfy the aforementioned property has a Lebesgue measure zero (with respect to ∂W) when either $W = \mathbb{R}_{++}^2$ or W is compact and ∂W is Lipschitz.

Shrinking B (if needed), we can ensure that w_{-i} is bounded away from 0 on B ; thus, $\nabla_{-i}v^*$ is again bounded away from 0 on B . It follows that every continuously differentiable ϕ supported on B such that $\nabla_i\phi = 0$ on $B \cap \{\mathbf{w} \in \mathbb{R}_+^2 | w_i = 0\}$ is admissible. In particular, ϕ of the following kind will do:

$$\phi(\mathbf{w}) = \begin{cases} \phi_1(w_{-i}) \phi_2\left(\frac{(w_i)^2}{r^2 - (w_{-i} - \widehat{w}_{-i})^2}\right) & \text{if } \|\mathbf{w} - \widehat{\mathbf{w}}\| < r, \\ 0 & \text{otherwise,} \end{cases}$$

where ϕ_2 is continuously differentiable, non-increasing satisfying $\phi_2(w) = 1$, $\phi_2'(w) = 0$ for $w \leq 0$, $\phi_2(w) = \phi_2'(w) = 0$ for $w \geq 1$ and ϕ_1 is smooth and supported on $[\widehat{w}_{-i} - r, \widehat{w}_{-i} + r]$. It is routine to verify that every ϕ of this form is admissible and $\phi(0, w_{-i}) = \phi_1(w_{-i})$ on $\partial W \cap B$, therefore (56) simplifies to

$$0 = \int_{\widehat{w}_{-i}-r}^{\widehat{w}_{-i}+r} \phi_1(w) \lambda_{-i}^*(0, w) \gamma_{-i} w f(0, w) dw. \quad (57)$$

The Fundamental Lemma of Calculus of Variations applied to (57) gives that $\lambda_{-i}^*(0, w) \gamma_{-i} w f(0, w)$ is 0 a.e. on $[\widehat{w}_{-i} - r, \widehat{w}_{-i} + r]$.

Combining two cases together, equation (50) ensues.

Necessity, equation (51). Since $\boldsymbol{\lambda}^*$ is weakly differentiable, equation (55) implies twice weak differentiability of ∇v^* . As discussed in Chapter 5.2.1 in Evans (2010), weak derivatives commute, i.e., $\nabla_{12}v^* = \nabla_{21}v^*$, which gives (51).

Sufficiency, step 1. We shall show first that v is in \mathcal{V} . By assumption, $\nabla v \geq \mathbf{0}$ on W^o , thus v is increasing on W due to continuity. By (55) and $\lambda_i(\mathbf{w}) \geq \underline{\lambda} > -1$, thus for some $m > 0$,

$$\nabla_i v(\mathbf{w}) w_i = \left(\frac{w_i}{1 + \lambda_i(\mathbf{w})} \right)^{1/(1-\gamma_i)} \leq m w_i^{1/(1-\gamma_i)},$$

which proves (vi) in the definition of \mathcal{V} due to (2). Property (v) can be shown similarly.

Finally, extending v to all of \mathbb{R}_+^2 by 0 (if necessary), obtain that for all $\mathbf{w}, \widehat{\mathbf{w}} \in W$,

$$\begin{aligned} v(\mathbf{w}) - v(\widehat{\mathbf{w}}) &= \int_0^1 \nabla v(t\mathbf{w} + (1-t)\widehat{\mathbf{w}}) \cdot (\mathbf{w} - \widehat{\mathbf{w}}) dt \leq \\ &\leq m \int_0^1 \sum_{i=1}^2 (tw_i + (1-t)\widehat{w}_i)^{\gamma_i-1} (w_i - \widehat{w}_i) dt = m \int_{\widehat{w}_i}^{w_i} \sum_{i=1}^2 w^{\gamma_i-1} dw. \end{aligned}$$

Letting $\widehat{\mathbf{w}}$ be the point of the minimum of v , which exists due to continuity of v , (2) gives

$$\int_W (v(\mathbf{w}) - v(\widehat{\mathbf{w}})) f(\mathbf{w}) d\mathbf{w} \leq m \sum_{i=1}^2 (1 - \gamma_i) \left(\mathbb{E} \left[w_i^{1/(1-\gamma_i)} \right] - \widehat{w}_i^{1/(1-\gamma_i)} \right) < \infty.$$

Sufficiency, step 2. In order to show optimality of v , we shall first rewrite the functional $\Upsilon(\widehat{v})$ as a function of derivatives of \widehat{v} . By equation (49),

$$\int_{W \cap B_r} \widehat{v}(\mathbf{w}) (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} = \int_{W \cap B_r} \widehat{v}(\mathbf{w}) \sum_{i=1}^2 \nabla_i (\lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) d\mathbf{w}. \quad (58)$$

Since λ is an element of $\mathscr{W}^{1,\infty}(W^o \cap B_r^o)$ and $(\widehat{v}(\mathbf{w}) \gamma_i w_i f(\mathbf{w}))_{i=1,2}$ is in $\mathscr{W}^{1,1}(W^o \cap B_r^o)$ due to (iv) and (vi) in the definition of \mathscr{V} , $(\widehat{v}(\mathbf{w}) \lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w}))_{i=1,2}$ is an element of $\mathscr{W}^{1,1}(W^o \cap B_r^o)$. By the Divergence Theorem,

$$\begin{aligned} &\int_{W \cap B_r} \sum_{i=1}^2 \nabla_i (\widehat{v}(\mathbf{w}) \lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) d\mathbf{w} =, \\ &= - \int_{W \cap \partial B_r} \widehat{v}(\mathbf{w}) \left(\sum_{i=1}^2 \lambda_i(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) \right) d\sigma(\mathbf{w}) \end{aligned} \quad (59)$$

where we used equation (50) to show that the integral on the boundary of W must be zero.

By the Chain Rule for weak derivatives,

$$\begin{aligned} \nabla_i (\widehat{v}(\mathbf{w}) \lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) &= \nabla_i (\widehat{v}(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) \lambda_i(\mathbf{w}) + \widehat{v}(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) \nabla_i \lambda_i(\mathbf{w}) = \\ &= \nabla_i \widehat{v}(\mathbf{w}) \lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) + \widehat{v}(\mathbf{w}) \nabla_i (\lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w})). \end{aligned}$$

Then, since $\widehat{v} \in \mathscr{V}$, $\gamma_i w_i f(\mathbf{w}) \in \mathscr{W}^{1,1}(W^o \cap B_r^o)$ and $\lambda_i(\mathbf{w}) \in \mathscr{W}^{1,\infty}(W^o \cap B_r^o)$, the brackets on the left-hand side in (59) can be unpacked to re-write (58) as follows:

$$\begin{aligned} \int_{W \cap B_r} \widehat{v}(\mathbf{w}) (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} &= - \int_{W \cap B_r} \sum_{i=1}^2 \nabla_i \widehat{v}(\mathbf{w}) \lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) d\mathbf{w} + \\ &+ \int_{W \cap \partial B_r} \widehat{v}(\mathbf{w}) \left(\sum_{i=1}^2 \lambda_i(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) \right) d\sigma(\mathbf{w}). \end{aligned} \quad (60)$$

We now study the limit of (60) as $r \rightarrow \infty$. Consider the expression on the left-hand side of equation (60). Since α is bounded, the Dominated Convergence Theorem applies, thus

$$\int_{W \cap B_r} \widehat{v}(\mathbf{w})(\alpha(\mathbf{w}) - 1)f(\mathbf{w})d\mathbf{w} \xrightarrow{r \rightarrow \infty} \int_W \widehat{v}(\mathbf{w})(\alpha(\mathbf{w}) - 1)f(\mathbf{w})d\mathbf{w}. \quad (61)$$

Consider the second term on the right-hand side of (60). First of all, note that ∂B_r is the set of points such $w_1^2 + w_2^2 = r^2$, therefore the outward unit normal is given by $n_i(\mathbf{w}) = \frac{w_i}{r} \geq 0$. Since $\boldsymbol{\lambda}$ is essentially bounded on W^o , we have the following chain of inequalities:

$$\begin{aligned} & \left| \int_{W \cap \partial B_r} \widehat{v}(\mathbf{w}) \left(\sum_{i=1}^2 \lambda_i(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) \right) d\sigma(\mathbf{w}) \right| \leq \\ & \leq \int_{W \cap \partial B_r} |\widehat{v}(\mathbf{w})| \left| \sum_{i=1}^2 \lambda_i(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) \right| d\sigma(\mathbf{w}) \leq \\ & \leq \int_{W \cap \partial B_r} |\widehat{v}(\mathbf{w})| \sum_{i=1}^2 |\lambda_i(\mathbf{w}) \gamma_i| |w_i n_i(\mathbf{w}) f(\mathbf{w})| d\sigma(\mathbf{w}) \leq \\ & \leq \bar{\lambda} \int_{W \cap \partial B_r} |\widehat{v}(\mathbf{w})| f(\mathbf{w}) r d\sigma(\mathbf{w}). \end{aligned} \quad (62)$$

where $\bar{\lambda}$ is an essential upper bound on $\max_i |\lambda_i \gamma_i|$. To obtain the last inequality in equation (62), we substituted for $\mathbf{n}(\mathbf{w})$, i.e., $\sum_{i=1}^2 w_i n_i(\mathbf{w}) = r$. The expected value of $|\widehat{v}|$ in the polar coordinates reads as follows (see the Coarea Formula or Theorem 4 in Appendix C in Evans (2010)):

$$\int_W |\widehat{v}(\mathbf{w})| f(\mathbf{w}) d\mathbf{w} = \int_0^\infty \frac{\int_{W \cap \partial B_r} |\widehat{v}(\mathbf{w})| f(\mathbf{w}) r d\sigma(\mathbf{w})}{r} dr. \quad (63)$$

Since $\widehat{v} \in \mathcal{V}$, the expected value of $|\widehat{v}|$ is finite, thus (63) implies that the denominator on the right-hand side goes to zero as $r \rightarrow \infty$. Therefore, by (62),

$$\int_{W \cap \partial B_r} \widehat{v}(\mathbf{w}) \left(\sum_{i=1}^2 \lambda_i(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) \right) d\sigma(\mathbf{w}) \xrightarrow{r \rightarrow \infty} 0. \quad (64)$$

Finally, since $\widehat{v} \in \mathcal{V}$, condition (vi) in the definition of \mathcal{V} gives that the first term on the right-hand side of (60) is bounded by an integrable function, so the Dominated Convergence Theorem applies and

$$\int_{W \cap B_r} \sum_{i=1}^2 \nabla_i \widehat{v}(\mathbf{w}) \lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) d\mathbf{w} \xrightarrow{r \rightarrow \infty} \int_W \sum_{i=1}^2 \nabla_i \widehat{v}(\mathbf{w}) \lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) d\mathbf{w}. \quad (65)$$

Taking all pieces together, using (60), (61), (64) and (65) we established

$$\int_W \widehat{v}(\mathbf{w})(\alpha(\mathbf{w}) - 1)f(\mathbf{w})d\mathbf{w} = - \int_W \sum_{i=1}^2 \nabla_i \widehat{v}(\mathbf{w}) \lambda_i(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) d\mathbf{w}.$$

Sufficiency, step 3. As shown in step 2, the value of $\Upsilon(\hat{v})$ can be re-written as

$$\begin{aligned}\Upsilon(\hat{v}) &= \int_W \left[\hat{v}(\mathbf{w}) (\alpha(\mathbf{w}) - 1) + \sum_{i=1}^2 \left(w_i^{1+\gamma_i} \left(\nabla_i \hat{v}(\mathbf{w}) \right)^{\gamma_i} - \gamma_i w_i \nabla_i \hat{v}(\mathbf{w}) \right) \right] f(\mathbf{w}) d\mathbf{w} = \\ &= \int_W \left[\sum_{i=1}^2 \left(w_i^{1+\gamma_i} \left(\nabla_i \hat{v}(\mathbf{w}) \right)^{\gamma_i} - \gamma_i w_i (1 + \lambda_i(\mathbf{w})) \nabla_i \hat{v}(\mathbf{w}) \right) \right] f(\mathbf{w}) d\mathbf{w}.\end{aligned}$$

Note that the term in the square brackets is strictly concave in $\nabla_i \hat{v}(\mathbf{w})$ and (pointwise) maximized at $(\nabla_i \hat{v}(\mathbf{w}))^{1/(\gamma_i-1)} w_i^{\gamma_i} = 1 + \lambda_i(\mathbf{w})$, which is nothing but the definition of ∇v . By construction, λ satisfies equation (22) and ∇v is a (weak) gradient of $v \in \mathcal{V}$. Therefore, $\Upsilon(\hat{v}) \leq \Upsilon(v)$. \square

We end this section by pointing out a certain well-known equivalence between v^* and λ^* . One can think equivalently of equations (49), (50) and (51) either as a second-order partial differential equation describing the solution to the relaxed problem v^* , or as a system of joint first order partial differential equations describing the optimal λ^* implied by that v^* . Formally: if $\left((1 + \lambda_i^*(\mathbf{w}))^{1/(\gamma_i-1)} w_i^{\gamma_i/(\gamma_i-1)} \right)_{i=1,2}$ is an element of $\mathcal{C}^1(W)$ and (51) holds. Then, there exists a unique (up to a constant) function $v \in \mathcal{C}^1(W^o)$ such that equation (18) holds for these v and λ . This function can be then extended by uniform continuity to W due to boundedness of ∇v on compact subsets of W .

C.4 On the first-order approach

C.4.1 Proof of Proposition 1

Proof. For v , which is a function of one variable, we use ∇v to denote the (weak) derivative of v with respect to that variable.

(1D). Take $\tilde{v} \in \mathcal{V}$. Since a path of integration is unique in the uni-dimensional case, we must have

$$\begin{aligned}\int_{\underline{w}}^{\overline{w}} \tilde{v}(w) (\tilde{\alpha}(w) - 1) g(w) dw &= \int_{\underline{w}}^{\overline{w}} (\tilde{v}(w) - \tilde{v}(\underline{w})) (\tilde{\alpha}(w) - 1) g(w) dw = \\ &= \int_{\underline{w}}^{\overline{w}} \nabla \tilde{v}(w) (G(w) - A(w)) dw,\end{aligned}$$

where the first equality follows from the normalization of α and the second equality is implied by integration by parts with $A(t) := \int_{\underline{w}}^t \tilde{\alpha}(w) g(w) dw$. Thus, $\Upsilon(\tilde{v})$ can be written as

$$\Upsilon(\tilde{v}) = \int_{\underline{w}}^{\overline{w}} \left((\nabla \tilde{v}(w))^{\gamma} w^{1+\gamma} - \gamma w \nabla \tilde{v}(w) \left(1 + \frac{A(w) - G(w)}{\gamma w g(w)} \right) \right) g(w) dw.$$

Pointwise optimization of this equation yields

$$(\nabla \tilde{v}(w))^{\gamma-1} w^\gamma = 1 + \frac{A(w) - G(w)}{\gamma w g(w)} = 1 + \tilde{\lambda}(w). \quad (66)$$

By construction, the planner's objective under \tilde{v} is higher than under any other alternative function in \mathcal{V} . Since $\tilde{\alpha}$ is decreasing, A is first-order stochastically dominated by G , thus $\tilde{\lambda} \geq 0$. By the same argument as in the proof of the sufficiency part in 10, $\tilde{v} \in \mathcal{V}$, thus $v^* = \tilde{v}$.

To find conditions for validity of the first-order approach, we use Lemma 3. The space X is convex, and by the chain rule, we have $\nabla v^{x,*}(x) = -\gamma \nabla v^*(x^{-\gamma}) x^{-(\gamma+1)}$. Monotonicity of $\nabla v^{x,*}$ on the interior of X is equivalent to convexity of $v^{x,*}$. The claim follows from $\nabla v^{x,*}(x) \propto x \cdot (1 + \tilde{\lambda}(x^{-\gamma}))$ due to (66).

(2D). Consider $v \in \mathcal{V}$ that is additively separable, i.e., $v(\mathbf{w}) = \tilde{v}_1(w_1) + \tilde{v}_2(w_2)$ for some functions $(\tilde{v}_1, \tilde{v}_2)$. If types are independently and identically distributed, then $f(\mathbf{w}) = g(w_1)g(w_2)$. Following the same steps as in the proof of the previous part,

$$\begin{aligned} \int_W v(\mathbf{w}) (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} &= \int_{\underline{w}}^{\bar{w}} \sum_{i=1}^2 \tilde{v}_i(w_i) \frac{1}{2} (\tilde{\alpha}(w_i) - 1) g(w_i) dw_i = \\ &= \sum_{i=1}^2 \int_{\underline{w}}^{\bar{w}} \nabla \tilde{v}_i(w_i) \frac{1}{2} (G(w_i) - A(w_i)) dw_i. \end{aligned}$$

It follows that the value of $\Upsilon(v)$ for an additively separable v can be re-written as

$$\Upsilon(v) = \sum_{i=1}^2 \int_{\underline{w}_i}^{\bar{w}_i} \left((\nabla \tilde{v}_i(w_i))^\gamma w_i^{1+\gamma} - \gamma w \nabla \tilde{v}_i(w_i) \left(1 + \frac{1}{2} \frac{A(w_i) - G(w_i)}{\gamma w_i g(w_i)} \right) \right) g(w_i) dw_i.$$

Pointwise optimization of this equation yields

$$(\nabla \tilde{v}_i(w_i))^{\gamma-1} w_i^\gamma = 1 + \frac{1}{2} \frac{A(w_i) - G(w_i)}{\gamma w_i g(w_i)} = 1 + \frac{1}{2} \tilde{\lambda}(w_i). \quad (67)$$

Since both limits $\lim_{w \rightarrow \underline{w}} \frac{G(w)}{wg(w)}$ and $\lim_{w \rightarrow \bar{w}} \frac{1-G(w)}{wg(w)}$ are finite, the vector field $(\frac{1}{2}\tilde{\lambda}(w_1), \frac{1}{2}\tilde{\lambda}(w_2))$ is a non-negative, bounded element of $\mathcal{W}^{1,\infty}(W^o)$ and satisfies (18), (22). In view of the sufficiency part of Proposition 10, we need to show that this vector field satisfies (49), (50). By definition, $A(\underline{w}) = G(\underline{w}) = 0$ and $A(\bar{w}) = G(\bar{w}) = 1$ that proves (50). To see the remaining equation, observe

$$\sum_{i=1}^2 \frac{\nabla_i \left(\frac{1}{2} \tilde{\lambda}(w_i) \gamma_i w_i g(w_1) g(w_2) \right)}{g(w_1) g(w_2)} = \sum_{i=1}^2 \frac{\tilde{\alpha}(w_i) - 1}{2} = \alpha(\mathbf{w}) - 1.$$

By the same argument as in the proof of the sufficiency part in 10, $v \in \mathcal{V}$, thus $v^* = v$.

To establish conditions for validity of the first-order approach, we again use Lemma 3. Observe that $v^{x,*}(\mathbf{x}) = v_1^*(x_1^{-\gamma}) + v_2^*(x_2^{-\gamma})$, but then $v^{x,*}$ is convex on X if and only if each $\nabla v_i^{x,*}$ is non-decreasing on the interior. It is easy to see that $\nabla_i v^{x,*}(x) \propto x \cdot \left(1 + \frac{1}{2}\tilde{\lambda}(x^{-\gamma})\right)$ due to (67).

The fact that the first-order approach is more likely to hold in the bi-dimensional model than in the uni-dimensional setting can be seen from

$$x \cdot \left(1 + \frac{1}{2}\tilde{\lambda}(x^{-\gamma})\right) - \hat{x} \cdot \left(1 + \frac{1}{2}\tilde{\lambda}(\hat{x}^{-\gamma})\right) \geq \frac{x}{2} \cdot \left(1 + \tilde{\lambda}(x^{-\gamma})\right) - \frac{\hat{x}}{2} \cdot \left(1 + \tilde{\lambda}(\hat{x}^{-\gamma})\right) \quad \forall x \geq \hat{x}.$$

□

C.4.2 Rawlsian planner (Section 3.1)

In this section, we look at the uni-dimensional model with the Rawlsian planner in more detail. Formally, the Rawlsian weights $\tilde{\alpha}^R$ corresponds to a linear functional on \mathcal{V} defined by

$$\int_{\underline{w}}^{\bar{w}} \tilde{v}(w) \tilde{\alpha}^R(w) g(w) dw = \tilde{v}(\underline{w}).$$

This functional is well-defined and is proportional to the Dirac delta function at $w = \underline{w}$ provided $g(\underline{w}) > 0$.

Even though $\tilde{\alpha}^R$ is not a continuous bounded function, optimal taxes can be obtained by the integration by parts as in the proof in Section C.4.1. Since $\int_{\underline{w}}^t \tilde{\alpha}^R(w) g(w) ds \equiv 1$, the corresponding $\tilde{\lambda}^{R,*}$ is given by

$$\tilde{\lambda}^{R,*}(w) = \frac{1 - G(w)}{\gamma w g(w)}.$$

The necessary and sufficient condition in Proposition 1 still applies here when $\underline{w} > 0$ so that $\tilde{\lambda}^{R,*}$ is finite at the bottom, and the set of G for which the FOA works is non-empty, i.e., $G(w) = 1 - (\underline{w}/w)^{-a}$ for some $a > 1$ will do.

We now show the first-order approach is necessarily violated irrespective of G for continuous and bounded Pareto weights that are “close” to $\tilde{\alpha}^R$. Specifically, let $(\tilde{\alpha}^n)$ be a sequence such that for every function $\tilde{v} \in \mathcal{V}$,

$$\int_{\underline{w}}^{\bar{w}} \tilde{v}(w) \tilde{\alpha}^n(w) g(w) dw \rightarrow_{n \rightarrow \infty} \tilde{v}(\underline{w}). \quad (68)$$

Let $\tilde{\lambda}^{n,*}$ be the optimum as defined in the proof in Section C.4.1, that is

$$\tilde{\lambda}^{n,*}(t) = \frac{\int_{\underline{w}}^t \tilde{\alpha}^n(w) g(w) dw - G(w)}{\gamma w g(w)}.$$

Clearly, \tilde{v} that equals to 1 for all $w \geq \hat{w}$ for some $\hat{w} \in (\underline{w}, \bar{w}]$ and 0 otherwise is an element of \mathcal{V} . As a result, (68) implies that $\tilde{\lambda}^{n,*}$ pointwise converges to $\tilde{\lambda}^{R,*}$ on $(\underline{w}, \bar{w}]$. But, since $\tilde{\lambda}^{n,*}(w) = 0$ for all n when $\underline{w} > 0$, the condition in Proposition 1 is always violated for sufficiently large n .

C.4.3 Proof of Proposition 2

Proof. Step 1. Since validity of the first-order approach is easier to check in the transformed space, we shall work mainly with it. Let (f^x, α^x) be transformations of density and Pareto weights on space X . Since W is compactly contained in \mathbb{R}_{++}^2 , and $\mathbf{w} \mapsto \mathbf{x}$ is a continuous diffeomorphism, the sequence $(f^x, \alpha^{x,\varepsilon})$ is a subset of $\mathcal{C}^{1,a} \times \mathcal{C}^{0,a}$.

Given some transformations of utility and Pareto weights $(v^x, \alpha^x) \in \mathcal{C}^{2,a} \times \mathcal{C}^{0,a}$ with $\nabla v^x < \mathbf{0}$ on X define $\Xi(v^x, \alpha^x) \in \mathcal{C}^{0,a} \times \mathcal{C}^{1,a}$ as follows:

$$\Xi_1(v^x, \alpha^x)(\mathbf{x}) := \sum_{i=1}^2 \nabla_i \left(\left(\left(-\frac{1}{\gamma_i} \nabla_i v^x(\mathbf{x}) \right)^{\gamma_i-1} - x_i \right) f^x(\mathbf{x}) \right) - (\alpha^x(\mathbf{x}) - 1) f^x(\mathbf{x}) \quad \forall \mathbf{x} \in X^o, \quad (69)$$

$$\Xi_2(v^x, \alpha^x)(\mathbf{x}) := \sum_{i=1}^2 \left(\left(-\frac{1}{\gamma_i} \nabla_i v^x(\mathbf{x}) \right)^{\gamma_i-1} - x_i \right) f^x(\mathbf{x}) n_i^x(\mathbf{x}) \quad \forall \mathbf{x} \in \partial X, \quad (70)$$

where \mathbf{n}^x is the outward unit normal to the boundary ∂X at \mathbf{x} . Here $\mathcal{C}^{1,a}(\partial X)$ stays for the space of restrictions of functions in $\mathcal{C}^{1,a}(X)$ to the boundary of X .

This operator Ξ is just a different way to record two conditions that are necessary for optimality, i.e., equations (49) and (50). To see it, note that $x_i = w_i^{-1/\gamma_i}$, thus $\frac{dx_i}{dw_i} = -\frac{1}{\gamma_i} w_i^{-(1+\gamma_i)/\gamma_i} = -\frac{1}{\gamma_i} x_i^{\gamma_i+1}$. Solve for $\nabla_i v^*$ as a function of $\nabla_i v^{x,*}$ and use (18) to obtain

$$\lambda_i^*(\mathbf{w}) = \left(\left(-\frac{1}{\gamma_i} \nabla_i v^{x,*}(\mathbf{x}) \right)^{\gamma_i-1} - 1 \right) x_i^{-1}.$$

By definition, $f(\mathbf{w}) = f^x(\mathbf{x}) \prod_{i=1}^2 \frac{1}{\gamma_i} x_i^{\gamma_i+1}$, thus

$$\nabla_i (\lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) = \nabla_i \left(\left(\left(-\frac{1}{\gamma_i} \nabla_i v^{x,*}(\mathbf{x}) \right)^{\gamma_i-1} - x_i \right) f^x(\mathbf{x}) \right) \prod_{j=1}^2 \frac{1}{\gamma_j} x_j^{\gamma_j+1}, \quad (71)$$

where on the left-hand (right-hand) side ∇_i stays for a derivative with respect to w_i (resp. x_i). Substitute (71) into (49) and divide each side by $\prod_{i=1}^2 \frac{1}{\gamma_i} x_i^{\gamma_i+1}$ to obtain that $v^{x,*}$ is a solution to $\Xi_1(\cdot, \alpha^{x,\varepsilon}) = 0$. Similarly, $v^{x,*}$ solves $\Xi_2(\cdot, \alpha^{x,\varepsilon}) = 0$.

By Proposition 10, for every $\varepsilon \geq 0$, if the solution $v^{\varepsilon,*}$ to the relaxed problem is sufficiently well-behaved, then (49) and (50) hold or, equivalently, the function $v^{x,\varepsilon,*}$ is a root of $\Xi(\cdot, \alpha^{x,\varepsilon})$.

Conversely, every sufficiently well-behaved root of $\Xi(\cdot, \alpha^{x,\varepsilon})$ defines the solution to the relaxed problem.

Step 2. For $\varepsilon = 0$, the Pareto weights are utilitarian, i.e., $\alpha^0 \equiv 1$, and $v^{x,0,*}$ can be found in a closed form from (69) and (70),

$$\left(-\frac{1}{\gamma_i} \nabla_i v^{x,0,*}(\mathbf{x})\right)^{\gamma_i-1} - x_i \equiv 0. \quad (72)$$

This suggests studying behavior of roots of the operator Ξ in a neighborhood of the utilitarian objective. Specifically, we shall establish existence of twice continuously differentiable solutions to $\Xi(\cdot, \alpha^x)$ and their continuity as a function of Pareto weights for economies with α^x that is close enough to α^0 . The main instrument is the Implicit Function Theorem, and to state it in our context, we need certain Banach spaces and auxiliary notations.

Let \mathcal{U} be the subspace of functions $v^x \in \mathcal{C}^{2,a}$ such that $\int_X v^x(\mathbf{x}) f^x(\mathbf{x}) d\mathbf{x} = 0$, \mathcal{A} be the subspace of functions $\alpha^x \in \mathcal{C}^{0,a}$ such that $\int_X \alpha^x(\mathbf{x}) f^x(\mathbf{x}) d\mathbf{x} = 1$ and Φ be the subspace of functions $\phi = (\phi_1, \phi_2) \in \mathcal{C}^{0,a}(X) \times \mathcal{C}^{1,a}(\partial X)$ such that

$$\int_X \phi_1(\mathbf{x}) d\mathbf{x} = \int_{\partial X} \phi_2(\mathbf{x}) d\sigma(\mathbf{x}).$$

Fix some closed ball B in $\mathcal{U} \times \mathcal{A}$ centered at $(v^{0,x,*}, \alpha^0)$ and small enough so that for some $m > 0$, we have

$$\max_{\mathbf{x} \in X} \max_{i=1,2} -\frac{1}{\gamma_i} \nabla_i v^x(\mathbf{x}) \geq m \quad \forall v^x \in B,$$

which is clearly possible due to compactness of $X \subset \mathbb{R}_{++}^2$ and (72). It follows that $\Xi : B^o \rightarrow \Phi$ is well-defined.

We claim that Ξ is continuously (Frechet) differentiable on B^o . To see it, observe that for every $v^x, \widehat{v}^x \in B$ and $n \geq 1$,

$$\begin{aligned} & \left(-\frac{1}{\gamma_i} \nabla_i v^x(\mathbf{x})\right)^{\gamma_i-n} - \left(-\frac{1}{\gamma_i} \nabla_i \widehat{v}^x(\mathbf{x})\right)^{\gamma_i-n} = \\ & = \left(\frac{n}{\gamma_i} - 1\right) \left(-\frac{1}{\gamma_i} \nabla_i \widehat{v}^x(\mathbf{x})\right)^{\gamma_i-(n+1)} (\nabla_i v^x(\mathbf{x}) - \nabla_i \widehat{v}^x(\mathbf{x})) + o(|\nabla_i v^x(\mathbf{x}) - \nabla_i \widehat{v}^x(\mathbf{x})|) \end{aligned} \quad (73)$$

and

$$\begin{aligned} & \nabla_i \left(\left(-\frac{1}{\gamma_i} \nabla_i v^x(\mathbf{x})\right)^{\gamma_i-n} - \left(-\frac{1}{\gamma_i} \nabla_i \widehat{v}^x(\mathbf{x})\right)^{\gamma_i-n} \right) = \\ & = \nabla_i \left(\left(\frac{n}{\gamma_i} - 1\right) \left(-\frac{1}{\gamma_i} \nabla_i \widehat{v}^x(\mathbf{x})\right)^{\gamma_i-(n+1)} (\nabla_i v^x(\mathbf{x}) - \nabla_i \widehat{v}^x(\mathbf{x})) \right) + o(|\nabla_i v^x(\mathbf{x}) - \nabla_i \widehat{v}^x(\mathbf{x})|), \end{aligned} \quad (74)$$

where we used compactness of $X \subset \mathbb{R}_{++}^2$, twice continuous differentiability and boundedness of $\nabla v^x, \nabla \widehat{v}^x$. Then, continuous (Frechet) differentiability of Ξ on B^o is due to (73), (74) and

our restrictions on f . In particular, the partial derivative with respect to v^x at $(v^{x,0,*}, \alpha^0)$ is given by

$$D\Xi_1(v^x)(\mathbf{x}) := \sum_{i=1}^2 \nabla_i \left(x_i^{(\gamma_i+1)/\gamma_i} \nabla_i v^x(\mathbf{x}) f^x(\mathbf{x}) \right) \quad \forall \mathbf{x} \in X, \quad (75)$$

$$D\Xi_2(v^x)(\mathbf{x}) := \sum_{i=1}^2 x_i^{(\gamma_i+1)/\gamma_i} \nabla_i v^x(\mathbf{x}) n_i^x(\mathbf{x}) f^x(\mathbf{x}) \quad \forall \mathbf{x} \in \partial X, \quad (76)$$

which is a mapping from \mathcal{U} to Φ .

Step 3. In view of Theorem 4.E in Zeidler (2012), if $D\Xi : \mathcal{U} \rightarrow \Phi$ is a bijection, then the equation $\Xi(\cdot, \alpha^x) = 0$ admits a unique solution on some closed ball B_α in \mathcal{A} , which is centered at α^0 . Moreover, this solution is a continuous function of α on B_α . Showing that $D\Xi$ is bijective amounts to solving a second-order (uniformly) elliptic partial differential equation with Neumann boundary conditions. Nardi (2015) pointed out that classical existence results for an oblique derivative problem cannot be directly applied to the Neumann problem like ours; thus, the latter needs a special treatment. We aren't aware of a textbook treatment of this existence result, and therefore follow Nardi, who suggested applying the Fredholm Alternative to a certain compact operator to show that $D\Xi$ is invertible on Φ . For completeness, we will sketch his argument.

To begin, consider the auxiliary problem:

$$D\Xi(v^x) - (v^x, 0) = \phi, \quad (77)$$

where, of course, ϕ is an element of Φ . This auxiliary problem admits a unique solution in $\mathcal{C}^{2,a}$ by Theorem 5.1 in Nardi (2015), because the homogeneous problem, i.e., $\phi \equiv \mathbf{0}$, admits only a trivial solution. The latter follows directly from Hopf's Maximum Principle. See, for example, Theorem 3.6 in Gilbarg et al. (1977).

Write $V(\phi)$ for the solution to (77) as a function ϕ . The solution to the original problem can be identified with $V(\phi)$ as follows: for every $v^x \in \mathcal{U}$,

$$D\Xi(v^x) = \phi \iff v^x - V(-v^x, 0) = V^x(\phi). \quad (78)$$

Moreover, it is easy to see that $V(\phi)$ is an element of \mathcal{U} , that is

$$\begin{aligned} \int_X V^x(\phi)(\mathbf{x}) d\mathbf{x} &= \int_X \sum_{i=1}^2 \nabla_i \left(x_i^{(\gamma_i+1)/\gamma_i} \nabla_i v^x(\mathbf{x}) f^x(\mathbf{x}) \right) d\mathbf{x} - \int_X \phi_1(\mathbf{x}) d\mathbf{x} = \\ &= \int_{\partial X} \sum_{i=1}^2 x_i^{(\gamma_i+1)/\gamma_i} \nabla_i v^x(\mathbf{x}) n_i^x(\mathbf{x}) f^x(\mathbf{x}) d\sigma(\mathbf{x}) - \int_X \phi_1(\mathbf{x}) d\mathbf{x} = \\ &= \int_{\partial X} \phi_2(\mathbf{x}) d\sigma(\mathbf{x}) - \int_X \phi_1(\mathbf{x}) d\mathbf{x} = 0, \end{aligned}$$

where the first line is due to (77), the second line follows from the Divergence Theorem and the last line is due to (77) again. As a result, $v^x \mapsto V(-v^x, 0)$ is a well-defined operator that maps \mathcal{U} to itself. Since the homogenous problem $D\Xi(v^x) = 0$ admits only a trivial solution, which is due to Hopf's Maximum Principle, the Fredholm Alternative (Theorem 5.B in Zeidler (2012)) implies that (78) has exactly one solution when $V(-v^x, 0)$ is compact.

Step 4. We shall now show compactness of $V(-v^x, 0)$ following Nardi (2015). First of all, we need to devise a certain bound.

Let $v^x = V(\phi)$, i.e., it solves (77), and suppose that $v^x(\mathbf{x}) \geq |v^x(\hat{\mathbf{x}})| \geq 0$ for all $\hat{\mathbf{x}} \in X$. If \mathbf{x} is in the interior, then optimality requires $\nabla_i v^x(\mathbf{x}) = 0$ and $\nabla_{ii} v^x(\mathbf{x}) \leq 0$. Conversely, if \mathbf{x} is on the boundary, then optimality requires $\nabla_1 v^x(\mathbf{x}) n_1(\mathbf{x}) - \nabla_2 v^x(\mathbf{x}) n_2(\mathbf{x}) = 0$. But, since $\nabla_1 v^x(\mathbf{x}) n_1(\mathbf{x}) + \nabla_2 v^x(\mathbf{x}) n_2(\mathbf{x}) = 0$, which is the boundary condition in (77), the gradient is identically 0 at \mathbf{x} . By Lemma 3.3 in Nardi (2015), we again have $\nabla_{ii} v^x(\mathbf{x}) \leq 0$. In either case, we obtain the following conclusion:

$$\|v\|_{\mathcal{C}^0} = v(\mathbf{x}) = D\Xi_1(v^x)(\mathbf{x}) - \phi_1(\mathbf{x}) \leq -\phi_1(\mathbf{x}) = |\phi_1(\mathbf{x})| \leq \|\phi_1\|_{\mathcal{C}^0}, \quad (79)$$

where the inequality is due to $D\Xi_1(v^x)(\mathbf{x}) = x_i^{(\gamma_i+1)/\gamma_i} \nabla_{ii} v^x(\mathbf{x}) f^x(\mathbf{x}) \leq 0$ and the next equality follows from $\phi_1(\mathbf{x}) = D\Xi_1(v^x)(\mathbf{x}) - v^x(\mathbf{x}) \leq 0$.

The reader can verify that the bound in (79) holds if $\|v^x\|_{\mathcal{C}^{0,a}}$ is attained at some point of minimum. As a result, (79) and the standard estimate (Theorem 6.30 in Gilbarg et al. (1977)) shows that there exists a number M , which is independent of v^x , such that for every $v^x \in \mathcal{U}$,

$$\|V(-v^x, 0)\|_{\mathcal{C}^{2,a}} \leq M \|v^x\|_{\mathcal{C}^{0,a}}. \quad (80)$$

Now, take some bounded sequence $(v^{x,n}) \subset \mathcal{U}$ in the $\mathcal{C}^{2,a}$ -norm. By (80), the corresponding sequence $(V(-v^{x,n}, 0))$ is bounded in the $\mathcal{C}^{2,a}$ -norm. Let $\frac{2}{2-a} < p < 2$, then clearly $(V(-v^{x,n}, 0))$ is a subset of the space $\mathcal{W}^{2,p}$ of twice weakly differentiable functions with derivatives that are p -integrable. By Theorem 6 in Section 5.6.3 of Evans (2010), $\mathcal{W}^{2,p}$ can be continuously embedded into the Hölder space $\mathcal{C}^{0,2-2/p}$. Since $a < 2 - 2/p$, the space $\mathcal{C}^{0,2-2/p}(X)$ can be compactly embedded into $\mathcal{C}^{0,a}(X)$, see for example Section 5.6.2 in Zeidler (2013).

Combining all facts, we obtain that $(V(-v^{x,n}, 0))$ admits a subsequence that converges in the $\mathcal{C}^{0,a}$ -norm, which proves compactness of $v^x \mapsto V(-v^x, 0)$ and verifies the condition for the Implicit Function Theorem.

Step 5. We now have all ingredients to prove that the FOA works for all sufficiently small $\varepsilon > 0$. First of all, we established that equations (69) and (70) have exactly one solution,

which turns out to be in $\mathcal{C}^{2,a}$. Since $W \subset \mathbb{R}_{++}^2$ is compact, the sufficiency part of Proposition 10 applies here, and our constructed solutions are, in fact, optimal in the relaxed problem.

As a consequence of the Implicit Function Theorem (Theorem 4.E in Zeidler (2012)), the sequence of solutions $(v^{x,\varepsilon,*})$ converges to $v^{x,0,*}$ in the $\mathcal{C}^{2,a}$ -norm. It is easy to see from (72) that the latter is strongly convex, i.e., $\nabla^2 v^{x,0} - \mu I$ is non-negative definite for some $\mu > 0$, where I is a 2×2 identity matrix and $\nabla^2 v^{x,0}$ is the Hessian of $v^{x,0}$. Since $\|v^{x,\varepsilon,*} - v^{x,0,*}\|_{\mathcal{C}^{2,a}} \rightarrow 0$, the elements of $(v^{x,\varepsilon,*})$ are strongly convex; thus, the FOA holds due to Lemma 3, for small $\varepsilon > 0$. \square

D Proofs for Section 4

Throughout this section, it is assumed that the first-order approach is valid, λ^* is in $\mathcal{C}^0(W) \cap \mathcal{C}^1(W^\circ)$, bounded on $W = \mathbb{R}_+^2$, and its derivatives are bounded as well.

D.1 Proof of Proposition 3

Proof. Observe that $\lambda^{*sg,*}$ is the same as $\tilde{\lambda}$ used in Proposition 1. In the proof of the first part of this proposition, we established that $\tilde{\lambda}$ corresponds to the solution to the relaxed problem. \square

D.2 Proof of Corollary 4

Proof. To apply the sufficiency part of Proposition 10, we need to show that the vector field in the claim satisfies (49), (50). Indeed,

$$\sum_{i=1}^2 \frac{\nabla_i \left(\beta_i \frac{1 - \mathbb{E}[\tilde{\alpha}_i(\tilde{w}_i) | \tilde{w}_i \geq w_i]}{\gamma_i \bar{\theta}_i(w_i)} \gamma_i w_i g_1(w_1) g_2(w_2) \right)}{g_1(w_1) g_2(w_2)} = \sum_{i=1}^2 \beta_i (\tilde{\alpha}_i(w_i) - 1) = \alpha(\mathbf{w}) - 1,$$

where we used independence of (w_1, w_2) and $\mathbb{E}[\tilde{\alpha}_1] = \mathbb{E}[\tilde{\alpha}_2] = 1$, $\beta_1 + \beta_2 = 1$. Since $\tilde{\alpha}_i \rightarrow 0$ and $t g_i(t) / \bar{\theta}_i(t) \rightarrow 1$ as $t \rightarrow 0$, equation (50) holds. It follows that

$$\left(\beta_1 \frac{1 - \mathbb{E}[\tilde{\alpha}_1 | w_1 \geq t]}{\gamma_1 \bar{\theta}_1(t)}, \beta_2 \frac{1 - \mathbb{E}[\tilde{\alpha}_2 | w_2 \geq t]}{\gamma_2 \bar{\theta}_2(t)} \right)$$

is the solution to the relaxed problem. By (19), optimal taxes are individual-based. \square

D.3 Conditional moments of optimal taxes

In this section, we formally state and prove equation (27) that computes averages of optimal marginal taxes conditional on level curves of some function Q .

Proposition 11. Let $Q : \mathbb{R}_{++}^2 \rightarrow \mathbb{R}_{++}$ be a continuous function with the range (\underline{q}, \bar{q}) and piecewise continuously differentiable level curves. Suppose that there exists $\mathbf{m} \in \mathbb{R}^2 \setminus \{0\}$ such that $\mathbf{t} \mapsto Q^t(\mathbf{t}) := Q\left(t_1^{1/m_1}, t_2^{1/m_2}\right)$ is an element of $\mathcal{W}^{1,1}(\mathbb{R}_{++}^2 \cap B_r^o)$, where B_r is a closed ball of radius $r > 0$ centered at $\mathbf{0}$, for each $r > 0$, and

$$\mathbb{E} \left[\sum_{i=1}^2 w_i |\nabla_i Q| \mid Q \leq q \right] < \infty. \quad (81)$$

Then,

$$\mathbb{E} \left[\sum_{i=1}^2 \gamma_i \lambda_i^* \frac{\partial \ln Q}{\partial \ln w_i} \mid Q = q \right] = \frac{1 - \mathbb{E}[\alpha | Q \geq q]}{-\partial \ln \mathbb{P}(Q \geq q) / \partial \ln q} = \frac{\mathbb{E}[\alpha | Q \leq q] - 1}{\partial \ln \mathbb{P}(Q \leq q) / \partial \ln q}. \quad (82)$$

Proof. Step 1. Since Q has piecewise continuously differentiable level curves, the boundary of $\{\mathbf{w} | Q(\mathbf{w}) \leq q\}$ is Lipschitz. Then, our assumptions on $\boldsymbol{\lambda}^*$ ensure that (49) can be transformed by the Divergence Theorem. Repeating the argument described in the proof of the sufficiency part of Proposition 10 and using (49), we obtain

$$\begin{aligned} \int_{Q^{-1}(q)} \sum_{i=1}^2 \lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w}) n_i(\mathbf{w}) d\sigma(\mathbf{w}) &= \int_{\{\mathbf{w} | Q(\mathbf{w}) \leq q\}} \sum_{i=1}^2 \nabla_i (\lambda_i^*(\mathbf{w}) \gamma_i w_i f(\mathbf{w})) d\mathbf{w}, \\ &= \mathbb{P}(Q \leq q) \mathbb{E}[\alpha - 1 | Q \leq q], \end{aligned} \quad (83)$$

where $\mathbf{n}(\mathbf{w})$ is the outward unit normal to $\{\widehat{\mathbf{w}} | Q(\widehat{\mathbf{w}}) \leq q\}$ at \mathbf{w} .

Step 2. To make further progress, it is convenient to work with the transformation of productivities given by $\mathbf{t} := (w_1^{m_1}, w_2^{m_2})$ so that the transformed density is given by $f^t(\mathbf{t}) := f\left(t_1^{1/m_1}, t_2^{1/m_2}\right) \prod_{i=1}^2 \frac{t_i^{(1-m_i)/m_i}}{|m_i|}$. By the Coarea Formula discussed in the mathematical appendix for every (Borel) measurable function $v : \mathbb{R}_{++}^2 \rightarrow \mathbb{R}$ that is integrable with respect to f on $\{\mathbf{t} | Q^t(\mathbf{t}) \leq q\}$,

$$\mathbb{P}(Q^t \leq q) \mathbb{E}[v | Q^t \leq q] = \int_{\underline{q}}^q \left(\int_{(Q^t)^{-1}(s)} v(\mathbf{t}) f^t(\mathbf{t}) \frac{d\sigma(\mathbf{t})}{\|\nabla Q^t(\mathbf{t})\|} \right) ds.$$

Then, by the definition of conditional expectation, for a.e. q ,

$$\mathbb{E}[v | Q^t = q] = \frac{d(\mathbb{P}(Q^t \leq q) \mathbb{E}[v | Q^t \leq q])}{d\mathbb{P}(Q^t \leq q)} = \frac{\int_{(Q^t)^{-1}(q)} v(\mathbf{t}) f^t(\mathbf{t}) \frac{d\sigma(\mathbf{t})}{\|\nabla Q^t(\mathbf{t})\|}}{d\mathbb{P}(Q^t \leq q) / dq}. \quad (84)$$

It is easy to see that $w_i \nabla_i Q(\mathbf{w}) = m_i t_i \nabla_i Q^t(\mathbf{t})$ and the unit normal vectors on the boundaries of $\{\mathbf{w} | Q(\mathbf{w}) \leq q\}$ and $\{\mathbf{t} | Q^t(\mathbf{t}) \leq q\}$ are given by $n_i(\mathbf{w}) = \frac{\nabla_i Q(\mathbf{w})}{\|\nabla Q(\mathbf{w})\|}$ and $n_i^t(\mathbf{t}) = \frac{\nabla_i Q^t(\mathbf{t})}{\|\nabla Q^t(\mathbf{t})\|}$,

resp. Then, thanks to assumption (81), $v(\mathbf{t}) := \sum_{i=1}^2 \lambda_i^* \left(t_1^{1/m_1}, t_2^{1/m_2} \right) m_i \gamma_i t_i \nabla_i Q^t(\mathbf{t})$ is integrable with respect to f^t on $\{\mathbf{t} | Q^t(\mathbf{t}) \leq q\}$. Thus, using (84) and changing variables from \mathbf{t} to \mathbf{w} we obtain the following:

$$\mathbb{E} \left[\sum_{i=1}^2 \lambda_i^* \gamma_i w_i \nabla_i Q \middle| Q = q \right] = \frac{\int_{(Q^t)^{-1}(q)} \lambda_i^*(\mathbf{w}) \gamma_i w_i n_i(\mathbf{w}) f(\mathbf{w}) d\sigma(\mathbf{w})}{d\mathbb{P}(Q \leq q) / dq}. \quad (85)$$

Divide each side of (85) by Q and combine it with (83) to establish the second part of equation (82).

Step 3. To see the first part of equation (82), use Bayes rule and $\mathbb{E}[\alpha] = 1$ to obtain

$$\mathbb{P}(Q \geq q) \mathbb{E}[1 - \alpha | Q \geq q] = \mathbb{P}(Q \leq q) \mathbb{E}[1 - \alpha | Q \leq q].$$

And, the result follows from the second part of equation (82). \square

D.4 Perturbation describing average taxes (equation (28))

As in the text, fix some smooth T and let $(y_1(\mathbf{w}), y_2(\mathbf{w}))$ be the optimal earnings of couple with productivities \mathbf{w} under this tax schedule. We shall (heuristically) construct a perturbation of T in which taxes on all couples with $w_i > t$ some small amount, say ε . A way to do is to increase marginal taxes on every couple with productivities \mathbf{w} such that w_i lies in some small band $[t, t + dt]$ by $(d\tau_1(w_{-i}), d\tau_2(w_{-i}))$. Of course, $d\boldsymbol{\tau}$ aren't arbitrary. First of all, differences in taxes between couples (t, w_{-i}) and $(t + dt, w_{-i})$ should be exactly ε . In addition, differences in taxes between couples (t, w_{-i}) and $(t, w_{-i} + dt)$ should be 0, thus

$$\varepsilon = d\boldsymbol{\tau}(w_{-i}) \cdot \nabla_i \mathbf{y}(t, w_{-i}) dt \quad \text{and} \quad 0 = d\boldsymbol{\tau}(w_{-i}) \cdot \nabla_{-i} \mathbf{y}(t, w_{-i}) dt. \quad (86)$$

To calculate the benefit from redistribution \mathfrak{B} , observe that this perturbation collects ε from each couple with $w_i > t$. So, ignoring the behavioral response, the perturbation redistributes $\mathbb{P}(w_i \geq t)$ in a lump-sum fashion so that

$$\mathfrak{B} = \varepsilon \cdot \int_t^\infty \int_0^\infty (1 - \alpha(\mathbf{w})) f(\mathbf{w}) dw_{-i} dw_i. \quad (87)$$

To calculate the costs of distortion \mathfrak{C} , let $(\delta_1(w_{-i}), \delta_2(w_{-i}))$ be the reduction of earnings of individuals with productivities \mathbf{w} in the band so that their tax bill increases by $\nabla T(\mathbf{y}(t, w_{-i})) \cdot \boldsymbol{\delta}(w_{-i})$. Since there are $g_i(t)dt$ couples in this band,

$$\mathfrak{C} = \nabla T(\mathbf{y}(t, w_{-i})) \cdot \boldsymbol{\delta}(w_{-i}) g_i(t) dt. \quad (88)$$

It remains to relate $\boldsymbol{\delta}(w_{-i})$, $d\boldsymbol{\tau}(w_{-i})$ and $\nabla_{\mathbf{y}}(t, w_{-i})$ to each other. Recall the first-order conditions defining earnings under the original tax, i.e., $1 - \nabla_j T(\mathbf{y}(\mathbf{w})) = \frac{y_j^{(1-\gamma_j)/\gamma_j}(\mathbf{w})}{w_j^{1/\gamma_j}}$ for $j = 1, 2$. First of all, totally differentiate the first-order conditions with respect to \mathbf{w} to obtain $\Sigma(\mathbf{w}) \nabla_i \mathbf{y}(\mathbf{w}) = \begin{bmatrix} \frac{1}{\gamma_i} (1 - \nabla_i T(\mathbf{y}(\mathbf{w}))) \\ 0 \end{bmatrix}$ and $\Sigma(\mathbf{w}) \nabla_{-i} \mathbf{y}(\mathbf{w}) = \begin{bmatrix} 0 \\ \frac{1}{\gamma_{-i}} (1 - \nabla_{-i} T(\mathbf{y}(\mathbf{w}))) \end{bmatrix}$, (89)

Second, adjust ∇T in these conditions by $d\boldsymbol{\tau}$ to obtain

$$\Sigma(t, w_{-i}) \boldsymbol{\delta}(w_{-i}) = d\boldsymbol{\tau}(w_{-i}), \quad (90)$$

where

$$\Sigma(\mathbf{w}) := \begin{bmatrix} \nabla_{ii} T(\mathbf{y}(\mathbf{w})) + \frac{1-\gamma_i}{\gamma_i y_i(\mathbf{w})} \nabla_i T(\mathbf{y}(\mathbf{w})) & \nabla_{12} T(\mathbf{y}(\mathbf{w})) \\ \nabla_{12} T(\mathbf{y}(\mathbf{w})) & \nabla_{-i-i} T(\mathbf{y}(\mathbf{w})) + \frac{1-\gamma_{-i}}{\gamma_{-i} y_{-i}(\mathbf{w})} \nabla_{-i} T(\mathbf{y}(\mathbf{w})) \end{bmatrix}.$$

Combine (86), (89) and (90) to obtain $\delta_i(w_{-i}) = \varepsilon \cdot \frac{\gamma_i t}{(1 - \nabla_i T(\mathbf{y}(\mathbf{w})))}$, $\delta_{-i}(w_{-i}) = 0$ and $d\tau_i(w_{-i}) = \Sigma_{ii} \delta_i(w_{-i})$, $d\tau_{-i}(w_{-i}) = \Sigma_{12} \delta_i(w_{-i})$. As a result, by equations (87) and (88),

$$\mathfrak{B} - \mathfrak{C} = \varepsilon \cdot \mathbb{P}(w_i \geq t) \left(1 - \mathbb{E}[\alpha | w_i \geq t] - \gamma_i \bar{\theta}_i(t) \frac{\nabla_i T(\mathbf{y}(\mathbf{w}))}{1 - \nabla_i T(\mathbf{y}(\mathbf{w}))} \right).$$

And, the expression for average taxes follows from equation (28) by setting $\mathfrak{B} - \mathfrak{C} = 0$.

D.5 Average distortions

D.5.1 Proof of Proposition 5

Proof. By $F^a \leq_{PQD} F^b$,

$$\mathbb{P}^a(w_{-i} \leq t_{-i} | w_i \leq t_i) \leq \mathbb{P}^b(w_{-i} \leq t_{-i} | w_i \leq t_i) \quad \forall t. \quad (91)$$

Since α is decreasing, the first-order stochastic dominance relationship in (91) gives $\mathbb{E}^a[\alpha | w_i \leq t] \leq \mathbb{E}^b[\alpha | w_i \leq t]$. Thus, by (28), $\mathbb{E}^a[\lambda_i^{a,*} | w_i = t] \leq \mathbb{E}^b[\lambda_i^{b,*} | w_i = t]$.

Finally, to obtain $\mathbb{E}^b[\lambda_i^{b,*} | w_i = t] \leq \lambda_i^{sng,*}(t)$, observe $\mathbb{E}[\alpha | w_i = t] = \tilde{\alpha}(w_i)$ and repeat the previous argument with $F^b \leq_{PQD} F^*$ □

D.5.2 Proof of Corollary 2

Proof. Due to separability of α ,

$$\mathbb{E}[\alpha | w_i = t] = \beta_i \tilde{\alpha}_i(t) + \beta_{-i} \mathbb{E}[\tilde{\alpha}_{-i}(w_{-i}) | w_i = t] = 1 + \beta_i (\tilde{\alpha}_i(t) - 1),$$

where we also used independence of (w_1, w_2) and $\mathbb{E}[\tilde{\alpha}_1] = \mathbb{E}[\tilde{\alpha}_2] = 1$, $\beta_1 + \beta_2 = 1$. The rest of the proof is identical to the proof of Proposition 5. □

D.5.3 Proof of Corollary 3

Proof. Observe that $(u_1, u_2) \in (0, 1)^2 \mapsto \alpha^a(G_1^{-1}(u_1), G_2^{-1}(u_2))$ is non-decreasing, thus it can be extended to a uniformly continuous function on $[0, 1]^2$. Then,

$$\lim_{t \rightarrow 0} \mathbb{E}^a[\alpha^a | w_i \leq t] = \lim_{t \rightarrow 0} \mathbb{E}^a[\alpha^a(0, w_{-i}) | w_i \leq t] \leq \lim_{t \rightarrow 0} \mathbb{E}^b[\alpha^a(0, w_{-i}) | w_i \leq t], \quad (92)$$

where the first equality is due to uniform continuity and the second inequality follows from (91).

Since $\alpha^a \sim \alpha^b$, these Pareto weights are related by $\alpha^a(0, w_{-i}) = \mathbb{E}^b[\alpha^a] \alpha^b(0, w_{-i})$, where

$$\mathbb{E}^b[\alpha^a] = \underbrace{\mathbb{E}^a[\alpha^a]}_{=1} + \int_W \alpha^a(\mathbf{w}) d\left(F^a(\mathbf{w}) - F^b(\mathbf{w})\right) = 1 + \int_W \underbrace{\left(F^b(\mathbf{w}) - F^a(\mathbf{w})\right)}_{\geq 0} d\alpha^a(\mathbf{w}). \quad (93)$$

And, equation (92) gives

$$\lim_{t \rightarrow 0} \mathbb{E}^a[\alpha^a | w_i \leq t] = \mathbb{E}^b[\alpha^a] \lim_{t \rightarrow 0} \mathbb{E}^b[\alpha^b(0, w_{-i}) | w_i \leq t] = \mathbb{E}^b[\alpha^a] \lim_{t \rightarrow 0} \mathbb{E}^b[\alpha^b | w_i \leq t],$$

where the last equality is due to the same uniform continuity argument as above. Note that for submodular α^a , the term $\mathbb{E}^b[\alpha^a]$ is less than one; thus, by (28),

$$\lim_{t \rightarrow 0} \frac{\mathbb{E}^b[\lambda_i^{b,*} \theta_i | w_i = t]}{\mathbb{E}^a[\lambda_i^{a,*} \theta_i | w_i = t]} = \lim_{t \rightarrow 0} \frac{\mathbb{E}^b[\alpha^b | w_i \leq t]}{\mathbb{E}^a[\alpha^a | w_i \leq t]} = \frac{1}{\mathbb{E}^b[\alpha^a]} \geq 1.$$

Clearly, $\mathbb{E}^b[\alpha^a]$ is strictly less than one when $F^b - F^a > 0$ is positive on a set of positive measure and α^a is strictly submodular, i.e., $d\alpha^a < 0$.

The rest of the proof can be shown similarly, therefore it is omitted. \square

D.5.4 Proof of Proposition 6

Proof. Consider first α^b that is more redistributive than α^a . As discussed in Chapter 6.E of Shaked and Shanthikumar (2007), the distribution with density $\alpha^a f$ first-order stochastically dominates the distribution with density $\alpha^b f$. Then,

$$\begin{aligned} \mathbb{E}[\alpha^a | w_i \geq t] &= \frac{\int_W \mathbf{1}_{[t, \infty)}(w_i) \alpha^a(\mathbf{w}) f(\mathbf{w}) d\mathbf{w}}{1 - G_i(t)} \geq \\ &\geq \frac{\int_W \mathbf{1}_{[t, \infty)}(w_i) \alpha^b(\mathbf{w}) f(\mathbf{w}) d\mathbf{w}}{1 - G_i(t)} = \mathbb{E}[\alpha^b | w_i \geq t], \end{aligned}$$

and $\mathbb{E}^a[\lambda_i^{a,*} | w_i = t] \leq \mathbb{E}^b[\lambda_i^{b,*} | w_i = t]$ follows from (28).

Consider next two benchmark economies such that $\tilde{\alpha}^b$ that is more redistributory than $\tilde{\alpha}^a$. Remark that the distribution with density $\tilde{\alpha}^a(w_j) f$ first-order stochastically dominates the

distribution with density $\tilde{\alpha}^b(w_j)f$ for $j = 1, 2$. Applying the same argument as in the first part of this proof,

$$\begin{aligned}\mathbb{E}[\alpha^a | w_i \geq t] &= \frac{1}{2}\mathbb{E}[\tilde{\alpha}^a(w_i) | w_i \geq t] + \frac{1}{2}\mathbb{E}[\tilde{\alpha}^a(w_{-i}) | w_i \geq t] \geq \\ &\geq \frac{1}{2}\mathbb{E}[\tilde{\alpha}^b(w_i) | w_i \geq t] + \frac{1}{2}\mathbb{E}[\tilde{\alpha}^b(w_{-i}) | w_i \geq t] = \mathbb{E}[\alpha^b | w_i \geq t].\end{aligned}$$

By (28), $\mathbb{E}^a[\lambda_i^{a,*} | w_i = t] \leq \mathbb{E}^b[\lambda_i^{b,*} | w_i = t]$. □

D.6 Average jointness

In this section, we study average jointness. It is assumed throughout that the following coefficients measuring tail-dependence and speed of convergence to are well-defined: $\bar{\chi} := \lim_{u \rightarrow 0} \frac{\bar{C}(u,u)}{u}$, $\chi := \lim_{u \rightarrow 0} \frac{C(u,u)}{u}$ and $\bar{\kappa} := \lim_{u \rightarrow 0} \frac{\ln u}{\ln \bar{C}(u,u)}$, $\kappa := \lim_{u \rightarrow 0} \frac{\ln u}{\ln C(u,u)}$. Moreover, both limiting distributions $F(\cdot|\infty)$ and $F(\cdot|0)$ exist.

D.6.1 Proof of Lemma 4 and Proposition 7

Proof. We start with a preliminary observation that will be useful to sign the average jointness at the extremes. Unpack $\mathbb{E}[\lambda_i^* | w_i = t]$ conditioning on spouse $-i$ being more and less productive than spouse i to get

$$\begin{aligned}\mathbb{E}[\lambda_i^* | w_i = t] &= \mathbb{E}[\lambda_i^* | w_i = t \leq w_{-i}] \mathbb{P}(w_{-i} \geq t | w_i = t) + \\ &+ \mathbb{E}[\lambda_i^* | w_i = t \geq w_{-i}] \mathbb{P}(w_{-i} \leq t | w_i = t).\end{aligned}\tag{94}$$

Observe that $\mathbb{P}(w_{-i} \leq t | w_i = t) = \frac{1}{2} \frac{d(1 - \mathbb{P}(\mathbf{w} \leq (t, t)))}{d(1 - \mathbb{P}(w_i \leq t))}$. By L'Hôpital's rule and symmetry,

$$\lim_{t \rightarrow \infty} \mathbb{P}(w_{-i} \leq t | w_i = t) = \lim_{t \rightarrow \infty} \frac{1}{2} \frac{1 - \mathbb{P}(\mathbf{w} \leq (t, t))}{1 - \mathbb{P}(w_i \leq t)} = 1 - \frac{1}{2} \lim_{t \rightarrow \infty} \mathbb{P}(w_{-i} \geq t | w_i \geq t),$$

which clearly equals to $1 - \frac{\bar{\chi}}{2}$. Then, dividing each side of (94) by $\mathbb{E}[\lambda_i^* | w_i = t]$ and taking t to ∞ , we obtain

$$1 = \lim_{t \rightarrow \infty} \bar{H}(t, t) \frac{\bar{\chi}}{2} + \lim_{t \rightarrow \infty} H(t, t) \left(1 - \frac{\bar{\chi}}{2}\right).\tag{95}$$

We are now in position to prove Lemma 4 and Proposition 7. Suppose first that F is right-tail dependent, i.e., $\bar{\chi} \neq 0$. As discussed in the text, $\bar{\kappa} = 1$, and (38) implies $\lim_{t \rightarrow \infty} \bar{H}(t, t) \geq 1$, where the inequality is strict when $F(\cdot|\infty)$ is non-degenerate as $\mathbb{E}[\tilde{\alpha}(w_{-i}) | w_i = \infty] > \tilde{\alpha}(\infty)$. By (95), $\lim_{t \rightarrow \infty} H(t, t) \leq 1$ with a strict inequality when $\lim_{t \rightarrow \infty} \bar{H}(t, t) > 1$, and the result follows.

Suppose next that F is right-tail independent, i.e., $\bar{\chi} = 0$, and either $\bar{\kappa} = \frac{1}{2}$ or $F(\cdot|\infty) \equiv 0$. It follows from (95) that $\lim_{t \rightarrow \infty} H(t, t) = 1$. Clearly, $\lim_{t \rightarrow \infty} \bar{H}(t, t) = \bar{\kappa} \leq 1$ when $F(\cdot|\infty)$ is degenerate, and

$$\lim_{t \rightarrow \infty} \bar{H}(t, t) = \frac{1 - \tilde{\alpha}(\infty)}{1 - \tilde{\alpha}(\infty) + 1 - \mathbb{E}[\tilde{\alpha}(w_{-i})|w_i = \infty]} \leq \frac{1 - \tilde{\alpha}(\infty)}{1 - \tilde{\alpha}(\infty) + 1 - \mathbb{E}[\tilde{\alpha}(w_{-i})]}$$

when $\bar{\kappa} = \frac{1}{2}$. The inequality is due to monotonicity of $\tilde{\alpha}$ and positive dependence of F so that $\mathbb{P}(\cdot|w_i \geq t)$ first-order stochastically dominates G . Then, since $\mathbb{E}[\tilde{\alpha}] = 1$, the value of $\lim_{t \rightarrow \infty} \bar{H}(t, t)$ is not more than one and strictly less than one if $F(\cdot|\infty)$ is first-order stochastically dominated by G . In both cases jointness is negative at the top, and it is strictly negative under the additional conditions mentioned in Lemma 4.

For the left corner, the analogues of (95) and (38) read as

$$1 = \lim_{t \rightarrow 0} \bar{H}(t, t) \left(1 - \frac{\chi}{2}\right) + \lim_{t \rightarrow 0} H(t, t) \frac{\chi}{2}, \quad \lim_{t \rightarrow 0} H(t, t) = \kappa \frac{\alpha(\mathbf{0}) - 1}{\mathbb{E}[\alpha|w_i = 0] - 1}.$$

And, the result follows from the same argument as above. \square

D.6.2 Proof of Corollary 4

Proof. The analog of equation (34) for the bottom, which is due to (27), reads as

$$\mathbb{E}[\lambda_i^*|w_i = t, w_{-i} \leq t] = \frac{1}{2} \frac{\mathbb{E}[\alpha|\mathbf{w} \leq (t, t)] - 1}{\gamma\theta(t)\eta(t, t)}.$$

It follows that the the sign of $J(t, t)$ is the same as the sign of $\phi(t)$, which is given by

$$\begin{aligned} \phi(t) &:= \gamma t g(t) (\mathbb{E}[\lambda_i^*|w_i = t, w_{-i} \geq t] - \mathbb{E}[\lambda_i^*|w_i = t, w_{-i} \leq t]) = \\ &= \frac{\frac{1}{2} \int_{\{\mathbf{w} \in W: \mathbf{w} \geq (t, t)\}} (1 - \alpha(\mathbf{w})) f(\mathbf{w}) d\mathbf{w}}{1 - F(t|t)} - \frac{\frac{1}{2} \int_{\{\mathbf{w} \in W: \mathbf{w} \leq (t, t)\}} (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w}}{F(t|t)}. \end{aligned}$$

We shall show that ϕ is positive (negative) below (resp. above) a certain threshold.

First of all, we note that $f(\mathbf{w}) = g(w_1)g(w_2)[1 + \rho(1 - 2G(w_1))(1 - 2G(w_2))]$ for the FGM copula; thus,

$$\begin{aligned} &\frac{1}{2} \int_{\{\mathbf{w} \in W: \mathbf{w} \geq (t, t)\}} (1 - \alpha(\mathbf{w})) f(\mathbf{w}) d\mathbf{w} = \int_{\{\mathbf{w} \in W: \mathbf{w} \geq (t, t)\}} \left(\frac{1}{2} - \tilde{\alpha}(w_i)\right) f(\mathbf{w}) d\mathbf{w} = \\ &= (1 - G(t)) \int_t^\infty \left(\frac{1}{2} - \tilde{\alpha}(w)\right) dG(w) - \rho G(t)(1 - G(t)) \int_t^\infty \left(\frac{1}{2} - \tilde{\alpha}(w)\right) (1 - 2G(w)) dG(w), \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{2} \int_{\{\mathbf{w} \in W: \mathbf{w} \leq (t, t)\}} (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} = \int_{\{\mathbf{w} \in W: \mathbf{w} \leq (t, t)\}} \left(\tilde{\alpha}(w_i) - \frac{1}{2}\right) f(\mathbf{w}) d\mathbf{w} = \\ &= G(t) \int_0^t \left(\tilde{\alpha}(w) - \frac{1}{2}\right) dG(w) + \rho G(t)(1 - G(t)) \int_0^t \left(\tilde{\alpha}(w) - \frac{1}{2}\right) (1 - 2G(w)) dG(w). \end{aligned}$$

It is convenient to define $\psi(t) := \int_0^t (\frac{1}{2} - \tilde{\alpha}(w)) dG(w)$, which is non-positive with $\psi(0) = \psi(\infty) = 0$ and $\int_0^\infty \psi(w) dG(w) < 0$ whenever $\tilde{\alpha}$ is non-constant. Using the fact $F(t|t) = G(t) + \rho(1 - 2G(t))G(t)(1 - G(t))$ and two above expressions, ϕ can be succinctly re-written as

$$\phi(t) = 2\rho G(t)(1 - G(t)) \left(\frac{\int_0^t \psi(w) dG(w)}{F(t|t)} - \frac{\int_t^\infty \psi(w) dG(w)}{1 - F(t|t)} \right).$$

The term in the brackets can be further unpacked as follows:

$$\begin{aligned} \frac{\int_0^t \psi(w) dG(w)}{F(t|t)} - \frac{\int_t^\infty \psi(w) dG(w)}{1 - F(t|t)} &= \frac{G(t) \int_0^\infty (-\psi(w)) dG(w)}{F(t|t)(1 - F(t|t))} \times \\ &\times \left(\frac{F(t|t)}{G(t)} - \frac{\int_0^t (-\psi(w)) dG(w) / G(t)}{\int_0^\infty (-\psi(w)) dG(w)} \right). \end{aligned}$$

By construction, there are numbers $\underline{t} \leq \bar{t}$ such that ψ is strictly decreasing on $(0, \underline{t})$, constant on (\underline{t}, \bar{t}) , and strictly increasing on (\bar{t}, ∞) . It follows that $t \mapsto \frac{\int_0^t (-\psi(w)) dG(w) / G(t)}{\int_0^\infty (-\psi(w)) dG(w)}$ first monotonically increases from 0 to some number above 1, and then it decreases to 1. On the other hand, $t \mapsto \frac{F(t|t)}{G(t)} = 1 + \rho(1 - 2G(t))(1 - G(t))$ first monotonically decreases from $1 + \rho > 1$ to some number below 1, and then it increases to 1. The last point follows from the fact that $\frac{F(t|t)}{G(t)} \geq 1$ if and only if $G(t) \leq \frac{1}{2}$, and

$$\frac{\partial (F(t|t) / G(t))}{\partial t} = \rho g(t)(4G(t) - 3) \geq 0 \iff G(t) \geq \frac{3}{4}.$$

Taking both pieces together, we conclude that there is a unique threshold \hat{t} such that

$$\frac{F(t|t)}{G(t)} - \frac{\int_0^t (-\psi(w)) dG(w) / G(t)}{\int_0^\infty (-\psi(w)) dG(w)} \geq 0 \iff t \leq \hat{t},$$

which concludes the proof. \square

D.6.3 Proof of Corollary 5

Consider the right tail and note that separability of α was used in the proof of Proposition 7 only when $\bar{\chi} = 0$ and $\bar{\kappa} = \frac{1}{2}$. However, under supermodularity (and symmetry) of α ,

$$\alpha(\infty) \geq -\mathbb{E}[\alpha] + 2\mathbb{E}[\alpha(\infty, w_{-i})].$$

As a result, since $\mathbb{E}[\alpha] = 1$ and $\mathbb{E}[\alpha(\infty, w_{-i})] \leq \mathbb{E}[\alpha|w_i = \infty]$ due to positive dependence of F ,

$$\lim_{t \rightarrow \infty} \bar{H}(t, t) = \frac{1 - \alpha(\infty)}{2 - 2\mathbb{E}[\alpha|w_i = \infty]} \leq 1.$$

The claim then follows.

As for the left tail, note that, under submodularity (and symmetry) of α ,

$$\alpha(\mathbf{0}) \leq -\mathbb{E}[\alpha] + 2\mathbb{E}[\alpha(0, w_{-i})],$$

which gives $\lim_{t \rightarrow 0} H(t, t) \leq 1$.

D.6.4 Proof of Corollary 6

Proof. To formally prove the corollary we need an auxiliary concept of tail order functions.

Following Hua and Joe (2011), define $\bar{b} : \mathbb{R}_{++}^2 \rightarrow \mathbb{R}_+$ as follows:

$$\bar{b}(s_1, s_2) := \lim_{u \rightarrow 0} \frac{\bar{C}(s_1 u, s_2 u)}{\bar{C}(u, u)}.$$

As explained in Hua and Joe (2011), under ultimate monotonicity of $\nabla \bar{C}$, the tail order function \bar{b} is well-defined, differentiable, homogeneous of order $\bar{\kappa}$ and satisfies

$$\lim_{u \rightarrow 0} \frac{\partial \bar{C}(s_1 u, s_2 u)}{\partial s_i} / \bar{C}(u, u) = \nabla_i \bar{b}(s_1, s_2). \quad (96)$$

In what follows we will only consider the right tail, the other tail can be studied along the same lines by defining $b(s_1, s_2) := \lim_{u \rightarrow 0} \frac{C(s_1 u, s_2 u)}{C(u, u)}$.

First of all, it follows from (96) that $\mathbb{P}(w_{-i} \geq t_{-i}^u | w_i \geq t_i^u) = \nabla_i \bar{C}(s u)$ converges to $\bar{\chi} \nabla_i \bar{b}(s)$ as $u \rightarrow 0$. Using (94), the analog of equation (95) along (\mathbf{t}^u) ensues:

$$1 = \lim_{u \rightarrow 0} \bar{H}_i(\mathbf{t}^u) \bar{\chi} \nabla_i \bar{b}(s) + H_i(\mathbf{t}^u) (1 - \bar{\chi} \nabla_i \bar{b}(s)). \quad (97)$$

Similarly, by (96), $\eta_i(\mathbf{t}^u) = \frac{(s_i u) \nabla_i \bar{C}(s u)}{\bar{C}(s u)}$ converges to $\frac{s_i \nabla_i \bar{b}(s)}{\bar{b}(s)}$ as $u \rightarrow 0$. Using (31) and (28), we obtain the following expression:

$$\sum_{i=1}^2 \lim_{u \rightarrow 0} \bar{H}_i(\mathbf{t}^u) \frac{s_i \nabla_i \bar{b}(s)}{\bar{b}(s)} (1 - \mathbb{E}[\alpha | w_i = \infty]) = 1 - \alpha(\infty). \quad (98)$$

We now have all ingredients to prove the corollary formally. Suppose first that $\bar{\chi} \neq 0$. Since \bar{b} is homogeneous of order $\bar{\kappa} = 1$, equation (98) implies that $\lim_{u \rightarrow 0} \bar{H}_i(\mathbf{t}^u) \geq 1$ for at least one spouse i . Remark that $\bar{\chi} \nabla_i \bar{b}(s) < 1$, because $\nabla_i \bar{C}(s u) \leq 1$ for all u and $\frac{d \nabla_i \bar{C}(s u)}{d u} > 0$ for all sufficiently small u due to our assumption of ultimate monotonicity, thus $\lim_{u \rightarrow 0} \bar{H}_i(\mathbf{t}^u) \leq 1$ by (97).

Suppose next that $\bar{\chi} = 0$ and either $\bar{\kappa} = \frac{1}{2}$ or $(F_1(\cdot | \infty), F_2(\cdot | \infty))$ are degenerate. In the latter case, $\mathbb{E}[\alpha | w_i = \infty] = \alpha(\infty)$; thus, by (98), $\lim_{u \rightarrow 0} \bar{H}_i(\mathbf{t}^u) \leq 1$ for some i . In the former case with $\bar{\kappa} = \frac{1}{2}$, if α is supermodular and F is positive dependent, then $\sum_{i=1}^2 \frac{1}{2} (1 - \mathbb{E}[\alpha | w_i = \infty]) \geq 1 - \alpha(\infty)$ due to the argument used in the proof of Corollary 5. It again follows $\lim_{u \rightarrow 0} \bar{H}_i(\mathbf{t}^u) \leq 1$ for some i . In both cases, since $\bar{\chi} = 0$, $\lim_{u \rightarrow 0} H_i(\mathbf{t}^u) = 1$, and the result ensues. \square

D.6.5 Relationship to Kleven et al. (2007)

In their working version of the paper, Kleven et al. (2007) (KKS for short) outlined how jointness can be signed at each productivity vector under the assumption of independent types. We now briefly (and heuristically) review their argument in the notations of the present paper focusing for simplicity on a symmetric economy with supermodular Pareto weights. The argument for submodular weights is identical.

When types are independent, equation (49) can be rewritten as

$$\sum_{i=1}^2 \frac{\nabla_i (\lambda_i^* (\mathbf{w}) \gamma w_i g(w_i))}{g(w_i)} = \alpha(\mathbf{w}) - 1. \quad (99)$$

Differentiate twice (99) to obtain

$$\sum_{i=1}^2 \nabla_i \left(\frac{\nabla_{12} (\lambda_i^* (\mathbf{w}) \gamma w_i g(w_i))}{g(w_i)} \right) = \nabla_{12} \alpha(\mathbf{w}). \quad (100)$$

Define the set of types U for which jointness is strictly positive as

$$U := \{\mathbf{w} \in W \mid \nabla_{-i} \lambda_i (\mathbf{w}) \gamma w_i g(w_i) > 0\}.$$

Due to the boundary conditions (50), the set U is contained in the interior of W .

KKS suggested to integrate (100) over U , assuming that it is non-empty, and then use the Divergence Theorem to obtain

$$\begin{aligned} \int_U \nabla_{12} \alpha(\mathbf{w}) d\mathbf{w} &= \int_U \sum_{i=1}^2 \nabla_i \left(\frac{\nabla_{12} (\lambda_i^* (\mathbf{w}) \gamma w_i g(w_i))}{g(w_i)} \right) = \\ &= \int_{\partial U} \sum_{i=1}^2 \left(\frac{\nabla_{12} (\lambda_i^* (\mathbf{w}) \gamma w_i g(w_i))}{g(w_i)} \right) n_i(\mathbf{w}) d\sigma(\mathbf{w}), \end{aligned} \quad (101)$$

where, as usual, $\mathbf{n}(\mathbf{w})$ is the outward unit normal to ∂U at \mathbf{w} . Clearly, the expression on the left-hand side is non-negative due to supermodularity of α . On the other hand, since on $\nabla_{-i} (\lambda_i^* (\mathbf{w}) \gamma w_i g(w_i)) > 0$ on U but $\nabla_{-i} (\lambda_i^* (\mathbf{w}) \gamma w_i g(w_i)) < 0$ on the interior of its complement, we must have $n_i(\mathbf{w}) \propto -\nabla_{12} (\lambda_i^* (\mathbf{w}) \gamma w_i g(w_i))$. It follows that the second line in (101) is non-positive. Conclude $\mathbf{n}(\mathbf{w}) = \mathbf{0}$, which a contradiction.

The KKS's argument is powerful but assumes quite a bit of smoothness and regularity. Using our techniques, under much milder smoothness, jointness can be shown to be negative on average when α is strictly supermodular. The reader can verify that under independence of \mathbf{w} , supermodularity of α translates into supermodularity of its conditional expectation, that is

$$\frac{\partial^2}{\partial t_1 \partial t_2} \mathbb{E}[\alpha | \mathbf{w} \leq \mathbf{t}] \geq 0 \quad \forall \mathbf{t}.$$

Then, since $\eta(t, t) \equiv 1$ and $\mathbb{E}[\alpha] = 1$,

$$H(t, t) = \frac{\frac{1}{2}\mathbb{E}[\alpha|\mathbf{w} \leq (t, t)] - 1}{\mathbb{E}[\alpha|w_i \leq t] - 1} \geq 1 \quad \forall t,$$

thus average jointness is non-positive.

D.7 Distortions within and across families

D.7.1 Proof of Proposition 8

Proof. (a). Since $\mathbb{E}[\alpha] = 1$, the Bayes rule gives

$$\mathbb{P}(I \leq \iota) (\mathbb{E}[\alpha|I \leq \iota] - 1) = \mathbb{P}(I \geq \iota) (1 - \mathbb{E}[\alpha|I \geq \iota]).$$

By the law of iterated expectations,

$$\lim_{\iota \rightarrow 0} \mathbb{E}[\alpha|I \leq \iota] = \mathbb{E}\left[\lim_{\iota \rightarrow 0} \mathbb{E}[\alpha|I \leq \iota, R = r]\right] = \lim_{\iota \rightarrow 0} \mathbb{E}[\alpha(R, 0)|I \leq \iota].$$

To see the first equality, interchange lim and expectation operators using the Dominated convergence theorem. To obtain the second one, first extend α by uniform continuity, then take ι to 0 and interchange lim and expectation operators again. As a result, by (40), if $\alpha(\cdot, 0) \geq 1$, then

$$\lim_{\iota \rightarrow 0} \mathbb{E}[\lambda_2^* - \lambda_1^*] \geq \lim_{\iota \rightarrow 0} \frac{L_\iota(\iota)}{\iota l_\iota(\iota)} \left(\lim_{\iota \rightarrow 0} \mathbb{E}[\alpha(R, 0)|I \leq \iota] - 1 \right) \geq 0.$$

(b) Since both limits $\lim_{r \rightarrow 0} \frac{L_r(r)}{r l_r(r)}$, $\lim_{r \rightarrow \infty} \frac{1 - L_r(r)}{r l_r(r)}$ are finite and $\tilde{\alpha}$ is decreasing, the vector field $(\tilde{\lambda}(R(\mathbf{w})), \tilde{\lambda}(R(\mathbf{w})))$ is a non-negative, bounded element of $\mathscr{W}^{1, \infty}(W^o)$ and satisfies (18), (22). In view of the sufficiency part of Proposition 10, we need to show that this vector field satisfies (49), (50). The reader can verify that $w_1 = r(1 + \iota^{1/(1-\gamma)})^{\gamma-1}$, $w_2 = \iota r(1 + \iota^{1/(1-\gamma)})^{\gamma-1}$ and the density l satisfies $l(r, \iota) r \iota = f(w_1, w_2) w_1 w_2$, which gives

$$\frac{\sum_{i=1}^2 \nabla_i \left(\tilde{\lambda}(r) \gamma w_i f(w_1, w_2) \right)}{f(w_1, w_2)} = \nabla \tilde{\lambda}(r) \gamma r + \tilde{\lambda}(r) \frac{\gamma r \nabla_r (l(r, \iota) r)}{l(r, \iota)}. \quad (102)$$

Since r and ι are independent, $\frac{\nabla_r (l(r, \iota) r)}{l(r, \iota)} = \frac{\nabla_r (l_r(r) r)}{l_r(r)}$ and (102) implies that (49) is equivalent to

$$\frac{\nabla_r \left(\tilde{\lambda}(r) r l_r(r) \right)}{l_r(r)} = \tilde{\alpha}(r) - 1,$$

which is exactly how $\tilde{\lambda}$ is defined. Since $\tilde{\lambda}$ is bounded and $n_i(\mathbf{w}) = 0$ whenever $w_i = 0$, equation (50) holds due to $w_i f(w_1, w_2) = 0$.

(c). By $L^a \leq_{PQD} L^b$,

$$\mathbb{P}^a (R \geq r | I \geq \iota) \leq \mathbb{P}^b (R \geq r | I \geq \iota) \quad \forall (r, \iota). \quad (103)$$

Since α is measurable only with respect to r and decreasing in this variable, the first-order stochastic dominance relationship in (103) gives $\mathbb{E}^a [\alpha | I \geq \iota] \geq \mathbb{E}^b [\alpha | I \geq \iota]$. Thus, by (40), $\mathbb{E}^a [\lambda_2^{a,*} - \lambda_1^{a,*} | I = \iota] \leq \mathbb{E}^b [\lambda_2^{b,*} - \lambda_1^{b,*} | I = \iota]$. We end this section with a remark about a general economy. For any (γ_1, γ_2) , let $\bar{\gamma} = \bar{\gamma}(\gamma_1, \gamma_2)$ be the average family elasticity of labor supply, e.g., $\bar{\gamma} = (\gamma_1 + \gamma_2) / 2$ or $\bar{\gamma} = \sqrt{\gamma_1 \gamma_2}$. Define the transformation $(r, \iota) \leftrightarrow (w_1^{\bar{\gamma}/\gamma_1}, w_2^{\bar{\gamma}/\gamma_2})$ as in the text, then formulas (39) and (40) extend as follows:

$$\mathbb{E} \left[\sum_{i=1}^2 \frac{(w_i^{\bar{\gamma}/\gamma_i})^{1/(1-\bar{\gamma})}}{(w_1^{\bar{\gamma}/\gamma_1})^{1/(1-\bar{\gamma})} + (w_2^{\bar{\gamma}/\gamma_2})^{1/(1-\bar{\gamma})}} \lambda_i^* \middle| R = r \right] = \frac{1 - \mathbb{E}[\alpha | R \geq r]}{\bar{\gamma} \bar{\theta}_r(r)} \quad (104)$$

and

$$\mathbb{E}[\lambda_2^* - \lambda_1^* | I = \iota] = \frac{1 - \mathbb{E}[\alpha | I \geq \iota]}{\bar{\gamma} \bar{\theta}_\iota(\iota)}. \quad (105)$$

Proposition 8 then apply directly. \square

D.8 Optimal individual- and family-based taxation

D.8.1 Proof of Lemma 5

Proof. (a). Consider some anonymous tax schedule T , i.e., $T(y_1, y_2) = T(y_2, y_1)$. When faced with such taxes, couples with productivities (w_1, w_2) and (w_2, w_1) obtain the identical utility $v(\mathbf{w})$, because

$$v(\mathbf{w}) = \max_{\mathbf{y} \geq \mathbf{0}} \sum_{i=1}^2 \left(y_i - T(\mathbf{y}) - \frac{1}{\gamma} \left(\frac{y_i}{w_i} \right)^{1/\gamma} \right) = \max_{\mathbf{y} \geq \mathbf{0}} \sum_{i=1}^2 \left(y_i - T(\mathbf{y}) - \frac{1}{\gamma} \left(\frac{y_i}{w_{-i}} \right)^{1/\gamma} \right).$$

Conversely, let (v, c, \mathbf{y}) be a direct mechanism that satisfies (6), (7), (8) and in which v is symmetric. By Lemma 2 and symmetry of v , $y_1(w_1, w_2) = y_2(w_2, w_1)$ for all \mathbf{w} . Then, it is easy to see that the tax function, which decentralizes (v, c, \mathbf{y}) , defined in Lemma 1 is anonymous.

(b). Consider some separable tax schedule T , i.e., $T(\mathbf{y}) = \tilde{T}_1(y_1) + \tilde{T}_2(y_2)$. In response to this schedule, each couple with productivities \mathbf{w} obtains utility $v(\mathbf{w})$ that is additively separable in \mathbf{w} , that is

$$v(\mathbf{w}) = \sum_{i=1}^2 \underbrace{\max_{y_i \geq 0} \left(y_i - \tilde{T}_i(y_i) - \frac{1}{\gamma} \left(\frac{y_i}{w_i} \right)^{1/\gamma} \right)}_{=: \tilde{v}_i(w_i)}.$$

Conversely, let (v, c, \mathbf{y}) be a direct mechanism that satisfies (6), (7), (8) and in which v is additively separable, i.e., $v(\mathbf{w}) = \tilde{v}_1(w_1) + \tilde{v}_2(w_2)$. By Lemma 2, the spouse i 's earnings is measurable only with respect to w_i . Define individual-based taxes $\tilde{T}_1(y_1)$ and $\tilde{T}_2(y_2)$ that decentralize (v, c, \mathbf{y}) as follows: $\tilde{T}_i(y_i) := y_i - \gamma \left(\frac{y_i}{w_i}\right)^{1/\gamma} - \tilde{v}_i(w_i)$ if $y_i = y_i(w_i)$ for some w_i and $\tilde{T}_i(y_i) := \infty$, otherwise. This assignment uniquely defines $T(\mathbf{y}) := \tilde{T}_1(y_1) + \tilde{T}_2(y_2)$ because $c(\mathbf{w}) = c(\hat{\mathbf{w}})$ whenever $\mathbf{y}(\mathbf{w}) = \mathbf{y}(\hat{\mathbf{w}})$ due to (6) and (8).

(c). Consider some tax schedule on total earnings T , i.e., $T(\mathbf{y}) = \tilde{T}(y_1 + y_2)$. In response to this schedule, each couple with productivities \mathbf{w} obtains utility $v(\mathbf{w})$ that is measurable w.r.t. r only, that is

$$v(\mathbf{w}) = \max_{Y \geq 0} \underbrace{\left(Y - \tilde{T}(Y) - \gamma \left(\frac{Y}{R(\mathbf{w})} \right)^{1/\gamma} \right)}_{=: \tilde{v}(R(\mathbf{w}))}, \quad (106)$$

where we used the fact that $\left(\frac{Y}{R(\mathbf{w})}\right)^{1/\gamma}$ is the minimal cost of obtaining aggregate earnings $Y = y_1 + y_2$ for such couples. To see it, let's minimize $\sum_{i=1}^2 \gamma \left(\frac{y_i}{w_i}\right)^{1/\gamma}$ subject to $Y = y_1 + y_2$. The reader can verify that the unique minimizer is $\frac{y_i}{Y} = w_i^{1/(1-\gamma)} / \left(w_1^{1/(1-\gamma)} + w_2^{1/(1-\gamma)}\right)$, and (106) ensues.

Conversely, let (v, c, \mathbf{y}) be a direct mechanism that satisfies (6), (7), (8) and in which v is measurable w.r.t. r only. Consider \mathbf{w} and $\hat{\mathbf{w}}$ with the same value of r , i.e., $R(\mathbf{w}) = R(\hat{\mathbf{w}})$, then, by (8),

$$\begin{aligned} & (y_1(\hat{\mathbf{w}}))^{1/\gamma} \left(\left(1 + \iota^{1/(1-\gamma)}\right)^{1/\gamma-1} - \left(1 + \hat{\iota}^{1/(1-\gamma)}\right)^{1/\gamma-1} \right) + \\ & + (y_2(\hat{\mathbf{w}}))^{1/\gamma} \left(\left(1 + \iota^{1/(\gamma-1)}\right)^{1/\gamma-1} - \left(1 + \hat{\iota}^{1/(\gamma-1)}\right)^{1/\gamma-1} \right) \geq 0, \end{aligned}$$

where $\iota = w_2/w_1$ and $\hat{\iota} = \hat{w}_2/\hat{w}_1$. Letting ι to $\hat{\iota}$ from below and above, we obtain that on the curve $R(\mathbf{w}) = r$

$$(y_1(\mathbf{w}))^{1/\gamma} \left(1 + \iota^{1/(1-\gamma)}\right)^{1/\gamma-2} \iota^{\gamma-1} = (y_2(\mathbf{w}))^{1/\gamma} \left(1 + \iota^{1/(\gamma-1)}\right)^{1/\gamma-2} \iota^{(2-\gamma)/(\gamma-1)}. \quad (107)$$

By (107), $y_i w_i^{1/(\gamma-1)}$ is the same for both $i = 1, 2$, thus $\frac{y_i}{Y} = w_i^{1/(1-\gamma)} / \left(w_1^{1/(1-\gamma)} + w_2^{1/(1-\gamma)}\right)$, where Y stays for aggregate earnings $y_1 + y_2$. As a result, Y and $\sum_{i=1}^2 \left(\frac{y_i}{w_i}\right)^{1/\gamma} = \left(\frac{Y}{R(\mathbf{w})}\right)^{1/\gamma}$ are both measurable only with respect to r . So, define family-based taxes \tilde{T} that decentralize (v, c, \mathbf{y}) as follows: $\tilde{T}(Y) := Y - \gamma \left(\frac{Y}{R(\mathbf{w})}\right)^{1/\gamma} - \tilde{v}(R(\mathbf{w}))$ if $Y = y_1(\mathbf{w}) + y_2(\mathbf{w})$ for some \mathbf{w} and $\tilde{T}(Y) := \infty$, otherwise. This assignment uniquely defines $T(\mathbf{y}) := \tilde{T}(y_1 + y_2)$ because $c(\mathbf{w}) = c(\hat{\mathbf{w}})$ whenever $\mathbf{y}(\mathbf{w}) = \mathbf{y}(\hat{\mathbf{w}})$ due to (6) and (8). \square

D.8.2 Proof of Proposition 9

Proof. (a). Let $v \in \mathcal{V}$ be symmetric. It is routine to verify that the functional $\Upsilon(v)$ satisfies

$$\begin{aligned} \Upsilon(v) &= \int_W v(\mathbf{w}) (\alpha^{anon}(\mathbf{w}) - 1) f^{anon}(\mathbf{w}) d\mathbf{w} + \\ &\quad + \int_W \sum_{i=1}^2 \left(w_i^{1+\gamma_i} (\nabla_i v(\mathbf{w}))^{\gamma_i} - \gamma_i w_i \nabla_i v(\mathbf{w}) \right) f^{anon}(\mathbf{w}) d\mathbf{w}. \end{aligned}$$

In other words, $v^{anon,*}$ solves the relaxed problem in the symmetric economy $(F^{anon}, \alpha^{anon}, \gamma)$. The claim then follows from our derivations in Section 4.2, i.e., equation (28).

(b). Consider $v \in \mathcal{V}$ that is additively separable, i.e., $v(\mathbf{w}) = \tilde{v}_1(w_1) + \tilde{v}_2(w_2)$ for some functions $(\tilde{v}_1, \tilde{v}_2)$. Then,

$$\begin{aligned} \int_W v(\mathbf{w}) (\alpha(\mathbf{w}) - 1) f(\mathbf{w}) d\mathbf{w} &= \int_0^\infty \sum_{i=1}^2 \tilde{v}_i(w_i) (\mathbb{E}(\alpha|w_i) - 1) g(w_i) dw_i = \\ &= \sum_{i=1}^2 \int_0^\infty \nabla \tilde{v}_i(w_i) (G(w_i) - A(w_i)) dw_i, \end{aligned}$$

where $A(t) := \int_0^t \mathbb{E}(\alpha|w_i = s) g(s) ds$. It follows that the value of $\Upsilon(v)$ for an additively separable v can be re-written as

$$\Upsilon(v) = \sum_{i=1}^2 \int_0^\infty \left((\nabla \tilde{v}_i(w_i))^{1/\gamma} w_i^{1+1/\gamma} - \frac{w_i}{\gamma} \nabla \tilde{v}_i(w_i) \left(1 + \frac{1}{2} \frac{A(w_i) - G(w_i)}{\gamma w_i g(w_i)} \right) \right) g(w_i) dw_i.$$

Pointwise optimization of this equation yields

$$(\nabla \tilde{v}_i(t))^{1/\gamma} t^{1/\gamma} = 1 + \frac{A(t) - G(t)}{\gamma t g(t)} = 1 + \lambda^{ind,*}(t). \quad (108)$$

Since $\lambda^{ind,*}$ is bounded from below by some $\underline{\lambda} > -1$, by the exact same argument as in the proof of the sufficiency part of Proposition 10, $\tilde{v}_1(w_1) + \tilde{v}_2(w_2)$, which are defined by (108), is an element of \mathcal{V} .

(c). Consider $v \in \mathcal{V}$ that is measurable with respect to r only, i.e., $v(\mathbf{w}) = \tilde{v}(R(\mathbf{w}))$ for some function \tilde{v} . Direct calculations give $\nabla_i v(\mathbf{w}) = \nabla \tilde{v}(r) \left(\frac{w_i}{r} \right)^{\gamma/(1-\gamma)}$. It follows that the value of $\Upsilon(v)$ for v that depends on \mathbf{w} only through $r = R(\mathbf{w})$ can be re-written as

$$\Upsilon(v) = \int_0^\infty \tilde{v}(r) (\mathbb{E}[\alpha|R=r] - 1) l_r(r) dr + \int_0^\infty \left((\nabla \tilde{v}(r))^\gamma r^{1+1/\gamma} - \nabla \tilde{v}(r) r \right) l_r(r) dr.$$

Integrate by parts as in the first part of this proof and pointwise optimize to obtain

$$(\nabla \tilde{v}(r))^{1/\gamma} r^{1/\gamma} = 1 + \frac{A(r) - L_r(r)}{\gamma r l_r(r)} = 1 + \lambda^{fam,*}(r), \quad (109)$$

where $A(r) := \int_0^r \mathbb{E}(\alpha | R = s) l_r(s) ds$. Since $\lambda^{fam,*}$ is bounded from below by some $\underline{\lambda} > -1$, by the exactly same argument as in the proof of the sufficiency part of Proposition 10, $\tilde{v}(R(\mathbf{w}))$, which is defined by (109), is an element of \mathcal{V} . \square

D.8.3 Proof of Corollary 7

Proof. To see the first part of the corollary, add up two equations (28) for $\mathbb{E}[\lambda_i | w_i = t]$ pre-multiplied by $\frac{g_i(t)}{g_1(t) + g_2(t)}$. The claim then follows from Part (a) of Proposition 9. As for the last two parts of the corollary, they directly follow from equations (39), (40) and Proposition (9). \square

E Quantitative analysis

E.1 Calibration

We use data from the 2020 CPS survey. In our dataset, we have pre-tax earnings of 11087 couples, each consisting of two individuals who (a) have a spouse in the same household, (b) worked for at least 20 weeks in 2020, (c) are 25-65 years old. Our measure of earnings includes only wage earnings. The sample is representative of approximately 42 million people.

We suppose that the data comes from a symmetric environment with $\gamma = 1/4$; thus, we symmetrize the dataset by creating one more copy of every household in which the identities of two spouses are interchanged. This gives us 2×11087 couples with identical distributions of earnings for each spouse and the same dependence patterns as before. We normalize earnings by 100 thousand so that the average value of individual earnings in the dataset equals 0.75.

Following Guner et al. (2014) and Heathcote et al. (2017) we assume that the data is generated with the following tax function: $T(y_1, y_2) = (y_1 + y_2) - \nu(y_1 + y_2)^{1-\tau}$. Guner et al. (2014) estimated (τ, ν) for married couples using the IRS data in which earnings are normalized by 53 thousand. Since we normalize earnings by 100 thousand, we adjust their estimate, which is $\tau = 0.06$ and $\nu = 0.91$, so that total tax bills in dollar terms are identical. The parameter τ doesn't need any adjustment but $\nu = 0.91 \times (\frac{53}{100})^\tau$.

Given the assumed log-linear tax schedule, each couple solves

$$\max_{(y_1, y_2) \geq \mathbf{0}} \nu (y_1 + y_2)^{1-\tau} - \sum_{i=1}^2 \gamma \left(\frac{y_i}{w_i} \right)^{1/\gamma},$$

which allows us to express unobserved productivities as a function of observed earnings (equation (44)) and construct the empirical distribution of productivities.

We calibrate a marginal distribution of productivities and their copula separately. Recall that the marginal G is assumed to follow a PLN distribution with parameters $(a, \mu, \sigma) \in \mathbb{R}_{++} \times \mathbb{R} \times \mathbb{R}_{++}$, that is

$$G(t) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right) - t^{-a} \exp(a\mu + a^2\sigma^2/2) \Phi\left(\frac{\ln t - \mu - a\sigma^2}{\sigma}\right).$$

Our first target moment is the Pareto statistic (computed with 183 observations at t that correspond to 99% percentile of the empirical cdf). In our sample this moment equals to 2.95, and since

$$\lim_{t \rightarrow \infty} \frac{\mathbb{E}[w_i | w_i \geq t]}{\mathbb{E}[w_i | w_i \geq t] - t} = a,$$

we set a to be 2.95. The second target moment is the Gini coefficient. It equals to 0.31 in the dataset. It can be shown (e.g., see Colombi (1990)) for a PLN distribution it is given by

$$2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1 + 2\frac{e^{a(a-1)\sigma^2}}{2a-1}\Phi\left(\frac{(1-2a)\sigma}{\sqrt{2}}\right),$$

where Φ is the standard normal distribution. This gives us $\sigma = 0.4$. Our final target moment is the mean value of individual productivities that equals 0.81 in the sample. Using the closed form expression

$$\mathbb{E}[w_i] = \frac{a}{a-1} e^{\mu + \sigma^2/2},$$

we get $\mu = -0.71$.

As for the copula of (w_1, w_2) , we calibrate it using the Kendall's tau dependence coefficient (see Chapter 5 in Nelsen (2006)), which is a rank measure of concordance, theoretically:

$$\mathbb{P}((w_1 - \tilde{w}_1)(w_2 - \tilde{w}_2) > 0) - \mathbb{P}((w_1 - \tilde{w}_1)(w_2 - \tilde{w}_2) < 0),$$

where (w_1, w_2) and $(\tilde{w}_1, \tilde{w}_2)$ are independent copies of productivities. Clearly, this statistic only depends on the underlying copula, not on G , and closed form expressions are available for many copulas. In our dataset, it equals to 0.21. We tried several copulas and found that the Gaussian one fits the data very well. For the Gaussian copula, Kendall's tau is given by $\frac{2 \arcsin \rho}{\pi}$, where ρ is its correlation parameter. This gives us $\rho = 0.33$. In Section E.4 of the appendix we re-calibrate the model to the FGM copula, i.e., $C(u_1, u_2) = u_1 u_2 [1 + \rho(1 - u_1)(1 - u_2)]$, for which Kendall's tau is given by $\frac{2\rho}{9}$, thus $\rho = 0.96$.

E.2 Numerical approach

In this section, we overview the numerical approach that we used to find the optimal taxes. First of all, we discretize the problem using a finite logarithmic grid of 399 equally spaced

productivities. The grid is logarithmic in the sense that a ratio of two consecutive points is constant. This allows to improve accuracy at the left tail and capture the thick right tail. Let $\{w^1, \dots, w^{400}\}$, where $w^1 = 0.12$ and $w^{400} = 10$, be this grid. The 400th point is added to ensure that our discretized relaxed problem can approximate the original relaxed problem in which the domain is unbounded. It will be convenient to also define $w^0 := 0$.

We numerically solve a relaxed problem that only contains downward incentive constraints, one for each spouse, that is

$$\begin{aligned} & \max_{v, \mathbf{y} \geq \mathbf{0}} \sum_{n_1, n_2=1}^{400} v(w^{n_1}, w^{n_2}) (\alpha(w^{n_1}, w^{n_2}) - 1) f(w^{n_1}, w^{n_2}) + \\ & + \sum_{i=1}^2 \sum_{n_1, n_2=1}^{400} \left(y_i(w^{n_1}, w^{n_2}) - \gamma \left(\frac{y_i(w^{n_1}, w^{n_2})}{w^{n_i}} \right)^{1/\gamma} \right) f(w^{n_1}, w^{n_2}) \end{aligned}$$

subject to the following set of incentive constraints: for all $n_i = 2, \dots, 400$, $n_{-i} = 1, \dots, 400$ and $i = 1, 2$,

$$v(w^{n_i}, w^{n_{-i}}) \geq v(w^{n_{i-1}}, w^{n_{-i}}) + \gamma y_i(w^{n_{i-1}}, w^{n_{-i}}) \left((w^{n_{i-1}})^{-1/\gamma} - (w^{n_i})^{-1/\gamma} \right).$$

In this problem f is set to be

$$f(w^{n_1}, w^{n_2}) = \begin{cases} \mathbb{P}(w^{n_{i-1}} < w_i \leq w^{n_i} \quad \forall i), & n_i, n_{-i} < 400; \\ \mathbb{P}(w^{n_{i-1}} < w_i, w^{n_{-i-1}} < w_{-i} \leq w^{n_{-i}}), & n_i = 400 > n_{-i}; \\ \mathbb{P}(w^{n_{i-1}} < w_i \quad \forall i), & n_i = n_{-i} = 400. \end{cases}$$

And, α is normalized so that $\sum_{n_1, n_2=1}^{400} \alpha(w^{n_1}, w^{n_2}) f(w^{n_1}, w^{n_2}) = 1$.

The solution to the relaxed problem is easy to find, and it is always the case that all incentive constraints are binding. Given this solution, we then numerically verify all remaining (global) incentive constraints. In all cases, we found that the first-order approach holds.

E.3 Comparison of distortions

Figure 9 plots the optimal distortions $\lambda_i^*(\cdot, w_{-i})$ when w_{-i} is fixed at its median value, i.e., 0.66, against the average distortions $\mathbb{E}[\lambda_i^* | w_i = t]$. For both the Gaussian and FGM copulas, the average distortions (red dashed lines) are very close to the optimal distortion with $w_{-i} = 0.66$ (solid blue lines).

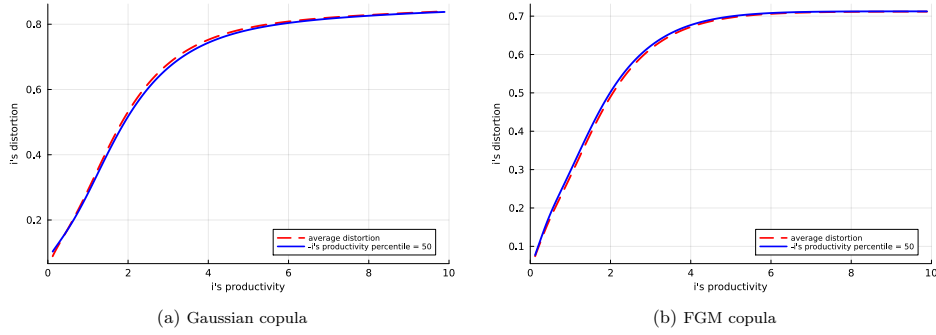


Figure 9: Comparison of average and pointwise distortions at 50th percentile

E.4 FGM copula

In this section, we report the optimal taxes when the empirical distribution of productivities is calibrated to the FGM copula. Recall that the FGM copula is defined to be $C(u_1, u_2) = u_1 u_2 [1 + \rho(1 - u_1)(1 - u_2)]$, where $\rho \in [-1, 1]$. We calibrate parameter ρ to be 0.96, matching the Kendall's tau statistic. Figure 10 shows the FGM copula fits isoquants of the empirical distribution fairly well (Panel (b)), but it does not match the speed of convergence to tail-independence (Panels (c) and (d)).

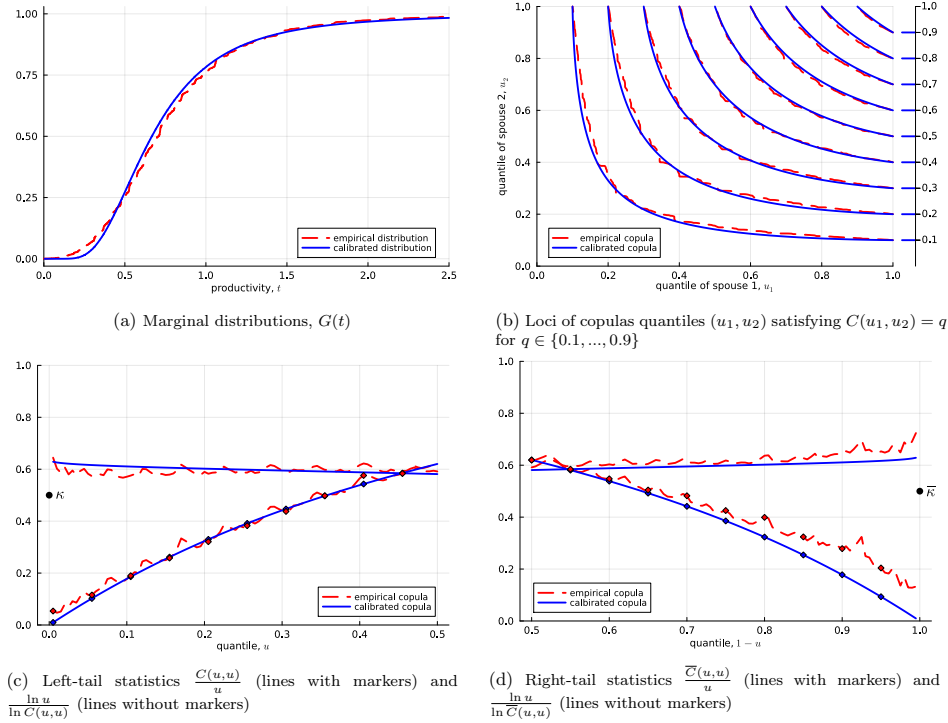


Figure 10: Empirical and calibrated joint distributions of productivities

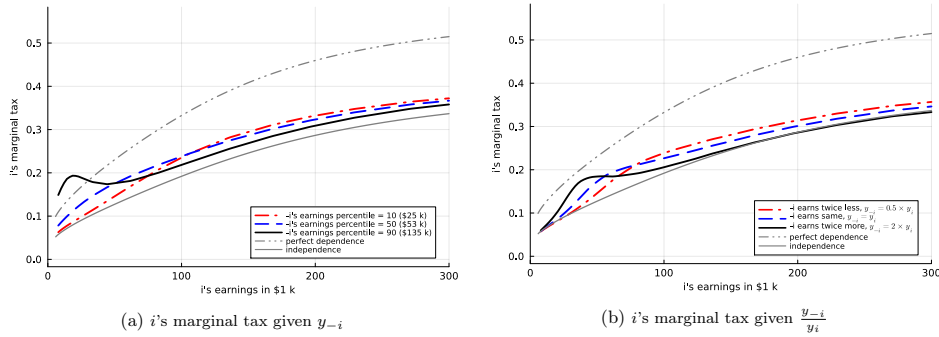


Figure 11: Optimal taxes, $m = 0.35$ and $k = 1$

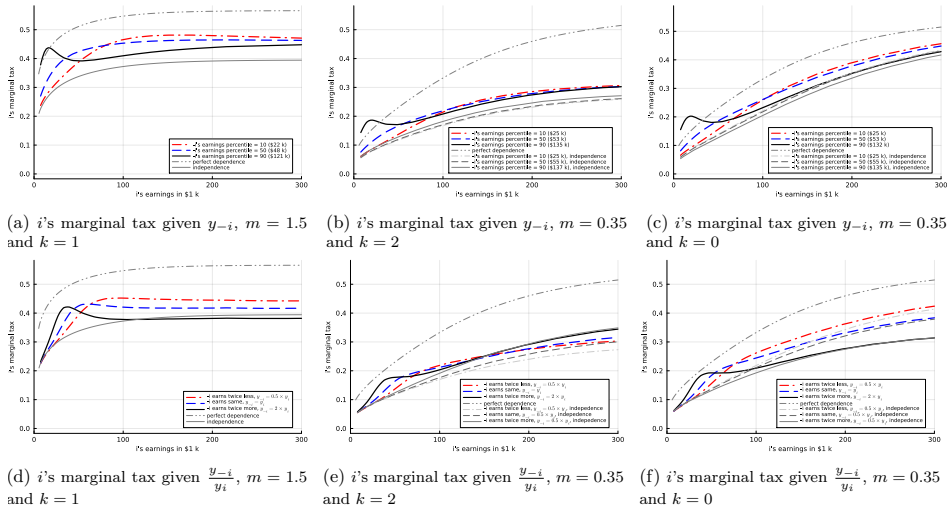


Figure 12: Optimal taxes, robustness to m and k

Figure 11 depicts the optimal taxes computed under the FGM copula in our baseline specification with separable weights and not too redistributive planner. The optimal taxes are much more negatively jointed at the top as compared to the optimal taxes computed under the Gaussian copula (Figure 6). Consistent with Proposition 4 the sign of jointness flips at a threshold, and it is negative (positive) for all large (small) w_i .

Figure 12 illustrates robustness of the optimal taxes. Qualitatively, redistributiveness and modularity of α have identical implications for the optimal taxes under the FGM copula and the Gaussian copula (Figure 7). First, if the planner is more redistributive, then the optimal taxes are larger and negative jointness occurs at lower levels. Second, supermodularity amplifies negative jointness at the top, and submodularity amplifies positive jointness at the bottom.

Figure 13 plots the optimal taxes as a function of the total family earnings and the share of the secondary earner. As in the case of the Gaussian copula (Figure 8), the optimal taxes vary substantially with the share of the secondary earner.

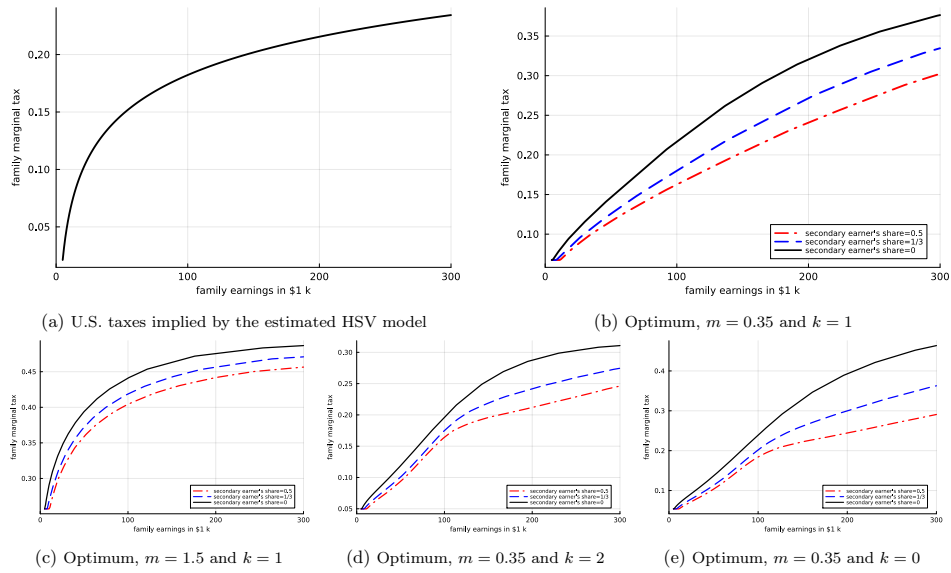


Figure 13: Marginal taxes on family earnings