Suplement to: Optimal Income Taxation and Commitment on the Labor Market

Paweł Doligalski

University of Bristol

November 30, 2018

Abstract

The supplementary material contains the proof of the Revelation Principle and associated results (Appendix A) as well as the details of the calibration of the quantitative model and the computational algorithm (Appendix B).

A. Revelation principle

Below I prove the revelation principle for the general dynamic single-agent mechanism design problem with hidden information and hidden action.

A.1. A general proof of the Revelation Principle

Consider two probability spaces $(R_1, \mathcal{B}_{R_1}, \mu_{R_1})$ and $(R_2, \mathcal{B}_{R_2}, \mu_{R_2})$. By Theorem 18.2 in Billingsley (2008) there exists a unique product probability space $(R_1 \times R_2, \mathcal{B}_{R_1 \times R_2}, \mu_{R_1 \times R_2})$, where $\mathcal{B}_{R_1 \times R_2}$ is a σ-algebra generated by $\mathcal{B}_{R_1} \times \mathcal{B}_{R_2}$ and $\mu_{R_1 \times R_2}(A \times B) = \mu_{R_1}(A) \cdot \mu_{R_2}(B)$ for all $A \in \mathcal{B}_{R_1}$ and $B \in \mathcal{B}_{R_2}$. In what follows, I will apply this theorem repeatedly.

The economy operates for $T \leq \infty$ periods. The fundamentals of the economy are given by three measurable spaces: a type space $(\Theta, \mathcal{B}_\Theta)$, an outcome space $(X, \mathcal{B}_X)$ and an action space $(A, \mathcal{B}_A)$, as well as a probability measure over full type histories μ_{Θ^T}, such that $(\Theta^T, \mathcal{B}_{\Theta^T}, \mu_{\Theta^T})$ is a probability space.

A randomization device is an arbitrary probability space. I will use three devices $(R_i, \mathcal{B}_{R_i}, \mu_{R_i})$ with $i \in \{x, m, a\}$, standing consecutively for the outcome randomization device, the reporting randomization device and the action randomization device.

A message space is a measurable space $(M, \mathcal{B}_M)$. A *reporting strategy* ρ consists of a reporting randomization device $(R_m, \mathcal{B}_{R_m}, \mu_{R_m})$ and a reporting function $m : \Theta^T \times R_m^T \times X^T \to M^T$, where m_t is $(\theta^t, r_m^t, x^{t-1})-$measurable for all $t \leq T$. An *action strategy* α contains an action randomization device $(R_a, \mathcal{B}_{R_a}, \mu_{R_a})$ and an action function: $a : \Theta^T \times R_m^T \times X^T \times R_a^T \to A^T$, where a_t is $(\theta^t, r_m^t, x^t, r_a^t)-$measurable for all $t \leq T$. The measurability assumptions imply that a report or an action can depend on the entire past history, but cannot depend on the future.[1] The measurability assumptions also indicate the order within a time period: first a report is sent, then an outcome is allocated by a mechanism and finally an action is taken.

A *mechanism* consists of *(i)* a message space $(M, \mathcal{B}_M)$, *(ii)* an outcome randomization device $(R_x, \mathcal{B}_{R_x}, \mu_{R_x})$, *(iii)* an outcome function $x : M^T \times R_x^T \to X^T$, where x_t is $(x^t \times r_x^t)-$measurable for all $t \leq T$ and *(iv)* a recommended action strategy α. In a single agent case, which is the sole focus of this paper, a recommended action strategy does not play any substantive role and is included in a mechanisms for convenience in setting up the planner's problem.[2]

Since outcomes depend on reports, which in turn depend on outcomes, it is useful to define an auxiliary function which disentangles this relationship. For an outcome function x and a reporting function m define a function $\xi_{x,m}$ such that

$$\xi_{x,m}(\theta^T, r_x^T, r_m^T) \equiv x(m(\theta^T, r_m^T, \xi_{x,m}(\theta^T, r_x^T, r_m^T)), r_x^T). \tag{1}$$

Since a report in period t can depend on outcomes up to period $t-1$, it is easy to see that $\xi_{x,m}$ has uniquely defined values. Denote the expectation of some function $g : X^T \times A^T \times \Theta^T \to \mathbb{R}$ with respect to mechanism Ψ, reporting strategy ρ and action strategy α, whenever the expectation is well defined, as

$$\mathbb{E}_{\Psi,\rho,\alpha}\left\{g(x, a, \theta^T)\right\} \equiv \int g(\xi_{x,m}(\theta^T, r_x^T, r_m^T), a(\theta^T, r_m^T, \xi_{x,m}(\theta^T, r_x^T, r_m^T), r_a^T), \theta^T) d\mu_{\Theta^T \times R_x^T \times R_m^T \times R_a^T}(\theta^T, r_x^T, r_m^T, r_a^T). \tag{2}$$

Conditional expectations are defined analogously.

I allow for additional restrictions on the equilibrium choices of agents by considering arbitrary equilibrium constraints. For instance, in the main body of this paper the relevant equilibrium constraints are the zero profit constraints and the limited commitment constraints. Define an equilibrium constraint as (g, t), where $g : X^T \times A^T \times \Theta^T \to \mathbb{R}$ and $t \in \{0, ..., T\}^4$. A mechanism Ψ with an associated strategies (ρ, α) satisfies the equilibrium constraint (g, t) if $\mathbb{E}_{\Psi,\rho,\alpha}\{g(x, a, \theta^T) \mid \theta^{t_1}, r_m^{t_2}, x^{t_3}, r_a^{t_4}\} \geq 0$. For any partic-

[1] For notational ease I do not allow reporting to depend on the action randomization device. Any joint randomization of reports and actions is governed by the reporting randomization device.

[2] In contrast, in the case with many agents the recommended action may be useful by coordinating actions of different agents (Myerson 1982).

ular set of equilibrium constraints, a set of strategies (ρ, α) satisfying these equilibrium constraints is given a mechanism Ψ is $C(\Psi)$.

Define a utility function as a continuous function $U : X^T \times A^T \times \Theta^T \to \mathbb{R}$. A reporting strategy ρ and an action strategy α constitute an *equilibrium of a mechanism* Ψ if

$$(\rho, \alpha) \in \arg \max_{(\hat{\rho}, \hat{\alpha}) \in \mathcal{C}(\Psi)} \mathbb{E}_{\Psi, \hat{\rho}, \hat{\alpha}} \left\{ U(x, \hat{a}, \theta^T) \right\}. \tag{3}$$

A resource function is a continuous function $H : X^T \times A^T \times \Theta^T \to \mathbb{R}$. A mechanism Ψ with an associated equilibrium (ρ, α) is feasible if $\mathbb{E}_{\Psi, \rho, \alpha} \left\{ H(x, a, \theta^T) \right\} \geq 0$. A *direct mechanism* is a mechanism with a message space $(\Theta, \mathcal{B}_\Theta)$. A *truthful reporting strategy*, denoted by ρ^*, is a reporting strategy with a reporting function $m^*(h) = \theta^T$ for all $h \in \mathcal{H}$. A direct mechanism is *incentive-compatible* if it has an equilibrium with the truthful reporting strategy and the recommended action strategy. A direct mechanism is *incentive-feasible* if it is incentive-compatible and feasible at the truthful equilibrium. Take a mechanism $((M, B_M), (R_x, \mathcal{B}_{R_x}, \mu_{R_x}), x, \alpha_\Psi)$ with associated equilibrium reporting strategy $\rho = ((R_m, \mathcal{B}_{R_m}, \mu_{R_m}), m)$ and action strategy $\alpha = ((R_a, \mathcal{B}_{R_a}, \mu_{R_a}), a)$. Define a *social choice function* as the implied equilibrium mapping from types and realizations of the randomization devices to outcomes and actions: $\theta^T \times r_x^T \times r_m^T \times r_a^T \to \xi_{x,m} \times a$. Two mechanism are *equivalent* if they have the same social choice function.

Lemma A.1 (Revelation Principle)**.** *For any feasible mechanism there exists an equivalent mechanism which is direct and incentive-feasible.*

Proof. Take some feasible mechanism $\Psi = ((M, B_M), (R_x, \mathcal{B}_{R_x}, \mu_{R_x}), x, \alpha_\Psi)$ and its equilibrium reporting strategy $\rho = ((R_m, \mathcal{B}_{R_m}, \mu_{R_m}), m)$ and action strategy $\alpha = ((R_a, \mathcal{B}_{R_a}, \mu_{R_a}), a)$. Note that the recommended action strategy α_Ψ can be different than the equilibrium action strategy α.

Take a probability space $(R'_m, \mathcal{B}_{R'_m}, \mu_{R'_m})$ with identical structure as the randomization device of the agent implied by ρ.[3] Construct a direct mechanism Γ by (i) setting the message space to $(\Theta, \mathcal{B}_\Theta)$, (ii) setting the outcome randomization device to $(R_x \times R'_m, \mathcal{B}_{R_x \times R'_m}, \mu_{R_x \times R'_m})$ and (iii) setting the outcome function to $x^*(\theta^T, (r_x, r'_m)^T) \equiv \xi_{x,m}(\theta^T, r_x^T, r_m'^T)$ for all $(\theta^T, r_x^T, r_m'^T) \in \Theta^T \times R_x^T \times R_m'^T$, (iv) setting the recommended action strategy to α. Then for any function $g(x^T, a^T, \theta^T)$ it is straightforward to see that

$$\mathbb{E}_{\Psi, \rho, \alpha} \left\{ g(x, a, \theta^T) \right\} = \mathbb{E}_{\Gamma, \rho^*, \alpha} \left\{ g(x^*, a, \theta^T) \right\} \tag{4}$$

as long as the original expectation is well defined.

It implies that the direct mechanism Γ evaluated at the truthful reporting strategy and action strategy α is feasible (replace g with the resource function H) and yields the

[3] $(R_m, \mathcal{B}_{R_m}, \mu_{R_m})$ and $(R'_m, \mathcal{B}_{R'_m}, \mu_{R'_m})$ can be understood as two independent, identically distributed lotteries.

same expected utility to the agent as the original mechanism (replace g with the utility function U). We can repeat the same reasoning with expectation of g conditional on any partial individual history to verify that all the equilibrium constraints are satisfied as well. What remains to be shown is that the mechanism Γ is incentive-compatible.

Assume that the direct mechanism Γ is not incentive-compatible. Specifically, suppose that there exist a reporting strategy $\hat{\rho} = ((\hat{R}_m, \mathcal{B}_{\hat{R}_m}, \mu_{\hat{R}_m}), \hat{m})$ and action strategy $\hat{\alpha} = ((\hat{R}_a, \mathcal{B}_{\hat{R}_a}, \mu_{\hat{R}_a}), \hat{a})$ which together with Γ satisfy all the equilibrium constraints and are such that $\mathbb{E}_{\Gamma,\hat{\rho},\hat{\alpha}} \left\{U(\hat{x}, \hat{a}, \theta^T)\right\} > \mathbb{E}_{\Gamma,\rho^*,\alpha} \left\{U(x^*, a, \theta^T)\right\}$. Then we can define a reporting strategy $\hat{\hat{\rho}}$ associated with the mechanism Ψ as

$$\hat{\hat{\rho}} = ((R_m \times \hat{R}_m, \mathcal{B}_{R_m \times \hat{R}_m}, \mu_{R_m \times \hat{R}_m}), \hat{\hat{m}} : \Theta^T \times R_m^T \times \hat{R}_m^T \times X^T \to M^T),$$

where $\hat{\hat{m}}(\theta^T, r_m^T, \hat{r}_m^T, x^T) \equiv m(\hat{m}(\theta^T, \hat{r}_m^T, x^T), r_m^T, x^T)$ for all $(\theta^T, r_m^T, \hat{r}_m^T, x^T) \in \Theta^T \times R_m^T \times \hat{R}_m^T \times X^T$. Note that we have

$$x(\hat{\hat{m}}(\theta^T, r_m^T, \hat{r}_m^T, x^T), r_x^T) = x(m(\hat{m}(\theta^T, \hat{r}_m^T, x^T), r_m^T, x^T), r_x^T) = x^*(\hat{m}(\theta^T, \hat{r}_m^T, x^T), (r_x^T, r_m^T)), \tag{5}$$

for all $(\theta^T, r_x^T, r_m^T, \hat{r}_m^T, x^T) \in \Theta^T \times R_x^T \times R_m^T \times \hat{R}_m^T \times X^T$, where the first equality follows from the definition of $\hat{\hat{m}}$ and the second from the definition of x^*. Then for any function $g(x^T, a^T, \theta^T)$, for which the expectation is well defined, we can repeat the reasoning above to show that $\mathbb{E}_{\Gamma,\hat{\rho},\hat{\alpha}} \left\{g(x^*, \hat{a}, \theta^T)\right\} = \mathbb{E}_{\Psi,\hat{\hat{\rho}},\hat{\alpha}} \left\{g(x, \hat{a}, \theta^T)\right\}$ and analogously for conditional expectations. It means that the mechanism Γ with the strategies $(\hat{\hat{\rho}}, \hat{\alpha})$ satisfies all the equilibrium constraints and

$$\mathbb{E}_{\Psi,\hat{\hat{\rho}},\hat{\alpha}} \left\{U(x, \hat{a}, \theta^T)\right\} = \mathbb{E}_{\Gamma,\hat{\rho},\hat{\alpha}} \left\{U(x^*, \hat{a}, \theta^T)\right\} > \mathbb{E}_{\Gamma,\rho^*,\alpha} \left\{U(x^*, a, \theta^T)\right\} = \mathbb{E}_{\Psi,\rho,\alpha} \left\{U(x, a, \theta^T)\right\}, \tag{6}$$

which contradicts the fact that (ρ, α) is an equilibrium of the mechanism Ψ. ■

A.2. Additional results for the specific framework of this paper

In order to make a mapping between the general framework from the previous subsection and the setting of the main text of this paper, I introduce the following assumptions and naming conventions. The outcome space is set to $X = \mathbb{R}_+ \times \mathbb{R}$ and the outcome function x is split into a consumption function c and labor income function y. The action space is $A = \mathbb{R}_+$ and I call the action function a the labor function n. The utility function is $U(c^T, y^T, n^T, \theta) = \sum_{t=1}^T \beta^{t-1}(u(c_t^T) - v(n_t^T))$. Finally, equilibrium constraints on the reporting strategy $\rho = ((R_m, \mathbb{B}_{R_m}, \mu_{R_m}), m)$ and the labor strategy $\nu = ((R_n, \mathbb{B}_{R_n}, \mu_{R_n}), n)$ are the zero profit constraints

$$\mathbb{E}_{\Psi,\rho,\nu} \left\{\pi_1(y, n, \theta^T) \mid \theta_1^T = \theta\right\} = 0 \text{ for all } \theta \in \Theta \tag{7}$$

and the limited commitment constraints

$$\kappa \geq \mathbb{E}_{\Psi,\rho,\nu}\left\{\pi_t(y,n,\theta^T) \mid \theta^t, \hat{r}_m^t, x^{t-1}, r_n^{t-1}\right\} \geq -\phi$$
$$\text{for all } t \leq T \text{ and all } (\theta^t, r_m^t, x^{t-1}, r_n^{t-1}) \in \Theta^t \times R_m^t \times X^{t-1} \times R_n^{t-1}. \quad (8)$$

Two additional lemmas below justifies not using introducing the labor (or action, as called in the previous subsection) randomization device or the outcome randomization device in the specific setting of this paper.

Lemma A.2. *Suppose that* $U(c^T, y^T, n^T, \theta) = \sum_{t=1}^T \beta^{t-1}(u(c_t^T) - v(n_t^T))$*, where* v *is increasing and strictly convex, and that the equilibrium constraints are given by (7) and (8). The equilibrium labor function does not depend on the labor randomization device.*

Proof. Consider a mechanism Ψ with an associated equilibrium reporting strategy ρ and the equilibrium labor strategy ν containing a labor randomization device $(R_n, \mathbb{B}_{R_n}, \mu_{R_n})$ and a labor function n. Construct a new labor function n' such that $n_t'(\theta^T, r_m^T, x^T) = \mathbb{E}_{\Psi,\rho,\alpha}\{n_t(\theta^T, r_m^T, x^T, r_n^T) \mid \theta^t, r_m^t, x^t\}$ for all $t \leq T$ and all (θ^T, r_m^T, x^T). Construct a labor strategy ν' containing the labor function n'. Since all the equilibrium constraints are linear in labor, it is straightforward to see that they are unaffected and $(\rho, \nu') \in \mathcal{C}(\Psi)$. Since the variability of labor supply is reduced and disutility from labor is strictly convex, the agent obtains strictly higher expected utility, which contradicts the initial claim that (ρ, ν) is an equilibrium. ■

Lemma A.3. *Suppose that* $U(c^T, y^T, n^T, \theta) = \sum_{t=1}^T \beta^{t-1}(u(c_t^T) - v(n_t^T))$*, where* u *is concave and* v *is strictly convex, both strictly increasing, and that the equilibrium constraints are given by (7) and (8). For any stochastic, incentive-feasible mechanism there is a deterministic, incentive-feasible mechanism which yields the same expected utility to all initial types.*

Proof. Consider some stochastic incentive-feasible mechanism $\Psi = ((R_x, \mathbb{B}_{R_x}, \mu_{R_x}), c, y, n)$ with associated equilibrium strategies (ρ^*, n), where n is a deterministic labor function. Construct a new, deterministic mechanism $\Psi' = (c', y', n)$ in the following way. First, for each period $t \leq T$ and each history of productivity reports θ^T find a constant consumption level $c_t'(\theta^T)$ which yields the same utility from consumption as the original mechanism:

$$u(c_t'(\theta^T)) = \mathbb{E}_{\Psi,\rho^*}\left\{u(c_t) \mid \theta^T\right\} \text{ for all } . \quad (9)$$

Since agents face less consumption risk and have the same utility from consumption, the planner saves some resources. Second, choose income function y' which for each time period $t \leq T$ and for each history of reports $\theta^T \in \Theta^T$ is equal to the expected income conditional on the history of reports:

$$y_t'(\theta^T) = \mathbb{E}_{\Psi,\rho^*}\left\{y_t \mid \theta^T\right\}. \quad (10)$$

I will show that a mechanism $\Psi' = (c', y', n)$ has an equilibrium (ρ^*, n). Conditional on reporting strategy ρ, agents choose labor strategy by solving

$$\min_{n} \mathbb{E}_{\Psi',\rho} \left\{ \sum_{t=1}^{T} \beta^{t-1} v(n_t) \right\} \text{ s.t. 7 and 8.} \tag{11}$$

Crucially, neither of the constraints depend on how labor income varies with r_x^T, what matters is the expected income conditional on a history of reports. For instance, the limited commitment constraints depend on $\mathbb{E}_{\Psi',\rho}\left\{\pi_t \mid \theta^t, r_m^t\right\}$, which can be expressed as

$$\mathbb{E}_{\Psi',\rho}\left\{\pi_t \mid \theta^t, r_m^t\right\} = \mathbb{E}_{\Psi',\rho} \left\{ \sum_{s=t}^{T} \beta^{s-t} \left(\theta_s^T n_s - y_s\right) \mid \theta^t, r_m^t \right\} \tag{12}$$

$$= \mathbb{E}_{\Psi',\rho} \left\{ \sum_{s=t}^{T} \beta^{s-t} \left(\theta_s^T n_s - \mathbb{E}_{\Psi',\rho}\left\{y_s \mid \theta^s, r_m^s\right\}\right) \mid \theta^t, r_m^t \right\} \tag{13}$$

and analogously for the zero profits constraints. Note that the expectation of income conditional on a history of reports is the same in Ψ and Ψ'. Therefore, conditional on a reporting strategy, agents would choose the same labor strategy in Ψ and Ψ'. Furthermore, by construction a given reporting strategy yields the same utility from consumption in Ψ and Ψ'. Therefore, if (ρ^*, n) is an equilibrium of the mechanism Ψ, it is also an equilibrium of the mechanism Ψ'.

Therefore, Ψ' is a incentive-compatible mechanism yielding identical expected utility to each initial type as Ψ. Since Ψ' involves more consumption insurance, it requires less resources than Ψ and, hence, is feasible. ■

B. Appendix to the quantitative exercise

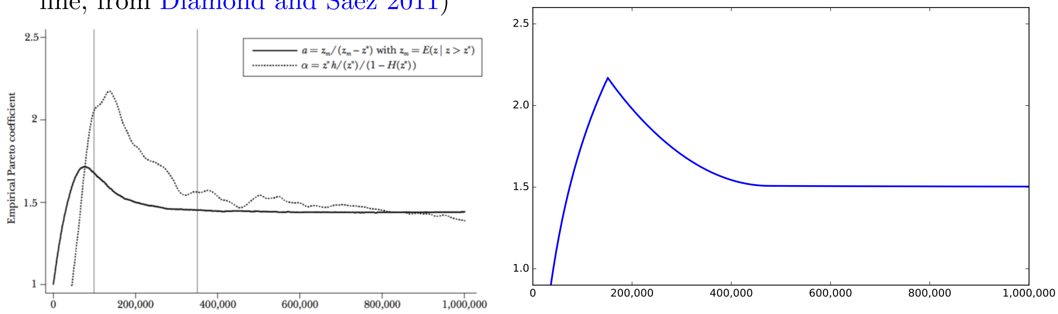
B.1. Matching the US income distribution

Denote the empirical cdf of income by $H(\cdot)$ and the empirical density of income by $h(\cdot)$. Incomes below \$150,000 are distributed according to the lognormal distribution with parameters μ_{LN} and σ_{LN}. Parameters μ_{LN} and σ_{LN} are chosen to match the mean income of \$64,000 as in Sachs, Tsyvinski, and Werquin (2016) and to ensure the continuity of the inverse hazard ratio at \$150,000. Incomes above \$150,000 are distributed according to the distribution that matches the pattern of the inverse hazard ratio $yh(y)/(1-H(y))$ from Diamond and Saez (2011). I approximate the inverse hazard ratio between \$150,000 and \$475,000 with a polynomial $\tilde{g}(y)$. First, I use the second degree polynomial matching the value of the inverse hazard ratio at \$150,000 and \$475,000 and having a local minimum at \$475,000. Define $\bar{H}(y) = 1 - H(y)$ and note that we can write the inverse hazard ratio as $-y\bar{H}'(y)/\bar{H}(y)$. Then it is easy to see that $\bar{H}(y) = Ce^{G(y)}$ in this income

Figure 1: Comparison of empirical and calibrated inverse hazard ratio of income

(a) Empirical inverse hazard ratio (dotted line, from Diamond and Saez 2011)

(b) Calibrated inverse hazard ratio



interval, where $G(y)$ is the antiderivative of $-\tilde{g}(y)/y$. The constant C is chosen such that $\bar{H}(y)$ is continuous at \$150,000. Then $1 - \bar{H}(y)$ is the corresponding cumulative distribution function of income. Second, above \$475,000 I set the inverse hazard ratio to a constant which amounts to a Pareto distribution for which I choose a support parameter to ensure the continuity of the cumulative density. Figure 1 demonstrates the fit of the calibrated distribution to the empirical hazard ratio of income.

B.2. Computing the optimal tax schedule under two-sided commitment with CRRA utility

Recall that the utility function is $u(c) = c^{1-\sigma}/(1-\sigma)$, $\sigma \geq 0$, for which the coefficient of the absolute risk aversion at the consumption level c is σ/c.

Risk neutral agents ($\sigma = 0$). The no-randomization constraint requires that the marginal tax rates are non-decreasing. We can apply this constraint in the manner similar to ironing from the standard screening model (see Mussa and Rosen (1978)). First, compute the Mirrleesian tax schedule $\mathcal{T}(\cdot)$ ignoring the no-randomization constraint. If the tax schedule is has non-decreasing tax rates everywhere, it is optimal under two-sided commitment. Otherwise, identify income intervals over which the tax rates are decreasing, flatten the tax rates in these intervals such that they are non-decreasing everywhere and optimize with respect to the tax rate in each interval. The optimal income interval $[y(\theta_1), y(\theta_2)]$ at which the tax rate is constant and equal $\mathcal{T}'(y(\theta_1))$ satisfies the following optimality condition:

$$\frac{\mathcal{T}'(y(\theta_1))}{1-\mathcal{T}'(y(\theta_1))}\varepsilon \int_{\theta_1}^{\theta_2} y(\theta) d\mu_\Theta(\theta)$$
$$= \int_{\theta_1}^{\infty} (\min\{y(\theta), y(\theta_2)\} - y(\theta_1)) d\mu_\Theta(\theta) - \int_{\theta_1}^{\infty} (\min\{y(\theta), y(\theta_2)\} - y(\theta_1)) d\tilde{\mu}_\Theta(\theta). \quad (14)$$

This optimality condition can be easily obtained perturbing the fixed tax rate at income interval $[y(\theta_1), y(\theta_2)]$ and requiring that the welfare gain from this perturbation is zero.

Risk averse agents ($\sigma > 0$). When agents are risk averse and have a coefficient of absolute risk aversion which varies with a consumption level, a change of a tax rate at some income level leads to a change of the coefficient of the absolute risk aversion at higher income levels, which leads to tightening or relaxation of the no-randomization constraint at higher income levels. This concern is not incorporated in the Mirrleesian tax formulas.[4] I avoid this complication by assuming that the no-randomization constraint is binding in the income interval $[0, \bar{y}]$ for some $\bar{y}$ and is slack elsewhere. It is a natural assumption to make since the empirical tax rates in the static Mirrlees model are U-shaped (Diamond 1998; Saez 2001). Under this assumption I can use the standard tax formula of Saez (2001) for incomes above $\bar{y}$. For incomes in interval $[0, \bar{y}]$ the tax rates can be extracted from the binding no-randomization constraint. A no-randomization constraint that is binding in $[0, \bar{y}]$ means that $\tilde{v}(y(\theta), \theta)$ is constant for all θ such that $y(\theta) \leq \bar{y}$. Using the agents' first-order condition, it implies that $(1 - \mathcal{T}'(y))u'(y - \mathcal{T}(y))$ is constant for all $y \leq \bar{y}$. Given $\mathcal{T}(0)$ and $\mathcal{T}'(0)$ we can recover the entire tax schedule over $[0, \bar{y}]$. $\mathcal{T}(0)$ and $\mathcal{T}'(0)$ are then chosen to maximize the planner's objective subject to the budget constraint.

References

Billingsley, P. (2008): *Probability and measure.* John Wiley & Sons.

Diamond, P., and E. Saez (2011): "The case for a progressive tax: from basic research to policy recommendations," *Journal of Economic Perspectives*, 25(4), 165–90.

Diamond, P. A. (1998): "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates," *American Economic Review*, pp. 83–95.

Guiso, L., and M. Paiella (2008): "Risk aversion, wealth, and background risk," *Journal of the European Economic Association*, 6(6), 1109–1150.

Mussa, M., and S. Rosen (1978): "Monopoly and product quality," *Journal of Economic Theory*, 18(2), 301–317.

Myerson, R. B. (1982): "Optimal coordination mechanisms in generalized principal-agent problems," *Journal of Mathematical Economics*, 10(1), 67–81.

[4] If instead the utility from consumption had a constant coefficient of absolute risk aversion, the Mirrleesian tax formulas would still be valid at income levels where the no-randomization constraint is not binding, as in the case of risk neutral agents. However, empirical evidence supports utility functions with the absolute risk aversion that is decreasing with consumption (Guiso and Paiella 2008).

Sachs, D., A. Tsyvinski, and N. Werquin (2016): "Nonlinear tax incidence and optimal taxation in general equilibrium," Discussion paper, National Bureau of Economic Research.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68(1), 205–29.