The Review of Economic Studies Advance Access published July 22, 2016

Review of Economic Studies (2016) **0**, 1–24 doi:10.1093/restud/rdw033

Taxing the Rich

AUGUSTIN LANDIER
Toulouse School of Economics

and

GUILLAUME PLANTIN
Sciences Po

First version received October 2011*; final version accepted June* 2016 (*Eds.*)

Affluent households can respond to taxation with means that are not economically viable for the rest of the population, such as sophisticated tax plans and international tax arbitrage. This article studies an economy in which an inequality-averse social planner faces agents who have access to a tax-avoidance technology with subadditive costs, and who can shape the risk profile of their income as they see fit. Subadditive avoidance costs imply that optimal taxation cannot be progressive at the top. This in turn may trigger excessive risk-taking. When the avoidance technology consists in costly migration between two countries that compete fiscally, we show that an endogenous increase in inequality due to risk-taking makes progressive taxation more fragile, which vindicates in turn risk-taking and can lead to equilibria with regressive tax rates at the top, and high migrations of wealth towards the smaller country.

Key words: Optimal taxation, Tax avoidance

JEL Codes: H21, H26

1. INTRODUCTION

The taxation of affluent households periodically comes to the forefront of the public debate. The view that the rich should pay more taxes than they currently do recently gained influence in the U.S. and in Europe. Likely reasons include the need for fiscal consolidation resulting from the financial and economic crisis that erupted in 2008, and long-term trends of increasing income inequality and decreasing top marginal tax rates, in particular in the U.S. The "Buffett rule" proposed by the Obama administration responded in particular to the spread of the sentiment that effective tax rates have become overly regressive.[1] The view is that low effective tax rates for the rich result not only from low nominal tax rates, but also from increased tax avoidance by the most affluent households.

This article develops a new theoretical framework to study the taxation of the rich. Our motivation is twofold. First, the taxation of the rich has first-order implications for public finances, simply because affluent households collect a significant fraction of aggregate income. In 2014, the top quintile and percentile of the U.S. income distribution, respectively, collected 36% and 20% of aggregate income (Piketty and Saez, 2003 updated in 2015). Secondly, taxing the rich

1. The proposed rule applied a minimum tax rate of 30% on individuals making more than a million dollars a year.

raises issues that are, in our view, quite different from that raised by the taxation of the rest of the population.

Taxing affluent households raises specific issues because the rich can respond to taxation with means that are by and large unavailable to the rest of the population. This article studies the taxation of a population of agents that can avail themselves of two such means: tax avoidance—the minimization of one's tax liabilities by legal or quasi-legal means, and income-risk shifting.

First, regarding tax avoidance, the saying that "the poor evade and the rich avoid" epitomizes that avoidance is concentrated at the top of the income distribution. We believe that this is so because common tax-avoidance techniques are profitable only when spread over sufficiently large pre-tax resources. We actually view this as the key economic distinction between tax avoidance and tax evasion. Evasion is an outright breach of tax law, which should naturally be thought of as displaying diseconomies of scale. It seems evident that concealing larger amounts from the tax authority, and converting them into secret consumption comes at a higher unit cost. Obvious reasons include the difficulty of settling large transactions with cash, and the lack of discretion entailed by an affluent lifestyle.

In contrast, tax-avoidance techniques entail costs that are not very sensitive to the income base to which they are applied. There are two main forms of tax avoidance. First, a major source of tax avoidance consists in tax plans that shape the timing, nature, and amount of taxable income so as to minimize taxes. Typical schemes consist in relabelling labour income as capital income, or in borrowing against capital gains instead of realizing them to consume. The ability of private equity and hedge fund managers to structure their pay as carried interest, which is taxed as dividends instead of labour income, is a simple example of such avoidance. Sophisticated tax planning involves significant fixed costs associated with the setup of complex legal structures and the remuneration of tax planners' human capital. Shackelford (2000) describes several widespread tax-avoidance plans, and notes that "these plans are restricted to the wealthiest of taxpayers because the implementation fees are so large that the income or transfer taxes saved must be enormous to justify purchasing the tax plan". That wealth managers and sophisticated tax planners impose high minimum accounts is consistent with such significant fixed costs. Also consistent with this, Lang *et al.* (1997) use detailed consumer survey data in Germany to show that the difference between legislated and effective tax rates increases with respect to income, and that a sizeable fraction of it is due to the exploitation of legal tax write-offs.

A second important form of tax avoidance consists in international tax arbitrage, by locating assets or establishing fiscal residence and/or citizenship in low-tax countries.[2] This form of tax avoidance also involves legal and transportation costs that are not very sensitive to income. Consistent with this, using data on the geographic mobility of soccer players, Kleven *et al.* (2013) document that it is only at the top of the income distribution that location choice is highly elastic to taxes. Studying the impact of the Danish preferential tax scheme for high-earning immigrants, Kleven *et al.* (2014) also document a high elasticity of migration of top earners. Another international tax arbitrage technique consists in keeping one's fiscal residence unchanged while making undeclared bank deposits in countries with strong bank secrecy. Strictly speaking, this pertains to evasion rather than avoidance, as it is illegal. But it it is a virtually undetectable fraud (holding international treaties fixed), and involves the same type of fixed legal and administrative costs as avoidance. Thus, we consider it to be part of the tax optimization techniques that we seek to model in this article. Exploiting inconsistencies in international accounts, Zucman (2013) estimates that 8% of total household financial wealth is held in tax havens. Recent attempts by the G20 at cracking down on this type of evasion may reduce its

2. An extreme form of such tax avoidance is that of "perpetual travelers"—individuals who spend sufficiently little time in any given country that they have no identified fiscal residence.

magnitude in the near future. Yet, Johannesen and Zucman (2014) offer suggestive evidence that rather than repatriating funds in response, evaders tend to relocate them to alternative less compliant havens. There is a large amount of anecdotal evidence suggesting that the wealthiest taxpayers manage to substantially decrease their tax bill by using sophisticated tax optimization techniques. For instance, the *New York Times* (NYT 30 December 2015) reports the common use by family offices of techniques such as investment in Bermuda-based reinsurers to transform short-term capital gains into long-term capital gains, which are taxed at a lower rate. The article mentions that despite political efforts, tax avoidance at the top of the income distribution remains effective: "From Mr. Obama's inauguration through the end of 2012, federal income tax rates on individuals did not change (excluding payroll taxes). But the highest-earning one-thousandth of Americans went from paying an average of 20.9 percent to 17.6 percent", thus suggesting a positive trend in use of tax optimization techniques by the wealthiest. In that same article, Victor Fleischer, a law professor at the University of San Diego says: "We do have two different tax systems, one for normal wage-earners and another for those who can afford sophisticated tax advice. At the very top of the income distribution, the effective rate of tax goes down, contrary to the principles of a progressive income tax system." Recently, The Panama Papers scandal (see *e.g.* NYT 4 April 2016) has exposed, by leaking confidential documents from the Mossack Fonseca law firm in Panama, an extensive use by the wealthiest people of offshore bank accounts and shell companies to avoid taxes. While it is unclear what fraction of the reported accounts correspond to legal versus illegal activities, the scandal sheds light on how prevalent the use of sophisticated tax optimization techniques seems to be among the wealthiest at a global level. "This is one way in which people with a lot of money step away from being average", comments Jack Blum (NYT 6 June 2016), a lawyer who served for more than a decade as a consultant to the Internal Revenue Service.

Consider now risk shifting. Average salaried workers have less discretion to manipulate the risk profile of their labour income than entrepreneurs or top executives, who have a larger scope for occupational choice or for corporate risk-taking that can directly affect their income: Hopenhayn and Vereshchagina (2009) study endogenous risk shifting by entrepreneurs; Shue and Townsend (2014) show empirically that top executives change the risk profile of their firm's income when their compensation package is more convex. Moreover, high-wealth households have access to a large set of sophisticated financial instruments with few risk-taking restrictions (*e.g.* investments in hedge funds), and, therefore, have a free hand at shaping the risk profile of their capital income, which is a sizeable fraction of their total income. By contrast, for the rest of the population, investment is limited to a more narrow and more regulated set of vehicles. Gambling options are available but typically come at a high expected loss.

We consider that tax avoidance and risk shifting are at least as plausible responses to taxation by the rich as the labour-supply decision studied in workhorse Mirrleesian models. We develop a parsimonious model aiming at studying the qualitative implications of these frictions for optimal taxation. Our article formalizes three main points within the same optimal taxation framework: (1) we show that in the presence of subadditive avoidance costs, the optimal tax scheme is non-progressive and does not depend on agents' nor governments' preferences; (2) if agents are allowed to gamble, such non-progressivity induces them to do so for concavification motives, which leads to higher pre-tax inequality and lower welfare; (3) when the avoidance technology consists in costly migration between asymmetric countries, we show that redistributive equilibria become fragile if the population in the large country has higher access to risky gambles, for example, because of a more sophisticated financial system, or because innovation and entry are easier. In such case, gambling equilibria appear, where taxes are more regressive, risk-taking takes place and raises pre-tax inequality, and migration flows of wealthy individuals towards the small country are larger.

We model tax avoidance as follows. We study the situation of a social planner who seeks to implement inequality-averse views in an endowment economy. The planner faces an informational friction. Agents privately observe their endowments, and can convert the fraction that they do not report to the planner into secret consumption at some cost. In line with the above discussion, we capture avoidance by assuming that this conversion comes at subadditive costs. The optimal redistribution scheme implemented by the social planner in the presence of such tax avoidance is simple. Net income has a fixed component and a variable one that increases with respect to the reported pre-tax income. This scheme is such that agents report their entire income: there is no avoidance in equilibrium. The fixed component equally splits among agents the total tax capacity of the planner, defined as the total resources that he could extract from the population if he was not redistributing any of it. The variable component makes every agent indifferent between reporting his entire income or reporting the lowest income level in the population. This simple scheme has two interesting properties. First, it does not depends on the exact preferences of the agents nor on that of the planner, only on the assumption that they all are concave. Secondly, taxation cannot be progressive.

We then add a risk-shifting friction to the tax-avoidance one. We introduce an initial stage during which agents can add any fair lottery to their endowments. We focus on the case in which the planner taxes them in an *ex post* optimal fashion after gambling has taken place. We find that the agents may gamble in equilibrium because they expect their utility to be non-concave over pre-tax income given the non-progressivity of *ex post* optimal taxation. Thus, subadditive avoidance costs cause larger pre-tax inequality at the top in the presence of endogenous risk-taking. Even when the planner has commitment power, it is still the case that agents extract risk-taking rents in addition to income-hiding rents, as the optimal tax scheme is typically less progressive in the presence of the risk-shifting friction.

Lastly, we endogenize avoidance costs by solving a model of fiscal competition between two asymmetric countries where citizens face a migration cost if they choose to switch fiscal residence. The avoidance cost is then the sum of this cost and of the (endogenously determined) taxes paid in the new fiscal residence. We show that an endogenous increase in pre-tax inequality caused by risk-taking in the larger country reduces these endogenous avoidance costs, and that this in turn spurs risk-taking and raises inequality. This mutual reinforcing of avoidance and risk-taking may lead to self-justified equilibria where (1) risk-taking occurs, (2) taxation becomes more regressive at the top, (3) wealthy individuals migrate towards the smaller country. The key intuition is that there are strong strategic complementarities in risk-taking by individuals: if the population of rich individuals in the large country ends up to be thick enough, they can expect the small country to offer attractive taxes at the top to poach them, which in turn makes risk-taking more attractive. This result suggests that limits to the accessibility of financial risk-taking and international fiscal cooperation may both be required in order to preserve equilibria with large levels of redistribution.

The article is organized as follows. Section 2 studies optimal taxation in the presence of a tax-avoidance friction. Section 3 adds the risk-shifting friction and studies its impact on pre- and post-tax inequality. Section 4 endogenizes the avoidance technology in a model of costly migration where two countries of different size compete fiscally. Section 5 discusses the related literature. Section 6 concludes. Proofs are relegated to an appendix. An Online Appendix details interesting extensions that are only briefly mentioned in the article.

2. SUBADDITIVE AVOIDANCE COSTS AND OPTIMAL TAXATION

Consider a one-date economy populated by a continuum of agents with unit mass. There is a single consumption good. The agents have identical preferences represented by a utility over consumption u that is increasing and strictly concave. Agents differ only with respect to their endowments of the consumption good—their "incomes". The cumulative income distribution F has support $[0,+\infty)$ and a finite mean:

$$\int_0^{+\infty} wdF(w) < +\infty.$$

We study the problem of a social planner who redistributes income in order to maximize the utilitarian welfare of the population. A direct implication of Jensen's inequality is that the first-best policy consists in ensuring that each agent consumes the same amount, equal to the average endowment $\int_0^{+\infty} wdF(w)$.

We depart from this first-best, and assume that the planner faces the following informational friction. Each agent privately observes his income. An agent with income x may report any amount $y \in [0,x]$ to the planner, and conceal the residual $x-y$. This concealed income $x-y$ can be converted into $g(x-y)$ units of secret consumption, where g is a continuous function that satisfies:

$$0 \leq g(z) \leq z. \tag{1}$$

This secret consumption adds up to the public one, which is the net transfer that the agent receives after the social planner redistributes aggregate reported income.

Under this general formulation, the friction facing the social planner could be interpreted either as tax evasion or as tax avoidance. As we explained in Section 1, we believe, however, that evasion and avoidance correspond to very distinct properties of the avoidance technology represented by g. In the case of evasion, which is an outright breach of tax law, the technology g should be thought of as displaying diseconomies of scale. In the case of tax avoidance, the technology g should feature economies of scale, at least within some range. Accordingly, we posit that avoidance costs are subadditive:

Assumption 1. (Subadditive avoidance costs). *The function g is superadditive. For all w, $w' \geq 0$*

$$g\left(w+w'\right) \geq g(w)+g(w'). \tag{2}$$

Notice that superadditivity and $g \geq 0$ imply that g is increasing. One interpretation of equation (2) is that a single affluent agent avoids more efficiently than a group of agents with the same aggregate tax base. A subadditive cost function is the defining feature of natural monopolies introduced by Baumol (1977). As is well known, this is a weaker property than scale economies.[3]

We now solve the planner's problem in the presence of this tax-avoidance friction. In application of the revelation principle, one can write down the planner's problem using only direct mechanisms. A direct mechanism is a pair of functions $(r(.), v(.))$ such that an individual with endowment w has the incentive to report $r(w) \in [0,w]$, and receives a net transfer $v(r(w))$ from the social planner after doing so.[4]

3. The Online Appendix shows sufficient conditions under which migrating at a fixed cost to a country with *progressive* taxes still yields such a superadditive function g.

4. We restrict the analysis to deterministic mechanisms. The Online Appendix gives sufficient conditions on u for this to be without loss of generality.

The social planner solves the programme $(\wp)$:

$$\max_{r,v} \int_0^{+\infty} u(v(r(w))+g(w-r(w)))\,dF(w) \tag{3}$$

$$\text{s.t.} \begin{cases} \int_0^{+\infty} v(r(w))dF(w) \leq \int_0^{+\infty} r(w)dF(w), \\ \forall w, w' \geq 0 \text{ s.t. } r(w') \leq w, \\ v(r(w))+g(w-r(w)) \geq v(r(w'))+g\left(w-r(w')\right). \end{cases}$$

The first constraint is the resource constraint of the planner. The other inequalities are incentive-compatibility constraints, ensuring that individuals report according to their types (which of course does not necessarily imply that they report their entire income). We show that the solution to this programme $(\wp)$ is very simple when tax avoidance comes at subadditive costs.

Proposition 2. (Optimal tax scheme). *Under Assumption* 1, *the optimal tax scheme is such that agents report their entire income and receive the same utility as if they were reporting the minimum income level in the population. Formally, the solution to* $(\wp)$ *is attained with* $\left(r^*, v^*\right)$ *defined as*

$$\begin{cases} r^*(w) = w, \\ v^*(w) = g(w) + \int_0^{+\infty} (t-g(t))\,dF(t). \end{cases} \tag{4}$$

Proof. See the Appendix. ||

Proposition 2 states that there is no tax avoidance in equilibrium: agents report their entire income. This is a direct consequence from the superadditivity of g. Any incentive-compatible tax scheme that implies some avoidance can be replaced with a more efficient one that does not entail any. To see this, suppose that a mechanism (r,v) implies $\int r(w)dF(w) < \int wdF(w)$. Then a scheme whereby an individual with income w reports w and receives $v(r(w))+g(w-r(w))+\varepsilon$ satisfies the resource constraint for $\varepsilon > 0$ sufficiently small. Further, it is incentive-compatible:

$$\begin{aligned} v(r(w))+g(w-r(w)) &\geq v(r(w'))+g(w-r(w')), \\ &\geq v(r(w'))+g(w'-r(w'))+g(w-w'). \end{aligned}$$

The first inequality stems from the incentive-compatibility of (r,v), the second one from the superadditivity of g. This second inequality means that this new mechanism is also incentive-compatible. It is strictly preferable to (r,v) because the income destruction induced by tax avoidance disappears.

Proposition 2 then elicits the most redistributive scheme among all "avoidance-free" ones. It simply consists in making every agent indifferent between reporting his entire income or none of it. Subadditive costs imply that an agent who is indifferent between reporting everything and reporting nothing also prefers a full report to any partial report.

The optimal tax scheme in Proposition 2 does not depend on the utility function u, and the proof of the proposition only uses that u is increasing and concave. Thus the scheme (4) is optimal as soon as agents' utility functions and the social welfare function are increasing and concave, and does not otherwise depend on these functions.[5] In fact, the constant term in v^*, $\int_0^{+\infty}(t-g(t))\,dF(t)$, is simply the *tax capacity* of the planner—the maximum revenue that he

5. In particular, the planner could pursue a Rawlsian objective.

can extract from the population. To see this, consider the following programme $(\wp')$:

$$\max_{r,\tau} \int_0^{+\infty} \tau(r(w))\,dF(w) \tag{5}$$

$$\text{s.t.} \begin{cases} \forall w \geq 0,\ \tau(r(w)) \leq r(w), \\ \forall w, w' \geq 0 \text{ s.t. } r(w') \leq w, \\ r(w) - \tau(r(w)) + g(w - r(w)) \geq r(w') - \tau\left(r(w')\right) + g\left(w - r(w')\right). \end{cases}$$

This programme formalizes (applying the revelation principle) the situation in which a planner seeks to extract as many resources as possible from the population for purposes that are outside the model. The function $r(w)$ describes the report of an agent with income w, while $\tau(r(w))$ describes by how much he is taxed. Notice that we impose that no agent be taxed beyond his reported income, which is plausible and ensures that the programme $(\wp')$ has a finite solution. We have

Corollary 3. (Tax capacity). *The maximum feasible taxation is the constant term in the optimal tax scheme (4). Formally, the solution to $(\wp')$ is $\int_0^{+\infty}(t - g(t))\,dF(t)$, and is attained with $\left(r^*, \tau^*\right)$ defined as*

$$\begin{cases} r^*(w) = w, \\ \tau^*(w) = w - g(w). \end{cases} \tag{6}$$

Therefore, optimal taxes are not strictly progressive and do not depend on individual preferences.

Proof. See the Appendix. ‖

The average tax rate

$$\frac{\tau^*(w)}{w} = 1 - \frac{g(w)}{w}$$

can of course be constant if g is linear, but cannot be strictly increasing. This would imply g strictly concave and thus strictly subadditive.

Overall, these results show that in the presence of scale economies in avoidance, the same tax scheme given by equation (4) addresses both the question of the *optimal* taxation—provided the objective is inequality-averse, and that of the *maximum feasible* taxation. This tax scheme depends only on pre-tax income distribution and on the avoidance technology. Pre-tax income distribution affects only $v^*(0)$, but not $v^*(w) - v^*(0)$. This contrasts with solutions to the standard Mirrlees problem, for which the optimal tax scheme is typically more sensitive to the fine details of the model. Also, note that, in empirical applications of our set-up, the avoidance technology g can be directly backed out from the observation of tax rates using equation (6).

3. TAX AVOIDANCE AND RISK SHIFTING

We argued in section 1 that a distinctive feature of affluent individuals is their superior ability to control the risk profile of their income. Occupational choices can serve to control the risk profile of labour income (as in Hopenhayn and Vereshchagina, 2009). The rich set of financial instruments available to high net worth investors gives them a free hand at selecting the risk profile of their capital income. This section introduces a second friction in our baseline model that formalizes this feature, and studies its interplay with tax avoidance as modelled in Section 2.

3.1. *Setup*

As in the baseline model of the previous section, a social planner seeks to maximize the utilitarian welfare of a continuum of agents with unit mass. The economy now has two dates, 0 and 1. Preferences, endowments, and the information structure are as follows.

Preferences. Agents value only date-1 consumption, over which they have CARA utility u.[6]

Endowments. Agents receive their entire endowment at date 0. The cumulative income distribution F_0 has support $[0,+\infty)$ and a finite mean. Agents need to store their income from date 0 to date 1 in order to consume. A risk-free storage technology with unit return is available. Agents may also enter into risk shifting in the following sense. Each agent may add to this risk-free return an idiosyncratic risky return. In this case, he has a free hand at choosing the unit-mean distribution of this risky return. Formally, an agent with initial income w_0 can choose a date-1 income with any cumulative distribution function with mean w_0 and support included in $[0,+\infty)$.

We assume that for each date-0 income level, an arbitrarily small measure of agents only has access to the risk-free storage and cannot gamble this way. This is a technical assumption meant to ensure that date-1 pre-tax income distribution has full support over $[0,+\infty)$.[7]

Information. Agents privately observe their date-0 income. The social planner does not observe their investment decisions, nor their resulting date-1 individual incomes. As in the previous section, agents can convert x units of concealed income into $g(x)$ units of secret consumption, where g is a continuous superadditive function that satisfies equation (1). The (exogenous) date-0 income distribution, that we denote F_0, is publicly observed. So is the (endogenous) date-1 income distribution, that we denote F_1.[8]

Thus, the social planner now faces two informational frictions. First, agents can divert income and secretly consume as in the previous section. Secondly, they can also secretly shift income risk. We model this risk-shifting ability as the possibility to add fair lotteries with arbitrary distribution to their income. This modelling choice has two advantages. First, excessive risk-taking is simply and clearly characterized in our model as the addition of non-rewarded risk to a safe endowment by a risk-averse agent. Secondly, this delivers sharp insights into the type of risk distributions that households willing to shift risk demand.

We suppose that the social planner lacks commitment power, and redistributes in an *ex post* optimal fashion at date 1, after agents have made their investment decision and received their date-1 income. More precisely, the timeline of this economy is as follows. At date 0, each of the agents decides on the distribution of the mean-preserving spread that he wishes to add to his date-0 endowment. He may of course also prefer to store at the risk-free rate. At date 1, agents receive their date-1 incomes. The resulting *ex post* income distribution F_1 is publicly observed. At this stage, the social planner announces a taxation mechanism (r,v) so as to maximize utilitarian welfare.

An equilibrium in this economy consists in an *ex post* income distribution F_1, a date-1 tax scheme (r^{**},v^{**}), and date-0 investment decisions such that:

- The tax scheme (r^{**},v^{**}) solves the programme $(\wp)$ described by equations (3) and (4) for an income distribution $F=F_1$.

6. The Online Appendix tackles the case of an arbitrary increasing concave utility function.

7. Alternatively, one could assume that agents only have access to distributions with full support over $[0,+\infty)$ when shifting risk. An agent interested in a discrete distribution could approximate it arbitrarily well with such a continuous one.

8. We only need that the date-0 income distribution be common knowledge. Assuming that so is F_1 slightly simplifies the exposition.

- Each agent makes a date-0 investment decision that maximizes his date-1 expected utility given his initial income and his beliefs about F_1 and (r^{**}, v^{**}).
- The distribution F_1 correctly aggregates the impact of these individual investment decisions on the initial income distribution F_0.
- Agents have correct beliefs about F_1 and (r^{**}, v^{**}).

We now characterize such equilibria. Notice first that since the planner maximizes *ex post* social welfare, Proposition 2 applies at date 1, and upon observing F_1, the social planner sets

$$\begin{cases} r^{**}(w) = w, \\ v^{**}(w) = g(w) + \int_0^{+\infty} (t - g(t))\, dF_1(t). \end{cases} \tag{7}$$

When facing his investment decision at date 0 and forming beliefs about F_1, an agent with initial income w_0 expects his date-1 consumption to be the random variable $v^{**}(\widetilde{w}_1)$, where $\widetilde{w}_1$ are his possibly random proceeds from investment and v^{**} is defined in equation (7). Thus he optimally chooses the distribution of $\widetilde{w}_1$ that maximizes his expected utility subject to the constraint that he expects to earn his initial endowment w_0 before taxes. Formally, this agent solves the following problem:

$$V(w_0, F_1) = \max_{G \in \Gamma} \int_0^{+\infty} u\left(g(w) + \int_0^{+\infty} (t - g(t))\, dF_1(t) \right) dG(w), \tag{8}$$

$$\text{s.t.} \quad \int_0^{+\infty} w\, dG(w) = w_0.$$

where Γ is the set of cumulative distribution functions with support included in $[0, +\infty)$. CARA preferences imply that the value function $V(w_0, F_1)$ of this programme satisfies

$$V(w_0, F_1) = \exp\left(-\alpha \int_0^{+\infty} (t - g(t))\, dF_1(t) \right) W(w_0), \tag{9}$$

where we adopt the convention $u(x) = -e^{-\alpha x}$, and $W(w_0)$ is the value function of the programme below that depends only on w_0:

$$W(w_0) = \max_{G \in \Gamma} \int_0^{+\infty} u(g(w))\, dG(w), \tag{10}$$

$$\text{s.t.} \quad \int_0^{+\infty} w\, dG(w) = w_0.$$

Thus, CARA preferences imply that the agent's beliefs about F_1 do not affect his investment choice. The reason is that F_1 only affects the constant term in v^{**}. The value of this constant has no impact on the attitude of a CARA agent towards the riskiness of his date-1 consumption. All that is left to complete the equilibrium characterization is to solve for equation (10) for all $w_0 \geq 0$. To do so, we introduce the concavification of the function $u \circ g$, and denote it $\overline{u \circ g}$. That is, the function $\overline{u \circ g}$ is the smallest concave function such that for all $w \geq 0$,

$$\overline{u \circ g}(x) \geq u(g(x)).$$

The function $\overline{u \circ g}$ exists and is unique (see, *e.g.* Aumann and Perles, 1965).

Lemma 4. (Optimal risk-taking). *We have*

$$W(w_0) = \overline{u \circ g}(w_0).$$

If $\overline{u \circ g}(w_0) = u(g(w_0))$*, then the agent stores at the risk-free rate.*

If $\overline{u \circ g}(w_0) > u(g(w_0))$*, then the agent takes additional risk. He may be indifferent among several distributions. In this case, the least risky one in the sense of second-order stochastic dominance is the binary distribution with support* $\{\underline{w}(w_0); \overline{w}(w_0)\}$*, where*

$$\begin{cases} \underline{w}(w_0) = \sup\{w \leq w_0 \ s.t.\ \overline{u \circ g}(w) = u(g(w))\}, \\ \overline{w}(w_0) = \inf\{w \geq w_0 \ s.t.\ \overline{u \circ g}(w) = u(g(w))\}. \end{cases} \tag{11}$$

Proof. See the Appendix. ‖

Lemma 4 formalizes that agents use lotteries in order to concavify their date-1 utility when the tax scheme given by equation (7) implies that their utility may be non-concave in their pre-tax income. This is formally related to the analysis in Hopenhayn and Vereshchagina (2009), where entrepreneurs, due to their outside options, shift risk for similar concavification purpose. In what follows, we assume that when indifferent among lotteries, agents pick the least risky lottery. We believe that this is a reasonable selection criterion.

The next proposition summarizes the equilibrium characterization above. In the remainder of the article, we denote $\widetilde{\rho}(w)$ the equilibrium lottery for an agent with initial income w defined in Lemma 4. That is, $\widetilde{\rho}(w) = w$ if the agent decides to store at the risk-free rate, and $\widetilde{\rho}(w)$ is the least risky lottery defined in Lemma 4 otherwise. By definition, for all $w \geq 0$,

$$E[u \circ g(\widetilde{\rho}(w))] = \overline{u \circ g}(w).$$

Proposition 5. (Optimal tax scheme in the presence of risk-taking). *In equilibrium, there is risk shifting by some agents if and only if*

$$\overline{u \circ g} \neq u \circ g. \tag{12}$$

An agent with initial income w *obtains a date-1 pre-tax income equal to* $\widetilde{\rho}(w)$*. His date-1 net income is*

$$v^{**}(\widetilde{\rho}(w)) = g(\widetilde{\rho}(w)) + v^{**}(0) \tag{13}$$

with

$$v^{**}(0) = \int_0^{+\infty} (t - g(t))\, dF_0(t) - \int_0^{+\infty} (E[g(\widetilde{\rho}(t))] - g(t))\, dF_0(t). \tag{14}$$

Proof. See above. ‖

Proposition 5 establishes a theoretical link between subadditive avoidance costs and the rise of *pre-tax* inequality at the top. Section 2 has shown that taxation cannot be progressive in the presence of subadditive avoidance costs. When considering whether gambling or not, each agent trades off the reduction in his expected tax bill due to non-progressive taxation with the disutility from adding risk to his consumption. The former benefit more than offsets the latter cost exactly when the simple condition (12) is satisfied. Each agent is negatively affected by gambling by the other agents because this reduces the fixed amount that he receives from the government by $\int_0^{+\infty}(E[g(\widetilde{\rho}(t))]-g(t))dF_0(t)$ from equation (14). The following corollary derives the implications of this equilibrium for utilitarian welfare and *ex ante* preferences towards risk-taking.

Corollary 6. (Risk-taking: welfare and political-economy implications).

- *Utilitarian welfare is strictly lower in the presence of risk-taking if and only if equation (12) holds.*
- *Suppose that agents can vote on risk-taking at the outset. It may be that some agents vote in favour of allowing risk-taking. An arbitrarily large fraction of agents vote for a ban on risk-taking if the initial income distribution is sufficiently egalitarian, however.*

Proof. See Appendix. ‖

If risk-taking occurs in equilibrium, this reduces utilitarian welfare for two reasons. First, it adds uncompensated risk to a risk-averse economy. Secondly, it reduces tax capacity and thus negatively affects redistribution through a reduction in the constant transfer made to each agent.

The second point in Corollary 6 discusses whether agents would favour an *ex ante* ban on risk-taking for the entire population. This is not necessarily the case. If the mass of agents for whom

$$\overline{u\circ g}(w)\neq u\circ g(w) \tag{15}$$

is sufficiently small, each of them would vote in favour of authorizing gambling as the resulting reduction in the planner's fixed payment is then smaller than the reduction in their own expected tax bills. Conversely, if the initial income distribution is sufficiently concentrated around a given income level w_0, then an arbitrarily large majority of agents would vote in favour of a ban on gambling even if equation (15) holds at w_0. The reason is that the reduction in the transfer from the planner then becomes arbitrarily close to the expected reduction in their individual tax bills. Thus, from an *ex ante* perspective, the option to gamble only adds white noise to their after-tax income and becomes unpalatable.

This simple result has the interesting political-economy implication that an initial increase in gross income inequality may induce more forceful lobbying in support of gambling—for example, in support of lax prudential regulation. The increased availability of gambling options would then in turn amplify the initial increase in inequality.

A simple example. To get a geometric intuition for the result, it is useful to use a simple particular case in which the tax-avoidance technology g is piecewise linear with a convex kink. Suppose that

$$g(x)=(1-\lambda)x+\mathbf{1}_{\{x\geq c\}}\Delta\lambda(x-c). \tag{16}$$

This corresponds to the case in which two tax-avoidance technologies are available. The first one dissipates a fixed fraction $\lambda\in(0,1)$ of each diverted unit of income. The second one wastes only $\lambda-\Delta\lambda\in(0,\lambda)$ out of each diverted income unit, but comes at a fixed cost $c\Delta\lambda>0$. An agent chooses the latter if and only if his date-1 income is larger than c. Figure 1 displays in this particular case the functions $u\circ g$ and $\overline{u\circ g}$.

The function $u\circ g$ and its concavification coincide over two intervals $[0,\underline{w}]$ and $[\overline{w},+\infty)$, where $\underline{w}<c<\overline{w}$.[9] The concavification $\overline{u\circ g}$ is strictly above $u\circ g$ over $(\underline{w},\overline{w})$, where it is equal to

9. It is possible that $\underline{w}=0$.

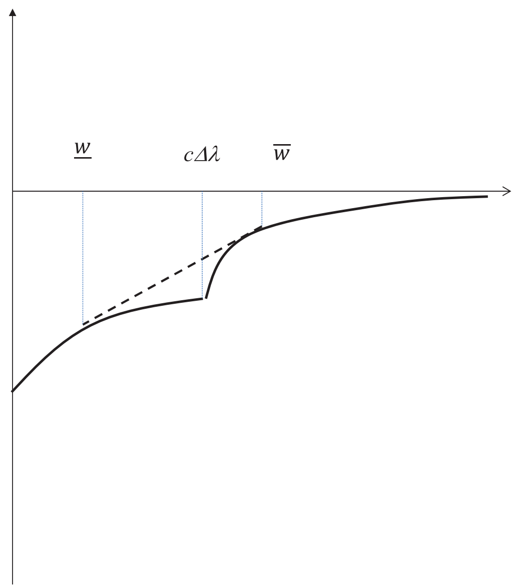

FIGURE 1

The solid curve represents the graph of $u\circ g$ and the dashed line its concavification over $[\underline{w},\overline{w}]$. The two functions coincide outside the segment

the chord linking the points $(\underline{w},u(g(\underline{w}))$ and $(\overline{w},u(g(\overline{w}))$. This chord is tangent to $u\circ g$ at these two points. From Lemma 4, any agent with an initial income $w_0\in[0,\underline{w}]\cup[\overline{w},+\infty)$ stores at the risk-free rate. If $w_0\in\left(\underline{w},\overline{w}\right)$, then the agent gambles and invests with a binary risky return so as to obtain a date-1 income equal to $\underline{w}$ with probability $\frac{\overline{w}-w_0}{\overline{w}-\underline{w}}$ or $\overline{w}$ with probability $\frac{w_0-\underline{w}}{\overline{w}-\underline{w}}$. The date-1 income distribution F_1 is thus riskier than F_0 in the sense of second-order stochastic dominance because the mass of F_0 between $\underline{w}$ and $\overline{w}$ is split into two atoms of F_1, in $\underline{w}$ (with mass $\int_{(\underline{w},\overline{w})}\frac{\overline{w}-w}{\overline{w}-\underline{w}}dF_0(w)$) and $\overline{w}$ (with mass $\int_{(\underline{w},\overline{w})}\frac{w-\underline{w}}{\overline{w}-\underline{w}}dF_0(w)$). This fully characterizes the equilibrium in this particular case.

Income distribution and demand for fake alpha. As is obvious in this simple example, the distribution of the risk taken by risk-shifting agents depends on their initial income. When they are relatively "poor", so that w_0 is on the right neighbourhood of $\underline{w}$, they purchase payoffs that are negative and small in absolute value with a large probability, and large and positive with a small probability—like a lottery ticket. Conversely, when income is in the left neighbourhood of $\overline{w}$, the investors favour trades that payoff a small excess return most of the time and generate rare, large losses. These risk profiles, labelled by Rajan (2010) as "fake alpha" strategies, are produced by collecting a fair premium for exposure to a large disaster risk. The profile of F_0 thus determines aggregate risk-taking. As the tail of F_0 becomes fatter, the demand for fake-alpha strategies increases.

With a general function g, there are two additional technical difficulties. Figure 2 illustrates them.

First, the set of income levels for which $u\circ g<\overline{u\circ g}$ is not necessarily a single interval. Secondly, it is possible to construct cases in which, unlike in the simple example above, the lottery that solves equation (10) for a given income level w_0 is no longer necessarily unique.

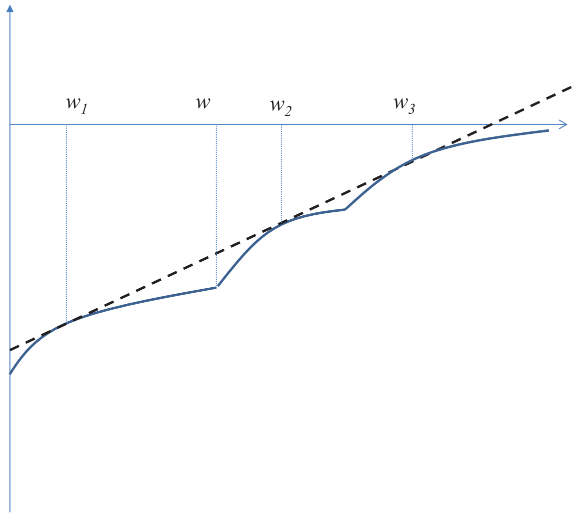

FIGURE 2

Here the straight line that concavifies $u \circ g$ in w has three points of contact with $u \circ g$, $\{w_1; w_2; w_3\}$. An agent with income w can concavify with a lottery that has support $\{w_1; w_2; w_3\}$, $\{w_1; w_3\}$, or $\{w_1; w_2\}$. The latter is less risky in the sense of second-order stochastic dominance

3.2. *Robustness*

3.2.1. The role of limited commitment. The degree of commitment power of a democratic government in terms of taxation is difficult to assess. Some commitment seems possible at the horizon of a mandate—as evidenced by tax amnesties. Commitment beyond a mandate seems difficult by construction—although some form of commitment is implicit in a pay-as-you-go pension system, for example. Here, the extreme assumption of limited commitment simplifies the analysis. It is important to stress, however, that the main results in this section do not live or die on the assumption that the planner taxes in an *ex post* optimal fashion. The case in which the planner can fully commit to a tax scheme announced before gambling takes place is fully treated in the Online Appendix. The optimal scheme simply consists in having the planner committing to a tax scheme whereby he collects $w - u^{-1} \circ \overline{u \circ g}(w)$ from agents with income w, and pays a fixed sum to each agent. In other words, the planner manufactures himself a concave expected utility over pre-tax income, instead of letting agents concavify themselves by means of risk shifting. This requires commitment power as soon as $\overline{u \circ g} \neq u \circ g$ because this requires the enforcement of an *ex post* inefficient tax scheme.

The gains from not having agents concavifying their utility themselves is that the date-1 income distribution does not get more diffuse as a result, which raises tax capacity, thereby generating a higher constant payment. Although actual gambling does not take place in equilibrium under full commitment, the optimal tax scheme still becomes less progressive than in the absence of risk-taking in order to prevent it. Thus, in addition to income-hiding rents, agents who would find it privately optimal to gamble still extract risk-taking rents. Even under full commitment, utilitarian welfare is still lower than absent the risk-shifting friction and this friction still affects after-tax inequality.

3.2.2. Dynamics. The assumption made throughout the article that consumption takes place at a single date is meant to focus on risk-taking decisions, and to abstract from consumption smoothing across dates. The Online Appendix studies an extension in which agents value consumption at several dates. Again, the main insights from the static case are unchanged, but two interesting new features emerge:

(1) *Distorted intertemporal allocations.* Agents now use a two-pronged strategy in order to concavify their utility at each date. They concavify utility over future consumption using risky storage as in the static case, and concavify current utility by distorting their consumption/saving decision.
(2) *Ratchet effect.* In the case in which the planner cannot commit to future tax schemes, agents do not report their entire current information about their future income because of the (rational) fear that the planner would use this information against them in the future.

4. ENDOGENOUS AVOIDANCE COSTS

This section endogenizes avoidance costs. As mentioned in Section 1, tax avoidance by the affluent takes two forms in practice. First, it consists in exploiting the incompleteness and/or loopholes of local tax law. It may alternatively consist in establishing fiscal residence in a jurisdiction with a more favourable tax regime. This section focusses on endogenizing the costs of this latter form of avoidance for two reasons. First, a given jurisdiction has arguably less control over it than over the exploitation of local loopholes through sophisticated schemes. Tax law can be clarified in order to reduce such local arbitrage in principle, unless political-economy or organizational frictions prevent it. Secondly, there is an already sizeable literature on the impact of migration on optimal taxation. In a recent contribution, Lehmann *et al.* (2014) extend the Mirleesian setting by allowing for migration between two jurisdictions that play Nash when setting their tax schemes.[10] In a parallel approach, we study the interplay of migration with the risk-shifting friction introduced here. The interaction between the two frictions is actually quite rich. We show that the (endogenous) gains from risk-taking and migration strongly reinforce each other in equilibrium, so much so that this can lead to multiple equilibria with varying levels of migration and pre- and post-tax inequality.

We proceed in two steps. In a first step, we study a simple model in which migration is the only force that puts an upper bound on taxing capacity. In a more general framework, we confirm the result on asymmetric country sizes and taxation established by Kanbur and Keen (1993).[11] In a second step, we study the interplay of this migration friction with the risk-shifting problem studied above.

4.1. *An elementary model of tax competition*

There are two countries, A and B. Country $C \in \{A; B\}$ is initially populated by a mass of μ_C agents such that

$$0 < \mu_B < \mu_A. \tag{17}$$

Each agent has a unit income. Agents have increasing utility over consumption. Each agent can decide to leave his initial residence and move to the other country where he is then taxed and consumes. He incurs a cost c from doing so. Migration costs are identically distributed across countries, with a c.d.f. F_m that admits a log-concave p.d.f. f_m with full support over

10. Their paper also offers a thorough review of the literature on tax-driven migration.
11. We relax their assumptions of risk neutrality and uniformly distributed migration costs.

$[0,+\infty)$. The tax authority within each country seeks to maximize its tax revenues. The exact sequence of events is as follows. First, countries simultaneously announce the tax that applies to their residents. Then agents make a choice of fiscal residence. Either they stay in their home country, or they migrate at a cost. Finally agents are taxed and consume within their residence.[12] For simplicity only, we do not impose that consumption be positive.

Remark. The assumption that taxes cannot be contingent on the initial residence of an agent is the case that has been studied in the literature thus far. On one hand, this prevents countries from attracting non-residents with tax rates lower than that applied to their initial residents.[13] On the other hand, countries cannot apply expatriation taxes to deter emigration. This is a limited deterrent, however, because exit taxes apply only to already accumulated wealth, and not to income earned after migration in practice. Whereas these are interesting extensions, the case that we focus on is better suited to the analysis of long-run relative fiscal attractiveness.

Note that there are no informational frictions such as unobservable income/productivity in this economy. We solve for subgame-perfect equilibria, and denote τ_A and τ_B the tax rates (or tax per capita here) of each country.

Lemma 7. (Kanbur–Keen result on asymmetric size and taxation). *There exists a unique equilibrium. It is such that $\tau_A > \tau_B$, so that migration takes place only from A to B. Further, τ_A and τ_B decrease with respect to μ_A/μ_B whereas $\tau_A - \tau_B$ increases, triggering more migration.*

Proof. See Appendix. ||

That the larger country taxes more admits a straightforward intuition. The small country has more to gain from poaching the other tax base and less to lose on its own population with lower tax rates. The comparative statics with respect to relative size are less obvious, and can be interpreted as follows. An authority's initial tax base serves as a commitment device to not lower taxes too much to attract non-residents. As countries become more asymmetric, this device is weaker and tax competition is more intense: Tax rates decrease in both countries. The tax rate in the small country is more sensitive to relative size than that of the large country, however. We now augment this model with the risk-shifting friction studied throughout the article.

4.2. *Inequality and avoidance options at the top: a two-sided interaction*

Suppose now that the population within each country is split across three incomes levels I_0, I_1, and I_2 such that

$$0 = I_0 < I_1 < I_2. \tag{18}$$

We interpret I_1 and I_2 as affluence levels ("rich" and "super-rich") and $I_0 = 0$ as a normalization for the rest of the population. Denote $\mu_{C,i}$ the mass of agents with income I_i in country C, and

$$\alpha_C = \frac{\mu_{C,2} I_2}{\mu_{C,1} I_1} \tag{19}$$

a measure of the thickness of the right tail of wealth distribution in country C.

12. For countries that tax worldwide income, such as the U.S., a fully effective change of fiscal residence involves citizenship renunciation.

13. Examples include the non-domiciled status in the U.K., or the Danish scheme discussed in Kleven *et al.* (2014).

We now suppose that agents with initial income I_1 can take fair gambles and move up or down the income ladder before taxes are announced in each country. Risk-taking abilities may differ across countries. If an agent from country C gambles, his income remains equal to I_1 with probability $1-x_c$. With probability x_C his income becomes either I_0 or I_2. The probability of I_2 conditional on a change in income is thus I_1/I_2. Gambles are pairwise independent.

Finally, we suppose that the moving cost of an agent with income I is cI, where c is privately observed by each agent at the outset and distributed as above. We also assume risk-neutral agents for simplicity.

The sequence of events is as follows. Agents decide to gamble or not. Tax authorities observe the resulting wealth distributions and announce taxes. Migrations take place. Agents are taxed and consume.

This setting deserves several comments.

- *Risk-taking.* The substantial assumption is that risk-taking abilities may differ across countries. Risk-taking takes place in practice through occupational choice and/or investment choices. In the former case, a higher risk-taking ability corresponds to an economy in which innovation and entry are easier.[14] In the latter, this corresponds to less regulated financial markets—at least as far as sophisticated investors are concerned. A larger x_C corresponds to more contestable rents and/or less regulated markets. We can easily extend the model to gambles with positive or negative NPV, and to correlated gambles.
- *Migration costs.* We do not claim that the assumption of moving costs that are linear in income is particularly plausible. In fact, little is known about the empirical structure of these costs to our knowledge. We use linear costs because this makes transparent that risk-taking and non-linear taxes are not driven by economies of scale in migration in our model. We can easily enrich the setup with non-linear costs that would affect risk-taking incentives.

For brevity, we suppose

$$(1-x_A)\mu_{A,1} > (1-x_B)\mu_{B,1}, \tag{20}$$

which means that risk shifting does not make the initially "large" country become "small" at the I_1 income level. We deem taxation in a given country progressive (regressive) if the tax rate on I_2 is higher (lower) than that on I_1.

Proposition 8. (Inequality and tax arbitrage reinforce each other). *Suppose*

$$\alpha_A < \alpha_B < \frac{1-x_B}{1-x_A}\alpha_A + \frac{x_A - x_B}{1-x_A}. \tag{21}$$

Then there are at least two equilibria. There exists an equilibrium in which no agent gambles and an equilibrium in which all agents gamble.

- *In the no-gambling equilibrium, taxation is progressive in both countries, even more so in the small one.*
- *In the gambling equilibrium, taxation is regressive in both countries, even more so in the small one. Compared with the no-gambling equilibrium, pre-tax inequality is larger in both countries.*

14. Aghion *et al.* (2015) offer evidence suggesting that innovativeness can explain some fraction of top income inequality.

Proof. See Appendix. ‖

Note that condition (21) is satisfied if the large country has a smaller right tail of wealth distribution before risk-taking takes place (first inequality: this ensures existence of the non-gambling equilibrium) and offers sufficiently stronger risk-taking abilities than the small one (second inequality: this warrants the existence of the gambling equilibrium).

The intuition behind Proposition 8 is as follows. From Lemma 7, the tax rate that each jurisdiction applies to a given income level is entirely determined by the relative sizes of the populations at this income level in each jurisdiction. Tax rates are lower in both countries when the populations are more asymmetric, and their difference increases with respect to asymmetry in size. Thus, if the large country has a thinner income tail than the small one, it means that tax rates are smaller at income I_1 than at I_2 because asymmetry in size is more pronounced at the I_1 level. Such progressive taxation discourages risk shifting. Conversely, a relatively more efficient risk-shifting technology implies that the tail of income distribution is thicker in the large country when all agents shift risk. This leads to regressive taxation, which vindicates in turn-risk shifting.

In other words, there is "safety in numbers" for the super-rich in the large country because if they are sufficiently many, then the small country is willing to aggressively lower taxes to attract them. But then, local taxes are also lower at the top and this in turn makes gambling attractive in the first place: Pre-tax inequality through risk-taking and avoidance options reinforce each other.

The possibility of multiple equilibria is not interesting *per se*, but it nicely illustrates complementarities between risk shifting and avoidance. The previous section suggested that subadditive avoidance costs affected not only after-tax but also pre-tax inequality in the presence of endogenous risk shifting. We now come full circle with Proposition 8, which shows that pre-tax inequality feeds back in turn on avoidance options when these come from the option to migrate. This gambling equilibrium offers a parsimonious theory that jointly explains the strong correlation between cuts in top tax rates and increases in top 1% income shares since 1975 established in Piketty *et al.* (2014), together with the rise of asset migration towards tax havens over the period established in Zucman (2013).

5. RELATED LITERATURE

This article studies how two frictions—avoidance and risk shifting—affect the tax and redistribution capacities of a social planner. As such, it blends ingredients that have been studied in distinct literatures, and from different angles. This section discusses how this article relates to the respective existing literatures on tax avoidance, on risk shifting, and on the effect of frictions on inequality.

First, there is a surprising contrast between the large evidence that taxpayers do take advantage of available legal methods of reducing their fiscal obligations, and the relatively sparse theoretical literature on this topic. The literature on tax avoidance is by and large descriptive (see, *e.g.* Stiglitz, 1985). Slemrod and Kopczuk (2002) and Piketty *et al.* (2014) capture avoidance in a reduced form, as an exogenous elasticity of taxable income to the tax rate. Like us, Casamatta (2011) and Grochulski (2007) adopt the alternative approach of modelling tax avoidance as a primitive informational friction, and then deriving optimal fiscal policy as an optimal mechanism. We share with these contributions the modelling of avoidance as an *ex post* moral-hazard problem of costly diversion. Grochulski (2007) establishes the result that increasing returns to avoidance imply the optimality of avoidance-free schemes. Casamatta (2011) shows that this no longer need be the case when the function g is concave.

Secondly, the risk-shifting friction is a form of *ex ante* moral hazard that has been thoroughly studied in financial economics. In their seminal paper, Jensen and Meckling (1976) show that overly leveraged firms may undertake value-destroying projects provided these are sufficiently risky. A large asset-pricing literature studies how non-concavities stemming from compensation schemes or career concerns create risk-shifting incentives for fund managers. Contributions include Basak *et al.* (2007), Carpenter (2000), Ross (2004), and Makarov and Plantin (2015). We borrow our formal modelling of risk shifting as a choice among arbitrary distributions from the latter.

Our focus on how risk-taking shapes the wealth distribution relates to a number of paper that study economies in which agents care not only for consumption but also for their status (see, *e.g.* Robson, 1992; Becker *et al*. 2003; or Ray and Robson, 2012) In these contributions, status may induce non-concavities in utility over endowment, so that only wealth distribution that are sufficiently inequal discourage agents from gambling. In contrast, agents care only about consumption in our economy.

A long-standing literature in public finance studies the impact of taxation on risk-taking. The point that taxation may be encouraging risk-taking can be traced back to Domar and Musgrave (1994). [15] Our mechanism connecting taxes and risk-taking differs from the one considered in that literature in two ways. First, we derive the tax system from an informational friction in the spirit of Mirrlees, whereas that literature assumes tax functions similar to those actually used in practice, in the spirit of Ramsey. Also, the existing literature focuses on compensated investment risk-taking, not on the use of lotteries to convexify the lower contour set of the value function as we do.

Our setup also relates to various forms of secret side-trading studied by the public-finance literature. Contributions include Cole and Kocherlakota (2001), or Golosov and Tsyvinski (2007). In Cole and Kocherlakota (2001), agents can secretly save at an exogenously given rate. Golosov and Tsyvinski (2007) endogenize the price of the assets that agents secretly trade. Broadly, the goal of this literature is to study how agents' ability to secretly trade affects efficient production and risk sharing in economies with asymmetric information. Our purpose is quite different. We study an economy that is trivially Pareto-efficient at the outset. There are no gains from social interaction between agents: They receive risk-free endowments of a single private good and do not produce. Tax avoidance and side trades matter only because of the presence of a social planner who uses taxation to implement inequality-averse social views. Our focus is on how tax avoidance and risk shifting stand in the way of this social planner.

Finally, the prediction that regressive taxation at the top endogenously increases pre-tax inequality because of a risk-shifting friction is novel to our knowledge, although Posner informally made this claim.[16] The growth and development literatures have shown that credit constraints may create poverty traps that amplify income inequality (see, *e.g.* Greenwood and Jovanovic, 1990; Banerjee and Newman, 1993; Galor and Zeira, 1993; or Aghion and Bolton, 1997). We suggest that another friction—risk shifting—may amplify inequality at the top of the income distribution. No systematic empirical test of this prediction has been carried out to our knowledge. Yet, Gentry and Gentry and Hubbard (2000) and Cullen and Gordon (2007) document the related fact that entrepreneurial risk-taking is reduced when taxation becomes more progressive.

6. CONCLUSION

This article develops an optimal taxation framework that aims at capturing the specific constraints faced by a state in the taxation of its high-wealth citizen. Specifically, we develop a model where

15. We are grateful to an anonymous referee for pointing out this reference.

16. http://www.becker-posner-blog.com/2006/12/should-we-worry-about-the-rising-inequality-in-income-and-wealth–posner.html

agents have access to a tax avoidance technology with subadditive costs and to a risk-shifting technology. Subadditive costs reflect the fact that there can be high fixed costs in setting up structures that allow tax avoidance. The ability to engage in risk shifting reflects the fact that agents can react to taxes by changing occupations or taking financial risk. These two constraints are particularly relevant for high-wealth individuals. We find that optimal taxes are not progressive at the top in this context and do not depend on preferences. These non-progressive taxes can lead in turn to inefficient risk-taking by agents. We endogenize avoidance costs as migration decisions by agents who arbitrage between different countries of residence. We show that redistribution in a large country can be severely impaired by the tax policy of a smaller country: we describe equilibria where the small country attracts a flow of high-wealth citizen from the large country, inducing inefficient risk-taking as well as lower and non-progressive taxes in the large country. We believe our model sheds light on the limits to the taxation of rich residents faced by developed economies. These limits are important to analyse the fiscal capacity of developed economies. While global tax governance could (as suggested by Piketty, 2014; Zucman, 2015) mitigate the effect of tax avoidance at the top, we show that, as long as mobility is not infinitely costly, an arbitrarily small non-compliant state can limit the ability of a country to tax its wealthiest residents. While we emphasize geographic mobility as an important source of endogenous avoidance costs, political economy forces (such as lobbying by the wealthiest individual) constitute an alternative source that could be studied in future research in the context of our model.

APPENDIX

A.1. *Proof of Proposition 2*

Step 1. We first show that we can without loss of generality restrict the analysis to mechanisms such that

$$\forall\, w \geq 0,\ r(w) = w.$$

Consider an arbitrary scheme (r,v) that satisfies constraints (4). Define the scheme (ρ,ν) as

$$\rho(w) = w,$$

$$\nu(w) = v(r(w)) + g(w - r(w)).$$

First, the scheme (ρ,ν) is incentive-compatible: For all $w \geq 0$ and w' s.t. $w' < w$, we have

$$\begin{aligned}\nu(w) &= v(r(w)) + g(w-r(w)) \geq v(r(w')) + g(w - r(w')),\\ &\geq v(r(w')) + g(w' - r(w')) + g(w - w'),\\ &= \nu(w') + g(w - w').\end{aligned}$$

The first inequality stems from the fact that (r,v) is incentive-compatible.

The second one follows from the fact that g is superadditive.

Secondly, the scheme (ρ,ν) is feasible:

$$\begin{aligned}\int_0^{+\infty} \nu(w)dF(w) &= \int_0^{+\infty} v(r(w))dF(w) + \int_0^{+\infty} g(w-r(w))dF(w)\\ &\leq \int_0^{+\infty} r(w)dF(w) + \int_0^{+\infty} (w-r(w))dF(w)\\ &\leq \int_0^{+\infty} wdF(w) = \int_0^{+\infty} \rho(w)dF(w).\end{aligned}$$

Finally, the scheme (ρ,ν) delivers the same utility as the scheme (r,v) for each income level. Thus, the restriction to avoidance-free schemes is without loss of generality. ■

Step 2. Consider the following auxiliary programme

$$\max_{\nu} \int_0^{+\infty} u(\nu(w))dF(w)$$

$$s.t. \begin{cases} \int_0^{+\infty} \nu(w)dF(w) \leq \int_0^{+\infty} wdF(w), \\ \forall w \geq 0,\ \nu(w) \geq g(w) + \nu(0). \end{cases} \tag{A.1}$$

This amounts to considering only the deviation of a zero-report $w'=0$ in the incentive-compatibility constraints of $(\wp)$. We will show that

$$V(w)=g(w)+\int_0^{+\infty}(t-g(t))dF(t)$$

solves this programme. It is easy to see that V satisfies constraints (A.1).

Consider a function v that solves this programme. Clearly, v must be (weakly) increasing. Thus, v admits a left limit $v(x^-)$ and a right limit $v(x^+)$ at each point $x\in(0,+\infty)$. Suppose that for some $x_0\in(0,+\infty)$, $v(x_0^-)<v(x_0^+)$. Then one could slightly increase v in the left neighbourhood of x, slightly decrease it in the right neighbourhood, and thus strictly increase utilitarian welfare while still satisfying constraints (A.1). Thus, v must be continuous over $(0,+\infty)$ almost surely (and with a similar argument also has a right-limit in 0).

Suppose now that for some $x_1\in(0,+\infty)$,

$$v(x_1)>g(x_1)+v(0). \tag{A.2}$$

Since v and g are continuous, inequality (A.2) actually holds over some neighbourhood Ω of x_1. Consider a bounded measurable function h with support within Ω s.t. $\int hdF=0$. The function

$$w\to v(w)+th(w)$$

satisfies constraints (A.1) for t sufficiently small. Thus, it must be that

$$\Phi(t)=\int_0^{+\infty}u(v(w)+th(w))dF(w)$$

has a local maximum in 0, or that

$$\Phi'(0)=\int_0^{+\infty}u'(v(w))h(w)dF(w)=0. \tag{A.3}$$

Since equation (A.3) holds for any function h, and u is strictly concave, it must be that v is constant over Ω. Clearly this implies that v must be constant over $[0,x_1)$, which cannot be unless g is equal to 0 over this interval. In any case, this contradicts equation (A.2). Thus $v=V$.

Since constraints (A.1) are necessary conditions for constraints (4) and V happens to satisfy constraints (4) by superadditivity, this concludes the proof.

A.2. *Proof of Corollary 3*

Clearly the tax scheme (r^*,τ^*) satisfies the constraints of $(\wp')$ by superadditivity of g. Any scheme (r,τ) that also satisfies these constraints must in particular satisfy:

$$\forall w\geq 0,\ \tau(r(w))\leq r(w)+g(w-r(w))-g(w).$$

The right-hand side is maximal for $r(w)=w$ from equation (1), in which case it is equal to $\tau^*(w)$, which establishes the result.

Taxation cannot be strictly progressive because this would imply that g is strictly concave and thus strictly subadditive.

A.3. *Proof of Lemma 4*

For $w_0>0$, define

$$W^D(w_0) = \min_{(z_1,z_2)\in\mathbb{R}^2} z_1+w_0z_2,$$
$$s.t.\ \forall w\ \geq\ 0,\ z_1+wz_2\geq u(g(w)). \tag{A.4}$$

The programme defining $W^D(w_0)$ is the dual of that defining $W(w_0)$. It has a simple graphical interpretation. It consists in finding, among all the straight lines above the graph of $u\circ g$, the one that takes the smallest value in w_0. Makarov and Plantin (2015) show that the solutions to the primal and dual problems coincide. It is graphically intuitive that W^D is the concavification of $u\circ g$. We now prove it formally.

Fix $w_0>0$. The function $(z_1,z_2)\to z_1+w_0z_2$ is continuous. Thus, there exists at least one $(z_1(w_0),z_2(w_0))$ satisfying equation (A.4) such that $W^D(w_0)=z_1(w_0)+z_2(w_0)w_0$. Clearly, $z_2(w_0)\geq 0$ since $u\circ g$ is strictly increasing. For such a pair $(z_1(w_0),z_2(w_0))$, let

$$S(w_0)=\{w\geq 0: z_1(w_0)+z_2(w_0)w=u\circ g(w)\}.$$

Continuity of $u\circ g$ implies that $S(w_0)$ is non-empty and closed. It is clearly bounded and, therefore, compact. Let

$$\underline{\sigma}(w_0)=\min S(w_0),\quad \overline{\sigma}(w_0)=\max S(w_0).$$

We have:

$$\underline{\sigma}(w_0)\leq w_0\leq\overline{\sigma}(w_0). \tag{A.5}$$

Proof. We prove that $w_0 \leq \overline{\sigma}(w_0)$. The proof that $\underline{\sigma}(w_0) \leq w_0$ is symmetric. Suppose the opposite that $w_0 > \overline{\sigma}(w_0)$ then for some $\varepsilon \in (0, w_0 - \overline{\sigma}(w_0))$, let

$$\eta(\varepsilon) = \min_{y \geq \overline{\sigma}(w_0)+\varepsilon} \left\{ \frac{z_1(w_0) - u \circ g(y)}{y} + z_2(w_0) \right\}.$$

Clearly, $\eta(\varepsilon) > 0$.

Define (z_1', z_2') as $z_1' = z_1(w_0) + (\overline{\sigma}(w_0)+\varepsilon)\eta(\varepsilon)$, $z_2' = z_2(w_0) - \eta(\varepsilon)$. The pair (z_1', z_2') satisfies equation (A.4). To see this, notice that $z_1' + yz_2' = z_1(w_0) + yz_2(w_0) + \eta(\varepsilon)(\overline{\sigma}(w_0)+\varepsilon - y)$. Thus, $z_1' + yz_2' > z_1(w_0) + yz_2(w_0) \geq u \circ g(y)$ for $y < \overline{\sigma}(w_0)+\varepsilon$. Further, $z_1' + yz_2' \geq z_1(w_0) + yz_2(w_0) - \eta(\varepsilon) y \geq u \circ g(y)$ for $y \geq \overline{\sigma}(w_0)+\varepsilon$ by definition of $\eta(\varepsilon)$. At the same time,

$$z_1' + w_0 z_2' = z_1(w_0) + w_0 z_2(w_0) + (\overline{\sigma}(w_0)+\varepsilon - w_0)\eta(\varepsilon) < z_1(w_0) + w_0 z_2(w_0),$$

which contradicts the definition of $(z_1(w_0), z_2(w_0))$. ■

Inequalities (A.5) imply that for each w_0, we can define

$$\begin{cases} \underline{w}(w_0) = \sup\left\{ w \leq w_0 \ s.t. W^D(w) = u(g(w)) \right\} \\ \overline{w}(w_0) = \inf\left\{ w \geq w_0 \ s.t. W^D(w) = u(g(w)) \right\} \end{cases}$$

because these sets are not empty: they respectively contain $\underline{\sigma}(w_0)$ and $\overline{\sigma}(w_0)$. It must be indeed that

$$\begin{cases} z_1(w_0) + z_2(w_0)\underline{\sigma}(w_0) = u \circ g(\underline{\sigma}(w_0)) = W^D(\underline{\sigma}(w_0)) \\ z_1(w_0) + z_2(w_0)\overline{\sigma}(w_0) = u \circ g(\overline{\sigma}(w_0)) = W^D(\overline{\sigma}(w_0)) \end{cases}.$$

Thus, for any $w_0 > 0$, if $W^D(w_0) \neq u(g(w_0))$, then $[\underline{w}(w_0), \overline{w}(w_0)]$ is not a singleton, and W_D is linear over it. We are now able to prove:

$$W^D(w) = \overline{u \circ g}(w). \tag{A.6}$$

‖

Proof. Notice first that by construction, $W^D \geq u \circ g$. Secondly, suppose that there exists a concave function θ such that

$$\theta \ \geq \ u \circ g,$$

$$\exists w_0 \ s.t. \ \theta(w_0) \ < \ W^D(w_0).$$

In this case, it must be that $u \circ g(w_0) < W^D(w_0)$. But then, this means that θ is above the line $y = z_1(w_0) + xz_2(w_0)$ in $\underline{w}(w_0)$ and $\overline{w}(w_0)$, and strictly below it in w_0: it cannot be concave. Thirdly, W^D is concave. Suppose otherwise that there exists $w_1 < w_2 < w_3$ such that the chord between $(w_1, W^D(w_1))$ and $(w_3, W^D(w_3))$ is strictly above $(w_2, W^D(w_2))$ in w_2. This contradicts that there exists a straight line that meets the graph of W^D in w_2 and that is above the graph of W^D, since such a straight line cannot be above both $(w_1, W^D(w_1))$ and $(w_3, W^D(w_3))$. ■

Equality (A.6) also defines the risk-taking choices of an individual with initial income w_0. If

$$u(g(w_0) = W^D(w_0) = W(w_0),$$

then the agent reaches $W(w_0)$ by investing at the risk-free rate. If

$$u(g(w_0) < W^D(w_0),$$

then we have

$$z_1(w_0) + \underline{w}(w_0) z_2(w_0) \ = \ u\big(g(\underline{w}(w_0))\big),$$

$$z_1(w_0) + \overline{w}(w_0) z_2(w_0) \ = \ u(g(\overline{w}(w_0))).$$

so that

$$W^D(w_0) \ = \ z_1(w_0) + z_2(w_0) w_0 = \frac{w_0 - \underline{w}(w_0)}{\overline{w}(w_0) - \underline{w}(w_0)} u(g(\overline{w}(w_0)))$$

$$+ \frac{\overline{w}(w_0) - w_0}{\overline{w}(w_0) - \underline{w}(w_0)} u\big(g(\underline{w}(w_0))\big).$$

Thus, the lottery that pays off $\underline{w}(w_0)$ with probability $\frac{\overline{w}(w_0)-w_0}{\overline{w}(w_0)-\underline{w}(w_0)}$ and $\overline{w}(w_0)$ with probability $\frac{w_0-\underline{w}(w_0)}{\overline{w}(w_0)-\underline{w}(w_0)}$ attains $W^D(w_0) = W(w_0)$. Other lotteries with support in $S(w_0)$ can also attain it. But in this case their support is on the left of $\underline{w}(w_0)$ and on the right of $\overline{w}(w_0)$. Thus, they are dominated by this minimum one in the sense of second-order stochastic dominance because their c.d.f. must single cross that of this minimum lottery. ‖

A.4. *Proof of Corollary 6*

The reduction in utilitarian welfare due to risk-taking is established in the Online Appendix. To prove the second point, straightforward computations show that

$$E\left[u\left(v^{**}(0)\right)+g(\widetilde{\rho}(w))\right]>u\left(v^{*}(0)+g(w)\right)$$

if and only if

$$E\left[u\left(g(\widetilde{\rho}(w))-g(w)-\int_0^{+\infty}(E[g(\widetilde{\rho}(t))]-g(t))dF_0(t)\right)\right]>u(0).$$

Notice that $\widetilde{\rho}(.)$ does not depend on F_0.

If an arbitrarily large mass of F_0 is concentrated on agents that do not shift risk, then the inequality is satisfied for an agent with income w who shifts risk, because $\int_0^{+\infty}(E[g(\widetilde{\rho}(t))]-g(t))dF_0(t)$ becomes arbitrarily small.

Let $t_0\in\arg\max_{t\geq 0}\{(E[g(\widetilde{\rho}(t))]-g(t))\}$. If the mass of F_0 is arbitrarily concentrated on t_0, then all agents are worse off because

$$\max_{w\geq 0}\left\{E[g(\widetilde{\rho}(w))]-g(w)-\int_0^{+\infty}(E[g(\widetilde{\rho}(t))]-g(t))dF_0(t)\right\},$$

which is positive, is arbitrarily close to 0. Thus, lotteries $g(\widetilde{\rho}(w))-g(w)-\int_0^{+\infty}(E[g(\widetilde{\rho}(t))]-g(t))dF_0(t)$ are arbitrarily close to fair in the limit and a risk-averse agent is unwilling to take any of them in the limit.

A.5. *Proof of Lemma 7*

A resident from country $C\in\{A:B\}$ with moving costs c moves to the other country $-C$ if and only if

$$\tau_C>\tau_{-C}+c. \tag{A.7}$$

(Indifferent agents are negligible.) Given tax rates τ_A, τ_B, the respective tax revenues of A and B are, therefore:

$$\tau_A[\mu_A(1-F_m(\tau_A-\tau_B))+\mu_B F_m(\tau_B-\tau_A)], \tag{A.8}$$

$$\tau_B[\mu_B(1-F_m(\tau_B-\tau_A))+\mu_A F_m(\tau_A-\tau_B)], \tag{A.9}$$

which yields first-order conditions:

$$\mu_A(1-F_m(\tau_A-\tau_B))+\mu_B F_m(\tau_B-\tau_A)$$
$$=\tau_A[\mu_A f_m(\tau_A-\tau_B)+\mu_B f_m(\tau_B-\tau_A)], \tag{A.10}$$

$$\mu_B(1-F_m(\tau_B-\tau_A))+\mu_A F_m(\tau_A-\tau_B)$$
$$=\tau_B[\mu_A f_m(\tau_A-\tau_B)+\mu_B f_m(\tau_B-\tau_A)]. \tag{A.11}$$

Suppose a solution to equations (A.10) and (A.11) is such that $\tau_B\geq\tau_A$. Substracting equation (A.10) from (A.11) then yields

$$0\leq\tau_B-\tau_A=\frac{1-\frac{\mu_A}{\mu_B}-2F_m(\tau_B-\tau_A)}{f_m(\tau_B-\tau_A)}<0, \tag{A.12}$$

a contradiction. It must, therefore, be that $\tau_A>\tau_B$, in which case equations (A.10) and (A.11) can be rearranged as

$$\tau_A=\frac{1-F_m(\tau_A-\tau_B)}{f_m(\tau_A-\tau_B)}, \tag{A.13}$$

$$\tau_A-\tau_B=\frac{1-\frac{\mu_B}{\mu_A}-2F_m(\tau_A-\tau_B)}{f_m(\tau_A-\tau_B)}. \tag{A.14}$$

There is a unique solution to equations (A.13) and (A.14) because $T(x)=[1-\mu_B/\mu_A-2F_m(x)]/f_m(x)$ is decreasing when positive. To see this, consider $x_1\leq x_2$ such that $T(x_i)\geq 0$. If $f_m(x_1)\leq f_m(x_2)$, then clearly $T_m(x_1)\geq T_m(x_2)$. Otherwise, note that

$$\frac{T(x)}{2}=\frac{1-F_m(x)}{f_m(x)}-\frac{1+\frac{\mu_B}{\mu_A}}{2f_m(x)}, \tag{A.15}$$

and both terms on the right-hand side are decreasing. Inspection of equations (A.13) and (A.14) yields the comparative statics properties of τ_A and τ_B.

A.6. *Proof of Proposition 8*

From the linearity of migration costs, tax rates at each income level are determined exactly as in Lemma 7.

We first show that no agent finds it optimal to gamble if he believes that the others do not. In this case, $\alpha_A < \alpha_B$ implies that taxes are progressive in both countries, with

$$0 < \tau_{A,2} - \tau_{B,2} < \tau_{A,1} - \tau_{B,1}. \tag{A.16}$$

This implies that a resident of the small country never moves (taxes are higher at every income level in the large country) and never gambles. This also implies that if a resident of the large country finds it optimal to move with income I_2, then he also finds it optimal to move with income I_1. Thus, he always faces an *ex ante* progressive schedule and never finds gambling optimal.

We then show that every agent finds it optimal to gamble if he believes that all other agents do so. The right-hand inequality in equality (21) can be re-arranged as

$$\frac{\mu_{A,1}(1-x_A)}{\mu_{B,1}(1-x_B)} < \frac{\mu_{A,2}+\mu_{A,1}x_A I_1/I_2}{\mu_{B,2}+\mu_{B,1}x_B I_1/I_2}, \tag{A.17}$$

meaning that income distribution has a thicker tail in the large country than in the small one after risk-taking. This implies that taxes are regressive in both countries, more so in the small one, and that

$$0 < \tau_{A,1} - \tau_{B,1} < \tau_{A,2} - \tau_{B,2}. \tag{A.18}$$

Again, a resident in the small country never finds it optimal to move. If a resident in the large country finds it optimal to move when he has income I_1, then it must also be the case when he has income I_2. Thus, he always faces regressive taxes and finds gambling optimal.

Acknowledgments. We thank Marco Ottaviani (the editor), five anonymous referees, Daron Acemoglu, Bruno Biais, Patrick Bolton, Philippe De Donder, Christian Hellwig, Thomas Mariotti, Thomas Piketty, Sven Rady, Sylvia Serfaty, Konrad Stahl, and seminar participants at Gerzensee (ESSET), Ecole Polytechnique, London School of Economics, Stockholm School of Economics, and Toulouse School of Economics for very helpful comments. Plantin benefitted from a European Research Council Starting Grant (N°263673 – RIFIFI). A.L. acknowledges financial support from a Scor Chair at the JJ Laffont foundation and from the European Research Council under the European Community Seventh Framework Programme (FP7/2007-2013) Grant Agreement N°312503 – SOLSYS.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

REFERENCES

AGHION, P., AKCIGIT U., BERGEAUD, A. *et al.* (2015), "Innovation and Top Income Inequality", (Working Paper).
AGHION, P. and BOLTON, P. (1997), "A Theory of Trickle-Down Growth and Development", *The Review of Economic Studies*, **64**.
AUMANN, R. and PERLES, M. (1965), "A Variational Problem Arising in Economics", *Journal of Mathematical Analysis and Applications*, **11**.
BANERJEE, A. and NEWMAN, A. (1977), "Occupational Choice and the Process of Development", *Journal of Political Economy*, **101**.
BASAK, S., PAVLOVA A. and SHAPIRO, A. (2007), "Optimal Asset Allocation and Risk Shifting in Money Management", *Review of Financial Studies*, **20**.
BAUMOL, W. (1977), "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry", *The American Economic Review* **67**.
BECKER, G., MURPHY K. and WERNING, I. (2005), "The Equilibrium Distribution of Income and the Market for Status", *Journal of Political Economy* , **113**.
CARPENTER, J. (2000), "Does Option Compensation Increase Managerial Risk Appetite?", *Journal of Finance*, **55**.
CASAMATTA, G. (2011), "Optimal Income Taxation with Tax Avoidance" (Working Paper).
COLE, H. and KOCHERLAKOTA, N. (2001), "Efficient Allocations with Hidden Income and Hidden Storage", *Review of Economic Studies*, **68**.
CULLEN, J. B. and GORDON, R. (2007), "Taxes and Entrepreneurial Risk-Taking: Theory and Evidence for the U.S.", *Journal of Public Economics*, **91**.
DOMAR, E. and MUSGRAVE, R. (1944), "Proportional Income Taxation and Risk-Taking", *The Quarterly Journal of Economics*, **58**.
GALOR, O. and ZEIRA, J. (1993), "Income Distribution and Macroeconomics", *The Review of Economic Studies*, **60**.
GENTRY, W. and HUBBARD, R. G. (2000), "Tax Policy and Entrepreneurial Entry", *American Economic Review*, **90**.

GOLOSOV, M. and TSYVINSKI, A. (2007), "Optimal Taxation with Endogenous Insurance Markets", *Quarterly Journal of Economics*, **122**.

GREENWOOD, J. and JOVANOVIC, B. (1990), "Financial Development, Growth, and the Distribution of Income", *Journal of Political Economy*, **98**.

GROCHULSKI, B. (2007), "Optimal Nonlinear Income Taxation with Costly Tax Avoidance", *Economic Quarterly - Federal Reserve Bank of Richmond*, **93**.

HOPENHAYN, H. and VERESHCHAGINA, G. (2009), "Risk taking by entrepreneurs", *The American Economic Review*, **99**.

JENSEN, M. and MECKLING, W. (1976), "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", *Journal of Financial Economics*, **3**.

JOHANNESEN, N. and ZUCMAN, G. (2014), "The End of Bank Secrecy? An Evaluation of the G20 Tax Haven Crackdown", *American Economic Journal: Economic Policy*, **6**.

KANBUR, R. and KEEN, M. (1993), "Jeux Sans Frontières: Tax Competition and Tax Coordination When Countries Differ in Size", *The American Economic Review*, **83**.

KLEVEN, H., LANDAIS C. and SAEZ, E. (2013), "Taxation and International Mobility of Superstars: Evidence from the European Football Market", *American Economic Review*, **103**.

KLEVEN, H., LANDAIS, C., SAEZ, E. *et al*. (2014), "Migration and Wage Effects of Taxing Top Earners: Evidence from the Foreigners' Tax Scheme in Denmark", *The Quarterly Journal of Economics*, **129**.

LANG, O., NORHASS, K. -H. and STAHL, K. (1997), "On Income Tax Avoidance; The Case of Germany", *Journal of Public Economics*, **66**.

LEHMANN, E., SIMULA L. and TRANNOY, A. (2014), "Tax Me if You Can! Optimal Nonlinear Income Tax between Competing Governments", *The Quarterly Journal of Economics*, **129**.

MAKAROV, I. and PLANTIN, G. (2015), "Rewarding Trading Skills Without Inducing Gambling", *Journal of Finance*, **70**.

PIKETTY, T. (2014), "Capital in the Twenty-First Century" (Belknap Press).

PIKETTY, T., and SAEZ, E. (2003), "Income Inequality in the United States, 1913–1998", *Quarterly Journal of Economics*, **118**, Table updates found at http://elsa.berkeley.edu/∼saez/.

PIKETTY, T., SAEZ, E. and STANTCHEVA, S. (2014), "Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities", *American Economic Journal: Economic Policy*, **6**.

RAJAN, R. (2010), "Bankers' Pay Is Deeply Flawed", *Financial Times*.

RAY, D. and ROBSON, A. (2012), "Status, intertemporal Choice, And Risk-Taking", *Econometrica*, **80**.

ROBSON, A. (1992), "Status, the Distribution of Wealth, Private and Social Attitudes to Risk", *Econometrica*, **60**.

ROSS, S. (2004), "Compensation, Incentives, and the Duality of Risk Aversion and Riskiness", *Journal of Finance* **59**.

SHACKELFORD, D. (2000), "The Tax Environment Facing the Wealthy", in Slemrod, J. B. (ed.) *Does Atlas Shrug? The Economic Consequences of Taxing the Rich*, (New York:Russell Sage Foundation and Harvard University).

SHUE, K., and TOWNSEND, R.R. (2014), "Swinging for the fences: Executive Reactions to quasi-random option grants" (Chicago Booth Research Paper No.13-03).

SLEMROD, J. and KOPCZUK, W. (2002), "The Optimal Elasticity of Taxable Income", *Journal of Public Economics*, **84**.

STIGLITZ, J. (1985), "The General Theory of Tax Avoidance", *National Tax Journal*, **38**.

ZUCMAN, G. (2013), "The Missing Wealth of Nations: Are Europe and the U.S. Net Debtors or Net Creditors?", *The Quarterly Journal of Economics*, **128**.

ZUCMAN, G. (2015), *The Hidden Wealth of Nations: The Scourge of Tax Havens* (University Of Chicago Press).