Optimal Income Taxation: 
Mirrlees Meets Ramsey*

Jonathan Heathcote  
Hitoshi Tsujiyama  
FRB of Minneapolis and CEPR  
Goethe University Frankfurt  
May, 2014 (Preliminary and Incomplete)

Abstract

What structure of income taxation can maximize the social benefits of redistribution and public insurance while minimizing the social harm associated with distorting the allocation of labor input? Many authors have advocated scrapping the current tax system, which redistributes primarily via marginal tax rates that rise with income, and replacing it with a flat tax system, in which marginal tax rates are constant and redistribution is achieved via non-means-tested transfers. In this paper we compare alternative tax systems in an environment with distinct roles for public and private insurance. We evaluate alternative policies using a social welfare function designed to capture the taste for redistribution reflected in the current tax system. In our preferred specification we find that moving to the optimal tax policy in the affine class is welfare reducing, while moving to the optimal fully non-linear Mirrlees policy generates only tiny welfare gains. These findings suggest that proposals for dramatic tax reform should be viewed with caution.

*We thank V.V. Chari, Mikhail Golosov, James Peck, Richard Rogerson, Ctirad Slavik, Kjetil Storesletten, Yuichi Waki and Tomoaki Yamada for helpful comments. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.
1 Introduction

In this paper we revisit a classic question in public finance: what structure of income taxation can maximize the social benefits of redistribution and public insurance while minimizing the social harm associated with distorting the allocation of labor input?

The tax and transfer systems currently in place in many countries, including the United States, achieve a measure of redistribution through a combination of tax rates that increase with income, coupled with means-tested transfers. However, many authors have argued that a more efficient way to redistribute would be to move to a flat tax system, in which marginal tax rates are constant across the income distribution, and redistribution is achieved via universal transfers. For example, Friedman (1962) advocated a "negative income tax", which effectively combines a lump-sum transfer with a constant marginal tax rate. Mirrlees (1971) pioneered the study of optimal tax design in environments with unobservable heterogeneity. He found that even if one imposes no constraints on the shape of the tax schedule, the optimal non-linear income tax is in fact close to linear, a finding that extends to many other papers in the subsequent literature: see Mankiw, Weinzierl, and Yagan (2009) for a recent survey.

In this paper we explore the welfare consequences of replacing the current US income tax system – which redistributes primarily via increasing marginal tax rates – with two alternatives tax schemes: the optimal one, and the best policy in the affine class. In our preferred model specification, we find that moving to the best affine policy is welfare reducing, while moving to the optimal fully non-linear Mirrlees policy generates only tiny welfare gains. These findings suggest that proposals for dramatic tax reform should be viewed with caution.

A natural starting point for characterizing the optimal structure of taxation is the Mirrleesian approach, which seeks to characterizes the optimal tax system subject only to the constraint that taxes must be a function of individual earnings. Taxes cannot be explicitly conditioned on individual productivity or individual labor input because these are assumed to be unobserved by the tax authority. The first step towards solving for the optimal tax schedule is to formulate a planner's problem in which the planner chooses allocations directly to maximize a social welfare function subject to resource feasibility, and subject to incentive constraints such that individuals have incentives to truthfully reveal their productivity types. Given a solution to this problem, an earnings tax schedule can be inferred, such that the same allocations are decentralized as a competitive equilibrium given those taxes. This Mirrleesian approach is attractive because it places no constraints on the shape of the tax schedule, and because the implied allocations are constrained efficient.

The alternative Ramsey approach to tax design is to restrict the planner to choose a tax schedule within a parametric class. While there are no theoretical foundations for
imposing ad hoc restrictions on the design of the tax schedule, the practical advantage
of doing so is that one can then consider tax design in richer models. We will contrast
the optimal Mirrlees policy to two particular parametric functional forms for the income
tax schedule $T(y)$. The first is an affine tax: $T(y) = \tau_0 + \tau_1 y$, where $\tau_0$ is a lump-sum
tax or transfer, and $\tau_1$ is a constant marginal tax rate. Under this specification, a
higher marginal tax rate $\tau_1$ translates into larger lump-sum transfers and thus more
redistribution. The second specification is borrowed from Heathcote, Storesletten, and
Violante (2014): $T(y) = y - \lambda y^{1-\tau}$. show that the following parametric function for
income taxes net of transfers closely approximates the current US tax and transfer
system: $T(y) = y - \lambda y^{1-\tau}$. In this HSV specification the parameter $\tau$ controls the
progressivity of the tax and transfer system: for $\tau = 0$ taxes are proportional to
income, while for $\tau > 0$ the marginal tax rate increases with income. Note that the
HSV specification rules out lump-sum transfers, since $T(0) = 0$. By comparing welfare
in the two cases we will learn whether it is more important to allow for lump-sum
transfers (as in the affine case) or to allow for marginal tax rates to increase with
income (as in the HSV case). We will also be interested in whether either affine or
HSV tax systems can come close to decentralizing constrained efficient allocations.

We want to provide quantitative guidance on the welfare-maximizing shape of the
tax function. With this goal in mind we extend the standard static Mirrlees framework
in two directions. First, rather than assuming that the government provides all the
insurance in the economy, we instead assume that agents are able to privately insure a
share of idiosyncratic labor productivity risk. Second, rather than taking a utilitarian
social welfare function as our baseline objective function, we instead evaluate alterna-
tive tax systems using a welfare function that is designed to be consistent with the
amount of redistribution embedded in the US tax code. In this environment we first
solve for constrained efficient allocations, following the Mirrlees approach, and then ex-
perience whether our parametric Ramsey-style tax functions come close to decentralizing
those allocations.

Private insurance operates as follows. Agents in our model are heterogeneous only
with respect to labor productivity, but idiosyncratic productivity has two orthogonal
components: $\log(w) = \alpha + \varepsilon$. When individuals first enter the economy they draw $\alpha$,
which is private information to the agent and cannot be privately insured – the standard
Mirrlees assumptions. Agents then purchase explicit private insurance indexed to the
second component of the wage $\varepsilon$.

We assume that the planner only observes total individual income, which comprises
labor earnings plus private insurance income. Because the planner cannot directly
observe insurable shocks, it cannot make individuals’ income or consumption a function
of their idiosyncratic insurable shocks. Thus our environment contains distinct roles
for both public and private insurance. For the purposes of practical tax design, the more risk agents are able to insure privately, the smaller is the role of the government in providing social insurance, and the less redistributive will be the resulting tax schedule.

The shape of the optimal tax schedule in any social insurance problem is sensitive to the social welfare function that the planner is assumed to be maximizing. We will consider a class of social welfare functions in which the weight on an agent with unobservable productivity component $\alpha$ takes the form $\exp(-\theta \alpha)$. Here the parameter $\theta$ determines the taste for redistribution: $\theta = 0$ corresponds to the utilitarian case.

What is the taste for redistribution in the United States? We argue that the degree of progressivity built into the actual US tax and transfer system is informative about the preferences of US voters and policymakers. Heathcote, Storesletten, and Violante (2014) show that the HSV functional form closely approximates the current US tax and transfer system: $T(y) = y - \lambda y^{1-\tau}$. We are able to characterize in closed-form the mapping between the taste-for-redistribution parameter $\theta$ in our class of social welfare functions and the progressivity parameter $\tau$ that maximizes welfare within this HSV class of tax / transfer systems. This mapping can be inverted to infer the US taste for redistribution $\theta^*$ that would lead a planner to choose precisely the observed degree of tax progressivity $\tau^*$. This empirically-motivated social welfare function will serve as our baseline objective function.

Another element of the environment that is known to be critical for the shape of the optimal tax function is the form of the distribution of uninsurable risk. Saez (2001) has forcefully argued that the heavy Pareto-like right tail for the distribution of labor earnings suggests that marginal tax rates should be increasing, at least over some portion of the earnings distribution. In our calibration we are careful to replicate observed dispersion in US wages. We assume that log labor productivity follows an exponentially-modified Gaussian (EMG) distribution, and we estimate the exponential parameter defining the weight of the right tail using cross-sectional data on the distribution of household earnings from the Survey of Consumer Finances. We decompose the overall variance of wages into the uninsurable and insurable components described above by adapting estimates from Heathcote, Storesletten, and Violante (forthcoming) on the fraction of idiosyncratic wage variation that is privately insured.

Our key findings are as follows.

First, in our baseline model, the welfare gains of moving from the current tax system – which we approximate using the HSV functional form – to the tax system that decentralizes the Mirrlees solution are very small: 0.1 percentage points of consump-

---

1For example, in the simplest static Mirrlees problem, one can construct a social welfare function in which Pareto weights increase with productivity at a rate such that the planner has no desire to redistribute, and thus (absent public expenditure) no desire to tax. Alternatively, a Rawlsian welfare objective that puts weight only on the least well-off agent in the economy will typically call for a highly progressive tax schedule.
tion. Moving to the optimal policy in the affine class would reduce welfare by around 0.6 percentage points.

Second, these results do not hinge critically on our baseline empirically-motivated social welfare function, which under-weights low productivity workers. When we re-compute optimal taxes assuming a standard utilitarian objective, the welfare gains from optimally increasing progressivity within the HSV class of tax functions exceed the welfare gains from moving to the optimal policy in the affine class.

Third, similar results extend to a specification in which there is no private insurance against idiosyncratic wage risk. Assuming away private insurance, while retaining our baseline social welfare function, leads to a larger role for government redistribution, and thus more progressive taxation. But the optimal policy in the HSV class again delivers higher welfare than the optimal policy in the affine class.

We conclude from these experiments that the presence of lump-sum transfers is not a critical component of optimal fiscal policy, but that it is important that marginal tax rates increase with income. This finding contrasts with much of the existing literature which has argued in favor of affine tax schedules.

Why is it important to have marginal tax rates that increase with income? Saez (2001) emphasized that the distribution for labor productivity is a critical determinant of the shape of the optimal tax schedule. With a heavy right tail to the wage distribution, there is a strong incentive to set high marginal tax rates at high income levels because the associated extra revenue from all higher-income individuals is large. When we counter-factually assume a log-normal distribution for wages, we find that an affine tax very nearly implements efficient allocations, while the best policy in the HSV class does less well.

Other elements also play a role in explaining why the efficient tax system involves relatively small transfers in our baseline model. First, our social welfare function puts relatively low weight on the utility of low productivity agents. Second, a portion of low wage draws reflect shocks that are privately insured and do not transmit to consumption. This reduces the role for public transfers in establishing a consumption floor.

In an extension to our baseline model we introduce a third component of idiosyncratic productivity, $\kappa$, which is privately uninsurable but observed by the planner. This component is designed to capture differences in wages related to observable characteristics like age and education. Because wages vary systematically by these characteristics, a constrained efficient tax system should explicitly index taxes to these observables (see, e.g., Weinzierl (2011)). In our model we assume that $\kappa$ is drawn before the agent can trade in financial markets, and therefore cannot be insured privately. We set the variance of this observable fixed effect to reflect the amount of wage dispersion that can
be accounted for by standard observables in a Mincer regression. We find that if the planner can condition taxes on the observable component of labor productivity it can generate large welfare gains, in part because it translates into lower marginal rates on average. Under an affine system with two observable types (think college versus high school), the type with higher observable productivity (college graduates) should face higher marginal tax rates, and receive smaller lump-sum transfers. Higher marginal rates on the more productive type allow the planner to redistribute across types, while smaller lump-sum transfers offset the disincentive effects of higher marginal tax rates on labor supply.

**Related Literature**

Seminal papers in the literature on taxation in the Mirrlees tradition include Mirrlees (1971), Diamond (1998), and Saez (2001). More recent work has focussed on extending the approach to dynamic environments: Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2013) are prominent examples. There are also many papers on tax design in the Ramsey tradition in economies with heterogeneity and incomplete private insurance markets. Recent examples include Conesa and Krueger (2006), who explore the Gouveia and Strauss (1994) functional form for the tax schedule, and Heathcote, Storesletten, and Violante (2014), who explore the function used by Feldstein (1969), Persson (1983), and Benabou (2000).

Chetty and Saez (2010) is one of the few papers to explore the interaction between public and private insurance in environments with private information. They consider a range of alternative environments, in most of which agents face a single idiosyncratic shock that can be insured privately or publicly. Section III of their paper explores a more similar environment to ours, in which there are two components of productivity, and differential roles for public versus private insurance with respect to the two components. Like us, they conclude that the government should focus on insuring the source of risk that cannot be insured privately. Relative to Chetty and Saez (2010) our contributions are two-fold: (i) we consider optimal Mirrleesian tax policy in addition to affine tax systems, and (ii) our analysis is more quantitative in nature.

Our interest in constructing social welfare functions that are broadly consistent with observed tax progressivity is related to a recent paper by Werning (2007). His goal is to characterize the Pareto efficiency or inefficiency of any given tax schedule, given an underlying skill distribution. Our paper is similar in spirit, in that we start with the current tax system, and then construct a social welfare function that makes the current system look as good as possible. However, our focus will be on quantifying the extent of inefficiency in the current system, rather than on a zero-one classification.
of efficiency.  

2 Environment

We now describe the static environment and define a competitive equilibrium.

Labor Productivity

There is a unit mass of agents. Agents differ only with respect to labor productivity \( w \) which has two orthogonal components

\[
\log w = \alpha + \varepsilon. \tag{1}
\]

These two idiosyncratic components differ with respect to whether or not they can be insured privately. The first component \( \alpha \in A \subset \mathbb{R} \) represents shocks that cannot be insured privately, while perfect private insurance exists against shocks to \( \varepsilon \in E \subset \mathbb{R} \). One interpretation for the differential insurance assumption is that \( \alpha \) represents fixed effects that are drawn before agents can participate in insurance markets, while \( \varepsilon \) captures productivity shocks later in the life cycle against which agents can purchase insurance. An alternative interpretation is that \( \varepsilon \) represents shocks that can be pooled within a family or other risk-sharing group, while \( \alpha \) is a common productivity component which is shared by all members of the group but which differs across groups.

We assume that \( \alpha \) and \( \varepsilon \) and hence \( w \) are not observed by the tax authority. We let the vector \((\alpha, \varepsilon)\) denote an individual’s type, and \( F_\alpha \) and \( F_\varepsilon \) denote the distributions for the two components. We assume \( F_\alpha \) and \( F_\varepsilon \) are differentiable with derivatives are given by \( f_\alpha \) and \( f_\varepsilon \), respectively.

Preferences

Agents have identical preferences over consumption \( c \), and work effort \( h \). The utility function is separable between consumption and work effort and takes the form

\[
u(c, h) = \frac{c^{1-\gamma}}{1-\gamma} - \frac{h^{1+\sigma}}{1+\sigma},
\]

where \( \gamma > 0 \) and \( \sigma > 0 \). Given this functional form, the Frisch elasticity of labor supply is \( 1/\sigma \). We denote by \( c(\alpha, \varepsilon) \) and \( h(\alpha, \varepsilon) \) consumption hours worked for an individual of type \((\alpha, \varepsilon)\).

Technology

\[\text{In our model environment, the distribution of productivity will be bounded above. It follows immediately that the current tax system is not Pareto efficient, since it violates the familiar zero-tax-at-the-top result.}\]
Aggregate output in the economy is simply aggregate effective labor supply

$$Y = \int \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon).$$

Aggregate output is divided between private consumption and a publicly-provided good that is allocated equally across all agents,

$$Y = \int \int c(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon) + G.$$

The resource constraint of the economy is then given by

$$\int \int c(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon) + G = \int \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\alpha(\alpha) dF_\varepsilon(\varepsilon). \tag{2}$$

**Insurance Markets**

After observing uninsurable component $\alpha$, agents trade in insurance markets in which they can purchase private insurance at actuarially fair prices against $\varepsilon$. The price of insurance claims that will pay one unit of consumption if and only if $\varepsilon \in E \subset \mathcal{E}$ is

$$Q(E) = \int_E dF_\varepsilon(\varepsilon)$$

for any Borel set $E$ in $\mathcal{E}$. The budget constraint for an agent with $\alpha$ is given by

$$\int B(\varepsilon; \alpha) Q(\varepsilon) d\varepsilon = 0, \tag{3}$$

where $B(\varepsilon; \alpha)$ denotes the quantity (positive or negative) of insurance claims purchased that pay a unit of consumption if and only if the draw for the insurable shock is $\varepsilon$.

**Government**

The tax authority does not directly observe $\alpha$ or $\varepsilon$, does not observe hours worked, and does not observe any of the trades $B(\cdot; \alpha)$ associated with private insurance against $\varepsilon$. The tax authority observes only total end of period income $y(\alpha, \varepsilon)$ which is the sum of labor earnings plus insurance payouts:

$$y(\alpha, \varepsilon) = \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) + B(\varepsilon; \alpha). \tag{4}$$

We let $T(y)$ denote the income tax schedule. Given that it observes income and taxes collected the authority also effectively observes consumption, since

$$c(\alpha, \varepsilon) = y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon)). \tag{5}$$

**Agent’s Problem**

The timing of events is as follows. Agents first draw $\alpha$. They then trade in insurance markets against $\varepsilon$. Next, each individual draws an $\varepsilon$, insurance pays out, and individuals choose how much to work. Finally, the government assesses taxes, and income less net
taxes is consumed.

Now we formally state the agent’s maximization problem. An agent with $\alpha$, before drawing $\varepsilon$, solves

$$\max_{c(\alpha,\varepsilon),h(\alpha,\varepsilon),B(\varepsilon;\alpha)} \int \frac{c(\alpha,\varepsilon)^{1-\gamma}}{1 - \gamma} - \frac{h(\alpha,\varepsilon)^{1+\sigma}}{1 + \sigma} dF_{\varepsilon}(\varepsilon),$$

subject to equations (3), (4) and (5).

**Equilibrium**

Given the income tax schedule $T$, a competitive equilibrium for this economy is a set of decision rules $\{c, h, B\}$, associated income $y$, and insurance prices $Q$ such that

1. The decision rules $\{c, h, B\}$ solve the agent’s maximization problem and the associated income $y$ is given by (4).
2. The insurance prices are actuarially fair, $Q(E) = \int_E dF(\varepsilon)$.
3. The resource feasibility constraint (2) is satisfied.
4. The government budget constraint is satisfied:

$$\int \int T(y(\alpha, \varepsilon)) dF_{\alpha}(\alpha) dF_{\varepsilon}(\varepsilon) = G.$$

While the model we describe here is static, it would be straightforward to develop a dynamic extension in which agents draw new values for the insurable shock $\varepsilon$ in each period. A more challenging extension would be to allow for persistent shocks to the unobservable non-insurable component of productivity $\alpha$.

### 3 Planner’s Problem

The planner maximizes a social welfare function characterized by weights $W(\alpha, \varepsilon)$ that potentially vary with $\alpha$ and $\varepsilon$. In Section 4, we develop a methodology for using the degree of progressivity built into the actual tax system to reverse engineer the nature of this variation.

#### 3.1 Ramsey Problem

The Ramsey planner chooses the optimal tax function in a given parametric class $T$.

For example, if we consider a class of affine functions, then $T = \{T : \mathbb{R}_+ \to \mathbb{R}_+ | T(y) =$

---

$^3$We will show later that the planner will be unable to offer contracts in which income or consumption vary with $\varepsilon$, and given that result we will focus on a social welfare function in which weights do not vary with $\varepsilon$. 

"
\[ \tau_0 + \tau_1 y \text{ for all } y \in \mathbb{R}_+, \tau_0 \in \mathbb{R}, \tau_1 \in \mathbb{R} \}. \] For HSV tax functions described in the introduction, \( T = \{ T : \mathbb{R}_+ \to \mathbb{R}_+ | T(y) = y - \lambda y^{1-\tau} \text{ for all } y \in \mathbb{R}_+, \lambda \in \mathbb{R}, \tau \in \mathbb{R} \} \).

The Ramsey problem is to choose the optimal income tax schedule in \( T \) subject to allocations being a competitive equilibrium:

\[
\max_{T \in T} \int \int W(\alpha, \varepsilon) u(c(\alpha, \varepsilon), h(\alpha, \varepsilon))dF_\alpha(\alpha)dF_\varepsilon(\varepsilon) \quad (8)
\]

subject to (2) and to \( c(\alpha, \varepsilon) \) and \( h(\alpha, \varepsilon) \) being solutions to the agent’s maximization problem (6).

We first simplify the agent’s problem (6) by substituting in constraints to give

\[
\max_{y(\alpha, \varepsilon), h(\alpha, \varepsilon)} \int \left( \frac{[y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon))]^{1-\gamma}}{1-\gamma} - \frac{h(\alpha, \varepsilon)^{1+\sigma}}{1+\sigma} \right) dF_\varepsilon(\varepsilon),
\]

subject to

\[
\int \exp(\alpha + \varepsilon)h(\alpha, \varepsilon) - y(\alpha, \varepsilon) dF_\varepsilon(\varepsilon) = 0.
\]

The first-order conditions for the agent’s problem are given by

\[
[y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon))]^{-\gamma} \left[ 1 - T'(y(\alpha, \varepsilon)) \right] - \mu(\alpha) = 0
\]

\[
-h(\alpha, \varepsilon)^{\sigma} + \exp(\alpha + \varepsilon)\mu(\alpha) = 0 \quad (9)
\]

where \( \mu(\alpha) \) is the multiplier on the constraint.

Our first proposition defines a property of the tax function that guarantees that income and consumption are independent of the insurable shock \( \varepsilon \).

**Proposition 1** If the tax function satisfies

\[
T''(y) > -\gamma \frac{(1 - T'(y))^2}{y - T(y)} \quad (10)
\]

for all feasible \( y \) then (i) the first order conditions are sufficient, and (ii) optimal income, consumption, and tax payments are independent of \( \varepsilon \).

**Proof.** The second order conditions for the agent’s problem are given by

\[
-\gamma [y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon))]^{-\gamma-1} \left[ 1 - T'(y(\alpha, \varepsilon)) \right]^2 - [y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon))]^{-\gamma} T''(y(\alpha, \varepsilon)) < 0
\]

\[
-\sigma h(\alpha, \varepsilon)^{\sigma-1} < 0
\]

The first equation is satisfied by (10), while the second equation is satisfied because \( \sigma > 0 \). Therefore, the first order conditions are sufficient.
The first equation in (9) is
\[ y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon)) - \gamma [1 - T'(y(\alpha, \varepsilon))] = \mu(\alpha). \]

The right-hand side is independent of the insurable shock \( \varepsilon \) and hence its derivative with respect to \( \varepsilon \) is zero. The derivative of the left-hand side of this equation with respect to \( \varepsilon \) is, by the Chain Rule,
\[
\frac{\partial}{\partial \varepsilon} \left\{ y(\alpha, \varepsilon) - T(y(\alpha, \varepsilon)) - \gamma [1 - T'(y(\alpha, \varepsilon))] \right\} = \frac{-\gamma (y - T(y))^{-\gamma - 1} [1 - T'(y)]^2 - (y - T(y))^{-\gamma} T''(y) \partial y(\alpha, \varepsilon)}{\partial \varepsilon}.
\]

The first term is nonzero by (10), which immediately implies that \( \frac{\partial y}{\partial \varepsilon} = 0 \).

Given this proposition, the first-order condition for labor supply can be written as
\[ h(\alpha, \varepsilon) = y(\alpha) - T(y(\alpha)) - \gamma [1 - T'(y(\alpha))] \exp(\alpha + \varepsilon) \frac{1}{\sigma}, \]
where \( y(\alpha) \) can be solved for from the budget constraint
\[ y(\alpha) = \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\varepsilon(\varepsilon). \]

We now show that (10) is satisfied for the tax functions in which we are particularly interested.

### 3.1.1 Affine Taxes

Suppose income taxes are of the affine class, \( T(y) = \tau_0 + \tau_1 y \). Then (10) is satisfied because
\[ T''(y) + \gamma \frac{[1 - T'(y)]^2}{y - T(y)} = \gamma \frac{(1 - \tau_1)^2}{y - T(y)} > 0. \]

With this class of tax functions, it is possible to characterize equilibrium allocations more sharply. We have the following explicit solution for hours worked, as a function of the wage and income
\[ h(\alpha, \varepsilon) = [(y(\alpha)(1 - \tau_1) - \tau_0)^{-\gamma} \exp(\alpha + \varepsilon)(1 - \tau_1)]^{\frac{1}{\gamma}}, \]
where \( y(\alpha) \) can be solved for from the budget constraint
\[ y(\alpha) = \int \exp(\alpha + \varepsilon) h(\alpha, \varepsilon) dF_\varepsilon(\varepsilon). \]
\subsection{3.1.2 HSV Taxes}

Suppose income taxes are of the HSV class, \( T(y) = y - \lambda y^{1-\tau} \). Then (10) becomes

\[
T''(y) + \gamma \frac{[1 - T'(y)]^2}{y - T(y)} = \lambda (1 - \tau) \tau y^{-\tau - 1} + \gamma \frac{(\lambda (1 - \tau) y^{-\tau})^2}{\lambda y^{1-\tau}}
\]

\[
= \lambda y^{-\tau - 1} (1 - \tau) [\tau + \gamma (1 - \tau)] > 0.
\]

This is satisfied for any progressive tax, \( \tau \in [0, 1) \), because

\[
\tau + \gamma (1 - \tau) > 0.
\]

It is also satisfied for any regressive tax, \( \tau < 0 \), if \( \gamma \geq 1 \), because

\[
\gamma \geq 1 > \frac{-\tau}{1 - \tau}.
\]

Therefore, for all relevant parameterizations, (10) is also satisfied for this class of tax functions.

The first-order condition for labor supply can be written as

\[
h(\alpha, \varepsilon) = [y(\alpha) - T(y(\alpha))]^{-\gamma} [1 - T'(y(\alpha))] \exp(\alpha + \varepsilon)^\frac{1}{\sigma}\]

\[
= \left[ \exp(\alpha + \varepsilon) (1 - \tau) \lambda^{1-\gamma} y(\alpha)^{-\gamma} \right]^\frac{1}{\sigma}
\]

where again \( y(\alpha) \) can be solved for from the budget constraint.

\section{3.2 Mirrlees Problem: Constrained Efficient Allocations}

The Mirrlees planner cannot observe either \( \alpha \) or \( \varepsilon \). The privately uninsurable \( \alpha \) component is standard in the literature, but the privately insurable component \( \varepsilon \) introduces a new complication. We now show that efficient allocations for consumption and income are independent of \( \varepsilon \), and that the planner’s problem collapses to a familiar form in which incentive constraints only pertain to \( \alpha \).

\subsection{3.2.1 Setup}

In the Mirrlees formulation of the program that determines constrained efficient allocations, rather than thinking of the planner choosing taxes, we will instead think of the planner as choosing both consumption \( c(\alpha, \varepsilon) \) and income \( y(\alpha, \varepsilon) \) as functions of the individual types \((\alpha, \varepsilon)\). It is clear that, by choosing taxes, the tax authority can choose the difference between income and consumption. It is less obvious that the planner can also dictate income \textit{levels} as a function of type. To achieve this, the Mirrlees formu-
lation of the planner’s problem includes incentive constraints that guarantee that for each and every type \((\alpha, \varepsilon)\) an agent of that type weakly prefer to deliver to the planner the value for income \(y(\alpha, \varepsilon)\) the planner intends for that type, thereby receiving (after-taxes) the value for consumption \(c(\alpha, \varepsilon)\), rather than delivering any alternative level of income. It is easiest to formulate the Mirrlees problem with the planner inviting agents to report their unobservable characteristics \(\alpha\) and \(\varepsilon\), and then assigning the agent an allocation for income \(y(\tilde{\alpha}, \tilde{\varepsilon})\) and consumption \(c(\tilde{\alpha}, \tilde{\varepsilon})\) on the basis of her reports \(\tilde{\alpha}\) and \(\tilde{\varepsilon}\). But since the planner is offering agents a choice between a menu of alternative pairs for income and consumption, it is clear that an alternative way to think about what the planner does is that it offers a mapping from any possible value for income to consumption. Such a schedule can be decentralized via a tax schedule on income \(y\) of the form \(T(y)\) that defines how rapidly consumption grows with income.\(^4\)

The timing within the period is as follows. In a first stage, agents draw \(\alpha \in A\). They make a report \(\tilde{\alpha} \in A\) to the planner. In a second stage, agents buy private insurance \(B(\varepsilon)\) against \(\varepsilon\). They then draw \(\varepsilon \in E\), and insurance pays out. Agents then make a report \(\tilde{\varepsilon} \in E\) to the planner. Finally, agents work sufficient hours to deliver \(y(\tilde{\alpha}, \tilde{\varepsilon})\), and receive consumption \(c(\tilde{\alpha}, \tilde{\varepsilon})\).

**Second stage agent’s problem**

As a first step towards characterizing efficient allocations, we start with the agent’s problem in the second stage, given the agent’s report \(\tilde{\alpha} \in A\) in the first stage and the optimal reporting strategy for \(\varepsilon, \hat{\varepsilon} : A \times A \times E \times B \to E\) where \(B\) is the functional space of all possible insurance purchase decision \(B\).\(^5\) Agents choose insurance purchases to solve the following problem

\[
\max_{B(\cdot)} \int \left( \frac{c(\tilde{\alpha}, \hat{\varepsilon}(\alpha, \tilde{\alpha}, \varepsilon; B))^{1-\gamma}}{1-\gamma} - \frac{y(\tilde{\alpha}, \hat{\varepsilon}(\alpha, \tilde{\alpha}, \varepsilon; B) - B(\varepsilon) \exp(\alpha + \varepsilon))^{1+\sigma}}{1+\sigma} \right) dF_\varepsilon(\varepsilon), \tag{12}
\]

\(^4\)Note that some values for income might not feature in the menu offered by the Mirrlees planner. Those values will not be chosen in decentralization with income taxes as long as income at those values is taxed sufficiently heavily. For example, suppose the lowest value for income in the Mirrlees solution is \(y\), with corresponding consumption value \(c\). Lower income values can be ruled in the competitive equilibrium by assuming that marginal tax rates are 100\% for income below \(y\) and that conditional on delivering at least \(y\) agents receive a lump-sum transfer equal to \(c\).

\(^5\)To obtain \(\hat{\varepsilon}\), we consider the following agent’s final stage problem, given the agent’s report \(\tilde{\alpha} \in A\) in the first stage and a generic insurance purchase decision \(B : E \to \mathbb{R}\). By the simple relationship between hours worked and income described in (4), the optimal reporting strategy \(\hat{\varepsilon} : A \times A \times E \times B \to E\) is given by

\[
\hat{\varepsilon}(\alpha, \tilde{\alpha}, \varepsilon; B) \in \arg \max_{\varepsilon \in E} \frac{c(\tilde{\alpha}, \hat{\varepsilon}(\alpha, \tilde{\alpha}, \varepsilon; B))^{1-\gamma}}{1-\gamma} - \frac{y(\tilde{\alpha}, \hat{\varepsilon}(\alpha, \tilde{\alpha}, \varepsilon; B) - B(\varepsilon) \exp(\alpha + \varepsilon))^{1+\sigma}}{1+\sigma}.
\]
subject to
\[ \int B(\varepsilon)Q(\varepsilon)d\varepsilon = 0. \] (13)

The solution to this problem is the optimal decision rule for the insurance purchases \(B(\varepsilon; \alpha, \tilde{\alpha})\).

Let \(v(\alpha, \tilde{\alpha}, \varepsilon, \tilde{\varepsilon})\) denote maximum utility after drawing \((\alpha, \varepsilon)\) and reporting \((\tilde{\alpha}, \tilde{\varepsilon})\).

### First stage planner’s problem

The planner maximizes the social welfare function characterized by the weights \(W(\alpha, \varepsilon)\) subject to the resource constraint, and incentive constraints that ensure that utility from reporting \(\alpha\) and \(\varepsilon\) truthfully and receiving the associated allocation is weakly larger than expected welfare from any alternative report and associated allocation. Formally, the planner solves

\[
\max_{\{c(\alpha, \varepsilon), y(\alpha, \varepsilon)\}} \int \int W(\alpha, \varepsilon)v(\alpha, \alpha, \varepsilon, \varepsilon)dF_\alpha(\alpha)dF_\varepsilon(\varepsilon),
\]

subject to

\[
\int \int c(\alpha, \varepsilon)dF_\alpha(\alpha)dF_\varepsilon(\varepsilon) + G = \int \int y(\alpha, \varepsilon)dF_\alpha(\alpha)dF_\varepsilon(\varepsilon),
\]

and subject to allocations being incentive compatible:

\[
\int v(\alpha, \alpha, \varepsilon, \varepsilon)dF_\varepsilon(\varepsilon) \geq \int v(\alpha, \tilde{\alpha}, \varepsilon, \varepsilon)dF_\varepsilon(\varepsilon) \quad \text{for all } \alpha \text{ and } \tilde{\alpha}, \quad (14)
\]

\[
v(\alpha, \tilde{\alpha}, \varepsilon, \varepsilon) \geq v(\alpha, \tilde{\alpha}, \varepsilon, \tilde{\varepsilon}) \quad \text{for all } \alpha, \tilde{\alpha}, \varepsilon \text{ and } \tilde{\varepsilon}. \quad (15)
\]

The second set of incentive constraints (15) imposes that agents weakly prefer to report \(\varepsilon\) truthfully for any report \(\tilde{\alpha}\) in the first stage. This ensures that truth-telling is incentive compatible for any first stage strategy. The first set of incentive constraints (14) impose that agents weakly prefer to report \(\alpha\) truthfully, assuming truthful reporting in the second stage. Thus truth-telling is incentive compatible at the first stage.

#### 3.2.2 Efficient Allocations cannot be conditioned on \(\varepsilon\)

We start by assuming agents follow arbitrary reporting strategies and compute the implied equilibrium allocations and utility. We show that for any reporting strategies, hours worked are not a function of the report \(\tilde{\varepsilon}\). It follows that a truth-telling reporting strategy is incentive compatible if and only if consumption is also independent of the report \(\tilde{\varepsilon}\). It follows further that there is nothing to be gained from the planner asking agents to report \(\varepsilon\) – since neither component of individual welfare (and hence social
Proposition 2 Income and consumption are independent of the insurable type \( \varepsilon \) in the constrained efficient allocation.

Proof. We begin with the second stage agent’s problem (maximize (12) subject to (13)). Let \( \mu \) denote the multiplier on the constraint. Note that \( \mu \) cannot be a function of \( \varepsilon \) because the constraint applies before \( \varepsilon \) is drawn. Applying the envelope theorem, the first-order condition becomes

\[
[\mu (\alpha, \tilde{\alpha}) \exp(\alpha + \varepsilon)]^{\frac{1}{\sigma}} = \frac{y(\hat{\alpha}, \hat{\varepsilon} (\alpha, \tilde{\alpha}, \varepsilon; B)) - B(\varepsilon)}{\exp(\alpha + \varepsilon)} \text{ for any } \varepsilon, \alpha \text{ and } \tilde{\alpha}.
\]

Because this is satisfied for any \( \varepsilon, \alpha \text{ and } \tilde{\alpha} \), we rewrite the last term using (4) and get the following

\[
h (\alpha, \tilde{\alpha}, \varepsilon) = [\mu (\alpha, \tilde{\alpha}) \exp(\alpha + \varepsilon)]^{\frac{1}{\sigma}} \text{ for any } \varepsilon, \alpha \text{ and } \tilde{\alpha}.
\]

The crucial thing to note here is that hours are independent of the reporting strategy \( \hat{\varepsilon} \). Because hours worked, and thus disutility from hours, are independent of the report \( \tilde{\varepsilon} \), it follows that the agent will choose whichever contract offers the highest value for \( c(\tilde{\alpha}, \tilde{\varepsilon}) \). Formally, the optimal reporting strategy \( \hat{\varepsilon} \) is given by

\[
\hat{\varepsilon} (\alpha, \tilde{\alpha}, \varepsilon) \in \arg \max \hat{\varepsilon} \in \varepsilon \left\{ \frac{c(\tilde{\alpha}, \hat{\varepsilon})^{1-\gamma}}{1 - \gamma} - \frac{h (\alpha, \tilde{\alpha}, \varepsilon)^{1+\sigma}}{1 + \sigma} \right\} = \arg \max \hat{\varepsilon} \in \varepsilon c(\tilde{\alpha}, \hat{\varepsilon}).
\]

Thus truth telling is incentive compatible if and only if consumption is independent of the individual report \( \tilde{\varepsilon} \). This means that neither consumption nor hours can be made contingent on the report of \( \varepsilon \) in any truth-telling allocation, and hence contracts must be of the form \( \{c(\tilde{\alpha}), y(\tilde{\alpha})\} \).

The intuition for this proposition is the following. Agents exploit insurance markets to deliver target income \( y(\tilde{\alpha}, \tilde{\varepsilon}) \) at minimum cost in terms of labor effort. That implies that they choose hours as a function of the draw for \( \varepsilon \) to make marginal disutility of hours \( h (\alpha, \tilde{\alpha}, \varepsilon)^{\sigma} \) proportional to the wage \( \exp(\alpha + \varepsilon) \) – a familiar complete markets result. The wage is obviously independent of \( \tilde{\varepsilon} \) and the constant of proportionality is too – it is the multiplier on the insurance purchases constraint, which applies prior to \( \varepsilon \) and thus \( \tilde{\varepsilon} \) being realized.

Note that because neither consumption nor income depend on \( \varepsilon \) in the constrained efficient allocation, the income or consumption tax system that decentralizes efficient allocations is also independent of \( \varepsilon \).
3.2.3 Restatement of Mirrlees problem

We now apply this result to produce a much simpler representation of the program that defines constrained efficient allocations. We drop the dependence of consumption and income on the insurable shock $\varepsilon$, i.e., \( \{c(\tilde{\alpha}), y(\tilde{\alpha})\} \). In addition, because the planner weights do not vary with $\varepsilon$, and because we do not need to worry about the second set of incentive constraints (15), we can also dispense with keeping track of private insurance markets altogether, and work with expected utility (conditional on $\alpha$) prior to drawing $\varepsilon$.

From the agent’s budget constraint we can solve for the multiplier on the insurance purchases constraint, $\mu$, and thus simplify the expression for hours $h(\alpha, \tilde{\alpha}, \varepsilon)$. Solving for $\mu$ and substituting the expression into (16) gives

$$h(\alpha, \tilde{\alpha}, \varepsilon) = \exp\left(\frac{\alpha + \varepsilon}{\sigma} y(\tilde{\alpha}) \right) \int \exp\left(\frac{\alpha + \varepsilon}{\sigma} dF(\varepsilon)\right).$$

Thus the utility, given type $(\alpha, \varepsilon)$ and report $\tilde{\alpha}$, is

$$v(\alpha, \tilde{\alpha}, \varepsilon) = \frac{c(\tilde{\alpha})^{1-\gamma}}{1-\gamma} - \frac{1}{1+\sigma} \left( \frac{\exp(\alpha + \varepsilon)\frac{1}{\sigma} y(\tilde{\alpha})}{\int \exp(\alpha + \varepsilon) \frac{1}{\sigma} dF(\varepsilon)} \right)^{1+\sigma}.$$

The expected utility, prior to drawing $\varepsilon$, is then

$$\int v(\alpha, \tilde{\alpha}, \varepsilon) dF(\varepsilon) = \int \left\{ \frac{c(\tilde{\alpha})^{1-\gamma}}{1-\gamma} - \frac{1}{1+\sigma} \left( \frac{\exp(\alpha + \varepsilon)\frac{1}{\sigma} y(\tilde{\alpha})}{\int \exp(\alpha + \varepsilon) \frac{1}{\sigma} dF(\varepsilon)} \right)^{1+\sigma} \right\} dF(\varepsilon)$$

$$= \frac{c(\tilde{\alpha})^{1-\gamma}}{1-\gamma} - \left( \int \exp(\varepsilon) \frac{1}{\sigma} dF(\varepsilon) \right)^{-\sigma} \frac{y(\tilde{\alpha})}{\exp(\alpha)} \left( \int \exp(\varepsilon) \frac{1}{\sigma} dF(\varepsilon) \right)^{1+\sigma}.$$

Let $U(\alpha, \tilde{\alpha}) \equiv \int v(\alpha, \tilde{\alpha}, \varepsilon) dF(\varepsilon)$. We can now restate the planner’s problem that defines constrained efficient allocations.

$$\max_{\{c(\alpha), y(\alpha)\}} \int W(\alpha) U(\alpha, \alpha) dF(\alpha) \quad \text{(17)}$$

subject to

$$\int c(\alpha) dF(\alpha) + G = \int y(\alpha) dF(\alpha), \quad \text{(18)}$$

and

$$U(\alpha, \alpha) \geq U(\alpha, \tilde{\alpha}) \quad \text{for all } \alpha \text{ and } \tilde{\alpha}. \quad \text{(19)}$$

There only one set of incentive constraints (19) that impose that agents weakly prefer to report $\alpha$ truthfully. Note that $\varepsilon$ no longer appears anywhere in this problem,
and nor does the second stage agent’s problem. The problem is identical to a standard static Mirrlees type problem, where the planner faces a distribution of agents with heterogeneous unobserved productivity $\alpha$. We will solve this problem numerically. Note, however, that the period utility function for each agent in this problem is not identical to the utility function we started with in the underlying problem, since the weight on hours worked is now $\Omega \equiv \left( \int \exp(\varepsilon) \frac{1+\sigma}{\sigma} dF(\varepsilon) \right)^{-\sigma}$.

### 3.2.4 Decentralization of constrained efficient allocations

Let $c^*(\alpha)$ and $y^*(\alpha)$ denote consumption and income in the constrained efficient allocation and let $h^*(\alpha, \varepsilon)$ denote the constrained efficient allocation rule for hours (all given a particular social welfare function $W(\alpha)$):

$$h^*(\alpha, \varepsilon) = \frac{\exp(\varepsilon)^{\frac{1}{\sigma}} y^*(\alpha)}{\exp(\alpha) \int \exp(\varepsilon)^{\frac{1+\sigma}{\sigma}} dF(\varepsilon)}.$$

Substituting constrained efficient allocations into the competitive equilibrium first-order condition for hours worked in the economy with income taxes (9) implicitly defines the marginal tax rates that decentralize constrained efficient allocations,

$$1 - T'(y^*(\alpha)) = \frac{y^*(\alpha)^{\sigma}}{c^*(\alpha)^{-\gamma} \exp(\alpha)^{1+\sigma} \left( \int \exp(\varepsilon)^{\frac{1+\sigma}{\sigma}} dF(\varepsilon) \right)^{\sigma}}.$$

Note that marginal tax rates do not vary with $\varepsilon$ because income (including insurance payouts) does not vary with $\varepsilon$.

### 3.3 First Best

If the planner can observe $\alpha$ directly, the welfare maximization problem is identical to the one described above, except that there are no incentive compatibility constraints (19). Formally, the planner’s problem is to maximize (17) subject to (18).

---

6It is clear that the Mirrlees solution could equivalently be decentralized using consumption taxes. In that case we would get

$$1 + T'(c^*(\alpha)) = \frac{c^*(\alpha)^{-\gamma} \exp(\alpha)^{1+\sigma} \left( \int \exp(\varepsilon)^{\frac{1+\sigma}{\sigma}} dF(\varepsilon) \right)^{\sigma}}{y^*(\alpha)^{\sigma}}.$$
4 Social Welfare Function

In this section we describe our methodology for using the degree of progressivity built into the actual tax system to infer social preferences.

We argue that the degree of progressivity built into the actual tax and transfer system is informative about the preferences of voters and policymakers. [XX add arguments from the probabilistic voting model]

We characterize in closed-form the mapping between the taste-for-redistribution parameter in a class of social welfare functions and the progressivity parameter that maximizes welfare within the HSV class of tax and transfer systems. This mapping can be inverted to infer the taste for redistribution that would lead a planner to choose precisely the observed degree of tax progressivity.

4.1 Empirically-Motivated Social Welfare Function

Social Preferences

Assume the social welfare function takes the form

$$W(\alpha) = \frac{\exp(-\theta \alpha)}{\int \exp(-\theta \alpha) dF_{\alpha}(\alpha)}. \quad (20)$$

The parameter $\theta$ controls the extent to which the planner puts relatively more or less weight on low productivity workers relative to high productivity workers. For example, with a negative $\theta$, the planner puts relatively high weights on the productive agents, while with a positive $\theta$ the planner cares more about the unfortunate agents. Moreover, the case that $\theta = 0$ corresponds to the Utilitarian social preferences, and this function converges to Rawlsian social preferences as $\theta \to \infty$. Hence, this specification is flexible enough to nest the standard social preferences in the literature.

We implicitly assume that the planner puts equal weight on agents with different realizations for $\varepsilon$. This is because, as explained in Section 3.2.2, constrained efficient allocations cannot be made to vary with $\varepsilon$.

Tax Function

Heathcote, Storesletten, and Violante (2014) argue that the following income tax function offers a reasonable approximation to the US tax and transfer system

$$T(y) = y - \lambda y^{1-\tau}. \quad (21)$$

See Section 5 for more details. The marginal tax rate on individual income is given by

$$T'(y) = 1 - \lambda (1 - \tau)(y)^{-\tau}. $$
Planner’s Problem

The planner solves a Ramsey problem (8) given the two-parameter class tax functions described by (21) and the social welfare function characterized by (20).

Note that while in principle the planner chooses two tax parameters, \( \lambda \) and \( \tau \), in (21), because it has to respect the government budget constraint, she effectively has a single choice variable, \( \tau \). Denote \( \tau^*(\theta) \) the welfare maximizing choice for \( \tau \) given a social welfare function indexed by \( \theta \).

Empirically-Motivated Social Welfare Function

While we do not directly observe \( \theta \), we do observe the degree of progressivity chosen by the US political system. Denote this degree by \( \tau^{US} \). Then the US social preference \( \theta^{US} \) must satisfy

\[ \tau^*(\theta^{US}) = \tau^{US}. \]  \hfill (22)

We use this fixed point problem (22) to reverse engineer \( \theta^{US} \) and obtain our empirically-motivated social welfare function indexed by \( \theta^{US} \).

4.2 Estimating the Social Preferences

We now describe how we operationalize this.

The budget constraint for an individual with \( (\alpha, \varepsilon) \) is

\[ c(\alpha) = \lambda y(\alpha, \varepsilon)^{1-\tau} \] for every \( \varepsilon \).

Since this applies for every \( \varepsilon \),

\[ c(\alpha)^{\frac{1}{1-\tau}} = \int \lambda^{\frac{1}{1-\tau}} y(\alpha, \varepsilon) dF_\varepsilon(\varepsilon), \]

which, using the definition for income (4) and substituting in the expression for hours (11), can be expressed as

\[ c(\alpha) = \lambda^{\frac{1+\sigma}{\sigma+\tau+\gamma(1-\tau)}} (1-\tau)^{\frac{1-\tau}{\sigma+\tau+\gamma(1-\tau)}} \exp \left( \frac{(1+\sigma)(1-\tau)}{\sigma+\tau+\gamma(1-\tau)} \alpha \right) \left[ \int \exp \left( \frac{1+\sigma}{\sigma} \varepsilon \right) dF_\varepsilon(\varepsilon) \right]^{\frac{\sigma(1-\tau)}{\sigma+\tau+\gamma(1-\tau)}}. \]  \hfill (23)

Given expressions (11) and (23), we use the government budget constraint to solve for \( \lambda \) and compute social welfare for any values for \( \tau \) and \( \theta \).

4.3 Analytical Form for \( \theta \) in Baseline Economy

As we explain in detail in Section 5, in our baseline economy, the utility is assumed to be logarithmic in consumption, so \( \gamma = 1 \). Also we assume that \( F_\alpha \) is Exponentially-
Modified Gaussian, \( EMG(\mu, \eta^2, a) \), and \( F_\varepsilon \) is Gaussian, \( N(-v_\varepsilon/2, v_\varepsilon) \). Then we have the following proposition.

**Proposition 3** The social preference parameter \( \theta^* \) is the solution to the following quadratic equation:

\[
\eta^2 \theta^* - \frac{1}{a + \theta^*} = -\frac{1}{\eta^2 (1 - \tau)} - \frac{1}{a - 1 + \tau} + \frac{1}{1 + \sigma} \left[ \frac{1}{(1 - g)(1 - \tau)} - 1 \right],
\]

where \( g \) is the ratio of government purchases to output, i.e., \( g = G/Y(\tau) \) and \( Y(\tau) \) denotes aggregate output.

**Proof.** See the Appendix. \( \blacksquare \)

The equation (24) is novel and useful because if the US political system is solving this problem, we can use (24) in conjunction with the observed choices for \( g^{US} \) and \( \tau^{US} \) to infer \( \theta^{US} \).

As a special case in which \( F_\alpha \) is also Gaussian and hence \( w \) follows a log-normal distribution, we get the following by taking \( a \to \infty \) in (24),

\[
\theta^* = -(1 - \tau) + \frac{1}{\eta^2} \frac{1}{1 + \sigma} \left[ \frac{1}{(1 - g)(1 - \tau)} - 1 \right].
\]

Using (24), we can do some comparative statics.

First, \( \theta^* \) is increasing in \( \tau \), as expected. Thus if we observe more progressive taxation, we can infer that the social planner puts less weight on higher wage individuals. Holding fixed observed \( \tau \), \( \theta^* \) is decreasing in \( \eta^2 \). Thus more uninsurable risk but the same tax progressivity means we can infer the planner has less desire to redistribute. Also \( \theta^* \) is decreasing in \( \sigma \), meaning that for the same observed \( \tau \), the less elastic we think labor supply is (and thus the smaller the distortions associated with progressive taxation) the less desire to redistribute we should attribute to the planner. Finally \( \theta^* \) is increasing in \( g \), meaning that, holding fixed progressivity, the larger the share of output devoted to public goods, the more the planner wants to redistribute. The logic here is that tax progressivity tends to reduce labor supply, making it more difficult to finance public goods, so governments that need to finance large expenditure will tend to choose less progressivity – unless they have a strong desire to redistribute.

[XX comparative statics wrt \( a ? \)]

---

7In this case, \( \theta^* \) is uniquely determined. Importantly, this provides a numerical guidance about which of the two roots that solves (24) is more reasonable.
5 Calibration

Preferences

We assume preferences are separable between consumption and labor effort, and logarithmic in consumption:

\[ u(c, h) = \log c - \frac{h^{1+\sigma}}{1 + \sigma}. \]

This specification is the same one adopted by Heathcote, Storesletten, and Violante (2014). The parameter \( \sigma \) controls the Frisch elasticity of labor supply, which is \( 1/\sigma \). We choose \( \sigma = 2 \) so that the Frisch elasticity is 0.5. This value is broadly consistent with the microeconomic evidence (see, e.g., Keane (2011)), and is also very close to the value estimated by Heathcote, Storesletten, and Violante (forthcoming). Note that because this preference specification is consistent with balanced growth, high and low wage workers will work equally hard in the absence of private insurance or public redistribution.

Tax and transfer system

Heathcote, Storesletten, and Violante (2014) argue that the US tax and transfer system is well approximated by the following function which expresses taxes net of transfers as a function of income \( y \):

\[ T(y) = y - \lambda y^{1-\tau}. \] (25)

This class of tax functions was first used by Feldstein (1969) and introduced into dynamic heterogeneous agent models by Persson (1983) and Benabou (2000). Heathcote, Storesletten, and Violante (2014) show that this functional form closely mimics actual effective tax rates in the United States.

They begin by noting that the functional form in (25) implies a linear relationship between \( \log y \) and \( \log(y - T(y)) \), with a slope equal to \((1 - \tau)\). Thus given micro data on household income before taxes and transfers, and income net of taxes and transfers, it is straightforward to estimate \( \tau \) by ordinary least squares. Using micro data from the Panel Study of Income Dynamics (PSID) for working age households over the period 2000 to 2006, Heathcote, Storesletten, and Violante (2014) estimate \( \tau = 0.151 \). Figure 1, borrowed from that paper, shows the relationship between income before and after taxes and transfers for fifty equal size bins of the distribution of household income, ranked from lowest to highest pre-government income. The x axis shows log of average pre-government income for each bin, while the y axis shows log of average income after taxes and transfers. The red line shows the least squares best fit through the underlying PSID micro data, with estimated slope \((1 - 0.151)\).
The remaining fiscal policy parameter, $\lambda$, is set such that aggregate revenue net of transfers is equal to 18.8 percent of GDP, which was the ratio of government purchases to output in the United States in 2005 ($\gamma = 0.188$).

**Wage distribution**

We need to characterize individual productivity dispersion, and we need to decompose this dispersion into an uninsurable component $\alpha$, and an orthogonal insurable component $\varepsilon$. For the variances of the two wage components, we adapt estimates from Heathcote, Storesletten, and Violante (forthcoming). They estimate a richer version of the model considered in this paper using micro data from the PSID and the Consumer Expenditure Survey, assuming that an individual’s labor productivity is equal to their reported earnings per hour. They are able to identify the relative variances of the two wage components by exploiting two key implications of the theory: a larger variance for insurable shocks will imply smaller cross-sectional variance in consumption, and a larger covariance between wages and hours worked. They find a variance for insurable shocks of $v^*_\varepsilon = 0.193$ in 2004 (adding the two solid lines in panels B and C of their Figure 3), which we adopt directly.\(^8\) Our target for variance of the uninsurable component is $v^*_\alpha = 0.273$, implying a total cross-sectional wage variance of $v^*_\alpha + v^*_\varepsilon = 0.466$.\(^9\) Given

\(^8\)As in this paper, the model in Heathcote, Storesletten, and Violante (forthcoming) features two independent sources of wage risk, one which is explicitly insurable, and one which remains uninsured in equilibrium. In the background the government runs the same type of tax-transfer system considered in this paper. The environment in Heathcote, Storesletten, and Violante (forthcoming) features explicit dynamics from which the present paper abstracts. In addition, Individuals in that model differ with respect to preferences as well as labor productivity.

\(^9\)Heathcote, Storesletten, and Violante (forthcoming) report a slightly lower total wage variance of 0.419 for 2004 (Figure 3). However, their estimate is net of wage variation reflecting differences in age. Rather than excluding this wage variation, we treat it as part of the uninsurable component of wages. In Section

---

Figure 1: Fit of HSV tax function
these estimates, 41% of the variance of log earnings is directly insurable via private markets.

Although we now have estimates for the variances of the two wage components, it is well known in the Mirrlees literature that higher moments of the shape of the productivity distribution have a large impact on the shape of the optimal tax schedule. In particular, Saez (2001) and others have emphasized that the shape of the right tail of the productivity distribution matters a lot, that there is more mass in the right tail than would be implied by a log-normal wage distribution, and that the right tail of the log wage distribution is well-approximated by an exponential distribution (so the right tail of the level wage distribution is Pareto). Unfortunately, as Mankiw, Weinzierl, and Yagan (2009) emphasize, it is difficult to sharply estimate the shape of the distribution given typical household surveys, such as the Current Population Survey (CPS). Part of the problem is that high income households tend to be under-represented in these samples. We therefore turn to the Survey of Consumer Finances (SCF) which uses data from the Internal Revenue Service (IRS) Statistics of Income program to ensure that wealthy households are appropriately represented.\footnote{The SCF has some advantages over the IRS data used by Saez (2001). First, the unit of observation is the household, rather than the tax unit. Second, the IRS data excludes those who do not file tax returns, or who file late. Third, people may have an incentive to under-report income to the IRS, while in principle they have no such incentive to do so for surveys like the SCF.}

Recall that we need distributions for both the insurable and uninsurable components of the wage. We will assume that the insurable component $\varepsilon$ is normally distributed, $\varepsilon \sim N(-v^*_\varepsilon/2, v^*_\varepsilon)$, and that the uninsurable component $\alpha$ follows an Exponentially-Modified Gaussian (EMG) distribution: $\alpha = \alpha_1 + \alpha_2$ where $\alpha_1 \sim N(\mu, \eta^2)$ and $\alpha_2 \sim Exp(a)$ so that $\alpha \sim EMG(\mu, \eta^2, a)$. This distributional assumption allows for a heavy right tail in the distribution for the uninsurable component of the log wage, which is heavier the smaller is the value for $a$. It is natural to attribute the heavy right tail in the log wage distribution to the uninsurable component for wages, since it seems unlikely that people are insured against the risk of becoming extremely rich.\footnote{For example, many individuals far in the right tail of the earnings distribution are entrepreneurs, and it is notoriously difficult to diversify entrepreneurial risk.} Note that given these assumptions on the distributions for $\alpha$ and $\varepsilon$, the distribution of the log wage ($\alpha + \varepsilon$) is itself EMG (so the level wage distribution is Pareto log-normal), since the sum of two independent normally distributed random variables ($\alpha_1$ and $\varepsilon$) is normal.

We will use the SCF data to estimate $a$. We do this by Maximum Likelihood, searching for the values of the three parameters in the EMG distribution that maximize the likelihood of drawing the observed distribution of log earnings. The resulting estimate
Figure 2: Fit of EMG Distribution

Figure 2 plots the empirical density against a normal distribution with the same mean and variance, and against the estimated EMG distribution. The density is plotted on a log scale, to magnify the tails. It is clear that the heavier right tail that the additional parameter in the EMG specification introduces delivers a substantially better fit.\(^\text{12}\)

One might be concerned that we are estimating \(a\) using an empirical distribution for earnings rather than wages (the SCF does not contain data on hours worked, so we cannot construct a wage measure in the usual way as earnings divided by hours worked). Fortunately, given our specification for the tax system and the fact that preferences in the model have the balanced growth property, we know that hours worked are independent of the uninsurable shock \(\alpha\), and thus the (uninsurable-risk-driven) right tails in the distributions for earnings and wages should be the same.\(^\text{13}\)

---

\(^\text{12}\)The empirical distribution for labor income in 2007 is constructed as follows. We define labor income as wage income plus two thirds of income from business, sole proprietorship, and farm. We then restrict our sample to households with at least one member aged 25-60, and with household labor income of at least $10,000 (for comparison, mean household labor income is $77,325).

\(^\text{13}\)Hours in our model do respond (positively) to insurable shocks. This implies that the variance of model earnings is larger than the variance of wages. But the distribution of the insurable component of log earnings remains normal, so the total distribution of earnings remains EMG. Similarly, one could introduce normally
Discretization

For computational purposes, as noted above, in solving the Mirrlees problem to characterize efficient allocations, the incentive constraints only apply to the uninsurable component of the wage $\alpha$, and the distribution for $\varepsilon$ appears only in the constant $\Omega$. Thus there is no need to approximate the distribution for $\varepsilon$, and we can assume these shocks are drawn from a continuous unbounded Normal distribution with mean $-v\varepsilon^*/2$ and variance $v\varepsilon^*$. 

[XX change this part] We take a discrete approximation to the continuous EMG distribution for $\alpha$ that we have discussed thus far. We start with the continuous EMG distribution, with parameters as follows: (i) exponential parameter $a = a^*$ (the SCF estimate), (ii) Normal variance $\eta^2 = v\varepsilon^* - 1/(a^*)^2$, so that the variance of $\alpha$ is $v\varepsilon^*$ (the HSV-based estimate), and (iii) Normal mean $\mu = \log\left(\frac{a-1}{a}\right) - \frac{\eta^2}{2}$, so that $E[\exp(\alpha)] = 1$. We then construct a grid of $N$ evenly-spaced values: $\alpha_1, \alpha_2, ..., \alpha_N$ with corresponding probabilities $\pi_1, \pi_2, ..., \pi_N$. We make the endpoints of the grid, $\alpha_1$ and $\alpha_N$ sufficiently widely spaced that only a tiny fraction of individuals lie outside these bounds in the true continuous distribution. In particular we set $\alpha_1$ such that $\exp(\alpha_1) = 0.12$. This is the ratio of the lower bound on labor income we imposed in constructing our sample ($10,000) to the average labor income in our sample. We set $\alpha_N$ such that $\exp(\alpha_N) = 74$, which corresponds to household labor income at the 99.99th percentile of the SCF labor income distribution relative to average income ($6.17 million). We read corresponding probabilities $\pi_i$ directly from the continuous EMG distribution, rescaling to ensure that $\sum_i \pi_i = 1$. For our baseline set of numerical results we set $N = 10,000$. In Section XX we report how the results change when we make increase or reduce $N$. 

The resulting model wage distribution $\exp(\alpha + \varepsilon)$ is plotted in Figure 3.

[TO GO LATER IN THE EXTENSION SECTION]

For the variance of observable shocks $v\kappa$, we use the estimate in Heathcote, Perri, and Violante (2010) who estimate the variance of cross sectional wage dispersion attributable to observables to be 0.108. We assume two-point equal-weight distribution for $\kappa$. This gives $\exp(\kappa_{\text{high}})/\exp(\kappa_{\text{low}}) = 1.93$. 

---

14 Assuming 2,000 household hours worked, this in turn corresponds to $5, which is less than the Federal minimum wage in 2007 ($5.85). Reducing the bound further would not materially affect any of our results since given the parameters for the EMG distribution, the probability of drawing $\alpha < \log(0.12)$ is vanishingly small.

15 Heathcote, Perri, and Violante (2010) report a total wage variance of 0.492 for 2004 (bottom right panel of Figure 5) using data from the Current Population Survey (CPS) and defining wages as annual earnings divided by annual hours. However, this variance pertains to the raw data and is thus inflated by measurement error in earnings and hours.
Figure 3: Wage Distribution
### Table 1: Benchmark

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>welfare</td>
</tr>
<tr>
<td>HSVUS</td>
<td>$\lambda : 0.836$, $\tau : 0.151$</td>
<td>$-$</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0 : -0.116$, $\tau_1 : 0.303$</td>
<td>$-0.58  $</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td>0.11</td>
</tr>
</tbody>
</table>

### 5.1 Social Welfare

We estimate the empirically motivated social welfare function. Specifically, we estimate the welfare parameter $\theta^*$ described in Section 4 using the U.S. tax data. The resulting weight is illustrated in Figure 4. It shows that the relative weights are increasing in wage, which might be intuitive given the low marginal weight in the U.S. economy.

### 6 Quantitative Analysis

**TO BE COMPLETED**

We focus on three cases:

1. HSV tax function: $T(y) = y - \lambda y^{1-\tau}$

2. Affine tax function: $T(y) = \tau_0 + \tau_1 y$

3. Mirrlees tax function (second best allocation)

**Baseline**

- Moving to affine tax system is welfare reducing
  
  $\Rightarrow$ Increasing marginal rates more important than lump-sum transfers

- Moving to fully optimal system generates only tiny gains (0.1%)  

- The optimal marginal tax rate is around 30%

- Almost no need for transfers

**Sensitivity**

What drives the results?
Figure 4: Social Welfare

![Graph showing the relationship between exp(\alpha) and Relative Pareto Weight.]

- X-axis: exp(\alpha)
- Y-axis: Relative Pareto Weight

The graph illustrates the increasing trend of the Relative Pareto Weight as \( exp(\alpha) \) increases.
Figure 5: HSV Tax Function
Figure 6: Affine Tax Function
Table 2: Utilitarian SWF

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>welfare</td>
</tr>
<tr>
<td>HSV\textsuperscript{US}</td>
<td>$\lambda : 0.836$ $\tau : 0.151$</td>
<td>–</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda : 0.821$ $\tau : 0.295$</td>
<td>1.38</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau_0 : -0.233$ $\tau_1 : 0.452$</td>
<td>0.45</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td>1.53</td>
</tr>
</tbody>
</table>

1. Empirically-motivated SWF $\rightarrow$ Utilitarian SWF: $w = 0$

2. Eliminate insurable shocks: $\tilde{v}_\alpha = v_\alpha + v_\varepsilon$ and $\tilde{v}_\varepsilon = 0$

3. Wage distribution has thin Log-Normal right tail: $\alpha \sim LN$

Sensitivity: Utilitarian SWF

- Utilitarian SWF $\Rightarrow$ stronger taste for redistribution
- Want higher tax rates and larger transfers
- Optimal HSV still better than optimal affine

Sensitivity: No Insurable Shocks

- No insurable shocks $\Rightarrow$ larger role for public redistribution
- Want higher tax rates and larger transfers
- Optimal HSV still better than optimal affine
- Utilitarian SWF + No insurable shocks $\Rightarrow$ Lump-sum transfers more important $\Rightarrow$ Optimal HSV worse than optimal affine

Sensitivity: Log-Normal Wage

- Log-normal distribution $\Rightarrow$ thin right tail
Table 3: No Insurable Shocks

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV&lt;sup&gt;US&lt;/sup&gt;</td>
<td>( \lambda : 0.836 \quad \tau : 0.151 )</td>
<td>(-\quad -)</td>
</tr>
<tr>
<td>HSV</td>
<td>( \lambda : 0.839 \quad \tau : 0.192 )</td>
<td>0.12</td>
</tr>
<tr>
<td>Affine</td>
<td>( \tau_0 : -0.156 \quad \tau_1 : 0.360 )</td>
<td>-0.21</td>
</tr>
<tr>
<td>Mirrlees</td>
<td>0.23</td>
<td>-2.11</td>
</tr>
</tbody>
</table>

Table 4: Log-normal Wage Distribution

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Tax Parameters</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV&lt;sup&gt;US&lt;/sup&gt;</td>
<td>( \lambda : 0.836 \quad \tau : 0.151 )</td>
<td>(-\quad -)</td>
</tr>
<tr>
<td>HSV</td>
<td>( \lambda : 0.826 \quad \tau : 0.070 )</td>
<td>0.27</td>
</tr>
<tr>
<td>Affine</td>
<td>( \tau_0 : -0.068 \quad \tau_1 : 0.250 )</td>
<td>0.34</td>
</tr>
<tr>
<td>Mirrlees</td>
<td>0.35</td>
<td>2.71</td>
</tr>
</tbody>
</table>

- Optimal HSV worse than optimal affine
- Optimal affine nearly efficient
- Want high marginal rates at the top when (i) few agents face those marginal rates, but (ii) can capture lots of revenue from higher-income households

**Extension**

Polynomial tax

Grid points
- Coarse grid \( \Rightarrow \) Mirrlees Planner can do much better
- Gives Mirrlees planner too much power if true distribution continuous

Type-Contingent Taxes
- Productivity partially reflects observable characteristics (e.g. education, age, gender)
- Some fraction of uninsurable shocks are observable: \( \alpha \rightarrow \alpha + \kappa \)
Figure 7: Saez Coefficient
Figure 8: Marginal Tax Rates
### Table 5: Polynomial Tax Functions

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Outcomes</th>
<th>welfare</th>
<th>Y</th>
<th>mar. tax</th>
<th>TR/Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline HSV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\lambda : 0.836$ $\tau : 0.151$</td>
<td>0.311</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polynomial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\tau_0$</td>
<td>$\tau_1$</td>
<td>$\tau_2$</td>
<td>$\tau_3$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>0.178</td>
<td>-</td>
<td>-</td>
<td>-1.31</td>
<td>5.61</td>
</tr>
<tr>
<td>-0.116</td>
<td>0.303</td>
<td>-</td>
<td>-</td>
<td>-0.58</td>
<td>0.41</td>
</tr>
<tr>
<td>-0.074</td>
<td>0.221</td>
<td>0.021</td>
<td>-</td>
<td>-0.10</td>
<td>0.61</td>
</tr>
<tr>
<td>-0.032</td>
<td>0.126</td>
<td>0.064</td>
<td>-0.003</td>
<td>0.05</td>
<td>0.79</td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td>0.11</td>
<td>0.82</td>
<td>0.287</td>
<td>0.003</td>
</tr>
<tr>
<td>First Best</td>
<td></td>
<td>8.80</td>
<td>15.79</td>
<td>0.000</td>
<td>0.208</td>
</tr>
</tbody>
</table>

### Table 6: Grid points

<table>
<thead>
<tr>
<th># of grid points</th>
<th>Affine</th>
<th>Mirrlees</th>
<th>First Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-0.58</td>
<td>3.82</td>
<td>8.80</td>
</tr>
<tr>
<td>100</td>
<td>-0.58</td>
<td>1.17</td>
<td>8.79</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.58</td>
<td>0.21</td>
<td>8.80</td>
</tr>
<tr>
<td>10,000</td>
<td>-0.58</td>
<td>0.11</td>
<td>8.80</td>
</tr>
<tr>
<td>100,000</td>
<td>-0.58</td>
<td>0.10</td>
<td>8.80</td>
</tr>
</tbody>
</table>
Figure 9: Marginal Tax Rates with different grid points
### Table 7: Type-contingent Taxes

<table>
<thead>
<tr>
<th>Tax System</th>
<th>Outcomes</th>
<th>welfare</th>
<th>Y</th>
<th>mar. tax</th>
<th>TR/Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV&lt;sub&gt;US&lt;/sub&gt;</td>
<td>$\lambda: 0.827$</td>
<td>$\tau: 0.151$</td>
<td>-</td>
<td>-</td>
<td>0.311</td>
</tr>
<tr>
<td>HSV</td>
<td>$\lambda^L: 1.040$</td>
<td>$\tau^L: 0.286$</td>
<td>1.34</td>
<td>4.64</td>
<td>0.212</td>
</tr>
<tr>
<td>$\lambda^H: 0.647$</td>
<td>$\tau^H: -0.054$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau^L_0: 1.040$</td>
<td>$\tau^L_1: 0.286$</td>
<td>1.39</td>
<td>5.20</td>
<td>0.199</td>
</tr>
<tr>
<td>Affine</td>
<td>$\tau^H_0: 0.647$</td>
<td>$\tau^H_1: -0.054$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mirrlees</td>
<td></td>
<td></td>
<td>1.46</td>
<td>5.20</td>
<td>0.200</td>
</tr>
</tbody>
</table>

- Heathcote, Perri, and Violante (2010) estimate variance of cross-sectional wage dispersion attributable to observables, $v_\kappa = 0.108$

- Planner should condition taxes on observables: $T(y; \kappa)$

- Consider two-point distribution for $\kappa$ (college vs high school)

- Significant welfare gains relative to non-contingent tax

- Conditioning on observables $\Rightarrow$ marginal tax rates of 20%

### 7 Conclusions

To be completed
Appendix: Proof of Proposition 3

In this appendix, we provide the proof of Proposition 3.

Given the HSV tax function (21), the decision rules become a function of $\tau$:

\[
c(\alpha; \lambda, \tau) = \lambda (1 - \tau)^{1+\sigma} \exp \left[ \frac{(1 - \tau)\alpha}{\sigma} \exp \left( \frac{1 - \tau v_\varepsilon}{2} \right) \right],
\]

\[
h(\varepsilon; \tau) = (1 - \tau)^{1+\sigma} \exp \left( \frac{-1 \cdot v_\varepsilon}{\sigma^2} \right) \exp \left( \frac{\varepsilon}{\sigma} \right).
\]

Plugging these to the resource constraint (2), we get

\[
\lambda(\tau) = \frac{(1 - \tau)^{1+\sigma} \exp \left( \frac{1 \cdot v_\varepsilon}{2} \right) - G}{(1 - \tau)^{1+\sigma} \exp \left( \frac{1 - \tau v_\varepsilon}{2} \right) \int \exp [((1 - \tau)\alpha)] dF_\alpha}.
\]

We can substitute these expressions into the planner’s objective function in order to get an unconstrained optimization problem with one choice variable, $\tau$. Therefore, differentiating the objective function with respect to $\tau$ gives a first-order condition that maps $\theta$ and $G$ into the optimal choice for $\tau$. Specifically, the planner’s objective function is

\[
\int W(\alpha) \left( \log (c(\alpha; \tau)) - \int \frac{h(\varepsilon; \tau)^{1+\sigma} \exp \left( \frac{1 \cdot v_\varepsilon}{\sigma^2} \right) \exp \left( \frac{\varepsilon}{\sigma} \right) dF_\varepsilon(\varepsilon)}{1+\sigma} dF_\alpha(\alpha) \right)
\]

and the government expenditure is given by

\[
G = g \int \int \exp(\alpha + \varepsilon) h(\alpha) dF_\alpha dF_\varepsilon.
\]

Substituting the above expressions gives an unconstrained optimization problem

\[
\max \tau \quad (1 - \tau) \int W(\alpha) dF_\alpha - \log(\int \exp [(1 - \tau)\alpha] dF_\alpha) + \log \left( (1 - \tau)^{1+\sigma} \exp \left( \frac{1 \cdot v_\varepsilon}{\sigma^2} \right) \exp \left( \frac{\varepsilon}{\sigma} \right) - G \right) - \frac{1 - \tau}{1+\sigma}
\]

and

\[
G = g(1 - \tau)^{1+\sigma} \exp \left( \frac{1 \cdot v_\varepsilon}{\sigma^2} \right) \exp \left( \frac{\varepsilon}{\sigma} \right).
\]

Note that the level of the government expenditure $G$ is fixed when the planner is solving the problem and hence it is not a function of $\tau$.

Since the moment generating function for the EMG distribution, $EMG(\mu, \eta^2, a)$, for $t \in \mathbb{R}$ is given by

\[
\int_a \exp (at) dF_\alpha = \frac{a}{a-t} \exp \left[ \mu t + \frac{\eta^2 t^2}{2} \right],
\]

given the social welfare function (20), the unconstrained optimization problem becomes
\[
\max_{\tau} \frac{\alpha}{\alpha + \theta} \exp\left(\frac{1 - \tau}{\alpha + \theta}\right) \int \alpha \exp(-\theta \alpha) dF_\alpha - \log \left(\frac{\alpha}{\alpha + 1 - \tau}\right) - \mu(1 - \tau) - \frac{\eta^2(1 - \tau)^2}{2} + \log \left[(1 - \tau)^{\frac{1}{\alpha + \theta}} \exp \left(\frac{1 + \eta^2}{2}\right) - G\right] - \frac{1 - \tau}{1 + \sigma}.
\]

(27)

Assume this problem to be well-defined, i.e., \(\int \alpha \exp(-\theta \alpha) dF_\alpha < \infty\). We want to further simplify this term.

Define

\[V(\alpha, \theta) \equiv \exp(-\theta \alpha) f_\alpha(\alpha),\]

where \(f_\alpha\) is the derivative of \(F_\alpha\), and we have

\[\frac{\partial V(\alpha, \theta)}{\partial \theta} = -\alpha \exp(-\theta \alpha) f_\alpha(\alpha)\].

**Lemma 4** Assume the support of \(\theta\) is compact, \([\bar{\theta}, \check{\theta}]\). Then the integral and the derivative of \(V\) are interchangeable, i.e.,

\[\int \frac{\partial}{\partial \theta} V(\alpha, \theta) d\alpha = \int \frac{\partial}{\partial \theta} V(\alpha, \theta) d\alpha.\]

**Proof.** It suffices to show that (i) \(V : \mathbb{R} \times [\bar{\theta}, \check{\theta}] \to \mathbb{R}\) is continuous and \(\frac{\partial V}{\partial \theta}\) is well-defined and continuous in \(\mathbb{R} \times [\bar{\theta}, \check{\theta}]\), (ii) \(\int V(\alpha, \theta) d\alpha\) is uniformly convergent, and (iii) \(\int \frac{\partial}{\partial \theta} V(\alpha, \theta) d\alpha\) is uniformly convergent. (i) is obvious since \(f_\alpha\) is continuous.

To prove (ii), we rely on Weierstrass M-test for uniform convergence. That is, if there exists \(\hat{V} : \mathbb{R} \to \mathbb{R}\) such that \(\hat{V}(\alpha) \geq |V(\alpha, \theta)|\) for all \(\theta\) and \(\hat{V}\) has an improper integral on \(\mathbb{R}\), then \(\int V(\alpha, \theta) d\alpha\) converges uniformly. Now define \(\hat{V}(\alpha) \equiv \sup_{\theta \in [\bar{\theta}, \check{\theta}]} |V(\alpha, \theta)|\). Then \(\hat{V}(\alpha) \geq |V(\alpha, \theta)|\) by construction. Also \(\hat{V}\) has an improper integral on \(\mathbb{R}\) because

\[\int_{-\infty}^{\infty} \hat{V}(\alpha) d\alpha = \int_{-\infty}^{0} V(\alpha, \check{\theta}) d\alpha + \int_{0}^{\infty} V(\alpha, \check{\theta}) d\alpha \leq \int_{-\infty}^{\infty} V(\alpha, \check{\theta}) d\alpha + \int_{-\infty}^{\infty} V(\alpha, \check{\theta}) d\alpha = \frac{a}{\alpha + \check{\theta}} \exp \left[-\mu \check{\theta} + \frac{\eta^2 \check{\theta}^2}{2}\right] + \frac{a}{\alpha + \check{\theta}} \exp \left[-\mu \check{\theta} + \frac{\eta^2 \check{\theta}^2}{2}\right] < \infty,
\]

where the first inequality comes from \(V(\alpha, \theta) \geq 0\) for any \(\alpha\) and \(\theta \in [\bar{\theta}, \check{\theta}]\). Thus \(\int V(\alpha, \theta) d\alpha\) is uniformly convergent.

We apply a similar logic to prove (iii) and find \(\tilde{V} : \mathbb{R} \to \mathbb{R}\) such that \(\tilde{V}(\alpha) \geq \left|\frac{\partial V(\alpha, \theta)}{\partial \theta}\right|\) for all \(\theta\) and \(\tilde{V}\) has an improper integral on \(\mathbb{R}\). Specifically, define \(\tilde{V}(\alpha) \equiv \sup_{\theta \in [\bar{\theta}, \check{\theta}]} \left|\frac{\partial V(\alpha, \theta)}{\partial \theta}\right|\). Then \(\tilde{V}(\alpha) \geq \left|\frac{\partial V(\alpha, \theta)}{\partial \theta}\right|\) by construction and \(\tilde{V}\) has an improper integral.
on $\mathbb{R}$, because the original Ramsey problem is assumed to be well-defined and hence $\int \alpha \exp(-\theta \alpha) dF_\alpha < \infty$ for any $\theta \in [\bar{\theta}, \hat{\theta}]$.

Applying this lemma, we get

$$\int \alpha \exp(-\theta \alpha) dF_\alpha = -\frac{\partial}{\partial \theta} \int \exp(-\theta \alpha) dF_\alpha = \frac{a}{a+\theta} \exp \left[ -\mu \theta + \frac{\eta^2 \theta^2}{2} \right] \left( \frac{1}{a+\theta} + \mu - \eta^2 \theta \right).$$

Substituting this expression into (27), the unconstrained optimization problem becomes

$$\max_{\tau} \; (1 - \tau) \left( \frac{1}{a+\theta} - \eta^2 \theta - \frac{1}{1+\sigma} \right) + \log (a - 1 + \tau) - \frac{\eta^2 (1-\tau)^2}{2} + \log \left[ (1 - \tau) \frac{1}{1+\sigma} \exp \left( \frac{1}{\sigma} \frac{\psi}{2} \right) - G \right].$$

The first order condition with respect to $\tau$ is

$$-\left( \frac{1}{a+\theta} - \eta^2 \theta - \frac{1}{1+\sigma} \right) + \frac{1}{a - 1 + \tau} + \eta^2 (1 - \tau) + \frac{\partial}{\partial \tau} \left\{ (1 - \tau) \frac{1}{1+\sigma} \exp \left( \frac{1}{\sigma} \frac{\psi}{2} \right) \right\} = 0.$$

Substituting (26) into this,

$$-\left( \frac{1}{a+\theta} - \eta^2 \theta - \frac{1}{1+\sigma} \right) + \frac{1}{a - 1 + \tau} + \eta^2 (1 - \tau) - \frac{1}{(1-g)(1+\sigma)(1-\tau)} = 0.$$

Therefore, the planner’s weight $\theta^*$ must solve

$$\eta^2 \theta^* - \frac{1}{a+\theta^*} = -\eta^2 (1 - \tau) - \frac{1}{a - 1 + \tau} + \frac{1}{1+\sigma} \left[ \frac{1}{(1-g)(1-\tau)} - 1 \right].$$

Q.E.D.

References


