On the Optimality of the Friedman Rule with Heterogeneous Agents and Non-Linear Income Taxation*

Carlos da Costa, Getulio Vargas Foundation†

Iván Werning, MIT, NBER and UTDT‡§

January 31, 2005

Abstract

This paper studies the optimality of the Friedman rule in a dynamic economy populated by heterogeneous agents subject to distortionary non-linear income taxation — an adaptation of Mirrlees’ framework. The conditions for the optimality of the Friedman rule are contrasted to those obtained in Ramsey frameworks — representative agent subject to proportional taxation. The conditions are generally met in our environment for two popular specifications of money demand. We examine the changes in welfare from inflation and show that the area under the demand curve generally provides a lower bound on the true change in social welfare.

JEL codes: D82, E40, E50, E60, H21.

Introduction

Friedman (1969) clarified the point that positive nominal interest rates represent a distorting tax on real money balances. In a first-best situation such distortions are

*We thank Fernando Alvarez, Luis Braido, Francisco Buera, Hugo Hopenhayn, Casey Mulligan and Pedro Teles for helpful comments and discussions.
†e-mail: cdacosta@fgv.br
‡e-mail: iwerning@mit.edu
§Corresponding author. Address: Department of Economics MIT, E52-251A, 50 Memorial Drive, Cambridge, MA 02142-1347.
unnecessary so the nominal interest rate should be set to zero. This prescription, known as the Friedman rule, is a cornerstone in monetary economics.\(^1\) Phelps (1973) countered that the second-best world we live in requires tolerating distortions due to government taxation, and that positive taxes are set on most goods. Why should money be treated any differently? What is so special about money? He concluded that money should generally be taxed, that nominal interest rates should be positive.

More recently, many authors have explored the conditions under which the Friedman rule remains optimal in a Ramsey (1927) optimal tax problem.\(^2\) Restrictions on preferences and transaction technologies—which in some cases amount to assuming that money is special—do imply the optimality of Friedman’s rule even in such a second-best world. These contributions study the optimal inflation tax assuming proportional taxation of labor income and a representative agent economy. It is well known in the public finance literature that tax prescriptions are sensitive to both assumptions.

In addition to ignoring distributional effects and arbitrarily restricting the available tax instruments, the Ramsey approach suffers from an inherent logical contradiction. In this approach lump-sum taxation would help achieve the first-best allocation, but it is ruled out. One common justification of this assumption is that it is meant to capture the undesirability of systems that are ‘too regressive’ without explicitly modeling such concerns. However, solutions to these problems yield prescriptions that attempt to emulate, however imperfectly, the missing desired lump-sum tax. The contradiction is clear, lump-sum taxes are ruled out as undesirable only to derive optimal tax prescriptions that imitate lump-sum tax. How can we judge whether such a tax system is preferable to the one generated directly by a lump-sum tax?

This paper reexamines the optimal inflation tax in a model that explicitly incorporates the concern for inequality. Following Mirrlees (1971, 1976), the conflict between redistribution and efficiency is captured by confronting a utilitarian planner with an informational asymmetry. One contribution of this paper is thus to combine a Mirrleesian optimal taxation framework with a standard dynamic-equilibrium model.\(^4\)

---

\(^1\) Another, completely unrelated, Friedman rule is the \(k\%\)-growth rule for monetary aggregates, advocated in Friedman’s 1968 A.E.A. presidential address.


\(^3\) A separate line of research studies the frictions that give rise to the demand for money, usually abstracting from public finance considerations. Some of these models have interesting implications for the optimality of the Friedman rule. See Lagos and Wright (2004) and the references therein.

\(^4\) Several papers have since fruitfully followed this approach, see especially Golosov, Kocherlakota and Tyvinski (2003) and Albanesi and Sleet (2003).
Our model’s backbone is the standard dynamic-equilibrium model commonly used to examine the optimality of the Friedman rule in a Ramsey framework. We depart by introducing individual heterogeneity in labor productivity and by not restricting the set of tax instruments arbitrarily. Instead, the planner is constrained by the primitive restriction that individual productivities are private information.

The allocations we consider can be implemented through non-linear taxation of income and a monetary policy determining the nominal interest rate. Individuals take as given the tax policy and make free choices regarding their labor, consumption and money balances. In particular, individuals saving or borrowing are not monitored nor taxed.\(^5\) Thus, the important difference between this paper and work based on the Ramsey paradigm is non-linear income taxation and individual heterogeneity.

We derive the optimality condition for the nominal interest rate within a money-in-the-utility-function framework. We then explore the conditions for the optimality of the Friedman rule. The condition turns out to be quite simple and intuitive. Money should not be taxed if individuals demand for money balances increases with work effort, holding expenditures constant.

To interpret this result, it is important to understand the auxiliary role played by the taxation of money.\(^6\) When a non-linear income tax is allowed, taxing money is only useful if it relaxes the incentive constraints that ensure individuals do not underreport their productivity. This turns out to be the case if an individual who considers deviating from the truth-telling equilibrium by underreporting productivity, demands higher money balances than those individuals they claim to be. The demand for money could differ because both agents would produce the same output, but the agent underreporting productivity would require less work effort. This explains why the critical feature is how the demand for money depends on work effort, holding expenditures constant.

We examine this property for two important specifications of money demand and find that it is met under very weak assumptions. In particular, for the shopping-time model (McCallum and Goodfriend, 1987 and Lucas, 2000) agents with lower work effort have more time to spare on shopping activities that substitute for money balances, so they demand less cash. As a result, if it weren’t for the non-negativity constraint on nominal interest rates, subsidizing money would be optimal, and the Friedman rule obtains as a corner solution.

For the cash-credit model introduced by Lucas and Stokey (1983) we find that

---

\(^5\) This situation is known in the contract theory literature as a case of ‘hidden-savings’.  
\(^6\) see the seminal contributions by Atkinson and Stiglitz (1976) and Mirrlees (1976).
the Friedman rule holds exactly when preferences are weakly separable between consumption goods and leisure since, in this case, money balances are unaffected by work effort. If, instead, leisure shifts consumption away from money-intensive goods then the Friedman rule holds as a corner solution.

We also examine the welfare costs of deviating from a zero-nominal-interest rate when the Friedman rule is optimal. We first show that in a first-best situation, without private information regarding productivity, the area under the economy’s compensated-demand curve represents aggregate welfare costs exactly. This explains Lucas’s (2000) finding that the area under the steady-state demand curve accurately approximates welfare costs. Income effects are small enough that the steady-state demand curve is close to the compensated-demand curve.

We then show that in our model, with private information, welfare costs of inflation are larger than in the first-best situation, without private information. It follows that the area under the compensated-demand curve is a lower bound for the true welfare costs. Indeed, the costs of positive nominal interest rate may be of first-order magnitude around a zero-nominal-interest rate, which is never possible in a first-best situation, or for an estimate based on the area under a demand curve. Finally, we perform a quantitative evaluation of welfare costs and compare them with Lucas’s calculations. These findings suggest that the welfare costs may be understated by as much as 50% in an analysis that ignores distortive taxation.

The next section introduces the model and the planner’s problem. Section 2 derives the optimality condition for the nominal interest rate. Section 3 verifies this condition for two specifications of money demand: the shopping-time and cash-credit models. Section 4 examines the welfare costs of inflation. Section 5 presents our conclusions. An appendix collects some proofs.

1 Model

Preferences and Technology. The economy is populated by a continuum measure one of individuals with identical preferences represented by

$$\sum_{t=0}^{\infty} \beta^t u(c_t, m_t, l_t),$$

(1)

where $c_t$, $m_t$ and $l_t$ represent consumption, real money balances and non-work time, respectively. We assume $\beta < 1$ and that $u$ is non-decreasing, concave and continuously differentiable. Real balances are $m_t \equiv M_t/p_t$ where $M_t$ is nominal money, and $p_t$ is
the price level.

Agents are indexed by their labor productivity, which is distributed in the population with distribution $F(w)$ for $w \in [\underline{w}, \bar{w}]$. The economy-wide resource constraint is

$$\int (Y_t(w) - c_t(w)) dF(w) \geq G \text{ for all } t = 0, 1, 2, ...$$

(2)

where $Y_t(w) = w(1 - l_t(w))$ represents the output produced by each agent with productivity $w$.

We adopt a utilitarian welfare function

$$\int \sum_{t=0}^{\infty} \beta^t u(c_t(w), m_t(w), l_t(w)) dF(w).$$

**Budget constraints.** Given a sequence of after-tax real income $\{y_t\}_{t=0}^{\infty}$, individuals face the sequence of nominal budget constraints,

$$p_t c_t + M_t + B_t \leq p_t y_t + M_{t-1} + (1 + i_t) B_{t-1} \text{ for all } t = 0, 1, 2, ...$$

where $B_t$ represents nominal bond holdings, and $i_t$ is the nominal interest rate. We also impose a standard No-Ponzi constraint. To simplify we initialize nominal wealth to zero, $M_{-1} + B_{-1} = 0$. This assumption makes the initial price level irrelevant, allowing us to focus on the determination of nominal interest rates.

The sequential budget constraints above and the No-Ponzi are equivalent to the present-value budget constraint,

$$\sum_{t=0}^{\infty} \psi_t(c_t + r_t m_t - y_t) \leq 0,$$

(3)

where

$$\psi_t \equiv \frac{p_t}{p_0} \prod_{s=0}^{t-1} \frac{1}{1 + i_s}$$

denotes the real price of consumption in period $t$, and $r_t \equiv i_t/(1 + i_t)$ represents the private cost of holding real money balances relative to consumption.

**Information.** Individual productivities are private information. Moreover, we assume that individual choices for consumption, leisure, money balances, and bond holdings are not observed by the planner: only output is observable by the planner.

---

7In our setup the free good aspect of money is of no importance for our results. In contrast, Correia and Teles (1996, 1999) have shown that in a Ramsey framework the free-good aspect of money is important for the optimality of the Friedman rule.
The assumption regarding the non-observability of consumption and savings is known in the contract-theory literature as a situation with hidden-savings. Without this assumption the cost of money would not necessarily be associated with the nominal interest rate. Allowing hidden-savings usually complicates the analysis of the constrained-efficient allocations. In our case the analysis remains tractable due to the simplifying assumption that individual productivity differences are permanent.

**Equilibrium.** We define a *policy* to be a sequence of income and output as a function of productivity, \( \{y_t(w), Y_t(w)\}_{t=0}^{\infty} \), and a sequence for the tax on money \( \{r_t\}_{t=0}^{\infty} \).

Given a policy \( \{y_t(w), Y_t(w), r_t\}_{t=0}^{\infty} \) a *competitive equilibrium* consists of sequences of real prices \( \{\psi_t\}_{t=0}^{\infty} \), real quantities \( \{c_t(w), m_t(w), l_t(w)\}_{t=0}^{\infty} \) and nominal money and price levels \( \{M_t, p_t\}_{t=0}^{\infty} \) such that:

(i) individuals optimize: \( \{c_t(w), m_t(w), l_t(w)\}_{t=0}^{\infty} \) maximizes (1) subject to (3) taking as given \( \{\psi_t, r_t\}_{t=0}^{\infty} \)

(ii) markets clear: (a) the resource constraint (2) holds; (b) money supply equals money demand: \( m_t = M_t/p_t \)

(iii) consistency of prices: \( \psi_t \equiv \prod_{s=0}^{t} (1 + i_s)^{-1} p_t/p_0 \) and \( r_t = i_t/(1 + i_t) \).

**Stationary Policies and Allocations.** Our focus is on stationary allocations for which \( y_t(w) = y(w), Y_t(w), r_t = r \). A stationary policy \( (y(w), Y(w), r) \) generates a stationary competitive equilibrium, with real prices \( \psi_t = \beta^t \), real allocation \( c_t(w) = c(y(w), Y(w), r, w) \), \( m_t(w) = m(y(w), Y(w), r, w) \) and \( l_t(w) = 1 - Y(w)/r \), and nominal variables that grow at a constant rate, i.e. \( M_t = ((1 + r)/\beta)^t M_0 \) with \( M_0 > 0 \) indeterminate.

In this environment, with permanent productivity differences, stationary policies and allocations are natural. Indeed, non-stationary allocations can only serve to mimic randomization schemes. Thus, when randomization is allowed, non-stationary allocations are without loss in generality. Appendix B makes explicit one connection between non-stationary allocations and randomization schemes. It is customary, except for technical considerations, to ignore randomization schemes. Given their abstract role we ignore both randomization schemes and non-stationary allocations.

The main advantage of stationary allocations is that it allows us to characterize the individual’s problem by the static subproblem

\[
V(y, Y, r, w) \equiv \max_{c,m} u(c, m, 1 - Y/w) \\
\text{s.t. } c + rm = y
\]
where \( Y \) and \( y \) represent output and and after-tax income, respectively. Let the uncompensated demands be \( c(y, Y, r, w) \) and \( m(y, Y, r, w) \). Similarly, define the expenditure function as \( e(v, Y, r, w) \equiv \min_{c,m}(c + rm) \) subject to \( u(c, m, 1 - Y/w) = v \), with corresponding compensated demands \( c^c(v, Y, r, w) \) and \( m^c(v, Y, r, w) \).

We assume the standard single-crossing condition that the marginal rate of substitution
\[
\frac{-V_Y(y, Y, r, w)}{V_y(y, Y, r, w)}
\]
is decreasing in \( w \) for all \((y, Y, r, w) \in \mathbb{R}^4_+\). This assumption ensures that more productive agents produce more. It is implied by weak conditions on the utility function \( u \), such as joint normality of consumption and money.

Given a policy \((y(w), Y(w), r)\) let \( v(w) \) denote the utility attained by an agent with productivity \( w \). Individuals choose their report to maximize utility,
\[
v(w) = \max_{w'} V(y(w'), Y(w'), r, w).
\]
Agents truthfully reveal their productivity only if \( w \in \arg\max_w V(y(w'), Y(w'), r, w) \), or equivalently, \( v(w) = V(y(w), Y(w), r, w) \). An allocation \((y(w), Y(w), r)\) that satisfies these constraints is termed incentive compatible.

**Planner Problem.** The planner selects an incentive compatible policy \((y(w), Y(w), r)\) so that the resulting competitive equilibrium allocation maximizes the utilitarian welfare function.

**2 Optimal Nominal Interest Rate**

Our first objective is to rewrite the planner’s problem in a tractable way. We begin with the incentive constraints. The envelope condition for \( v(w) \) (Milgrom and Segal, 2002) and truth-telling imply the local incentive constraint:
\[
v(w) = V(y(w), Y(w), r, w) + \int_{\tilde{w}}^{w} V_w(y(\tilde{w}), Y(\tilde{w}), r, \tilde{w})d\tilde{w}.
\]
(4)

Single-crossing also requires \( Y \) to be non-decreasing. Thus, incentive compatibility implies (4) and the monotonicity of \( Y \). The converse is also true, that is (4) and monotonicity of \( Y \) are sufficient to ensure incentive compatibility (see Fudenberg and Tirole, 1991).
Substituting \( y(w) = e(v(w), Y(w), r, w) \), the planner’s problem is,

\[
\max_{v, Y, r \geq 0} \int v(w) dF(w)
\]

subject to the resource constraint,

\[
\int (Y(w) - c^e(v(w), Y(w), r, w)) dF(w) \geq G,
\]

the local incentive constraint,

\[
v(w) = V(e(v(\tilde{w}), Y(\tilde{w}), r, \tilde{w}), Y(w), r, w) + \int_{\tilde{w}}^{w} V_w(e(v(\tilde{w}), Y(\tilde{w}), r, \tilde{w}), Y(\tilde{w}), r, \tilde{w}) d\tilde{w}
\]

and the monotonicity condition, that \( Y \) be non-decreasing. The analysis now follows Mirrlees (1976) closely.\(^8\)

The associated Lagrangian is

\[
L \equiv \int (v + \lambda (Y - c^e)) dF + \int (\mu - v + \int w V_w d\tilde{w}) d\mu(w)
\]

where \( \lambda > 0 \) is the scalar Lagrangian multiplier on the resource constraint and \( \mu(w) \) is the Lagrangian multiplier for constraint (4), normalized so that \( \mu(\tilde{w}) = 0 \). Integrating the second term by parts yields

\[
L = \int (v + \lambda (Y - c^e)) dF - \int \mu V_w dw + \int (\mu - v) d\mu.
\]

The first-order necessary condition for an interior optimum is then

\[
\frac{\partial L}{\partial r} = -\int \mu (V_{wr} + V_{wy} e_r) dw - \lambda \int c_r^e dF \leq 0,
\]

with equality if \( r > 0 \). Differentiating Roy’s identity, \( V_r + V_y m \equiv 0 \), with respect to \( w \), and using the envelope condition \( e_r = m^c \), gives \( V_{wr} + V_{wy} e_r = -V_y m_w \). Homogeneity and symmetry of conditional demands implies that \( c_r^e = -r m_r^e \). Substituting we obtain

\[
r \int m_r^e dF \leq -\frac{1}{\lambda} \int V_y m_w dw, \tag{5}
\]

\(^8\)If the monotonicity condition is binding over some interval \([w_1, w_2]\) then agents in this interval receive the same bundle \((y(w), Y(w)) = (y^*, Y^*)\) for \( w \in [w_1, w_2] \). This situation is known as bunching. It is common in the literature to ignore bunching and not impose the monotonicity constraint on the planner’s problem. In the analysis that follows we take full account of this constraint.
with equality if \( r > 0 \). Note that \( V_y > 0 \) and \( \lambda > 0 \). It is critical to understand the sign of \( \mu \) and \( m_w \).

The sign of the multiplier \( \mu(w) \) is determined by the direction the local incentive constraint binds at \( w \) — whether agents are tempted to under- or over-report. In particular, incentive constraints bind downward, high productivity agents tempted to underreport, when \( \mu(w) \geq 0 \). This represents the situation where the planner would like to redistribute at the margin from higher- to lower-productivity agents but is limited by incentives. Indeed, for \( r = 0 \), the marginal tax rate on income shares the sign with \( \mu \), indicating in a different way that \( \mu(w) > 0 \) represents the case where redistribution is desired.

**Proposition 1** If leisure is a normal good and \( r = 0 \) then incentive constraints bind downwards and \( \mu(w) \geq 0 \) for all \( w \).

**Proof.** Since \( r = 0 \) we can rewrite the planner’s problem as a standard two good problem in terms of consumption and labor only for the utility function \( U(c,l) = \max_m u(c,m,l) \). This is the original problem studied by Mirrlees (1971). For this model, Brunner (1993) and Ebert (1992) have show that normality of leisure is sufficient for incentive constraints to bind downwards and for \( \mu(w) \) to be positive for all \( w \).

This implies that in a neighborhood of \( r = 0 \) the sign of the right hand side of (5) is determined by the sign of \( m_w \). Combining this result with the optimality condition (5), it follows that, increasing the nominal interest rate from zero reduces welfare if and only if \( m_w \leq 0 \). In this case \( r = 0 \) satisfies the first order condition for \( r \) given by (5) above.

Although when \( r > 0 \) no condition on primitives is known that ensures that incentive constraints bind downward, for the rest of the paper we limit attention to the cases of economic interest where redistribution does takes place from high- to low-productivity individuals so that \( \mu(w) \geq 0 \) for all \( w \).

The following proposition then follows immediately from the optimality condition (5) and \( m^c_r \leq 0 \).

**Proposition 2** If \( m_w(y(w),Y(w),r,w) \leq 0 \) for all \( w \) the Friedman rule is optimal, i.e. \( r^* = 0 \). Conversely, if \( m_w(y(w),Y(w),r,w) > 0 \) for all \( w \), then taxing money is optimal, i.e. \( r^* > 0 \).

---

9Numerical simulations show that even for \( r > 0 \) this is indeed the relevant case, see Section 4.2 below.
This result has a strong intuition that can be explained as follows. Consider the case where \( m_w > 0 \) so that taxing money is optimal. In this case an agent considering underreporting productivity would demand more money than the agent he claims to be. Thus, a tax on money relaxes the incentive constraint by making such deviations more costly. When \( m_w \leq 0 \) exactly the opposite is true, taxing money makes incentive constraints worse.

An agent under-reporting his true productivity has more non-work time than the agents he impersonates since he produces the same output, \( Y \), with higher productivity, \( w \). This additional time is what determines the sign of \( m_w \) and the optimality of the Friedman rule. Note that the condition that \( m_w \leq 0 \) does not imply that equilibrium money holdings are lower for more productive agents. Typically, the composed function \( m(y(w), Y(w), r, w) \) representing equilibrium money demand as a function of \( w \) will be increasing even if \( m_w \leq 0 \).

When preferences are such that consumption and money are weakly separable from leisure it follows that \( m_w(y, Y, r, w) = 0 \) and the Friedman rule it optimal. Indeed, in this case the direction incentive constraints binds is irrelevant. This corresponds to the case considered in Atkinson and Stiglitz’s (1976) celebrated uniform-taxation result. Uniform taxation of consumption and money requires money not to be taxed.

3 Two Models of Money Demand

This section examines the optimality condition derived above for two important models that can be seen as specializations of the money-in-the-utility-function framework.

3.1 Shopping-Time Model

In the shopping-time model (McCallum and Goodfriend, 1987, Lucas, 2000) a primitive utility function, \( U \), is defined over consumption and leisure only. Consumption requires shopping time, and money serves to economize on this time. Let \( s(c, m) \) denote the shopping time required to obtain consumption \( c \) with money balances \( m \). This maps into the money-in-the-utility function by setting:

\[
    u(c, m, l) \equiv U(c, l - s(c, m)).
\]

We now show that for this model \( m_w \) is negative. This occurs because an increase in non-work time lowers the need for time-saving money balances. Thus, the Friedman rule holds as a corner solution.
Proposition 3 In the shopping-time model, if consumption and money are normal goods then $m_w(y, Y, r, w) \leq 0$ and the Friedman rule is optimal.

The proof is contained in the appendix. Note that the result requires no assumptions on the shopping technology and only very natural normality assumptions. In contrast, in a representative agent Ramsey framework, Correia and Teles (1996) show that the optimality of the Friedman rule requires the shopping time function to be homogenous.

3.2 Cash-Credit Model

In the cash-credit model (Lucas and Stokey, 1983) utility, $\tilde{U}$, is defined over two consumption goods, $c^1$ and $c^2$, and leisure, $l$. The credit-good, $c^1$, can be purchased with credit while the cash-good, $c^2$, requires money up-front so that $c^2 \leq m$.

This model can be mapped into the money-in-the-utility-function framework. Let $c \equiv c^1 + c^2$ and define

$$u(c, m, l) \equiv \tilde{U}(c - m, m, l).$$

It follows that if preferences over consumption goods $c^1$ and $c^2$ are weakly separable from $l$, then the demand for money balances does not depend on the leisure and $m_w = 0$. Thus, the Friedman rule is optimal without any qualification regarding the direction the incentive constraints bind.

Proposition 4 In the cash-credit model if preferences over consumption goods are weakly separable so that

$$\tilde{U}(c^1, c^2, l) = u(h(c^1, c^2), l)$$

for some functions $u$ and $h$, then $m_w(y, Y, r, w) = 0$ and the Friedman rule is optimal.

If preferences are not separable but instead leisure is complement with the credit good, $c^1$, then it is optimal to subsidize money. In this case the Friedman rule would obtain as a strict corner solution. Note that in a representative-agent Ramsey framework weak separability does not ensure the optimality of the Friedman rule.

4 Welfare Costs of Inflation

In this section we examine the welfare costs of inflation. First, we define an aggregate welfare-cost measure and relate it to the area under the demand curve. We then perform a quantitative exercise and contrast the results to Lucas’ (2000) calculations.
In our model with heterogenous agents, the planner’s welfare function provides a natural aggregate-welfare measure. We reexpress changes in the welfare function in terms of consumption by introducing an additional endowment \( \omega \) of goods owned by the planner,

\[
G + \int c^e(y(w), Y(w), r, w)dF(w) \leq \int Y(w)dF(w) + \omega,
\]

and using \( \omega \) as a compensating-differential for changes in \( r \). Consider the problem of maximizing over \( y \) and \( Y \) given \( r \) and \( \omega \) and denote by \( W(r, \omega) \) the maximized utilitarian welfare function. We define \( \omega(r) \) to be the level of the endowment that allows the planner to obtain the same welfare level as \( r = 0 \) and \( \omega = 0 \).

\[
W(r, \omega(r)) = W(0, 0).
\]

Differentiating the planner’s problem and using the envelope condition,

\[
\frac{\partial W}{\partial r} = \frac{\partial L}{\partial r} = \lambda r \int \frac{\partial m_c}{\partial r} dF + \int V_y \mu m_w dw
\]

where \( \lambda \) is the multiplier on the resource constraint. Similarly,

\[
\frac{\partial W}{\partial \omega} = \lambda.
\]

A compensated change requires,

\[
\frac{\partial W}{\partial r} + \frac{\partial W}{\partial \omega} \frac{d\omega}{dr} = 0
\]

implying

\[
\frac{d\omega}{dr} = -r \int \frac{\partial m_c}{\partial r} dF + \left( -\frac{1}{\lambda} \int V_y \mu m_w dw \right).
\]

The first term in (6) represents the usual marginal compensated variation which, over any small change in \( r \), is associated with the change in the area under the average compensated demand curve, which is one way to measure the change in welfare that results in a first-best world.

The second term in (6) represents the effect of a change in the nominal interest rate on the incentive constraints. If this second term is zero the welfare cost change is

\footnote{In representative-agent economies a standard measure of welfare changes is given by the equivalent increase in consumption. In our model with a representative agent such a measure coincides with the welfare measure \( \omega(r) \) for small enough changes.}
exactly equal to the change computed using the area under the compensated demand curve. This is the case when the Friedman rule holds exactly, as in a cash-credit model with weakly separable utility. In the shopping-time model the second term is always positive so the welfare cost change is always greater than the area under the compensated demand curve. We summarize this result in the next proposition.

**Proposition 5** The welfare cost from a small change in the nominal interest rate is bounded below by the change in the area under the compensated demand curve for real money balances.

We now perform a numerical exercise to investigate quantitatively how much larger the welfare losses may be. We adopt the shopping-time model specification used by Lucas (2000):

\[ U(c, l) = \log c + \gamma \log(l) \]
\[ s(c, m) = \frac{c}{km} \]

for some \( k > 0 \). Conditional on work time \( n = 1 - l - s \), the steady state demand for money is homogenous with respect to consumption and is approximately log-log with elasticity of one-half

\[ \frac{m}{c} \approx \sqrt{\frac{1}{krg(n)}}. \]

The increasing function \( g(n) \) captures the effect of work time on money demand.

For our exercise we take \( w \) to be log-normally distributed so that \( \log w \sim N(\mu, \sigma) \), and following Tuomala (1990) set \( \mu = -1 \) and \( \sigma = 0.39 \). The constant \( k \) affects the level of money demand, we calibrated it so that it was consistent with the US experience. We set \( \gamma = 1 \) and initially set \( G = 0 \), so that all taxation is for redistribution.

The figure plots the normalized welfare cost

\[ \frac{\omega(r)}{\int c(w)f(w)dw} \]

against \( r \). The welfare cost of a 5% nominal interest rate in this exercise is around 1.2%, while a nominal interest rate of 20% has a cost of 2%. The gains from moving from a nominal interest rate of 6% to 3% are about 0.4%. These welfare costs are larger but close to those obtained by Lucas. They are also very close to the area under the compensated demand curve.\(^{11}\)

\(^{11}\)For all values of \( r \) the marginal labor income tax rate is an inverted-U shape and is consistent
The difference between the welfare costs and the area under the demand curve term becomes larger with a more concave welfare function and with positive government expenditure. Both changes tend to increase the optimal marginal income taxes. Simulations show that the welfare costs increase by about 5-10% for parameterization that imply average marginal income tax rates between 30-35%.

5 Concluding Remarks

Macroeconomists have modeled agent heterogeneity and incorporated elements of contract theory to address a number of important issues. The optimal taxation literature started by Mirrlees (1971) was at the forefront of both developments. However, for the dynamic taxation issues that macroeconomists have explored, the representative agent Ramsey paradigm still reigns supreme. Our paper is the first attempt to bridge this gap by incorporating a Mirrleesian setup in an otherwise standard, dynamic-equilibrium model used by macroeconomists.

One advantage of such an approach is that the source and intuition for results appears more transparent. In Ramsey settings tax prescriptions are somewhat mysterious, intuitions often elusive and endlessly debatable. A clear intuition for results is extremely important to understand elements the model has or has not captured.

\[ \text{with other numerical exercises reported by Mirrlees (1971), Tuomala (1990) and others. Average tax rates are increasing in } w. \text{ For } G = 0, \text{ the average marginal tax in our exercise is low.} \]
We concede that for some questions a Ramsey approach may be a tractable way of obtaining the right insights. However, the difficulty lies in knowing \textit{a priori} when this will be the case. For now, it appears that the only way of being sure is by carrying out a more primitive analysis, with heterogeneity and asymmetric information, as we have done here.

**Appendix A: Proposition 3**

We derive a preliminary result which characterizes the sign of $m_w(y, Y, r, w)$ in terms of the properties of a standard demand system. Consider the following consumer problem,

$$\max_{c, m, l} u(c, m, l)$$

subject to,

$$c + rm + wl = I$$

where $I$ is ‘full-income’. Let the demands for this problem be $\tilde{c}(r, w, I)$, $\tilde{m}(r, w, I)$ and $\tilde{l}(r, w, I)$. Let $\eta$ and $\varepsilon$ represent income and price elasticities and $s_i$, the share of expenditure good $i = c, m, l$.

**Lemma 1** Assume expenditure is a normal good, so that $s_c \eta_c + s_m \eta_m > 0$, then $m_w \leq 0$ if and only if,

$$\eta_c \varepsilon_{mw} \geq \eta_m \varepsilon_{cw}.$$  \hfill (7)

**Proof.** Define the shadow wage and income, $w(l, r, y)$ and $I(l, r, y)$, by the identities

\begin{align*}
l &\equiv \tilde{l}(r, w(l, r, y), I(l, r, y)), \\
y &\equiv \tilde{c}(r, w(l, r, y), I(l, r, y)) + r\tilde{m}(r, w(l, r, y), I(l, r, y))
\end{align*}

Define,

\begin{align*}c(l, r, y) &\equiv c(r, w(l, r, y), I(l, r, y)), \\
m(l, r, y) &\equiv m(r, w(l, r, y), I(l, r, y)).\end{align*}
Differentiating these four identities with respect to \( l \) yields:

\[
\begin{align*}
1 &= \epsilon_{lw}\varepsilon_{wl} + \eta_l\varepsilon_{Il}, \\
0 &= s_c\varepsilon_{cl} + s_m\varepsilon_{ml}, \\
\varepsilon_{cl} &= \varepsilon_{cw}\varepsilon_{wl} + \eta_c\varepsilon_{Il}, \\
\varepsilon_{ml} &= \varepsilon_{mw}\varepsilon_{wl} + \eta_m\varepsilon_{Il},
\end{align*}
\]

where \( \varepsilon_{ly} \equiv \partial \log l / \partial \log y \), and \( s_c = c/I \) and \( s_m = rm/I \) are the expenditure shares.

Note that:

\[
\varepsilon_{Il} = -s_c\varepsilon_{cw} + s_m\varepsilon_{mw} s_c\eta_c + s_m\eta_m \varepsilon_{wl}.
\]

We then obtain:

\[
\varepsilon_{cl} = \left( \varepsilon_{cw} - \eta_c \frac{s_c\varepsilon_{cw} + s_m\varepsilon_{mw}}{s_c\eta_c + s_m\eta_m} \right) \varepsilon_{wl}.
\]

Given the assumption of joint normality it can be shown that \( \varepsilon_{wl} < 0 \). To see this note that the budget constraint \( y + lw(l, y, r) = I(l, y, r) \) implies that

\[
s_l(\varepsilon_{wl} + 1) = \varepsilon_{Il}
\]

substituting this into the first equation and rearranging yields:

\[
\varepsilon_{wl}(\varepsilon_{lw} + \eta_l s_l) = 1 - \eta_l s_l,
\]

\( \varepsilon_{lw} + \eta_l s_l < 0 \) by the Slutsky equation and \( 1 - \eta_l s_l = s_c\eta_c + s_m\eta_m > 0 \), implying \( \varepsilon_{wl} < 0 \).

It follows that, \( \varepsilon_{cl} \geq 0 \) if and only if,

\[
\varepsilon_{cw} - \eta_c \frac{s_c\varepsilon_{cw} + s_m\varepsilon_{mw}}{s_c\eta_c + s_m\eta_m} \geq 0
\]

which upon rearranging yields the condition in the Lemma. ■

**Proof of Proposition 3.** To verify the hypothesis of the Lemma we consider the related subproblem that conditions on a level for consumption:

\[
\begin{align*}
\max_{m,l} & \quad U(c, l - s(c, m)) \\
\text{s.t.} & \quad c + rm + wl = I
\end{align*}
\]
with solution $M(c, I, r, w)$ and $L(c, I, r, w)$. The first order condition w.r.t. $m$ is,

$$-v_m(c, m) = \frac{r}{w},$$

and the second order condition requires $v_{mm} \geq 0$. This implies that $\partial M/\partial I = 0$ and $\partial M/\partial w \geq 0$, or in terms of elasticities $\eta_M = 0$ and $\varepsilon_{Mw} \geq 0$. Since $m(r, w, I) \equiv M(c(r, w, I), I, r, w)$ this implies that,

$$\eta_m = \varepsilon_{Mc} \eta_c,$$

$$\varepsilon_{mw} = \varepsilon_{Mc} \varepsilon_{cw} + \varepsilon_{Mw}.$$ 

Substituting these expressions into condition (7) yields,

$$\eta_c \left( \varepsilon_{Mc} \varepsilon_{cw} + \varepsilon_{Mw} \right) \geq \left( \eta_c \varepsilon_{Mc} \right) \varepsilon_{cw},$$

which is equivalent to $\eta_c \varepsilon_{Mw} \geq 0$. Since $\eta_c > 0$ by assumption and $\varepsilon_{Mw} > 0$ the result follows.

**Appendix B: Non-Stationary and Randomized Allocations**

This appendix shows a connection between non-stationary allocations for $\{y_t, Y_t\}_{t=0}^\infty$ and randomization schemes when $r_t = 0$.12 The planner’s problem must satisfy the resource constraints (recall that with $r_t = 0$ we have that $c_t = y_t$)

$$\int (Y_t(w) - y_t(w)) dF(w) \geq G,$$

which implies that

$$\sum_{t=0}^\infty \beta^t \int (Y_t(w) - y_t(w)) dF(w) \geq G.$$

Now consider the relaxed planner problem that substitutes the resource constraints with this single intertemporal constraint.

---

12By considering the relaxed problem without hidden-savings, that is, where agents cannot trade intertemporally among themselves and the planner chooses the expenditure of agents directly, one can show that $r_t = 0$ if incentive constraints bind downward. The analysis is identical to that in Section 2 and is omitted.

The hidden-saving constraints can then be shown to be satisfied for a stationary allocation. So in this case we have a solution to the original problem.
Non-Stationary Problem

\[
\max \int \sum_{t=0}^{\infty} \beta^t V(y_t(w), Y_t(w), 0, w) dF(w)
\]

subject to the resource constraint

\[
\int (Y_t(w) - c(y_t(w), Y_t(w), 0, w)) dF(w) \geq G
\]

for \( t = 0, 1, 2, \ldots \) and the incentive constraints

\[
\sum_{t=0}^{\infty} \beta^t V(y_t(w), Y_t(w), 0, w) \geq \sum_{t=0}^{\infty} \beta^t V(y'_t(w'), Y'_t(w'), 0, w)
\]

for all \( n = 2, \ldots, n \).

Now consider the following static problem with randomization. The planner selects random variables \( Y(w, \varepsilon) \) where \( \varepsilon \) represents the outcome of a randomization device and has distribution \( \mu \).

**Randomization Problem**

\[
\max \int \int V(y(w, \varepsilon), Y(w, \varepsilon), 0, w) dF(w) d\mu(\varepsilon)
\]

subject to the resource constraint

\[
\int \int (Y(w, \varepsilon) - y(w)) d\mu(\varepsilon) dF(w) \geq G
\]

and the incentive constraints

\[
\int (V(y(w, \varepsilon), Y(w, \varepsilon), 0, w) - V(y(w', \varepsilon), Y(w', \varepsilon), 0, w)) d\mu(\varepsilon) \geq 0
\]

for all \( n = 2, \ldots, n \).

It follows that the non-stationary-relaxed-problem is a restricted version of the randomization-problem where \( \varepsilon \) is allowed to take a countable number of realizations \( t = 0, 1, 2, \ldots \) with probability \( \mu \) putting weights \( \beta^t / (1 - \beta) \). Thus, if randomization is undesirable the solution to the non-stationary-relaxed problem yields a stationary allocation. Since a stationary allocation also satisfies the resource constraints in each period it is a solution.

One can also show in this context that if randomization schemes are allowed then
the allocation can be made stationary without loss in generality.

References


