Tax arbitrage and labor supply

Jonas Agell\textsuperscript{a,\,*}, Mats Persson\textsuperscript{b}

\textsuperscript{a}Department of Economics, Uppsala University, Box 513, SE-751 20 Uppsala, Sweden
\textsuperscript{b}Institute for International Economic Studies, Stockholm University, SE-106 91 Stockholm, Sweden

Abstract

We examine how tax avoidance in the form of trade in well-functioning asset markets affects the basic labor supply model. We show that models that integrate tax arbitrage and labor supply decisions may shed light on a number of positive and normative questions concerning modern systems of income taxation. Such models also appear to have striking implications for empirical research. Studies that ignore tax avoidance may easily come up with biased estimates of the tax responsiveness of the labor supply of high-wage individuals. Also, because of tax avoidance, international comparisons of income inequality will exaggerate the redistributive achievements of high-tax countries like Sweden.

\textcopyright 2000 Elsevier Science S.A. All rights reserved.

Keywords: Labor supply; Tax avoidance; Tax progression

\textit{JEL classification:} H21; H23; H24

1. Introduction

In the 1970s and 1980s, Sweden imposed the industrialized world’s most steeply progressive income tax schedule. Due to the compound impact of income taxes, value added taxes, and payroll taxes, the marginal tax wedges for some broad categories of employees was in the range 80–90\%. The gut reaction of most economists would certainly be that tax distortions of such magnitudes ought to imply a high and easily detectable cost in the form of lost work hours. Even so,
recent studies suggest that the response in work effort was modest; see e.g. Agell et al. (1998).

Do Swedes bother less about incentives than people at large? We believe the answer is no. There are many ways of rationalizing a small labor supply response to high tax wedges, like the influence of social norms (Lindbeck, 1995), or the effects of corporatist wage bargaining systems (Summers et al., 1993). Sandmo (1991) discusses the possibility that the welfare state mitigates tax disincentives by public spending programs that promote labor supply.

But there might be an even simpler reason for the modest labor supply response. There is good reason to believe that the statutory rate of tax progression exaggerated the labor supply distortions that confronted Swedes with high incomes. Two decades ago, Nobel Laureate Gunnar Myrdal (1978) complained that high marginal tax rates had created so strong incentives for high-income individuals to exploit a variety of tax avoidance schemes that the tax system no longer redistributed income. The tax system looked egalitarian, but in practice taxpayers restored the incentive structure by perfectly legal asset transactions. A decade later this perception had reached the highest political circles. At a press conference in 1988, the Social Democratic Minister of Finance and the leader of the powerful confederation of blue-collar workers agreed that 30 years of egalitarian tax policy had come to a dead end. Due to the pervasive nature of tax avoidance, spurred by financial deregulation, high-income individuals escaped their fair share of the tax burden. Two years later, as an acknowledgement of the discrepancy between statutory and actual tax rates, the former were drastically lowered.

The observation that the study of labor supply may require an integrated analysis of how decisions on tax planning and asset trade affect budget sets is not novel. In an overview of the lessons from the US Tax Reform Act of 1986, Auerbach and Slemrod (1997, p. 627) conclude that ‘...the statutory tax rate is not a reliable measure of how the tax system affects the opportunities of individuals and firms, and the true budget sets reflect not only the apparent relative prices that would prevail in the absence of avoidance, but also how real behavior facilitates avoidance and vice versa’.

From observations such as these it is surprising that the labor supply literature abstracts from issues of tax planning and tax arbitrage. The econometric analysis of labor supply typically treats an individual’s asset income as exogenous, and determined independently of the supply of hours; see e.g. Hausman (1981), MaCurdy et al. (1990), and Blomquist (1996). The positive analysis of how tax changes affect labor supply implicitly assumes that the effective marginal tax rate changes in tandem with the statutory marginal tax rate. The normative analysis of income taxation following Mirrlees (1971) rests on the assumption that the optimal income tax schedule is some nonlinear function applied to true labor income.

We examine how tax avoidance in the form of asset trade interacts with the
basic labor supply model. We show that unlimited trade in perfect asset markets has a number of striking implications. Tax arbitrage will continue until all individuals have the same marginal tax rates and taxable incomes. In the process the gains from tax arbitrage will be concentrated to the tails of the wage distribution, while individuals with average wages might lose out. When we introduce various — seemingly realistic — limitations on the arbitrage process, marginal tax rates and taxable incomes will no longer be the same for all individuals. But the distribution of taxable income will still be more compressed than the wage distribution, and the formal rate of tax progression will still exaggerate the labor supply disincentives of high-wage individuals. An important insight is that a failure to account for tax arbitrage in econometric work on labor supply and in applied work on income distribution may lead to serious problems.

The kind of mechanisms that we examine here seems most relevant in countries with high marginal tax rates, non-uniform capital taxation, and well-developed financial markets. As we discuss in the next section, Sweden seems to fit these conditions quite well — at least before the 1990–91 tax reform. In countries with less developed financial markets, lower marginal tax rates, and/or uniform capital taxation, there is much less scope for tax arbitrage. Even so, the rapid pace of innovation in financial markets seems to suggest that our topic is of more general interest.

2. Labor income versus taxable income: some stylized facts

The general strategy of tax planners is to claim deductible expenses against fully taxable income, and report income in forms granted preferential tax treatment. The incentive for doing so is of course greater in countries with high statutory tax rates. High-income earners can then exploit a number of asset transactions to escape taxation. In Sweden before the 1990–91 tax reform these transactions ranged from sophisticated schemes of transforming corporate source income into low-taxed capital gains, to much simpler operations that exploited the fact that until 1985 net negative asset income was subtracted without limitation from labor income when calculating taxable income. These schemes of buying low-taxed assets using borrowed money were not confined to traditional tax shelters like housing and works of art. There was also the opportunity to invest, within limits, borrowed

---

1 In this context we should mention two related strands of literature. Firstly, following the seminal work of Allingham and Sandmo (1972), several contributions study the interaction of labor supply and illegal tax evasion (e.g. Sandmo, 1981; Cowell, 1990). Secondly, there is the literature on the theory of labor supply in the presence of a general tax avoidance technology; see Mayshar (1991), Slemrod (1994), and Feldstein (1995). Our analysis differs from the former body of literature in that we focus on legal tax avoidance, and from the latter in that we focus on asset trade and equilibrium asset pricing.
money in untaxed pension plans or tax favored savings accounts, or to engage in
tax avoidance through straightforward intra-family debt transactions (see Agell et al., 1998).

People seem to have responded to these incentives. According to empirical
studies, individuals with high labor income, and high marginal tax rates, were
much more prone to own tax-favored assets, and to go into debt; see Agell and
Edin (1990), and Edin et al. (1995). Also, the revenue statistics indicate that
people claimed deductions to the extent that the personal capital income tax caused
a substantial aggregate revenue loss. Further evidence is given in Table 1, where
we reproduce the results of Malmer and Persson (1994), based on the 1980
cross-section of the income distribution survey of Statistics Sweden. Individuals
are sorted into deciles by reported labor income. Column 1 reports the average
labor income in each decile. Column 2 shows the tax (‘simulated tax’) that should
have materialized if people’s only source of taxable income had been the labor
income reported in the first column (allowing for the 1980 tax brackets, basic
deductions, etc.). Column 3 shows the tax (‘final tax’) that materializes when one
considers the actual taxes on all kinds of taxable asset income. Columns 4 and 5,
finally, show the implied average tax rates.

Starting at the bottom row, which shows the figures for the average individual in

<table>
<thead>
<tr>
<th>Labor income group</th>
<th>Labor income</th>
<th>Simulated tax on labor income</th>
<th>Final tax, allowing for all taxes on personal capital income</th>
<th>Tax in percent of labor income</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Simulated</td>
<td>Final</td>
</tr>
<tr>
<td>1</td>
<td>2936</td>
<td>362</td>
<td>561</td>
<td>12.1</td>
</tr>
<tr>
<td>2</td>
<td>22,725</td>
<td>5971</td>
<td>5937</td>
<td>25.8</td>
</tr>
<tr>
<td>3</td>
<td>36,104</td>
<td>10,378</td>
<td>10,052</td>
<td>28.2</td>
</tr>
<tr>
<td>4</td>
<td>46,827</td>
<td>14,643</td>
<td>14,015</td>
<td>30.7</td>
</tr>
<tr>
<td>5</td>
<td>56,335</td>
<td>18,389</td>
<td>17,596</td>
<td>31.9</td>
</tr>
<tr>
<td>6</td>
<td>62,554</td>
<td>21,100</td>
<td>20,038</td>
<td>33.0</td>
</tr>
<tr>
<td>7</td>
<td>68,744</td>
<td>24,540</td>
<td>22,906</td>
<td>34.5</td>
</tr>
<tr>
<td>8</td>
<td>74,901</td>
<td>28,213</td>
<td>25,656</td>
<td>36.3</td>
</tr>
<tr>
<td>9</td>
<td>85,518</td>
<td>35,531</td>
<td>31,265</td>
<td>39.8</td>
</tr>
<tr>
<td>10</td>
<td>130,579</td>
<td>71,599</td>
<td>57,361</td>
<td>52.6</td>
</tr>
<tr>
<td>Total</td>
<td>58,722</td>
<td>23,078</td>
<td>20,543</td>
<td>38.1</td>
</tr>
</tbody>
</table>


There could of course be many reasons apart from tax incentives that explain why high income
people were so heavily indebted (e.g. indebtedness because of educational expenses). However, the
marked decrease in indebtedness that took place after the implementation of the 1990–91 tax reform
seems to suggest that tax incentives play an important role; see Agell et al. (1998).
the sample, we see that the deductions for interest expenses, realized capital losses, and private pension savings exceeded reported asset income by a substantial amount. As a consequence, the final tax falls short of the simulated one, and the average tax on labor income is lowered from 38.1 to 33.9%. But these tax reductions are very unevenly distributed. While the discrepancy between simulated and final tax rates is modest in the lower deciles, it is huge in the upper ones. Through various capital transactions the average individual in decile 10 lowered the average tax on labor income from 52.6 to 42.1%. Even more strikingly, Malmer and Persson report that a third of the individuals in the highest decile managed to reduce the final tax on labor income by more than 25%, and that a tenth of them halved the same tax rate.

From figures such as these it is hard to avoid the conclusion that Myrdal was more right than wrong, and that the formal rate of tax progression very much exaggerated the effective one. This interpretation is given further support by a study of Hansson and Norrman (1986), which concludes that the 1982 Swedish income tax can be characterized as de facto proportional. The Swedish experience does not seem to be unique. In their study of income tax avoidance in Germany, Lang et al. (1997) conclude that various legal and semi-legal tax write-off opportunities led to a dramatic reduction of the effective marginal tax rates of high-income households. For a discussion of the less clear-cut evidence for the US, see Gordon and Slemrod (1988).

3. The simple model

3.1. Labor supply

As a point of reference, assume first that no tax arbitrage is possible. Consider an individual with wage rate \( w \), labor supply \( \ell \), and a time endowment normalized at unity. Given some tax schedule \( T(w\ell) \), the individual solves the optimization problem

\[
\begin{align*}
\text{Max} & \quad u(c, 1 - \ell) \\
\text{s.t.} & \quad c = w\ell - T(w\ell).
\end{align*}
\]

Assuming that the problem is well-behaved, this gives rise to a labor supply function of the form

\[
\ell = \ell(\tau, w) \tag{1}
\]

where \( \tau \) is a vector of parameters summarizing various properties of the tax system. Appended with an exogenous wealth term this is a standard formulation used in empirical studies of labor supply, that exploit sophisticated techniques to deal with the non-linearities of the tax system.
Consider now the implications of introducing tax avoidance in the form of trade in a perfect capital market. More specifically, we consider an economy with two types of assets, tax-exempt and taxable claims. We assume that no other form of wealth exists, and that the claims are *inside assets*, that is, people can freely issue both types of claims. This last assumption is a strong simplification (relaxed in Section 4), but it serves the useful purpose of illuminating the essence of the problem.

Denote by $r$ the risk-free interest rate on taxable claims, and by $\rho$ the risk-free interest rate on tax-exempt claims. As no individual has any initial wealth, a positive holding of taxable claims must be balanced by a negative holding of tax-exempt claims of the same amount, and vice versa. As a matter of convention, we let $X$ denote borrowing against taxable claims. In line with a conventional global income tax we assume that the interest expense $rX$ is fully deductible when calculating taxable income. Taxable income $B$ is thus $(\ell w - rX)$, and the tax paid\(^3\) is $T(\ell w - rX)$. Throughout we will assume that the tax schedule is continuously differentiable, that $0 \leq T'(\cdot) \leq 1$, and that the tax schedule is progressive\(^4\) in the sense that $T''(\cdot) > 0$. An individual with a wage rate $w$ solves the following optimization problem:

\[
\begin{align*}
\text{Max}_{c, X} & \quad u(c, 1 - \ell) \\
\text{s.t.} & \quad c = \rho X - rX + w\ell - T(\ell w - rX).
\end{align*}
\]

The first-order conditions are

\[
\frac{u_s(c, 1 - \ell)}{u_t(c, 1 - \ell)} = w \cdot (1 - T'(w\ell - rX)),
\]

\[
\frac{\rho}{r} = 1 - T'(w\ell - rX).
\]

Eq. (4) is our arbitrage condition, saying that the after-tax interest rate on taxable claims must equal the interest rate on tax-exempt claims. Since all individuals face the same relative asset yield $\rho/r$ in a perfect capital market, Eq. (4) implies that all individuals will have the same marginal tax rate $T'(w\ell - rX)$. Tax arbitrage will therefore transform the previously non-linear statutory tax schedule into an effective *linear* schedule with slope $1 - \rho/r$. Also, since the marginal tax rate is a monotone function of taxable income, everybody will report the same taxable income:

\[
w\ell - rX = T'^{-1}(1 - \rho/r) = B(\rho/r).
\]

\(^3\)If the individual invests in taxable claims, $-rX$ will be a positive number.

\(^4\)This assumption is in stark contrast to the result of Mirrlees (1971) that marginal tax rates should decrease at the highest levels of income. What interests us here is not the optimal income tax, but rather some implications of the tax structures that have been in place in most countries for a long time.
This is intuitively reasonable. With unlimited tax arbitrage, high-income individuals issue taxable claims and hold tax-exempt claims, while low-income individuals hold taxable claims and issue tax-exempt claims. This process continues until all taxable incomes, and hence all marginal tax rates, are equalized. From an efficiency point of view, we may also note that this implies that

\[
\frac{u_s(c,1-\ell)}{u_s(c,1-\ell)} = w^\rho r.
\]

In the presence of tax arbitrage, individual marginal rates of substitution between consumption and leisure are strictly proportional to the marginal rate of transformation (i.e. the wage rate). Since \( \rho \) and \( r \) are unequal, the factor of proportionality is to be regarded as a tax wedge.

Combining (3), (4), (5), and the budget constraint in (2), gives us labor supply \( \ell \) and asset demand \( X \) (or, rather, the demand for interest deductions \( rX \)):

\[
\ell = \ell(\rho/r, w), \quad \text{(6)}
\]

\[
rX = R(\rho/r, w). \quad \text{(7)}
\]

Since all individuals will be located at the same point on the tax schedule, we do not enter the vector \( \tau \) as a separate argument in (6) and (7). However, it will still affect the form of the \( \ell \) and \( R \) functions. In the no-arbitrage case, however, individuals will be located at different points on the tax schedule. This is the reason why we explicitly enter \( \tau \) as a separate argument in (1).

Here, one might of course ask why the government would choose a non-linear tax schedule if arbitrage would automatically make the effective tax schedule linear. The answer is either that the non-linear statutory schedule might be the result of some public choice mechanism that is outside the present model; or that the government, having not (yet) read this paper, incorrectly believes that it can impose a non-linear schedule.

The relative asset yield is determined by invoking equilibrium in the asset market. We assume that all agents have identical preferences and differ only with respect to the wage rate \( w \) which, assuming a linear production technology, is the individual’s marginal and average productivity. In a Miller (1977) equilibrium, with purely inside assets, if follows that

\[
\int XdF(w) = 0 \Rightarrow r \int XdF(w) = \int R(\rho/r, w)dF(w) = 0
\]

where \( F(w) \) is the cumulative distribution function of wages. This condition determines \( \rho/r \) and thereby the effective marginal tax rate in the linearized tax

\[\text{While there are two assets, there is only one independent equilibrium condition in asset markets. As a consequence, we obtain solutions for relative rather than absolute asset returns.}\]
schedule. It follows readily that an individual’s labor supply in general equilibrium, \( \ell^* \), depends not only on her own wage but also on various moments of the overall wage distribution. Asset trade tends to link the labor supply decision to the wage structure, much in the same way as considerations of envy and interdependent utility can produce demand and supply functions that depend on measures of relative income.\footnote{See e.g. Persson (1995) and the references cited therein.}

We can obtain some characterizations of the equilibrium. Integrating (5) over all individuals, and using the Miller equilibrium condition (8), yields \( B(\rho/r) = w\ell^* \), where \( w\ell^* \) is the average labor income in the economy. Substituting this back into (5), we obtain the individual’s demand for tax deductions in general equilibrium:

\[
\begin{align*}
    rX^* &= w\ell^* - w\ell^*.
\end{align*}
\]

As we should expect, individuals with a higher-than-average labor income have a positive demand for tax deductions, while the opposite holds for those with lower-than-average income. Individuals with average labor income will not participate in the arbitrage process. Also, when the tax system is purely redistributive, and when the government’s budget is balanced, there is a simple relation between the curvature of the tax schedule and the relative asset yield. An often adopted measure of the degree of tax progression is the measure of residual tax progression:

\[
\begin{align*}
    v(B) &= \frac{1 - T'(B)}{1 - T(B)}.
\end{align*}
\]

Our assumptions about the tax system imply that \( 0 < v(B) < 1 \). In equilibrium everyone has the same taxable income \( B = w\ell^* \). With a redistributive tax system and a balanced budget, it follows that \( T(B) = T(w\ell^*) = 0 \). Thus, \( v(w\ell^*) = 1 - T'(w\ell^*) \), and by (4)

\[
\begin{align*}
    v(w\ell^*) &= \frac{\rho}{r}.
\end{align*}
\]

In equilibrium the relative asset yield equals the degree of tax progression of the formal bracket schedule, evaluated at the average labor income.

A logarithmic example. So far our analysis is quite general. It turns out, however, that we can obtain surprisingly simple closed-form solutions for a particular parameterization of the model, namely a logarithmic utility function of the form

\[
\begin{align*}
    u(c, 1 - \ell) &= \ln c + \alpha\ln (1 - \ell).
\end{align*}
\]

We also assume that the tax system is characterized by constant residual
progression in the sense that \( v \) in (10) is a constant. As shown by Jakobsson (1976), this means that disposable income out of a taxable income \( B \) is an exponential function of \( B \):

\[
B - T(B) = \beta B^v. \tag{11}
\]

Here, \( 0 < v < 1 \), and a lower value of \( v \) means a higher degree of progressivity, in the sense of disposable income being a more concave function of taxable income. For a given \( v \), we use the parameter \( \beta \) to ensure a balanced budget, and for simplicity, we only deal with a purely redistributive tax system. Within our arbitrage model, where everybody has the same taxable income, this implies

\[
T(B) = 0 \iff \beta = B^{1-v}. \tag{12}
\]

In an economy without arbitrage we get the labor supply function

\[
\ell = \frac{v}{v + \alpha}. \tag{13}
\]

This corresponds to the supply function of the standard model (1), and it exhibits the well-known property that with logarithmic preferences and no initial wealth, labor supply is independent of the wage rate \( w \).

In the arbitrage-economy, individual labor supply becomes

\[
\ell = \frac{1}{1 + \alpha} \left[ 1 - \frac{\alpha}{w} \left( \frac{1 - v}{v} \right) \left( \frac{\rho}{\beta v} \right)^{1/(v-1)} \right] \tag{14}
\]

in partial equilibrium. We see that in an arbitrage-economy, modeling labor supply as (13) gives an incorrect functional form for the labor supply function. Using the general equilibrium condition (8), we can also solve for endogenous taxable income \( B(\rho/r) \):

\[
B = \frac{v}{v + \alpha} \bar{w},
\]

where \( \bar{w} \) is the average wage rate, and where we have used the fact that \( \rho/r = v \). Substituting this into the partial equilibrium supply function (14), and using (12), gives us labor supply \( \ell^* \) in general equilibrium:

\[
\ell^* = \frac{1}{1 + \alpha} \left( 1 - \frac{\alpha}{v + \alpha} \frac{\bar{w}}{w} \right). \tag{15}
\]

Now, labor supply does depend on the wage rate even with logarithmic preferences and no other wealth. Comparing labor supply in the arbitrage and no-arbitrage economies, it follows that tax arbitrage leads to lower labor supply for less able individuals \((\bar{w}/w > 1)\), and to higher labor supply for the relatively able ones.
3.2. Efficiency and tax incidence

The introduction of tax arbitrage could be the result of a conscious policy decision by the government, like deregulating the capital market, or making certain transactions legal that were not permitted under the earlier legislation. Or it could be the result of technological developments in financial markets that make asset trade possible at low transaction costs. In either case, it is interesting to know who would gain and who would lose from the introduction of tax arbitrage.

Let us first consider a regime where no tax arbitrage is possible. The indirect utility of an individual with a wage rate \( w \), i.e. the utility resulting from labor supply (1) and its corresponding consumption, is denoted by \( V(w) \). Similarly, the indirect utility of the same individual under a regime where tax arbitrage is possible (i.e. the utility resulting from labor supply \( \ell^a \) and its corresponding consumption) is denoted by \( V'_a(w) \), where the subscript stands for 'arbitrage'. The change in utility from introducing arbitrage is thus \( \Delta = V'_a(w) - V(w) \).

Assume now that the tax schedule is the same for the two cases (this assumption will be relaxed shortly). Then the introduction of tax arbitrage can make nobody worse off, since everybody can abstain from engaging in asset trade and supply exactly the same amount of labor as before. If someone nevertheless would choose to engage in asset trade, it would be only because this would yield higher utility.

The change in relative price between consumption and leisure, from \( w(1 - T'(w\ell)) \) to \( w\ell/r \), would be greatest for people at the tails of the wage distribution. Thus these people would have the strongest incentives to trade with each other, high-income earners issuing taxable claims and purchasing tax-exempt ones, and low-income earners doing the converse. People with average income \( w\ell \) would have nobody to trade with, and no incentive to do so. Thus they would supply the same amount of labor as before, and set \( X = 0 \).

The utility gain from introducing tax arbitrage would therefore be greatest at the tails of the wage distribution,\(^7\) and zero at the mean, as is illustrated by the middle curve in Fig. 1. Note, however, that this reasoning is based on the assumption of an unchanged tax schedule.\(^8\) This is obviously unrealistic. Since the very purpose of tax arbitrage is to reduce taxes, it is reasonable to assume that the introduction of arbitrage would lead to a government budget deficit. In order to maintain budget balance, the government would thus have to raise taxes. Thus the average individual would not be unharmed by the introduction of tax arbitrage. Even if she does not engage in asset trade, she would be harmed by the tax increase and thus suffer a fall in utility. The distribution of \( \Delta \) over income groups would therefore

---

7 Other papers suggesting that tax arbitrage might benefit individuals at the tails of the income distribution include Gordon and Slemrod (1988), and Agell and Persson (1990).
8 It is also based on the assumption of an exogenously determined wage structure. If wages adjust in response to the implied changes in labor supply, people with average income might be harmed by the introduction of tax arbitrage.
Fig. 1. Distribution of utility gains from introducing tax arbitrage.

look like the lower curve in Fig. 1, showing a utility loss in the middle of the income distribution.

On the other hand, we cannot rule out that the introduction of tax arbitrage leads to a budget surplus rather than a deficit. This happens if the reduction of marginal tax rates for people with high wage rates leads to such a large increase in their labor supply that it outweighs the fall in labor supply among those with low wage rates. Then the tax base would increase, and the government could reduce taxes. Thus the average individual would gain from the introduction of tax arbitrage, as would everybody else. The distribution of utility gains would then look like the upper curve in Fig. 1.

How likely are these different cases? This is of course an empirical question. For the particular parameterization employed above (logarithmic preferences, iso-elastic tax schedule), the introduction of tax arbitrage will always necessitate a tax increase. Thus, people in the middle of the wage distribution will always lose. This presumes, however, that an average worker exists. What happens if this is not the case? Clearly, removal of people who do not exploit the scope for tax arbitrage, but pay additional taxes, increases the probability that there will be only winners from the introduction of tax arbitrage.\(^9\)

Finally, two words of caution are warranted. Firstly, the reason why tax arbitrage might be a good thing, producing net gains for everyone, is simply that the initial non-linear tax schedule is arbitrarily chosen (possibly as the result of

\(^9\)Agell and Persson (1998) present a numerical analysis of this possibility for the case of a discontinuous wage distribution. In a model with two groups of workers, logarithmic utility, and an iso-elastic tax system, we show that the introduction of tax arbitrage may benefit both groups, in spite of a deficit that requires a tax hike.
some public-choice mechanism outside the model). It is therefore hardly surprising
that an equally arbitrary linear tax schedule might result in a Pareto improvement.
Secondly, in practice there are often real resource costs associated with many tax
avoidance schemes, ranging from transactions costs and legal fees to costs
associated with an inefficient allocation of risk, consumption, and productive
capital. For example, since much tax avoidance activities involve real estate, an
important cost probably consists of an oversized housing sector. Once we allow for
such secondary excess burdens, the possibility that the introduction of tax arbitrage
leads to a Pareto improvement gets smaller.

4. Introducing limitations to the arbitrage process

In the preceding section, we have seen how a simple arbitrage technology leads
to labor supply responses that are quite different from the ones commonly
analyzed in the literature. The most striking implication is that unlimited arbitrage
leads to complete equalization of taxable incomes, and of the effective marginal
tax rates on labor income. Although data for some countries indicate that tax
avoidance leads to a substantial compression of the distribution of taxable income
relative to the distribution of labor income, incomes are of course not completely
equalized — a fact that suggests that there are limits to the arbitrage process.
Obvious explanations are that imperfections in the capital market make it difficult
to completely avoid taxation, or that the government is efficient in closing the
loopholes that open up for arbitrage.\(^\text{10}\) In this section we will explore two ways of
incorporating such considerations in the analysis, limitations on short sales, and
tax arbitrage in an asset that yields a direct utility service. We discuss the
implications for applied work on labor supply and income distribution in Section
5.

4.1. Outside assets and constraints on short sales

As in the previous section, we model a situation when the only reason for asset
trade is the desire to avoid taxation when marginal tax rates differ across
individuals. But we now assume that the tax exempt asset is an outside asset in
fixed supply, and that this asset (for simplicity referred to as land) cannot be sold
short. We assume that land is productive, in the sense that each land unit produces
\( \rho \) units of the consumption good.\(^\text{11}\) The introduction of land means that there is an
additional layer of heterogeneity. Each individual is now identified by a vector \((w, X)\), where \(X\) is the initial land endowment.

\(^{10}\)For a general discussion of how capital market imperfections and government regulations affect the
\(^{11}\)We thus maintain the notion of a two-good economy, where people consume goods and leisure.
Asset trade proceeds as follows. At the beginning of the period, individuals exchange land endowments for taxable consumption loans. This exchange takes place at the per unit land price $P$. For any individual the proceeds from sales/purchases of land can then be written as $P(X - \bar{X})$, where $X$ is the holding of land at the end of the period, when production occurs. Due to the restriction on short sales, $X \geq 0$. As there can be no net savings in our static model, any change in an individual’s land value is matched by a transaction of the opposite sign in the taxable asset. Specifically, an individual who spends a positive amount $P(X - \bar{X})$ on land increases tax deductible debt by the same amount. As before, $r$ is the interest rate on the taxable asset.

The individual maximizes

$$\max_{\ell, X} u(c, 1 - \ell)$$

s.t. $c = \rho X - \bar{r}(X - \bar{X}) + w \ell - T(w \ell - \bar{r}(X - \bar{X}))$

$X \geq 0$

where $\bar{r} = rP$. The first-order conditions become

$$\frac{u_s(c_1 - \ell)}{u_i(c, 1 - \ell)} = w(1 - T'(w \ell - \bar{r}(X - \bar{X})))$$

$$\rho - \bar{r}(1 - T'(w \ell - \bar{r}(X - \bar{X}))) \leq 0$$

$$= 0 \text{ if } X > 0$$

The complementary slackness condition (18) has important implications for the shape of the labor supply function. Individuals who face a binding quantity constraint in the credit market react to different labor supply incentives than individuals who are at an interior portfolio optimum. Consider first the case of an interior solution, when (18) holds as an equality. For individuals with $X > 0$, tax arbitrage is driven to the point where the after-tax return on the taxable asset equals $\rho$. For these workers, the constraint on short sales is not an issue, and labor supply is determined in a way that parallels our previous analysis. The relevant marginal tax rate is the yield ratio $\rho / \bar{r}$, the same for everyone. It is easy to show that the behavioral functions take the form

$$\ell = \ell(\rho \bar{r}, w, \bar{r}X)$$

$$\bar{r}X = X(\rho \bar{r}, w, \bar{r}X)$$

i.e. the tax system is linearized. Compared with the analysis in Section 3, the only novelty is that (19) and (20) include a wealth effect from the initial land endowment.

To derive the labor supply function for workers who are at a corner solution, we simply set $X = 0$ in (17). We then have

$$\ell = \ell(\rho \bar{r}, w, \bar{r}X).$$
For this group the labor supply function is quite similar to the one that materializes from the standard labor supply model in the presence of a nonlinear tax system, and an exogenous income from wealth, $\bar{r}X$. In particular, the labor supply function includes a vector of parameters, $\tau$, describing global properties of the tax system. Although individuals at corners engage in intra-marginal asset trade, what matters for the form of the labor supply function is the extent of tax arbitrage on the margin.

To characterize the determination of equilibrium asset prices we need to derive the selection criterion that allocates individuals between the arbitrage and no-arbitrage regimes. Consider an individual for whom the non-negativity constraint is binding. For this individual we can write (18) as

$$\frac{\rho}{\bar{r}} - 1 + T'(w\ell(\tau, w, \bar{r}X) + \bar{r}X) = \phi(\tau, w, \bar{r}X, \rho/\bar{r}) < 0.$$  

In the limit the equation $\phi(\tau, w, \bar{r}X, \rho/\bar{r}) = 0$ then gives us the curve in $(w, \bar{X})$ space that separates those who are at an interior optimum from those who are at a corner solution. For the particular case of a logarithmic utility function and a tax system with constant residual tax progression, analyzed in the previous section, the separating surface can be given a nice intuitive interpretation. After straightforward but tedious manipulations, one can show that the separating hyperplane can be written as

$$w + \bar{r}X = \frac{v + \alpha}{v} \left( \frac{\rho}{\bar{r}} \cdot \frac{1}{\beta v} \right)^{\frac{1}{v-1}}.$$  

(22)

Everybody with a full income $w + \bar{r}X$ less than the right-hand side of (22) will be in a constrained optimum, while everybody with a full income that is greater than or equal to the right-hand side will be in an unconstrained optimum. Inherently poor individuals will face a binding credit constraint and a nonlinear tax system, and they will supply labor according to (21). Inherently rich individuals will face a perfect capital market and a linearized tax system, and they will supply labor according to (19).

We can now integrate (20) over all unconstrained individuals to obtain aggregate land demand. In equilibrium, this has to be equal to the total quantity of land available (which we for simplicity set equal to unity):

$$\int_{\phi(.)>0} X(\rho/\bar{r}, w, \bar{r}X) dF(w, \bar{X}) = \bar{r} \cdot 1,$$  

(23)

where $F(w, \bar{X})$ now is a joint cumulative distribution function. Eq. (23) gives a solution for the endogenous variable $\bar{r}$, or rather for the price of land, since $\bar{r} = rP$.

To sum up, when we assume that the tax-exempt asset is not an inside asset, but an outside one which can be held in positive amounts only, some of the more striking predictions from the simple model in Section 3 disappear. Taxable
incomes and effective marginal tax rates will now be the same only for that segment of the population which is at an interior portfolio optimum — an affluent subgroup consisting of those with high wages and/or large land endowments. Less affluent individuals do not hold the tax-exempt asset, and their taxable incomes vary across individuals.

Another observation is that constraints on short sales make it less likely that the introduction of tax arbitrage (via a policy decision, or as the consequence of technological innovation in financial markets) produces utility gains at both tails of the wage distribution. Although high-wage and low-wage individuals have strong incentives to engage in asset trade, the quantity constraint prevents low-wage individuals from going short in the tax-exempt asset. Thus, only high-wage individuals can reap the full reward from tax arbitrage. Finally, and very much in line with the analysis of the preceding section, we cannot rule out the possibility that tax arbitrage leads to such a boost to the labor supply of high-wage individuals, that the government can lower taxes. When the income tax distorts the labor supply decision of high-ability individuals to a great extent, everyone may benefit from the introduction of arbitrage.

4.2. When the tax-exempt asset is an argument in the utility function

In this subsection we briefly discuss a permutation, which turns out to have an interesting connection to the simple model of Section 3. Assume that the outside asset \( X \) appears as an argument in the utility function. Formally, this means a slight reformulation of (16), with \( X \) as an argument in the utility function, and with \( \rho = 0 \):

\[
\begin{align*}
\text{Max } & u(c, 1 - \ell, X) \\
\text{s.t. } & c = w \ell - \tilde{r}(X - \tilde{X}) - T(w \ell - \tilde{r}(X - \tilde{X}))
\end{align*}
\]  

(24)

where \( X \) is now an ordinary consumption good. In practice, we may think of \( X \) as housing, or some consumer durable. This model gives a representation of how housing, or durables, appear in the labor supply decision of an individual in an economy with full interest deductibility. Model (16) above has a closer resemblance to a world where \( X \) is raw land (which gives a constant yield of the consumption good, but which does not appear in the utility function). Working out the first-order conditions gives us solutions for the supply of labor, \( \ell(\tau, w, \tilde{r}, \tilde{X}) \), and for the demand for housing, \( X(\tau, w, \tilde{r}, \tilde{X}) \). These solutions will differ from those in Section 3 in the sense that both first-order conditions of (24) will involve marginal utilities. Several implications follow.

Firstly, since marginal utilities are different for different individuals, taxable incomes will no longer be equalized. But for standard utility functions high-wage individuals will consume more housing than they would have done if interest
expenses had not been tax deductible. Thus the distribution of taxable income will be compressed relative to the wage distribution.

Secondly, consider an econometrician who ignores the fact that there are two consumption goods \((c \text{ and } X)\), and incorrectly specifies the model as

\[
\max_{\ell} u(c, 1 - \ell)
\]
\[
s.t. \quad c = w\ell - \bar{r}(X - \bar{X}) - T(w\ell - \bar{r}(X - \bar{X})),
\]

where \(\bar{r}(X - \bar{X})\) is regarded as exogenously given non-labor income. Clearly, the implied labor supply function \(\ell(z, w, \bar{r}(X - \bar{X}))\) will be mis-specified, too. The problems associated with this are not really due to the econometrician’s neglect of tax arbitrage per se, but rather to neglecting the fact that there are two consumption goods instead of one. The implications for the econometrics of labor supply of introducing a tax deductible good in the utility function is discussed by Triest (1992).

Finally, we have assumed that everyone consumes the same type of housing. In reality, people also make a tenure choice decision: low-income earners tend to consume rental housing, while high-income earners tend to live in their own houses. This suggests an arbitrage mechanism reminiscent of the simple model of Section 3. In that model, low-income earners issue tax-exempt claims and hold taxable ones, while high-income earners hold tax-exempt claims and issue taxable ones. In a housing market with tenure choice, a typical low-income earner pays a non-deductible rent to a landlord and has some taxable savings in the bank. The typical landlord is a high-income earner who has borrowed (tax-deductible) money from the bank in order to buy an apartment house, the income of which is in general favorably taxed. The payment streams are thus very similar: in both cases, low-income earners pay non-deductible and non-taxable (or lightly taxed) money to high-income earners, while high-income earners pay tax-deductible and taxable money to low-income earners.

5. Implications for empirical research

The analysis of the previous sections suggests that there are few general results on the incidence and efficiency effects of tax arbitrage. The answers to questions such as ‘Who gains and who loses from introducing tax arbitrage?’, and ‘Does tax arbitrage lead to a Pareto improvement?’, depend on the exact specification of the model. However, more robust insights appear to emerge when we turn to the implications of tax arbitrage and asset trade for empirical research. All our models — irrespective of their precise assumptions — seem to suggest that empirical studies that fail to account for tax arbitrage may come up with estimates that are seriously biased. Let us see why.
5.1. Econometrics of labor supply

In one form or another, the static labor supply function (1) — extended to include a measure of exogenous ‘non-labor income’ — has been at the center of attention in a large number of microeconometric studies that try to estimate how labor supply reacts to changes in economic incentives. Much of this literature has been preoccupied with the design of econometric techniques capable of dealing with nonlinear tax systems. In this process, the typical individual is modeled as maximizing utility with respect to work hours, subject to a nonlinear budget constraint that treats reported capital income as a given constant. This procedure creates three important problems.

Firstly, there is the issue of self-selectivity. As suggested by our model of constraints on short sales in Section 4.1, people will choose between two labor supply functions. For arbitrageurs (i.e. the inherently rich), labor supply is given by a functional form like (19). For individuals who face a binding constraint in the asset market (i.e. the inherently poor), some other functional form, like (21), applies. Econometric studies that build on the assumption that labor supply has the same functional form for everyone will come up with biased estimates.

Secondly, there is the issue of endogeneity of asset income. The standard procedure is to trace out individuals’ budget sets by changing the supply of hours, holding asset income constant. Under certain assumptions this will give a correct local estimate of the marginal tax rate at the observed equilibrium point of the individual (provided that taxable and nontaxable asset incomes are correctly measured). Thus, even in the presence of tax arbitrage one may obtain a reliable estimate of how disposable income increases with an infinitesimal increase in individual labor supply. Since, however, asset income adjusts endogenously, the estimated marginal tax rate will exceed the effective marginal tax rate for any discrete change in labor supply. This is illustrated in Fig. 2. The concave curve shows consumption as a function of labor supply, given the statutory tax schedule, and holding asset income constant. The straight line shows the effective budget set, given that the individual engages in tax arbitrage. At the equilibrium point, E, the econometrician may derive a (local) estimate of the true marginal incentives for labor supply. But the global incentives are shown by the straight line, rather than by the concave curve.

Thirdly, there is the issue of measurement error in asset income. The prime evidence about labor supply in Sweden, and in many other countries, stems from studies that rely on data from individuals’ tax returns. But the essence of tax arbitrage is that people invest their wealth in ways that are imperfectly measured or registered by the tax authorities. Our analysis suggests that the implied error in

---

12In most practical applications, the statutory budget set is not continuously differentiable, but piece-wise linear. With a piece-wise linear tax system, treating asset income as exogenous will no longer lead to a locally correct estimate of the marginal tax rate.
the measurement of asset income will be correlated with the gross wage. The income tax returns of high-wage individuals will to a much greater extent than the income tax returns of low-wage individuals understate true asset income. Ceteris paribus, this systematic measurement error will tend to create a downward bias in the estimate of the gross wage coefficient. What the econometrician interprets as evidence of a backward bending labor supply function may simply reflect the influence of unobservable asset income.

Of course, the severity of these problems will differ quite a lot across countries, and between different time periods. In a situation with binding credit constraints, and low tax rates, it might be quite innocent to ignore the implications of tax arbitrage. Also, the problems may not be equally severe for all subgroups in the labor market. The econometric issues that we raise are likely to be of particular importance for the study of the labor supply decisions of the affluent. These individuals have much to gain from lowering their taxable income, and there is reason to believe that they have easier access to the credit market than other people.\footnote{For a recent analysis of the labor supply decisions of the affluent, see Moffitt and Wilhelm (1998).}

5.2. Income redistribution in the welfare state

A final observation is that our analysis casts strong doubt on the common use of tax return data in applied work on income distribution. According to several
studies Sweden ranks, together with some other high-tax countries, as one of the most egalitarian countries in the western industrialized world; see Gottschalk and Smeeding (1997). The Gini coefficient for reported disposable income is much smaller in Sweden than in the USA, a finding that is often attributed to the equalizing impact of progressive income taxes and transfers. But for reasons discussed in Section 5.1, the reliability of tax return data as an indicator of true economic income is likely to be a decreasing function of the tax rate. We would expect disposable income of the affluent, computed from the income tax returns, to underestimate actual disposable income. Because Swedish high-income earners are confronted with much higher tax rates than their US counterparts, this bias is likely to be larger in the Swedish than in the US data, thereby creating an impression of a much higher degree of equality in Sweden.

6. Concluding remarks

Our analysis rests on a number of strong assumptions. The analysis in Section 3 relies on the existence of a perfect asset market, and the analysis of capital market imperfections in Section 4 is specific rather than general. There are other ways of introducing assumptions that soften the stark prediction of equal taxable income, the most obvious one being perhaps to introduce uncertainty and risk aversion. This would make people less prone to engage in tax arbitrage to a full extent. Also, our models do not account for the fact that there is an intertemporal dimension to many tax arbitrage strategies (think of the postponed taxation associated with pension plans). Developing models that explore such alternative mechanisms seem like a worthwhile exercise.

We believe that research on how tax arbitrage interacts with labor supply decisions may shed light on a number of important questions concerning modern systems of income taxation. Firstly, what makes people work, given the very high marginal tax rates that can be observed for some countries during some periods? The traditional answer has been that labor supply is rather inelastic. Our proposed answer is different. With the tax avoidance technologies that became increasingly available during the 1980s, those who care about incentives need not pay those high tax rates.

Secondly, how come so many countries turned their tax systems into simple, linear ones during the 1980s and 1990s? Our answer is that during this period, credit markets became increasingly well-functioning. Thus the effective tax schedules became more linear, and the reforms that were undertaken can be seen as a recognition of this fact.

Thirdly, the presence of tax arbitrage has potentially dramatic implications for the analysis of the optimal income tax along the lines of Mirrlees (1971). Tax

\[\text{14}\text{For discussions of the Nordic experience, see Sørensen (1994), and Nielsen and Sørensen (1997).}\]
arbitrage implies that a large class of nonlinear optimal tax schedules will be hard to implement. Exploring models of optimal income taxation that are explicit about asset trade seems like an exciting research topic.

Fourthly, models that allow for tax arbitrage and asset trade have important implications for the econometric analysis of labor supply, and for applied work on income distribution. Studies that ignore the effect of tax arbitrage and asset trade on labor supply incentives will come up with biased estimates of important elasticities, and international comparisons of income inequality will exaggerate the redistributive achievements of high-tax countries like Sweden.

Fifthly, how does tax avoidance affect estimates of the deadweight burden of the income tax? Feldstein (1995) argues that models that account for tax avoidance will produce calculations of excess burdens that are many times larger than those implied by standard Harberger-calculations. A contrary view, from the perspective of a high-tax country, is that tax avoidance may be seen '...as a means of blowing off some steam in order to save the engine from exploding' (Lindencrona, 1993, p. 166). Our analysis shows that it is easy to develop models where asset trade reduces the efficiency cost of income taxation. In the end, whether tax avoidance leads to smaller or greater excess burdens is an empirical question, that depends on country-specifics, like the structure of the tax system, the type of available arbitrage technologies, etc.

Sixthly, from a public choice perspective tax arbitrage and a highly progressive tax system can be viewed as an ingenious way of reconciling incompatible political ambitions. High marginal tax rates convey the message that politicians care about the less well off, while a generous attitude towards tax avoidance prevents the very same tax system from destroying the incentives of the rich and the highly educated. The political economy of tax design seems like an important research topic for the future.

Acknowledgements

We thank Dan Anderberg, Axel Börsch-Supan, Roger Gordon, Ángel López Nicolás, Sören Bo Nielsen, Shlomo Yitzhaki, and two referees for comments. We have benefited from presentations at the 1998 Trans-Atlantic Public Economics Seminar, Swedish Ministry of Finance, and the Universities of Aarhus, Bergen, and Umeå. We thank Annika Persson of the National Tax Board for providing us with the data in Table 1.

References


Sørensen, P.B., 1994. From the global income tax to the dual income tax: recent tax reforms in the Nordic countries. International Tax and Public Finance 1, 57–79.


