# Quantifying the Inefficiency of the US Social Security System

Mark Huggett and Juan Carlos Parra*

June 15, 2005

**Abstract**

How far is the US social insurance system from an efficient system? We answer this question within a model where agents receive idiosyncratic, labor-productivity shocks that are privately observed. When social security and income taxation comprise the social insurance system, the maximum possible efficiency gain is equivalent to a 10.5 percent increase in consumption. This occurs when labor productivity differences are set to the permanent differences estimated in US data.

JEL Classification: D80, D90, E21

Keywords: Social Security, Idiosyncratic Shocks, Efficient Allocations, Private Information

*

Affiliation: Georgetown University
Address: Economics Department; Georgetown University; Washington DC 20057- 1036
E-mail: mh5@georgetown.edu and jcp29@georgetown.edu
Homepage: http://www.georgetown.edu/faculty/mh5 and www.georgetown.edu/users/jcp29
Phone: (202) 687- 6683
Fax: (202) 687- 6102

# 1 Introduction

One rationale for a social security system is the provision of social insurance for risks that are not easily insured in private markets. The Economic Report of the President (2004, Ch. 6) articulates this view. It claims that the provision of social insurance for labor income risk over the life cycle is one of the main problems that justifies a government role in old-age entitlement programs. It argues that labor income is risky but is not easily insured. One reason given for why insurance is difficult is that labor income is partly under an individual's control by the choice of unobserved effort or unobserved labor hours. It then argues that the US social security system provides valuable insurance through a progressive retirement benefit based on lifetime earnings.

Given this argument, we find it natural to ask how far a stylized version of the US social security system is from an efficient system? An answer would give the maximum potential efficiency gain to superior social insurance arrangements. This can also be viewed as a measure of inefficiency. There are at least two reasons why this question is difficult to answer. First, there are many sources of risk to consider and social security systems have distinct benefits tailored to these risks. Second, there are other mechanisms in the US economy, such as income taxation, that are important sources of state-contingent taxes and that may have an important insurance role. Thus, analyzing the inefficiency of social security quickly becomes an analysis of the inefficiency of the tax-transfer system as a whole.

This paper provides a simple benchmark analysis. This simplification is gained by (i) analyzing one component of the US social security system, the retirement component, in isolation, (ii) treating social security together with income taxation as the entire tax-transfer system and (iii) focusing on a single but very important source of risk. The risk that is examined here is idiosyncratic labor-productivity risk. We focus on this risk for two reasons. First, individual workers experience substantial variation in wage rates which are not related to systematic life-cycle variation or to aggregate fluctuations.[1] Second, this risk is a natural way to model labor income as risky but difficult to insure.

The degree of inefficiency of the US social security system together with the income tax system is determined by comparing an agent's ex-ante, expected utility in the model of the US social insurance system to the maximum ex-ante, expected utility that a planner could achieve for the agent. In the model of the US economy it is assumed that there is a risk-free asset for transferring resources over time and that social

[1]Heathcote, Storresletten and Violante (2004) examine annual wage data for US males. They divide (log) wage rates into components capturing life-cycle, business-cycle and idiosyncratic wage variation. They further divide the idiosyncratic component into permanent, persistent and transitory subcomponents and find substantial variation in each subcomponent. See Card (1994) for related work.

security together with income taxation are the only means for transferring resources across states (i.e across an agent's labor-productivity histories). The planner faces two constraints. Allocations must use no more resources in expected present value terms than are used in the US system and must be incentive compatible. The incentive problem arises from the fact that the planner only observes an agent's earnings. Earnings equal the product of labor productivity and labor hours. Thus, the planner does not know whether earnings of an agent are low because labor productivity is low or because labor hours are low. Under these circumstances, the Revelation Principle implies that the allocations between an agent and a planner that can be achieved are precisely those that are incentive compatible.[2]

Preview of Results:

1. The maximum efficiency gain that can be achieved in moving from the allocation under the US social insurance system to the utility possibility frontier is equivalent to a 6.8 percent increase in consumption at each age when there is no labor-productivity risk but a 10.5 percent increase with risk. This occurs when labor-productivity differences are set to the permanent differences estimated in US data. Only a small part of this efficiency gain can be achieved by changing the allocation of consumption, fixing the labor allocation.

2. In the absence of risk, standard intuition (see Feldstein (1996, p. 4) among others) is that inefficiency is increasing in the magnitude of distortionary taxation as efficient allocations equate marginal rates of substitution and transformation. This is consistent with our finding that, in the absence of risk, inefficiency is 0.1 percent when there is social security and no income taxation but is 6.8 percent when social security and income taxation are both present. Marginal earnings taxes in the model are positive at all ages when only social security is analyzed, but are substantially higher when social security and income taxation are both present. Social security leads to positive net marginal tax rates as the proportional social security tax rate on earnings exceeds the present value of marginal social security benefits incurred from additional earnings at all ages.

3. In the presence of permanent labor-productivity risk, efficient allocations do not equate marginal rates of substitution and transformation. Thus, the standard intuition for the inefficiency of social security and income taxation needs to be reconsidered. When utility functions are additively separable between consumption

[2] It is also useful to view this model as one where a planner faces a cohort of ex-ante identical agents that are large in number and that experience idiosyncratic but not aggregate risk. The planner either extracts a prespecified present value of resources from the cohort or gives a prespecified present value of resources to the cohort. Section 3.3 of the paper discusses assumptions such that these are two interpretations of the same model.

and labor, then in an efficient allocation the intertemporal marginal rate of substitution of consumption equals the intertemporal marginal rate of transformation (i.e. the gross interest rate), but the marginal rate of substitution of consumption for labor is below labor productivity for all agent's but the agent with the highest productivity shock. The model of the US social insurance system distorts the intertemporal marginal rate of substitution of consumption below the gross interest rate because marginal income tax rates are positive. The model also distorts the marginal rate of substitution of consumption for labor below an agent's labor productivity. The magnitude of this wedge typically increases at each age as the productivity shock increases. This is because marginal income tax rates increase with income and because marginal net social security tax rates increase with earnings for all agents who are below the maximum taxable earnings level in the social security system. The magnitude of the wedges in an efficient allocation have precisely the opposite pattern.

4. The Economic Report of the President (2004) takes the view that a social security system with a progressive benefit formula and with roughly proportional tax rates is likely to provide valuable social insurance. In the model economies with permanent labor-productivity risk, social security and income taxation are progressive in the sense that the present value of net-taxes paid as a fraction of the present value of earnings increases as the present value of earnings increases. We find that efficient allocations also have progressive average lifetime tax rates. In fact, these tax rates are substantially more progressive in an efficient allocation than under the model of the US social insurance system.

The paper is organized as follows. Section 2 highlights related research. Section 3 presents the modeling framework. Section 4 sets model parameters. Section 5 presents the main results of the paper. Section 6 discusses the results and highlights some important extensions.

# 2 Related Research

This paper builds upon the social security and optimal contract theory literatures. We highlight the papers from these literatures which are most closely related to our work.

To address the role of social security in the provision of social insurance, one needs a model with some risk that is not easily insured in private markets. Imrohoroglu et al (1995), Huang et al (1997), Huggett and Ventura (1999) and Storesletten et al (1999) were among the early works to quantitatively analyze social security systems in the presence of idiosyncratic earnings risk. This paper shares much in common with these papers in that it uses computational methods and adopts the modeling of the US social security system developed in Huggett and Ventura (1999).[3]

Our work is also related to the efficiency gains literature. This literature determines whether or not specific policy changes produce Pareto improvements and calculates the magnitude of any efficiency gains. For example, the classic work by Auerbach and Kotlikoff (1987, Ch. 10) computes efficiency gains from more closely linking marginal social security benefits to marginal social security taxes in a model which abstracts from aggregate and idiosyncratic risk. Our work computes efficiency gains in a model with idiosyncratic, labor-productivity risk that is privately observed. In fact, we compute the maximum efficiency gain, which we label the inefficiency of the social insurance system. Relatively few papers in the efficiency gains literature calculate how far social insurance systems are from efficient allocations.[4]

This paper also builds upon the optimal contract theory literature that emphasizes privately-observed, labor-productivity risk. This literature began with Mirrlees (1971). Diamond and Mirrlees (1978, 1986) extended this framework to consider the optimal disability insurance problem.[5] Two papers from this literature which are similar in spirit to our work are Hopenhayn and Nicolini (1998), who consider the optimal unemployment insurance problem, and Golosov and Tsyvinski (2004), who consider the optimal disability insurance problem. These papers are similar to ours in that allocations produced by stylized models of US institutions are compared to efficient allocations both at a theoretical and at a quantitative level. Our work differs in that we analyze the retirement component of the US social security system and in that we consider the interaction of social security with the US income-tax system.

This paper is also related to work in dynamic contract theory, such as Green (1987), Spear and Srivastava (1987), Thomas and Worrall (1990), Atkeson and Lucas (1992) and Fernandes and Phelan (2000). In this work recursive methods are used to characterize solutions to dynamic contracting problems. An important issue is the nature of tax-transfer systems that implement solutions to dynamic contracting problems

[3]Imrohoroglu et al (2000) survey the social security literature that emphasizes idiosyncratic earnings risk.

[4]Lindbeck and Persson (2003) review the literature on efficiency gains and social security reform. We mention four papers from this literature which differ in the risk analyzed. Hubbard and Judd (1987) determine whether social security improves upon no social security system when there is mortality risk and private markets do not provide annuities. Krueger and Kubler (2003) determine whether there are efficiency gains to adopting a pay-as-you-go social security system in place of private pensions when there is aggregate productivity risk. Huang et al (1997) ask whether there are efficiency gains from two specific changes in social security when agents face idiosyncratic mortality and earnings risk. Nishiyama and Smetters (2004) ask whether there are efficiency gains in moving from the US system to an individual accounts system when agents face idiosyncratic wage risk.

[5]Salanie (1997) reviews the contract theory literature. Diamond (2003) relates work in contract theory to the design of social security systems.

with labor-productivity risk. Albanesi and Sleet (2003), Battaglini and Coate (2004), Golosov and Tsyvinski (2004) and Kocherlakota (2003) present early results which are useful for thinking about this issue.

# 3 Framework

## 3.1 Preferences

An agent's preferences over consumption and labor allocations over the life cycle are given by a calculation of ex-ante, expected utility.

$$E[\sum_{j=1}^{J}\beta^{j-1}u(c_j,l_j)] = \sum_{j=1}^{J}\sum_{s^j\in S^j}\beta^{j-1}u(c_j(s^j),l_j(s^j))P(s^j)$$

Consumption and labor allocations are denoted $(c,l)=(c_1,...,c_J,l_1,...,l_J)$. Consumption and labor at age $j$ are functions $c_j : S^j \rightarrow R_+$ and $l_j : S^j \rightarrow [0,1]$ mapping j-period shock histories $s^j \equiv (s_1,...,s_j) \in S^j$ into consumption and labor decisions. The set of possible j-period histories is denoted $S^j = \{s^j = (s_1,...,s_j) : s_i \in S, i = 1,...,j\}$, where $S$ is a finite set of shocks. $P(s^j)$ is the probability of history $s^j$. An agent's labor productivity in period $j$, or equivalently at age $j$, is given by a function $\omega(s_j,j)$ mapping the period shock $s_j$ and the agent's age $j$ into labor productivity.

## 3.2 Incentive Compatibility

It is assumed that labor productivity is observed only by the agent. The principal observes the output of the agent which equals the product of labor productivity and work time. In this context, the Revelation Principle (see Mas-Colell et al (1995, Prop. 23.C.1)) implies that the allocations $(c,l)$ that can be achieved between a principal and an agent are precisely those that are incentive compatible.

We now define what it means for an allocation to be incentive compatible. For this purpose, we define the report function $\sigma \equiv (\sigma_1,...,\sigma_J)$, which is composed of period report functions $\sigma_j$ that map shock histories $s^j \in S^j$ into $S$. The truthful report function is denoted $\sigma^*$ and has the property that $\sigma_j^*(s^j)=s_j$ in any period for any j-period history. An allocation $(c,l)$ is *incentive compatible* (IC) provided that the truthful report function always gives at least as much expected utility to the agent as any other feasible report function.[6] The expected utility of an allocation

[6] A report function $\sigma$ is feasible for an allocation $(c,l)$ provided that in any period in any history an agent's true labor productivity $\omega(s_j,j)$ is always large enough to produce the output required by a report

$(c,l)$ under a report function $\sigma$ is denoted $W(c,l;\sigma,s_1)$.[7] This is defined below, where $\hat{s}^j \equiv (\sigma_1(s^1),...,\sigma_j(s^j))$ denotes the j-period reported history when the true history is $s^j$. Using this notation, $(c,l)$ is IC provided $W(c,l;\sigma^*,s_1) \geq W(c,l;\sigma,s_1), \forall s_1, \forall \sigma$.

$$W(c,l;\sigma,s_1) \equiv \sum_{j=1}^{J} \sum_{s^j \in S^j} \beta^{j-1} u(c_j(\hat{s}^j), \frac{l_j(\hat{s}^j)\omega(\sigma_j(s^j),j)}{\omega(s_j,j)}) P(s^j|s_1)$$

## 3.3 Decision Problems

This paper focuses on two decision problems: the social security problem and the private information planning problem. These problems have the same objective but different constraint sets. $V^{ss}$ and $V^{pp}$ denote the maximum ex-ante, expected utility achieved in these problems.

$$V^{ss} \equiv \max_{(c,l)\in\Gamma^{ss}} E[\textstyle\sum_{j=1}^{J} \beta^{j-1} u(c_j,l_j)]$$

$$\Gamma^{ss} = \{(c,l) : \textstyle\sum_{j=1}^{J} \frac{c_j}{(1+r)^{j-1}} \leq \sum_{j=1}^{J} \frac{(\omega(s_j,j)l_j - T_j(x_j,\omega(s_j,j)l_j))}{(1+r)^{j-1}} \text{ and } x_{j+1} = F_j(x_j,\omega(s_j,j)l_j,c_j), x_1 \equiv 0\}$$

$$V^{pp} \equiv \max_{(c,l)\in\Gamma^{pp}} E[\textstyle\sum_{j=1}^{J} \beta^{j-1} u(c_j,l_j)]$$

$$\Gamma^{pp} = \{(c,l) : E[\textstyle\sum_{j=1}^{J} \frac{(c_j - \omega(s_j,j)l_j)}{(1+r)^{j-1}}] \leq Cost \text{ and } (c,l) \text{ is IC } \}$$

The constraint set $\Gamma^{ss}$ for the social security problem is specified by a tax function $T_j$ and a law of motion $F_j$ for a vector of state variables $x_j$. The tax function states the agent's tax payment at age $j$ as a function of period earnings $\omega(s_j,j)l_j$ and the state variables $x_j$. A negative tax is a transfer. The social security problem requires that the present value of consumption is no more than the present value of labor earnings less net taxes for any labor-productivity history.[8] The next section demonstrates that this abstract formulation is able to capture important features of the US social security and income tax system.

The constraint set $\Gamma^{pp}$ for the planning problem has two restrictions. First, the expected present value of consumption less labor income cannot exceed some specified

(i.e. $0 \leq l_j(\hat{s}^j)\omega(\sigma_j(s^j),j) \leq \omega(s_j,j), \forall j, \forall s^j$, where $\hat{s}^j \equiv (\sigma_1(s^1),...,\sigma_j(s^j))$).

[7] $W(c,l;\sigma,s_1)$ is defined only for $\omega(s_j,j) > 0$. Later in the paper, we will set labor productivity to zero beyond a retirement age. It is then understood that labor supply is set to zero at those ages.

[8] The budget set can equivalently be formulated as a sequence of budget restrictions where the agent has access to a risk-free asset, starts life with zero units of this asset and must end life with non-negative asset holding.

value, denoted $Cost$. Present values are computed with respect to an exogenous real interest rate $r$. Second, allocations $(c, l)$ must be incentive compatible (IC).

Ex-ante expected utility can be ordered in these problems so that $V^{pp} \geq V^{ss}$. This occurs when $Cost$ in the planning problem is selected to equal the expected present value of taxes incurred in a solution $(c^{ss}, l^{ss})$ to the social security problem (i.e. $Cost \equiv E[\sum_{j=1}^{J} -T_j(x_j, \omega(s_j, j)l_j^{ss})/(1+r)^{j-1}]$). The argument is based on showing that if the allocation $(c^{ss}, l^{ss})$ achieves the maximum in the social security problem, then $(c^{ss}, l^{ss})$ is also in $\Gamma^{pp}$. Since $(c^{ss}, l^{ss})$ satisfies the present value condition in $\Gamma^{ss}$, then it also satisfies the expected present value condition in $\Gamma^{pp}$. Thus, it remains to argue that $(c^{ss}, l^{ss})$ is incentive compatible. However, the fact that $(c^{ss}, l^{ss})$ is an optimal choice for the agent in the social security problem implies that it is incentive compatible.

To conclude this section, we raise two issues concerning how to interpret solutions to the planning problem. First, is a solution to the planning problem a Pareto efficient allocation? Solutions to the planning problem are Pareto efficient allocations between a risk-averse agent and a risk-neutral principal with discount factor $1/(1+r)$ when the utility possibility frontier is downward sloping. It is straightforward to show that the frontier is downward sloping when the agent's period utility function $u(c_j, l_j)$ is additively separable and is strictly increasing and continuous in consumption. Second, does a solution to the planning problem also solve the problem of maximizing ex-ante, expected utility of a large cohort of ex-ante identical agents subject to incentive compatibility and to the requirement that the realized present value cost to the planner not exceed some prespecified level? The assumption here is that agents experience idiosyncratic but not aggregate risk. The contract theory literature mentioned in section 2 imposes the requirement that a present value condition or a market clearing condition must hold in equilibrium but not necessarily for any conceivable (non-equilibrium) reports that these agents could make.[9] Under this requirement, a solution to the planning problem is a solution to the planning problem with a large cohort of ex-ante identical agents.

## 3.4 US Tax-Transfer System

The tax function and law of motion $(T_j, F_j)$ are now specified to capture features of the US social security system together with the US federal income tax system. Specifically, the tax function $T_j$ is the sum of social security taxes $T_j^{ss}$ and income taxes $T_j^{inc}$. The state variable $x_j = (x_j^1, x_j^2)$ in $T_j$ has two components: $x_j^1$ is an agent's average earnings

[9]See Mas-Colell and Vives (1991) for a discussion of this issue and for results on implementation in exchange economies with a continuum of agents.

up to period $j$ and $x_j^2$ is an agent's asset holdings.

$$T_j(x_j, \omega(s_j, j)l_j) = T_j^{ss}(x_j^1, \omega(s_j, j)l_j) + T_j^{inc}(x_j^1, x_j^2, \omega(s_j, j)l_j)$$

### 3.4.1 Social Security

The model social security system taxes an agent's labor income before a retirement age $R$ and pays a social security transfer at and after the retirement age. Specifically, taxes are proportional to labor earnings $(\omega(s_j, j)l_j)$ for earnings up to a maximum taxable level $e_{max}$. The social security tax rate is denoted by $\tau$. Earnings beyond the maximum taxable level are not taxed. After the retirement age, a transfer $b(x^1)$ is given that is a fixed function of an accounting variable $x^1$. The accounting variable is an equally-weighted average of earnings before the retirement age $R$ (i.e. $x_{j+1}^1 = [\min\{\omega(s_j, j)l_j, e_{max}\} + (j-1)x_j^1]/j$). The earnings that enter into the calculation of $x_j^1$ are capped at a maximum level $e_{max}$. After retirement, the accounting variable remains constant at its value at retirement.

$$T_j^{ss}(x_j^1, \omega(s_j, j)l_j) = \begin{cases} \tau \min\{\omega(s_j, j)l_j, e_{max}\} & : \quad j < R \\ -b(x_j^1) & : \quad j \geq R \end{cases}$$

The relationship between average past earnings $x^1$ and social security benefits $b(x^1)$ in the model is shown in Figure 1. Benefits are a piecewise-linear function of average past earnings. Both average past earnings and benefits are normalized in Figure 1 so that they are measured as multiples of average earnings in the economy. The first segment of the benefit function in Figure 1 has a slope of .90, whereas the second and third segments have slopes equal to .32 and .15. Thus, the benefit function bends over. The bend points in Figure 1 occur at 0.21 and 1.29 times average earnings in the economy. The variable $e_{max}$ is set equal to 2.42 times average earnings. The bend-points and the maximum earnings $e_{max}$ are set at the actual multiples of mean earnings used in the US social security system. The slopes of the benefit function are also set to those in the US social security system.[10] Figure 1 says that the social security

[10] Under the US Social Security system, a person's monthly retirement benefit is based on a person's averaged indexed monthly earnings (AIME). For a person retiring in 2002, this benefit equals 90% of the first $592 of AIME, plus 32% of AIME between $592 and $3567, plus 15% of AIME over $3567. Dividing these "bend points" by average earnings in 2002 and multiplying by 12 gives the bend points in Figure 1. Bend points change each year based on changes in average earnings. The maximum taxable earnings from 1998- 2002 averaged 2.42 times average earnings. All these facts, as well as average earnings data, come from the Social Security Handbook (2003). The retirement benefit above is for a single-person household. The US system offers a spousal benefit that we abstract from.

retirement benefit payment is about 45 percent of average earnings for a person whose average earnings over the lifetime equals mean earnings in the economy.

The specification of the model social security system captures many features of the old-age component of the US social security system. Two differences are the following:

(i) The accounting variable in the actual US system is an average of the 35 highest earnings years, where the yearly earnings measure which is used to calculate the average is capped at a maximum earnings level.[11] In the model, earnings are capped at a maximum level just as in the US system, but earnings in all pre-retirement years are used to calculate average earnings.

(ii) In the actual US system the age at which benefits begin can be selected within some limits with corresponding "actuarial" adjustments to benefits. In the model the age $R$ at which retirement benefits are first received is fixed.

### 3.4.2 Income Taxation

Income taxes in the model economy are determined by applying an income tax function to a measure of an agent's income. The empirical tax literature has calculated effective tax functions (i.e. the empirical relationship between taxes actually paid and economic income).[12] We use tabulations from the Congressional Budget Office (2004, Table 3A and Table 4A) for the 2001 tax year to specify the relation between US average effective federal income tax rates and income. Figure 2 plots average effective tax rates for two types of households: head of household is 65 or older and head of household is younger than 65. The horizontal axis in Figure 2 measures income in 2001 dollars. Figure 2 shows that US average federal income tax rates increase strongly in income.

In the model economy, we choose income taxes $T_j^{inc}(x_j^1, x_j^2, \omega(s_j, j)l_j)$ before and after the retirement age $R$ to approximate the average tax rates in Figure 2. We proceed in three steps. First, we approximate the US data in 2001 dollars with a continuous function. Specifically, we use the quadratic function passing through the origin that minimizes the squared deviations of the tax function from data. This gives average tax functions before and after the retirement age. Second, we express model income in 2001 dollars.[13] Third, the average tax rates on model income are

[11]The 35 highest years are calculated on an indexed basis in that indexed earnings in a given year equal actual nominal earnings multiplied by an index. The index equals the ratio of mean earnings in the economy when the individual turns 60 to mean earnings in the economy in the given year. In effect, this adjusts nominal earnings for inflation and real earnings growth.

[12]See, for example, Gouveia and Strauss (1994).

[13]This is done using the ratio between the average US economy earnings and average model earnings. The figure for US average earnings is $32,922. This comes from the benefit calculation section of the Social Security Handbook (2003).

given by the function estimated in the first step after expressing model income in 2001 dollars. Model income equals the sum of labor income $\omega(s_j, j)l_j$, asset income $x_j^2 r$ and social security transfer income $b_j(x_j^1)$. Asset income is calculated as follows: $x_{j+1}^2 = \omega(s_j, j)l_j + x_j^2(1 + r) - T_j(x_j^1, x_j^2, \omega(s_j, j)l_j) - c_j$.

# 4 Parameter Values

The benchmark results of the paper are based on the parameter values in Table 1. There are $J = 61$ model periods in an agent's lifetime. This corresponds to real-life ages 20 to 80. The retirement age (i.e. age at which retirement benefits are first received) occurs in model period $R = 46$ which corresponds to a real-life retirement age of 65. This is the current age at which full benefits are received in the US system. The social security tax rate $\tau$ is set to equal 10.6 percent of earnings. This is the combined employee-employer tax for old-age and survivor's insurance component of social security. The social security benefit function $b(x)$ and the income tax function $T_j^{inc}$ are given by Figure 1 and Figure 2. The previous section discussed how these functions were selected.

An agent's labor productivity is given by a function $\omega(s_j, j) = \mu_j s_j$. The term $\mu_j$ captures the systematic variation in mean labor productivity with age. We set $\mu_j$ equal to the US cross-sectional, mean-wage profile for males from Heathcote et al (2004). This is displayed in Figure 3, where we normalize $\mu_1$ to equal 1. We impose that an agent is not able to work at age 65 or afterwards. The term $s_j$ captures idiosyncratic variation in labor productivity. We consider two possibilities for the stochastic structure of shocks: perfectly permanent shocks and purely temporary shocks. In the case of permanent shocks, an agent is "born" at age $j = 1$ with a realization of the permanent shock which remains with the agent over the life cycle. The agent receives no subsequent shocks. In the case of temporary shocks, an agent draws a shock each period independently from a fixed distribution. In both cases the distribution of shocks is a discrete approximation to a lognormal distribution (i.e. $\log(s_j) \sim N(-\sigma^2/2, \sigma^2)$).[14]

Heathcote et al (2004) have decomposed the idiosyncratic component of variation of log wages of US males into the sum of permanent, persistent and purely temporary components. They estimate that the variance of the purely temporary component of log wage shocks is $\sigma^2 = 0.074$ and that the variance of the permanent component of log wage shocks is $\sigma^2 = 0.109$.[15] These estimates will lie in the range of the variances

[14] We approximate the lognormal distribution with 5 equally-spaced points in logs in the interval $[-3\sigma, 3\sigma]$. Probabilities are set to the area under the normal distribution, where midpoints between the approximating points define the limits of integration. This follows Tauchen (1986).

[15] The estimates cited in the text are the average values of the variances of the respective shock components.

$\sigma^2$ for temporary and permanent shocks that we consider in the next section.

Table 1: Parameter Values

| Definition | Symbol | Value |
|---|---|---|
| Model Periods | $J$ | $J = 61$ |
| Retirement Period | $R$ | $R = 46$ |
| Social Security Tax | $\tau$ | $\tau = .106$ |
| Benefit Function | $b(x)$ | Figure 1 |
| Income Tax Function | $T^{inc}$ | Figure 2 |
| Labor Productivity | $\omega(s_j, j)$ | $\omega(s_j, j) = \mu_j s_j$<br>$\log(s_j) \sim N(-\sigma^2/2, \sigma^2)$ |
| Mean Productivity Profile | $\mu_j$ | Figure 3 |
| Interest Rate | $r$ | $r = 0.042$ |
| Discount Factor | $\beta$ | $\beta = 1.0/(1+r)$ |
| Preferences | $u(c,l)$ | $u(c,l) = \frac{c^{(1-\rho)}}{(1-\rho)} + \phi \frac{(1-l)}{(1-\gamma)}^{(1-\gamma)}$<br>$\rho = 1, \gamma = 3.2856$<br>$\phi$ see text |

One important restriction on the utility function $u(c,l)$ is the assumption of additive separability. Much of the literature on dynamic contract theory with a labor decision referenced in section 2 is based on this assumption. We make use of this assumption when we design a procedure to compute solutions to the planning problem.[16] The discount factor $\beta$ and the real interest rate $r$ are set so that $\beta(1+r) = 1$. Under these assumptions, the consumption profile over the life cycle is flat in a solution to the

These values come from Heathcote et al (2004, Table 2) after weighting the variance in 1967 by the average factor loadings from 1967-1996.

[16]It is used in Theorem A.3 in the Appendix to establish which incentive constraints bind and to develop a two-stage approach to solve the recursive-dual problem. The algorithm to compute solutions to the social security problem does not make use of additive separability.

planning problem, when there is no labor-productivity risk. We set the real interest rate equal to 4.2 percent. This is the average real return over the period 1946- 2001 to an equally-weighted portfolio of stock and long-term bonds (see Siegel (2002, Tables 1-1 and 1-2)).

In the benchmark model, we set $u(c,l) = c^{(1-\rho)}/(1-\rho) + \phi \frac{(1-l)}{(1-\gamma)}^{(1-\gamma)}$. This choice implies a constant elasticity of intertemporal substitution of consumption equal to $\epsilon = -1/\rho$ and a constant Frisch elasticity of leisure with respect to the wage equal to $\epsilon_{leisure} = -1/\gamma$.

We now discuss how we set the parameters $\rho$ and $\gamma$. We make use of estimates based on micro data and the assumption that the period utility function for consumption and labor is additively separable. The estimates of $\epsilon$ surveyed in Browning et al (1999, Table 3.1) range from $-0.25$ to $-1.56$. This would suggest a coefficient $\rho$ ranging from below 1.0 to 4.0. In the benchmark model we set $\rho = 1$ (i.e. $u(c) = \log(c)$) and later examine the sensitivity of the results to higher values. We note that log utility is widely used in general equilibrium models for balanced growth considerations. The literature that estimates the Frisch elasticity of labor supply (see Browning et al (1999, Table 3.3)) is useful for setting $\gamma$. For the preferences under consideration, the Frisch elasticities of labor and leisure are related as follows: $\epsilon_{labor} = -\epsilon_{leisure}(1-l)/l$. We set the parameter $\gamma$ to match an estimate of the Frisch elasticity of male labor supply. Domeij and Floden (2004, Table 5) estimate that $\epsilon_{labor} = 0.49$, using annual data for US males.[17] We choose $\gamma = 3.2856$ to match this estimate of the labor elasticity when labor $l$ in the model equals the average fraction of time worked in the US.[18] The remaining parameter $\phi$ is set so that, given all other model parameters, the average fraction of time worked in the model equals the average value in the US economy. When the variances for the permanent and temporary shocks are set to the point estimates discussed above, the value $\phi$ equals 0.5887 for the permanent shock case and 0.5481 for the temporary shock case.

[17] Using US data, they find that the estimated elasticity is larger when the data set excludes households with small amounts of liquid assets. The estimate in the text is for households with liquid assets equal to at least one month's wages. This estimate is higher than many in the literature but still within the range of estimates surveyed by Browning et al (1999, Table 3.3).

[18] The average fraction of time worked in the US is 0.3832. This equals average hours worked divided by available work time. Average hours worked comes from Heathcote et al (2004, Table 1). Available work time equals 16 hours per day times 365 days per year.

# 5 Results

## 5.1 Quantitative Assessment of Inefficiency

This section quantifies the inefficiency of the US system when labor-productivity shocks are temporary or permanent. The magnitude of the inefficiency is the percentage increase $\alpha$ in consumption in the allocation $(c^{ss}, l^{ss})$ for the social security problem so that ex-ante expected utility is the same as in an allocation $(c^{pp}, l^{pp})$ solving the private information planning problem, holding the expected present value of resources equal in both problems. This calculation is shown below. This calculation also has the interpretation of the maximum efficiency gain that can be achieved in moving to the utility possibility frontier, fixing the resources given to the planner. The results of this section are based on computing solutions to the social security problem and the planning problem. Computational methods are described in detail in Appendix A.

$$E[\sum_{j=1}^{J} \beta^{j-1} u(c_j^{ss}(1+\alpha), l_j^{ss})] = E[\sum_{j=1}^{J} \beta^{j-1} u(c_j^{pp}, l_j^{pp})] \equiv V^{pp}$$

Figure 4 highlights the inefficiency of the US social insurance system in the benchmark model for a range of values for the variance of log labor-productivity shocks. This measure of inefficiency is increasing in the variance of the shocks. This holds both when one considers social security without an income tax system and when social security and income taxation are combined together.

To quantify the size of the inefficiency of the US system, one would need an estimate of the variance of the shocks to log wages. As described in the previous section, Heathcote et al (2004) estimate that $\sigma^2 = 0.074$ for temporary shocks and that $\sigma^2 = 0.109$ for permanent shocks. Thus, a one standard deviation shock increases wages by about 33 percent in the permanent shock case and 27 percent in the temporary shock case. Using these estimates, the inefficiency of the combined social security and income tax system is 10.5 percent of consumption in the permanent shock case and 8.2 percent in the temporary shock case. Figure 4 also shows that the inefficiency measure is substantially larger when income taxation and social security are analyzed together than when income taxation is abstracted from.

## 5.2 Decomposing Inefficiency

We now decompose the inefficiency measure in Figure 4. Specifically, we contrast the inefficiency of the allocation $(c^{ss}, l^{ss})$ under the US social insurance system with the inefficiency of the allocation $(c^{*}, l^{ss})$. The $(c^{*}, l^{ss})$ allocation maximizes the agent's ex-ante expected utility, holding labor fixed at $l^{ss}$ and imposing both the incentive compatibility and the resource constraints from the planning problem.[19] This decomposition highlights the degree to which inefficiency can be reduced by choosing consumption optimally given an allocation of labor. It also highlights the remaining efficiency gains that can be obtained by adjusting labor and further adjusting consumption all the way to an efficient allocation. Expressed more figuratively, this decomposition describes the reduction in inefficiency from better distributing a cake of fixed size versus the reduction in inefficiency resulting from changing the size of the cake and changing the mix of labor used to produce it.

Figure 5 presents the results of this decomposition for the case of permanent shocks. We find that a change in the labor allocation is essential in order to achieve the bulk of the potential efficiency gains. They cannot be obtained by superior consumption allocations, given the labor allocation produced by the model of the US social insurance system. Intuitively, the gains from a superior consumption allocation come from compressing the distribution of consumption across labor-productivity histories at each age and from superior consumption smoothing over time for a given history. Since the utility function is concave in consumption at each age, compressing consumption will lead to an increase in expected utility. To stay within the class of incentive compatible allocations, this compression can only occur up to the point where the incentive constraints bind. Thus, for example, perfect equality of consumption will not be incentive compatible when the labor allocation calls for output differences across agents.

## 5.3 Sources of Inefficiency

We now try to further understand what lies behind the results in Figure 4. We focus first on the case of no labor productivity risk and then consider productivity risk.

### 5.3.1 No Labor-Productivity Risk

In an efficient allocation marginal rates of substitution and transformation are equated in the absence of risk. Given additive separability (i.e. $u(c,l) = u(c) + v(l)$), two of these necessary conditions can be rewritten as follows: $u'(c_j) = \beta u'(c_{j+1})(1+r)$ and

[19] We decompose inefficiency by computing the value $\alpha$ solving the equation below. The allocation $c^*$ solves the maximization problem below.

$$E[\sum_{j=1}^{J} \beta^{j-1} u(c_j^*(1+\alpha), l_j^{ss})] = E[\sum_{j=1}^{J} \beta^{j-1} u(c_j^{pp}, l_j^{pp})]$$

$$max_{c \in \Gamma(l^{ss}, Cost)} E[\textstyle\sum_{j=1}^{J} \beta^{j-1} u(c_j, l_j^{ss})]$$

$$\Gamma(l^{ss}, Cost) \equiv \{c : (c, l^{ss}) \text{ is } IC \text{ and } E[\textstyle\sum_{j=1}^{J} \frac{(c_j - \omega(s_j, j) l_j^{ss})}{(1+r)^{j-1}}] \leq Cost\}$$

$-v'(l_j)/u'(c_j) = \omega(s_j, j)$. In the allocation under the model social insurance system, an optimizing agent equates marginal rates of substitution to the after-tax marginal rates of transformation. Thus, in the absence of risk, non-zero marginal tax rates on asset or labor income lead to inefficient allocations in the model.

We now highlight the marginal tax rate on labor earnings. The focus on earnings is motivated by Figure 5 which shows that changing the labor allocation is key to achieving the bulk of the efficiency gains in Figure 4. Figure 6 graphs this marginal tax rate over the life cycle. Consider first the model without an income tax. In this model the marginal tax on earnings decreases with age.[20] Why is this? The marginal tax rate equals the social security tax rate $\tau$ less the present value of the marginal benefits incurred from an extra unit of earnings. Thus, the marginal tax rate decreases with age because the present value of marginal social security benefits incurred by an extra unit of earnings increases as an agent ages. This occurs for two reasons. First, since the retirement benefit in the model is based on average earnings, a one unit increase in earnings in any period raises the social security retirement payment by the same amount.[21] Second, since the real interest rate in the model is positive, the present value of these marginal benefits incurred is greater towards the end of the working life cycle than at the beginning.

Now consider the model with income taxation and social security. Figure 6 shows that the income tax substantially increases the marginal tax rate on labor earnings. In fact, the marginal tax rate on earnings now increases with age over the early part of the working life cycle. Intuitively, this occurs when income over the life cycle is hump-shaped since average and marginal income tax rates increase with income (see Figure 2).

Figure 7 highlights labor allocations over the life cycle. In an efficient allocation labor is hump-shaped. This follows directly from the first-order necessary conditions. Specifically, the assumption that $\beta(1+r) = 1$ implies that consumption is flat over the life cycle. The necessary condition for labor (i.e. $-v'(l_j)/u'(c_j) = \omega(s_j, j)$) then implies that labor is hump-shaped when labor productivity is hump-shaped (see Figure 3) and when the disutility of labor $v$ is concave. Under the model social insurance system, the labor allocation is lower early in life than in the efficient allocation. This occurs not because marginal tax rates on earnings are relatively high early in life. Instead, this

[20]Clearly, introducing other features of the US social security system (e.g. the spousal benefit or the fact that benefits are based on the 35 highest earnings years) would affect the patterns in Figure 6. The results in Figure 6 for the case of no income tax are similar to the marginal social security tax rates calculated by Feldstein and Samwick (1992, Table 1).

[21]Clearly, this is sensitive to abstracting from growth in average, economy-wide earnings as discussed in section 3.4.1. Appendix B extends the model to allow for growth in average, economy-wide earnings and the indexing of individual earnings to economy-wide earnings.

can be attributed to the effect that positive marginal income tax rates have in inducing consumption to fall with age by making future consumption relatively more expensive in terms of current resources.

### 5.3.2 Labor-Productivity Risk

We now try to understand the nature of the inefficiency displayed in Figure 4 for economies where labor-productivity risk is permanent. We start by looking at marginal rates of substitution between consumption and labor. Figure 8 highlights the wedge between this marginal rate of substitution and the corresponding marginal rate of transformation. More specifically, Figure 8 graphs the ratio of the marginal rate of substitution to the agent's labor productivity at each age over the life cycle for each of the five possible values of the permanent shock. Figure 8a shows that in an efficient allocation this marginal rate of substitution is below an agent's labor productivity for all agents but the agent with the highest permanent shock. Furthermore, within age groups the magnitude of this wedge between rates of substitution and transformation is greatest for agents with the lowest labor productivity and decreases as labor productivity increases.

The pattern of the wedges in the model social insurance system is quite different. Figure 8b shows that this marginal rate of substitution is below an agent's labor productivity for all agents at all ages. In addition, within age groups the wedge typically increases as an agents labor productivity increases. This is precisely the opposite of the results from Figure 8a.

The wedge is smallest for low productivity agents for two reasons. First, these agents have realtively low incomes and marginal income tax rates are relatively low at low income levels (see Figure 2). Second, these agents expect to be on the steep part of the social security benefit function (see Figure 1) but face the same social security tax rate on earnings as all agents who are below the maximum taxable earnings level. As earnings increase both the marginal income and marginal net social security tax rate increase. This holds until earnings hit the maximum taxable earnings under social security. This occurs in Figure 8b for the highest productivity shock agent at age $j = 14$. Beyond this age the highest shock agent faces a zero marginal net social security tax.

It seems plausible that the increases in inefficiency as labor productivity risk increases are related to the pattern of wedges. Specifically, as risk increases from the non risk case, the efficient wedge is always zero for the high shock agent but the wedge in the model is increasing as risk increases. This leads high shock agents to not work enough compared to efficient hours of work.

### 5.4 Tax Implementing Efficient Allocations

- Tax implemtation
- Ave Lifetime Tax Rates

Figure 9 plots average lifetime net-tax rates. The net-tax rate equals the ratio of the present value of earnings less consumption to the present value of earnings. Thus, this tax rate is positive when in present-value terms the agent is a net source of resources to the social insurance system. Recall from section 4 that there are exactly five equally-spaced values of the log of the permanent productivity shock $\log(s)$ on the interval $[-3\sigma$ to $3\sigma]$. Thus, there are five points in Figure 9. In the allocations we compute it is always the case that labor earnings (i.e. the product of productivity and work time) are increasing in the level of an agent's permanent productivity shock.

We highlight three properties from Figure 9. First, high productivity agents work more under an efficient allocation than under the US system with the opposite pattern holding for low productivity agents. This is implied by the fact that the present value of earnings increase in an efficient allocation for high productivity agents compared to the US system. Clearly, the opposite pattern occurs for low productivity agents. Second, the net-tax rate is increasing in the agent's productivity both in the US system and in an efficient allocation. Thus, both allocations offer insurance in the sense that the net-present value of transfers decreases as labor productivity increases. Third, the net-tax rate increases much more sharply in an efficient allocation than under the US social insurance allocation. In fact, the net-tax rate under the US system is always positive but under an efficient allocation it varies from less than $-150$ percent to about 33 percent.

# 6 Discussion

# References

Albanesi, S. and C. Sleet (2003), Dynamic Optimal Taxation with Private Information, manuscript.

Atkeson, A. and R. Lucas (1992), On Efficient Distribution with Private Information, *Review of Economic Studies*, 59, 427-453.

Auerbach, A. and L. Kotlikoff (1987), *Dynamic Fiscal Policy,*(Cambridge University Press, Cambridge).

Battaglini, M. and S. Coate (2004), Pareto Efficient Income Taxation with Stochastic Abilities, manuscript.

Browning, M., Hansen, L. and J. Heckman (1999), Micro Data and General Equilibrium Models, in *Handbook of Macroeconomics,* ed. J. Taylor and M. Woodford, (Elsevier Science B.V., Amsterdam).

Card, D. (1994), Intertemporal Labor Supply: An Assessment, in C. Sims (ed) *Advances in Econometrics: Sixth World Congress.*

Congressional Budget Office (2004), Effective Federal Tax Rates: 1979- 2001, available at http://www.cbo.gov/.

Diamond, P. (2003), *Taxation, Incomplete Markets, and Social Security,* (The MIT Press, Cambridge).

Diamond, P. and J. Mirrlees (1978), A Model of Social Insurance with Variable Retirement, *Journal of Public Economics*, 10, 295-336.

Diamond, P. and J. Mirrlees (1986), Payroll-Tax Financed Social Insurance with Variable Retirement, *Scandinavian Journal of Economics*, 88, 25-50.

Domeij, D. and M. Floden (2004), The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased, manuscript.

Economic Report of the President (2004), (United States Government Printing Office, Washington).

Feldstein, M. (1996), The Missing Piece in Policy Analysis: Social Security Reform, *American Economic Review,* 85(2), 1-14.

Feldstein, M. and A. Samwick (1992), Social Security Tax Rules and Marginal Tax Rates, *National Tax Journal,* 45, 1-22.

Fernandes, A. and C. Phelan (2000), A Recursive Formulation for Repeated Agency with History Dependence, *Journal of Economic Theory*, 91, 223-247.

Golosov, M., Kocherlakota, N. and A. Tsyvinski (2003), Optimal Indirect and Capital Taxation, *Review of Economic Studies*, 70, 569-88.

Golosov, M. and A. Tsyvinski (2004), Designing Optimal Disability Insurance: A Case for Asset Testing, NBER Working Paper No.10792.

Gouveia, M. and R. Strauss (1994), Effective Federal Income Tax Functions: An Exploratory Analysis, *National Tax Journal*, 47, 317- 39.

Green, E. (1987), Lending and the Smoothing of Uninsurable Income, in *Contractual Arrangements for Intertemporal Trade*, editors N. Wallace and E. Prescott, (Minneapolis: University of Minnesota Press).

Huang, H., Imrohoroglu, S. and T. Sargent (1997), Two Computations to Fund Social Security, *Macroeconomic Dynamics,* 1, 7-49.

Heathcote, J., Storresletten, K. and G. Violante (2004), The Cross-Sectional Implications of Rising Wage Inequality in the United States, manuscript.

Hubbard, G. and K. Judd (1987), Social Security and Individual Welfare: Precautionary Savings, Liquidity Constraints and the Payroll Tax, *American Economic Review*, 77, 630- 46.

Hopenhayn, H. and J. Nicolini (1998), Efficient Unemployment Insurance, *Journal of Political Economy,* 105, 412- 38.

Huggett, M. and G. Ventura (1999), On the Distributional Effects of Social Security Reform, *Review of Economic Dynamics*, 2, 498- 531.

Imrohoroglu, A., Imrohoroglu, S. and D. Joines (1995), A Life Cycle Analysis of Social Security, *Economic Theory*, 6, 83-114.

Imrohoroglu, A., Imrohoroglu, S. and D. Joines (2000), Computational Models of Social Security: A Survey, in *Computational Methods for the Study of Dynamic Economies*, R. Marimon and A. Scott, eds., (Oxford University Press, Oxford).

Kocherlakota, N. (2003), Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation, Federal Reserve Bank of Minneapolis Staff Report.

Krueger, D. and F. Kubler (2003), Pareto Improving Social Security Reform when Financial Markets are Incomplete!?, manuscript.

Lindbeck, A. and M. Persson (2003), The Gains from Pension Reform, *Journal of Economic Literature*, XLI, 74-112.

Lucas, R. (2003), Macroeconomic Priorities, *American Economic Review,* 93, 1-14.

Mas-Colell, A. and X. Vives (1993), Implementation in Economies with a Continuum of Agents, *Review of Economic Studies*, 60, 613- 29.

Mas-Colell, A., Whinston, M. and J. Green (1995), Microeconomic Theory, (Oxford University Press, Oxford).

Mirrlees, J. (1971), An Exploration into the Theory of Optimum Income Taxation, *Review of Economic Studies*, 38, 175- 208.

Nishiyama, S. and K. Smetters (2004), Does Social Security Privatization Produce Efficiency Gains?, manuscript.

Prescott, E. (2003), Why Do Americans Work So Much More Than Europeans?, Staff Report 321, Federal Reserve Bank of Minneapolis.

Press, W., Teukolsky, S., Vetterling, W., and B. Flannery (1994), *Numerical Recipes in Fortran: The Art of Scientific Computing,* 2nd ed., Cambridge University Press.

Salanie, B. (1997), The Economics of Contracts, (MIT Press, Cambridge).

Siegel, J. (2002), Stocks for the Long Run, Third Edition, (McGraw Hill, New York).

Social Security Handbook (2003), available at www.socialsecurity.gov.

Spear, S. and S. Srivastava (1987), On Repeated Moral Hazard with Discounting, *Review of Economic Studies*, 54, 599-617.

Storesletten, K., Telmer, C. and A. Yaron (1999), The Risk-Sharing Implications of Alternative Social Security Arrangements, *Carnegie-Rochester Conference Series on Public Policy,* 50, 213- 59.

Tauchen, G. (1986), Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions, *Economics Letters*, 20, 177-81.

Thomas, J. and T. Worrall (1990), Income Fluctuations and Asymmetric Information: An Example of a Repeated Principal-Agent Problem, *Journal of Economic Theory*, 51, 367-90.

# A Appendix: Computational Methods

Appendix A contains three sections. Section A.1 provides theory for computing solutions to the private information planning problem. Section A.2 describes our methods for computing solutions to the planning problem and the US social security problem. FORTRAN programs that compute solutions to these problems are available upon request. Section A.3 proves all Theorems from section A.1.

## A.1 Private Information Planning Problem: Theory

Theory for analyzing the private information planning problem is laid out in three steps. Step 1 states a dual problem with the feature that solutions to the dual problem are solutions to the original planning problem. Step 2 provides an equivalent formulation of incentive compatibility that is useful for a recursive statement of the dual problem. Step 3 formulates the dual problem as a dynamic programming problem and indicates how to further simplify this problem for computational purposes. Throughout this Appendix, we specialize the labor-productivity function to be $\omega(s_j, j) = s_j$ solely to simplify expressions.

### A.1.1 Primal and Dual Problems

Primal Problem: $\max E[\sum_j \beta^{j-1} u(c_j, l_j)]$

subject to (1) $(c, l)$ is IC and (2) $E[\sum_j (c_j - s_j l_j)/(1+r)^{j-1}] \leq Cost$

Dual Problem: $\min E[\sum_j (c_j - s_j l_j)/(1+r)^{j-1}]$

subject to (1) $(c, l)$ is IC and (2) $E[\sum_j \beta^{j-1} u(c_j, l_j)] \geq u^*$

*Theorem A1: Assume $u(c,l) = u(c) + v(l)$, $u(c)$ is continuous on $R^1_{++}$, $u(c)$ is strictly increasing and $u(0) = -\infty$. If $(c,l)$ solves the Dual Problem, given $u^* > -\infty$, then $(c,l)$ solves the Primal Problem, given $Cost \equiv E[\sum_j \frac{(c_j - s_j l_j)}{(1+r)^{j-1}}]$.*

Proof: See Appendix A.3

Theorem A2 provides conditions which are equivalent to the incentive compatibility conditions.[22]

*Theorem A2: For the case of independent shocks, $(c,l)$ is IC if and only if $\exists \{w_j(s^{j-1})\}_{j=2}^{J+1}$ such that restrictions (a)-(b) hold:*

*(a)* $u(c_j(s^{j-1}, s_j), l_j(s^{j-1}, s_j)) + \beta w_{j+1}(s^{j-1}, s_j) \geq$
$u(c_j(s^{j-1}, s'_j), l_j(s^{j-1}, s'_j)(s'_j/s_j)) + \beta w_{j+1}(s^{j-1}, s'_j), \forall (s^{j-1}, s_j), \forall s'_j$

*(b)* $w_j(s^{j-1}) = E[u(c_j(s^j), l_j(s^j)) + \beta w_{j+1}(s^j) | s^{j-1}], \forall j$ *and* $w_{J+1}(s^J) = 0$

*where $s^j$ denotes the history of truthful reports up to period $j$.*

Proof: See Appendix A.3.

[22] These results are adaptations of Green (1987, Lemma 1-2).

### A.1.2 Recursive Formulation of the Dual Problem

Following Green (1987) and Spear and Srivastava (1987), a recursive formulation of the Dual Problem is provided below for the case of temporary shocks. The function $C_j(w)$ is the minimum expected discounted cost of obtaining utility $w$. The notation $(c_i, l_i, w_i)$ describes period consumption, labor and future utility delivered when shock $s_i$ occurs and the agent tells the truth. There are $N$ shock values that occur with probability $(\pi_1, ..., \pi_N)$. Shocks are ordered so that $s_1 < s_2 < ... < s_N$.

Recursive Dual Problem: $C_j(w) = \min \sum_i [c_i - l_i s_i + (1+r)^{-1} C_{j+1}(w_i)]\pi_i$

subject to

(i) $w = \sum_i [u(c_i, l_i) + \beta w_i]\pi_i$,

(ii) $u(c_i, l_i) + \beta w_i \geq u(c_j, l_j(s_j/s_i)) + \beta w_j, \forall i, j$

Theorem A3 establishes some basic properties of the incentive constraints.[23] These properties are used to simplify the computation of the Recursive Dual Problem. The following compact notation is used: $C_{ij} \equiv u(c_i, l_i) + \beta w_i - [u(c_j, l_j(s_j/s_i)) + \beta w_j]$. $C_{ii-1} \geq 0$ is called a local downward incentive constraint, whereas $C_{i-1i} \geq 0$ is called a local upward incentive constraint. Theorem A3 says that (a) the local upward and downward constraints convey all the IC restrictions (Theorem A3(ii)), (b) if all the local downward constraints bind then all local upward constraints also hold (Theorem A3(iii)) and (c) in a solution to the Recursive Dual Problem all local downward constraints bind (Theorem A3(iv)). Theorem A3(i) also delivers the standard insight that incentive compatibility alone implies that "earnings" or "output" $l_i s_i$ increases as the shock index $i$ increases. This relies on additive separability.

*Theorem A3: Assume $u(c, l) = u(c) + v(l)$, $u$ and $v$ are strictly concave, $u$ is increasing, $v$ is decreasing and that shocks are independent. Then*

*(i) Incentive compatibility implies that $l_i s_i$ is increasing in $i$.*

*(ii) $C_{ii-1}, C_{i-1i} \geq 0, i = 2, ..., N$ imply that $C_{ij} \geq 0 \quad \forall i, j$.*

*(iii) $C_{ii-1} = 0, i = 2, ..., N$ imply that $C_{i-1,i} \geq 0, i = 2, ..., N$ and that $C_{i-1,i} > 0$ whenever $l_i s_i > l_{i-1} s_{i-1}$.*

*(iv) In a solution to the Recursive Dual Problem all local downward constraints bind.*

Proof: See Appendix A.3.

To compute solutions to the Recursive Dual Problem it is useful to solve two subproblems: DP 1 and DP 2. These problems reduce the dimensionality of the choice variables by making use of additive separability of the objective. Dimensionality can be further reduced by solving DP 1' in place of DP 1. DP 1' solves out for utility $z_i$ in terms of promised utility $w$ and the labor plan $(l_1, ..., l_N)$. This uses the fact, established in Theorem A3, that in a solution to the

[23] Theorem A3 is parallel to results which hold w/o a labor-leisure decision (e.g. Thomas and Worrall (1990)) and to results in the literature following Mirrlees (1971).

recursive dual problem all local downward constraints hold with equality and that when these constraints hold with equality they imply all the restrictions of incentive compatibility.

Subproblems:

(DP 1) $C_j(w) = \min \sum_i [-l_i s_i + \hat{C}_j(z_i)]\pi_i$
(1) $w = \sum_i [v(l_i) + z_i]\pi_i$
(2) $v(l_i) + z_i \geq v(l_j(s_j/s_i)) + z_j, \forall i, j$

(DP 1') $C_j(w) = \min \sum_i [-l_i s_i + \hat{C}_j(f_i(l_1, ..., l_N; w)))]\pi_i$

(DP 2) $\hat{C}_j(z) = \min_{\{(c,w'):z=u(c)+\beta w'\}} c + (1+r)^{-1} C_{j+1}(w')$

The functions $z_i = f_i(l_1, ..., l_N; w)$ in problem DP 1' are constructed in the two equations below. The first equation holds for $i > 1$ by repeated substitutions from the downward IC constraint. This equation says that promised utility $z_i$ to a person with shock $i$ is the utility to the person with the lowest shock $z_1$ plus the sum of the utility differences when one lies downward one shock. These utility differences are positive by Thm. A.3(i). The second equation holds by substituting the first equation into the promise keeping constraint. This then states $z_1$ in terms of the labor choices and promised utility $w$. These two equations define the functions $z_i = f_i(l_1, ..., l_N; w)$ for $i = 1, ..., N$.

$$z_i = z_1 + \sum_{j=2}^{i} [v(l_{j-1}(s_{j-1}/s_j)) - v(l_j)]$$

$$z_1 = w - \sum_{i=1}^{N} v(l_i)\pi_i - \sum_{i=2}^{N} [\sum_{j=2}^{i} (v(l_{j-1}(s_{j-1}/s_j))) - v(l_j)]\pi_i$$

## A.2 Computation

### A.2.1 Social Security Problem

The social security problem is stated below as a dynamic programming problem. This involves reformulating the present value budget constraint as a sequence of budget constraints where resources are transfered across periods with a risk-free asset. Risk-free asset holding must then always lie above period and shock specific borrowing limits: $\underline{a}_j(s)$.[24] The state variable is $(a, s, z)$ where $a$ is asset holdings, $s$ is the period productivity shock and $z$ is average past earnings. The functions $T_j$ and $F_j$ describe the tax system and the law of motion for average past earnings. Labor productivity is a Markov process with transition probability $\pi(s'|s)$.

$V_j(a, s, z) = \max_{(c,l,a')} u(c, l) + \beta \sum_{s'} V_{j+1}(a', s', z')\pi(s'|s)$
(1) $c + a' \leq a(1+r) + \omega(s, j)l - T_j(a, z, \omega(s, j)l)$
(2)$c \geq 0, a' \geq \underline{a}_j(s); l \in [0, 1]$
(3) $z' = F_j(z, \omega(s, j)l)$

[24]These limits are the maximum present value of labor earnings plus social security benefits in the worst labor-productivity history. This assumes that one can borrow against future social security benefits.

This problem is solved computationally by backwards induction. The value function $V_j(a, s, z)$ is computed at selected grid points $(a, s, z)$ by solving the right-hand-side of Bellman's equation. We use the simplex method (see Press et al (1994)). Evaluating the right-hand-side of Bellman's equation involves a bi-linear interpolation of the function $V_{j+1}(a', s', z')$ over the two continuous variables $(a', z')$. We set the borrowing limit to a fixed value $\underline{a}$ in each period. We then relax this value so that it is not binding. This is a device for imposing period and state specific limits $\underline{a}_j(s)$. To use this device, penalties are imposed for states and decisions implying negative consumption.[25]

We compute ex-ante, expected utility $V^{ss}$ and the expected cost of running the social security system, denoted $Cost$, by simulation, under the assumption that an agent starts out with no assets. Specifically, we draw a large number (10,000) of lifetime labor-productivity profiles, compute realized utility and realized cost for each profile, using the computed optimal decision rules, and then compute averages.

### A.2.2 Planning Problem

We describe how we compute the optimized value $V^{pp}$, given the value of $Cost$. The algorithm for the temporary shock case is presented first.

(DP 1') $C_j(w) = \min \sum_i [-l_i s_i + \hat{C}_j(f_i(l_1, ..., l_N; w)))]\pi_i$
(DP 2) $\hat{C}_j(z) = \min_{\{(c,w'):z=u(c)+\beta w'\}} c + (1+r)^{-1} C_{j+1}(w')$

ALGORITHM:

1. Set terminal value function on grid points $w \in \{w_1, ..., w_M\}$: $\hat{C}_J(w) \equiv u^{-1}(w)$
2. For each $w \in \{w_1, ..., w_M\}$, we use amoeba from Press et al (1994) to solve the right-hand-side of DP 1' to compute $C_j$. This involves a linear interpolation of $\hat{C}_j$.
3. Given $C_j$, compute $\hat{C}_{j-1}$ at grid points by solving DP 2. This is done by grid search.
4. Repeat steps 2-3 for all ages $j$ back to age 1.
5. Find $V^{pp}$ solving $C_1(V^{pp}) = Cost$. This is done by simulation using the optimal decision rules.

We now indicate how to compute $V^{pp}$ for the case of permanent shocks. The permanent shock problem is restated below.

$V^{pp} \equiv \max_{(l_j(s),c_j(s))} \sum_s [\sum_j \beta^{j-1}(u(c_j(s)) + v(l_j(s)))]P(s)$ s.t.
(i) $\sum_s [\sum_j (c_j(s) - l_j(s)s)/(1+r)^{j-1}]P(s) \leq Cost$

[25]The backward induction procedure takes as given a value for average earnings in the economy. This is used to determine the tax function $T_j$. Thus, an additional loop is needed so that guessed and implied values of average earnings coincide. To compute solutions, we use 500 evenly spaced grid points on assets $a$ and 25 grid points on average earnings $z$ over the interval $[0, e_{max}]$. Recall from section 4 that there are 5 shocks values $s$.

(ii) $\sum_j \beta^{j-1}(u(c_j(s)) + v(l_j(s)) \geq \sum_j \beta^{j-1}(u(c_j(s')) + v(l_j(s')s'/s)), \forall s, \forall s'$

We analyze a "relaxed" problem which is the same as the problem above except that we require that only the local downward incentive constraints hold rather than all the incentive constraints. The local downward incentive constraints are the constraints implied by requiring that truth telling from shock $s_i$ dominate claiming to be one shock lower (i.e. claiming shock $s_{i-1}$). It is straightforward to show two results. First, in a solution to the relaxed problem all the local downward incentive constraints bind. Second, if an allocation $(c, l)$ has the property that all the local downward incentive constraints bind and $l_j(s)s$ is increasing in $s$ for all $j$, then all the incentive constraints hold. The argument for this follows the proof of similar claims in Theorem A3. Our computational strategy is therefore to compute solutions to the relaxed problem AND to verify ex-post that in all periods $l_j(s)s$ is increasing in $s$.[26]

We compute solutions to the relaxed problem by solving the equivalent problem below. This equivalent problem is useful for computational purposes as it reduces the dimension of the control variables by substituting out all binding constraints. This equivalence follows from two observations. First, additive separability of $u(c, l)$ implies that consumption is chosen in the relaxed problem without distortion. Using this fact, maximization could then be done over labor and the lifetime utility of consumption $u(s)$. This eliminates the choice of consumption from the problem. The relevant cost constraint is written in the second equation below, where $COST(u(s))$ is a known function, derived from the first order conditions to the relaxed problem, describing the minimum resource cost of obtaining lifetime utility of consumption $u(s)$.[27] Second, we also eliminate maximizing over $u(s)$ by expressing $u(s) = g(l, s, Cost)$ as a function of the labor plan and other data. To do this, we solve for $u(s)$ from the relevant binding local downward incentive constraints and the binding cost constraint. The third and fourth equations below are intermediate steps towards computing $u(s) = g(l, s, Cost)$. They use the fact that shocks are ordered so that $s_1 < s_2 < ... < s_N$.

$$\max_{(l)} \sum_s [\sum_j \beta^{j-1} v(l_j(s)) + g(l, s, cost)] P(s)$$

$$\sum_s [COST(u(s)) - \sum_j s l_j(s)/(1+r)^{j-1}] P(s) = Cost$$

$$\sum_j \beta^{j-1} v(l_j(s_i)) + u(s_i) = \sum_j \beta^{j-1} v(l_j(s_{i-1}) s_{i-1}/s_i)) + u(s_{i-1})$$

$$u(s_n) = u(s_1) + \sum_{i=2}^{n} [\sum_j \beta^{j-1} v(l_j(s_{i-1}) s_{i-1}/s_i)) - \sum_j \beta^{j-1} v(l_j(s_i))]$$

We use amoeba from Press et al (1994) to solve the relaxed problem. This involves maximizing over labor choices $(l_1(s), ..., l_{R-1}(s))$. These choices lie in an $R - 1 \times N$ dimensional

[26]In actual computations we verify that $l_j(s)\omega(s, j)$ is increasing in $s$, since $\omega(s, j) = s$ is used in Appendix A soley to reduce notational clutter.

[27]When $\beta(1+r) = 1$, $COST(u(s))$ has a simple form as consumption is constant. When $\beta < 1$ and $r > 0$ then $COST(u(s)) = u^{-1}[(1-\beta)u(s)/(1-\beta^J)][1 - (1/(1+r))^J](1+r)/r$.

space as there are $R-1$ labor periods and $N$ possible permanent shocks. Each evaluation of the objective requires the computation of $g(l,s,Cost)$. This involves finding a value $u(s_1)$ solving the second and fourth equations above, given $(l_1(s),...,l_{R-1}(s))$ and Cost.

## A.3 Proofs of Theorems A1-3

*Theorem A1: Assume $u(c,l)=u(c)+v(l)$, $u(c)$ is continuous on $R^1_{++}$, $u(c)$ is strictly increasing and $u(0)=-\infty$. If $(c,l)$ solves the Dual Problem, given $u^* > -\infty$, then $(c,l)$ solves the Primal Problem, given $Cost \equiv E[\sum_j \frac{(c_j - s_j l_j)}{(1+r)^{j-1}}]$.*

Proof: Suppose not. Thus, there exists $(\bar{c},\bar{l})$ that is IC and costs no more than $(c,l)$ but that delivers strictly more expected utility than $(c,l)$. Construct $(c^*,l^*)$ that satisfies constraints (1)-(2) in the Dual Problem but that delivers strictly lower cost than $(c,l)$.

Set $l^*_j \equiv \bar{l}_j, \forall j$ and $c^*_j \equiv \bar{c}_j, \forall j \geq 2$. Set $c^*_1(s)$ to solve $u(c^*_1(s)) = u(\bar{c}_1(s)) - \epsilon$. Thus, $c^*_1(s)$ produces a uniform decrease in utility in period 1 of $\epsilon > 0$. If $\bar{c}_1(s) > 0, \forall s$, then by continuity there exists $\epsilon > 0$ such that $c^*_1(s) \geq 0, \forall s$ and $E[\sum_j \beta^{j-1} u(c^*_j, l^*_j)] \geq u^*$. Clearly, $u(0)=-\infty$ implies that $\bar{c}_1(s) > 0$. Since $(\bar{c},\bar{l})$ is IC and the utility decrease is uniform regardless of reports, $(c^*,l^*)$ is also IC. This is a contradiction since $(c^*,l^*)$ costs strictly less than $(c,l)$. □

*Theorem A2: For the case of independent shocks, $(c,l)$ is IC if and only if $\exists \{w_j(s^{j-1})\}_{j=2}^{J+1}$ such that restrictions (a)-(b) hold:*

*(a)* $u(c_j(s^{j-1},s_j), l_j(s^{j-1},s_j)) + \beta w_{j+1}(s^{j-1},s_j) \geq$
$u(c_j(s^{j-1},s'_j), l_j(s^{j-1},s'_j)(s'_j/s_j)) + \beta w_{j+1}(s^{j-1},s'_j), \forall (s^{j-1},s_j), \forall s'_j$

*(b)* $w_j(s^{j-1}) = E[u(c_j(s^j), l_j(s^j)) + \beta w_{j+1}(s^j) | s^{j-1}], \forall j$ *and* $w_{J+1}(s^J) = 0$

*where $s^j$ denotes the history of truthful reports up to period $j$.*

Proof:

($\Rightarrow$) Backward induction on restriction (b) defines the function $w_{j+1}$ uniquely. Substitute $w_{j+1}$ into restriction (a). The resulting inequality is then a direct implication of $(c,l)$ being IC. Specifically, it is implied by truth telling being superior to a feasible report $\sigma$ where one reports truthfully at all ages and histories except $(s^{j-1},s_j)$ where the report is $s'_j$ rather than $s_j$. [Independence used here.]

($\Leftarrow$) Suppose not. Then restriction (a)-(b) hold but there is a report $\sigma$ that strictly improves over truth telling, given $(c,l)$. Let $\sigma$ have the smallest number of false reports at distinct age-histories $s^j$ among those report functions $\sigma$ that strictly improve over truth telling. Such a $\sigma$ exists since the number of age-histories is finite. Choose $j$ as large as possible so that $\sigma$ involves a false report (i.e. $\sigma_j(s^j) \neq s_j$) at some $s^j$. Then restriction (a)-(b) implies that given that $\sigma$ has been used in the past, telling the truth in period j and subsequently leads to at least as much conditional expected utility at $s^j$ as using $\sigma$. Thus, there is another feasible report function that strictly improves over truth telling and that has a smaller number of false reports. Contradiction. □

*Theorem A3: Assume $u(c,l)=u(c)+v(l)$, $u$ and $v$ are strictly concave, $u$ is increasing, $v$ is decreasing and that shocks are independent. Then*

*(i) Incentive compatibility implies that $l_i s_i$ is increasing in $i$.*

*(ii)* $C_{ii-1}, C_{i-1i} \geq 0, i = 2, ..., N$ *imply that* $C_{ij} \geq 0 \quad \forall i, j$.

*(iii)* $C_{ii-1} = 0, i = 2, ..., N$ *imply that* $C_{i-1,i} \geq 0, i = 2, ..., N$ *and that* $C_{i-1,i} > 0$ *whenever* $l_i s_i > l_{i-1} s_{i-1}$.

*(iv) In a solution to the Recursive Dual Problem all local downward constraints bind.*

Proof:

(i) Assume that it is feasible to claim to have received shock $i$ when one has shock $i-1$. If not, then $l_i s_i \geq l_{i-1} s_{i-1}$ holds trivially. Thus, we have that $C_{ii-1}, C_{i-1i} \geq 0$. Adding these inequalities and using the fact that $u(c, l) = u(c) + v(l)$ implies the first equation below. The second equation rearranges the first. The second equation and $v$ concave then implies that $l_i s_i \geq l_{i-1} s_{i-1}$ must hold.

$$v(l_i) - v(l_{i-1} s_{i-1}/s_i) \geq v(l_i s_i/s_{i-1}) - v(l_{i-1})$$

$$v(l_i s_i/s_i) - v(l_{i-1} s_{i-1}/s_i) \geq v(l_i s_i/s_{i-1}) - v(l_{i-1} s_{i-1}/s_{i-1})$$

(ii) Show first that $C_{ij} \geq 0, \forall j < i$. As a first step show that $C_{ii-2} \geq 0$. This follows from the three lines below. The first line is $C_{i-1i-2} \geq 0$. The second line follows from line one and the fact that $v(l_{i-1} s_{i-1}/s) - v(l_{i-2} s_{i-2}/s)$ increases as $s$ increases for $s \geq s_{i-1}$. The last fact holds since $l_i s_i$ increases as $i$ increases (Thm. A3(i)) and since $v$ is concave. Line three follows from line two and $C_{ii-1} \geq 0$.

$$u(c_{i-1}, l_{i-1}) + w_{i-1} \geq u(c_{i-2}, l_{i-2} s_{i-2}/s_{i-1}) + w_{i-2}$$

$$u(c_{i-1}, l_{i-1} s_{i-1}/s_i) + w_{i-1} \geq u(c_{i-2}, l_{i-2} s_{i-2}/s_i) + w_{i-2}$$

$$u(c_i, l_i) + w_i \geq u(c_{i-1}, l_{i-1} s_{i-1}/s_i) + w_{i-1} \geq u(c_{i-2}, l_{i-2} s_{i-2}/s_i) + w_{i-2}$$

To show that $C_{ij} \geq 0$ holds for all $j < i$, proceed by induction repeating the three steps above, where the first step is the induction step.

It remains to show that $C_{ij} \geq 0, \forall j > i$ if any of these upward lies are feasible. As a first step show that $C_{ii+2} \geq 0$. This follows from the three lines below for essentially the same reasons as in the argument above. The remainder of the proof follows by an induction which is parallel to that given above.

$$u(c_{i+1}, l_{i+1}) + w_{i+1} \geq u(c_{i+2}, l_{i+2} s_{i+2}/s_{i+1}) + w_{i+2}$$

$$u(c_{i+1}, l_{i+1} s_{i+1}/s_i) + w_{i+1} \geq u(c_{i+2}, l_{i+2} s_{i+2}/s_i) + w_{i+2}$$

$$u(c_i, l_i) + w_i \geq u(c_{i+1}, l_{i+1} s_{i+1}/s_i) + w_{i+1} \geq u(c_{i+2}, l_{i+2} s_{i+2}/s_i) + w_{i+2}$$

(iii) Implied by $v$ being strictly concave.

(iv) (Rough argument) Suppose not. Let $(c_i, l_i, w_i)$ be a solution in which a downward constraint is not binding. Construct $(c_i^*, l_i^*, w_i^*)$ so that labor and future utility are the same as before but consumption is different. Squeeze the consumption distribution so that (a) mean consumption is lower, (b) all downward constraints still hold and (c) mean $u(c)$ unchanged. This lowers the objective and satisfies all constraints. Contradiction.

[Note: Argument involves lowering consumption in some state. Thus, one needs strictly positive consumption. A sufficient condition for this to hold for states $w > -\infty$ is $u(0) = -\infty$.]

$\square$

# A Appendix: Tax Implementation

The following claims are useful in tax implementing solutions to the planning problem.

*Claim 1: Assume that $u(c_j, l_j) = u(c_j) + v(l_j)$, where $u$ and $v$ are strictly concave and differentiable. Assume also that shocks are permanent. If $(c, l)$ is an interior solution to the planning problem, then the following hold:*

*(i)* $u'(c_j(s))/\beta u'(c_{j+1}(s)) = 1 + r, \forall s$, *for* $j = 1, ..., J - 1$.

*(ii)* $-v'(l_j(s))/u'(c_j(s)) = B_j(s)\omega(s, j), \forall s$, *for* $j = 1, ..., R - 1$, *where* $B_j(s) = 1$ *for* $s = N$ *and* $B_j(s) < 1$ *otherwise.*

Proof: Claim 1 follows from reorganizing the first order conditions (i.e. $dL/dc_j(s) = 0$ and $dL/dl_j(s) = 0$) associated with the Lagrangean function below. [To use this Lagrangean, one needs a proof that only the local downward constraints are relevant.] In what follows, $s^+$ denotes one higher shock than $s$, whereas $s^-$ denotes one lower shock than $s$.

$$L = \sum_s \sum_j \beta^{j-1} u(c_j(s), l_j(s))P(s) + \lambda[Cost - \sum_s \sum_j \frac{(c_j(s) - \omega(s, j)l_j(s))P(s)}{(1 + r)^{j-1}}]$$

$$+ \textstyle\sum_{s=2}^{N} \gamma(s) \sum_j \beta^{j-1}[u(c_j(s), l_j(s)) - u(c_j(s^-), l_j(s^-)\frac{\omega(s^-,j)}{\omega(s,j)})]$$

Claim 1(i) follows directly from $dL/dc_j(s) = 0$ and additive separability.

Claim 1(ii) follows from $dL/dc_j(s) = 0$ and $dL/dl_j(s) = 0$. These conditions imply that $-v'(l_j(s))/u'(c_j(s)) = B_j(s)\omega(s, j)$. It is straightforward to see that $B_j(s) = 1$ for $s = N$. For $s < N$, the term $B_j(s)$ is given below. What remains to be shown is that $B_j(s) < 1$ for $s < N$. This follows from two observations. First, the Lagrange multiplier $\gamma(s)$ is strictly positive. Second, the term multiplying $\gamma(s^+)$ in the denominator is less than one. The first observation can be proven by contradiction. The second follows from the fact that $\omega(s, j)/\omega(s^+, j) < 1$ and that $v$ is strictly concave.

$$B_j(s) \equiv \frac{[P(s) + \gamma(s) - \gamma(s^+)]}{[P(s) + \gamma(s) - \gamma(s^+)\frac{v'(l_j(s)\frac{\omega(s,j)}{\omega(s^+,j)})\frac{\omega(s,j)}{\omega(s^+,j)}}{v'(l_j(s))}]}$$

□

Could try to say something about intertemporal labor distortion. Using $dL/dl_j(s) = 0$ it is true that $v'(l_j(s))/\beta v'(l_{j+1}(s)) = A_j(s)(1+r)$, where $A_j(s)$ is given below. Highest shock guy has labor moving with labor productivity (assuming $\beta(1 + r) = 1$). Thus, in general $a_j(s) \neq 1$.

$$A_j(s) \equiv \frac{[P(s) + \gamma(s) - \gamma(s^+)\frac{v'(l_{j+1}(s)\frac{\omega(s,j+1)}{\omega(s^+,j+1)})\frac{\omega(s,j+1)}{\omega(s^+,j+1)}}{v'(l_{j+1}(s))}]}{[P(s) + \gamma(s) - \gamma(s^+)\frac{v'(l_j(s)\frac{\omega(s,j)}{\omega(s^+,j)})\frac{\omega(s,j)}{\omega(s^+,j)}}{v'(l_j(s))}]}$$

The goal is to establish conditions under which solutions to the planning problem are tax implementable. The definition below provides a notion of tax implementation. Claim 2 then provides a first result on tax implementation. Claim 2 can be paraphrased as saying that if one restricts earnings profiles to those that occur in a realization of a solution $(c^*, l^*)$, then one can impose taxes on earnings histories and implement $(c^*, l^*)$ via simple borrowing and lending markets.[28]

*Definition: An allocation $(c^*, l^*)$ is tax implementable if there exists a function $Tax : \Gamma \to R$ such that $(c_j^*(s), l_j^*(s))$ is a solution to the Tax Problem below for all $s \in S$.*

*Tax Problem:* $\max_{(c_j,l_j)\in\Gamma(Tax,s)} \sum_j \beta^{j-1}u(c_j, l_j)$

$\Gamma(Tax, s) = \{\{c_j, l_j\}_{j=1}^J : \sum_j \frac{(c_j-\omega(s,j)l_j)}{(1+r)^{j-1}} \leq -Tax(\{\omega(s, j)l_j\})$ *and* $\{\omega(s, j)l_j\} \in \Gamma\}$

$\Gamma = \{\{z_j\}_{j=1}^J : \exists s \in S$ *such that* $z_j = l_j^*(s)\omega(s, j), j = 1, ..., J\}$.

*Claim 2: Assume that the assumptions of Claim 1 hold. If $(c^*, l^*)$ is an interior solution to the planning problem, then $(c^*, l^*)$ is tax implementable. Furthermore, the following tax function does the job: $Tax(\{z_j\}) = \sum_j \frac{(\omega(s,j)l_j^*(s))-c_j^*(s))}{(1+r)^{j-1}}$ when $z_j = \omega(s, j)l_j^*(s), \forall j$ for some $s \in S$.*

Proof: The proof is based on showing that the following two inequalities hold. The result then follows since by construction $(c_j^*(s), l_j^*(s)) \in \Gamma(Tax, s)$ for all $s \in S$.

$$\sum_j \beta^{j-1}u(c_j^*(s), l_j^*(s)) \geq \max_{\hat{s}\in S} \sum_j \beta^{j-1}u(c_j^*(\hat{s}), l_j^*(\hat{s})\frac{\omega(j,\hat{s})}{\omega(j,s)}) \geq \max_{(c_j,l_j)\in\Gamma(Tax,s)} \sum_j \beta^{j-1}u(c_j, l_j)$$

The leftmost inequality above follows since $(c^*, l^*)$ is IC. The rightmost inequality is implied by the fact that, setting labor earnings $(l_j^*(\hat{s})\omega(\hat{s}, j))$ to any profile allowed, consumption $(c_j^*(\hat{s}))$ solves the Tax problem. To establish this latter fact, note that because $u$ is concave the first order conditions and budget constraint are sufficient for a solution to the Tax problem, given labor earnings. Claim 1(i) then implies that the first order conditions for this problem hold when evaluated at $(c_j^*(\hat{s}))$. □

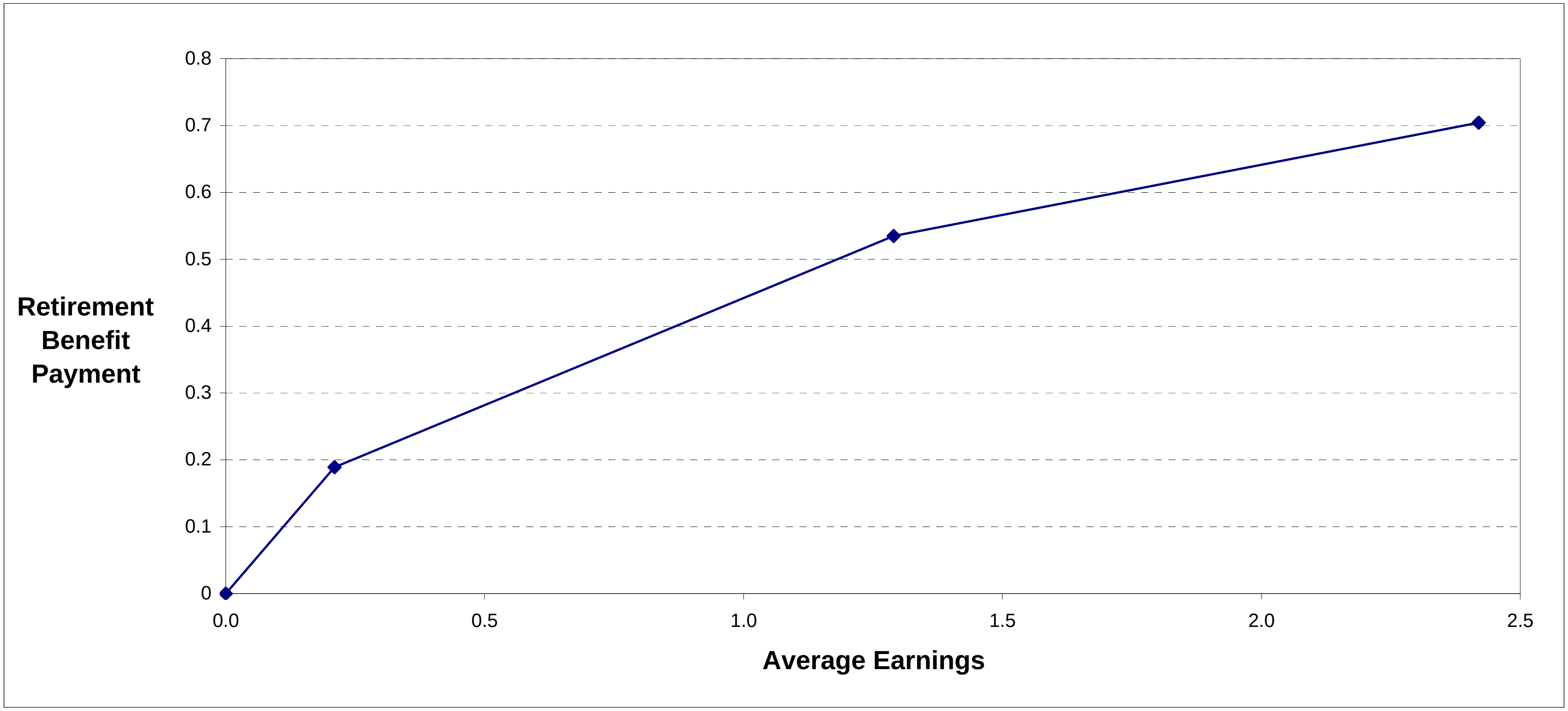
One could hope to extend the result in Claim 2. For example, one could hope to relax the domain restriction on the tax function or to say something more about the structure of the tax functions that implement solutions to the planning problem. Both extensions involve knowing something more about the labor allocations that solve the planning problem.

*Claim 3: Assume that the assumptions of Claim 1 hold. If $(c^*, l^*)$ is an interior solution to the planning problem, then $(c^*, l^*)$ is tax implementable. Furthermore, the tax function is a tax*

[28] If one were to specify the budget constraint in the Tax Problem as a sequence of budget restrictions, then the timing of tax collection would be unimportant. In this context only the net-present value of taxation along an earnings profile is important.

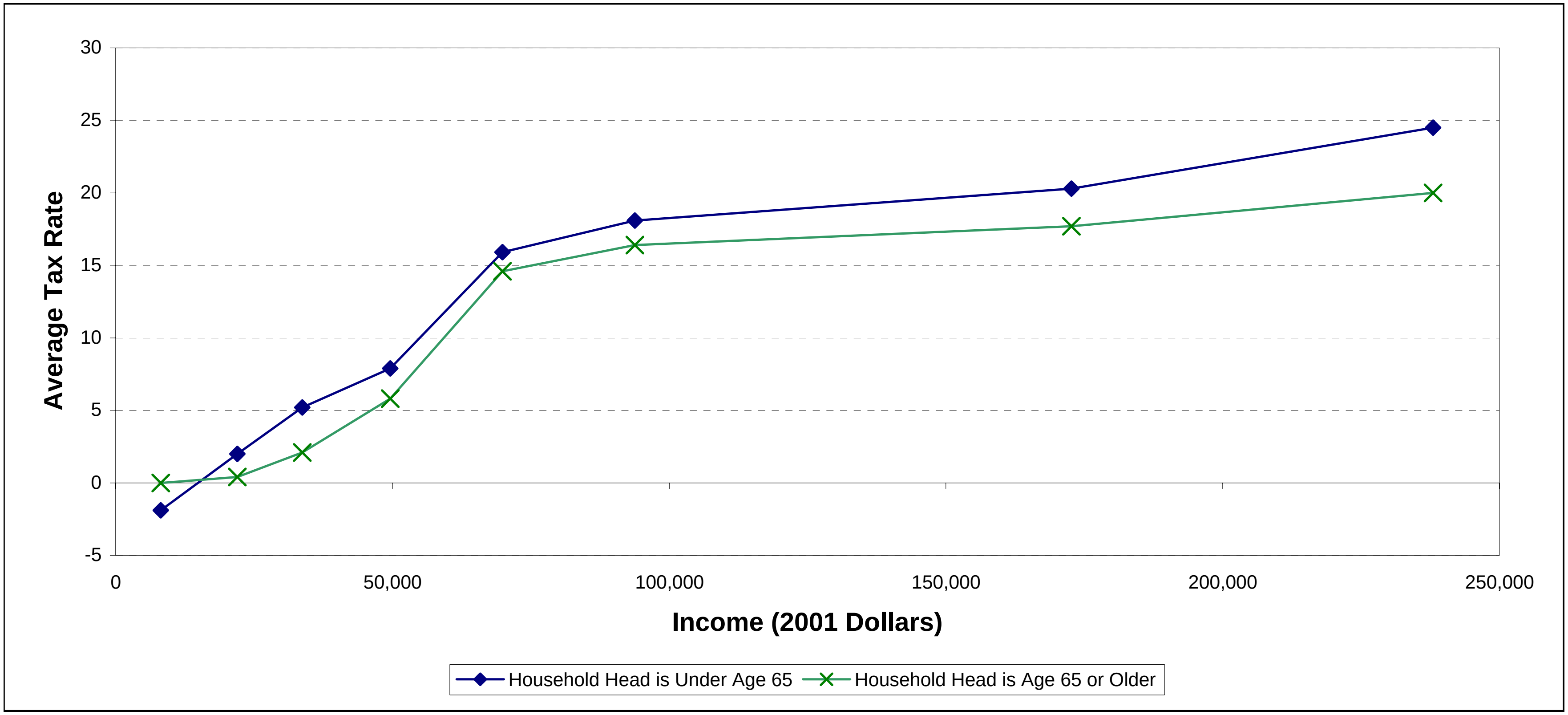
*on the present value of earnings rather than simply a function of the entire history of earnings. [Domain of present values??]*

## Figure 1: US Social Security Benefit Formula



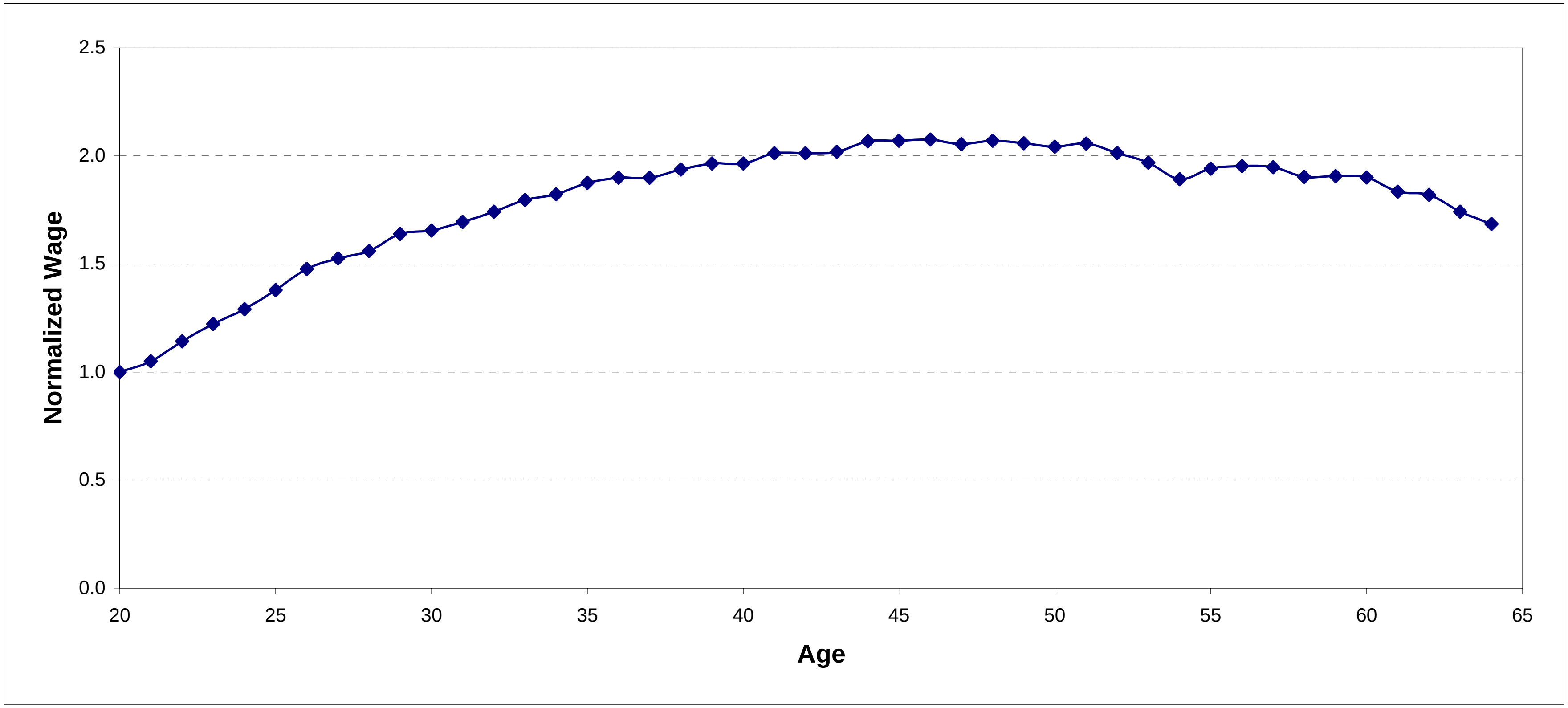

Average earnings and benefit payments are both expressed as a multiple of average economy wide earnings.

## Figure 2: Average Tax Rates



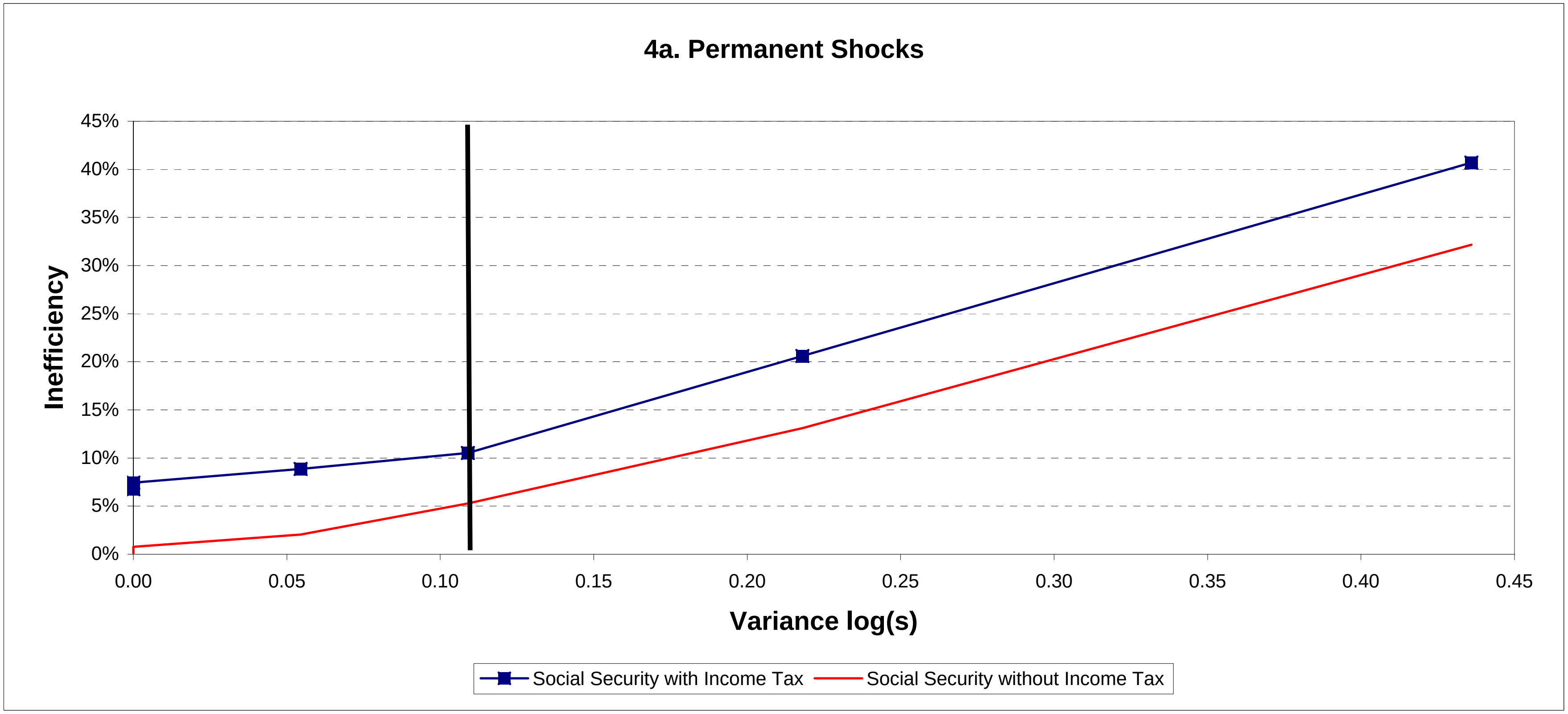

Source: Congressional Budget Office (2004)

## Figure 3: US Wage Profile



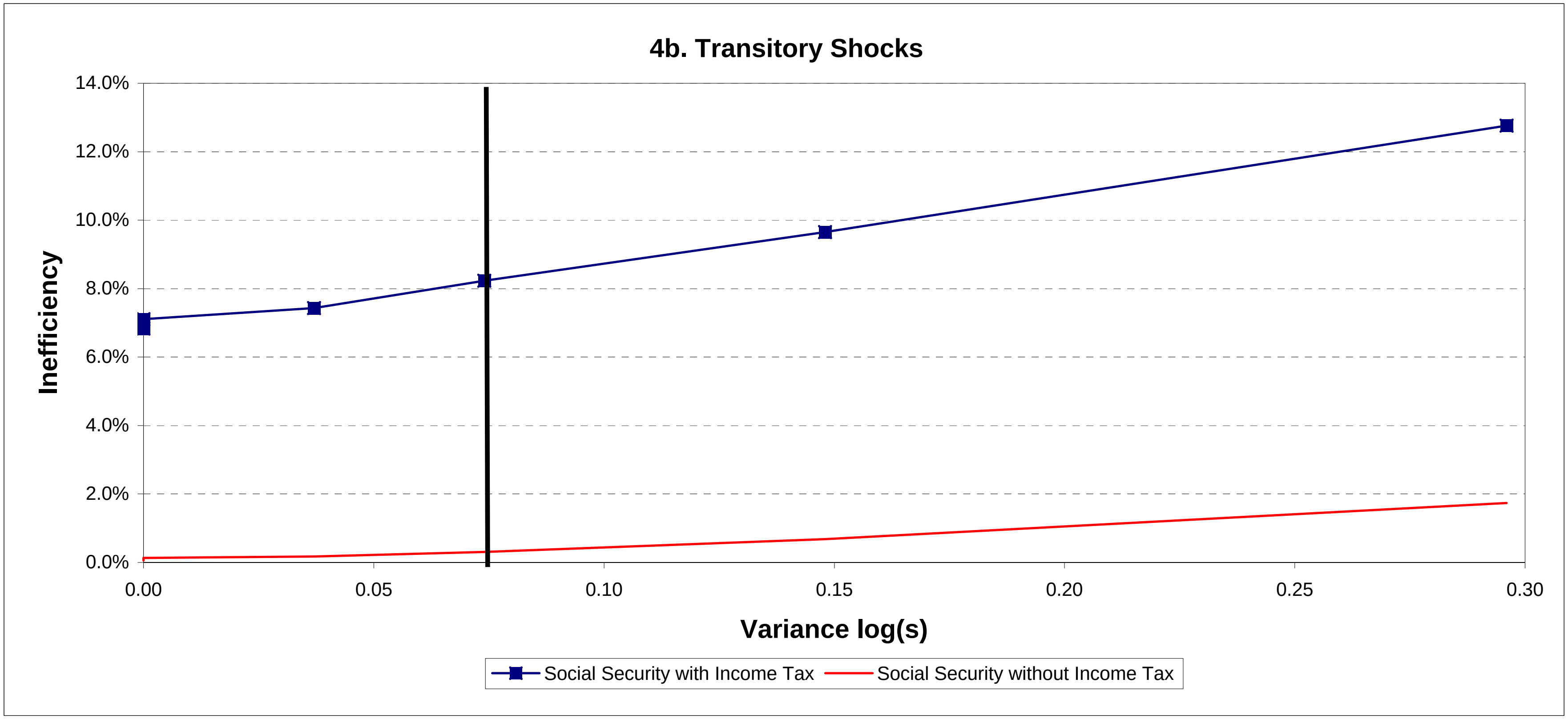

Source: Heathcote et al (2004). Panel Study of Income Dynamics data on wages for males from age 20 to 64.

# Figure 4: Inefficiency



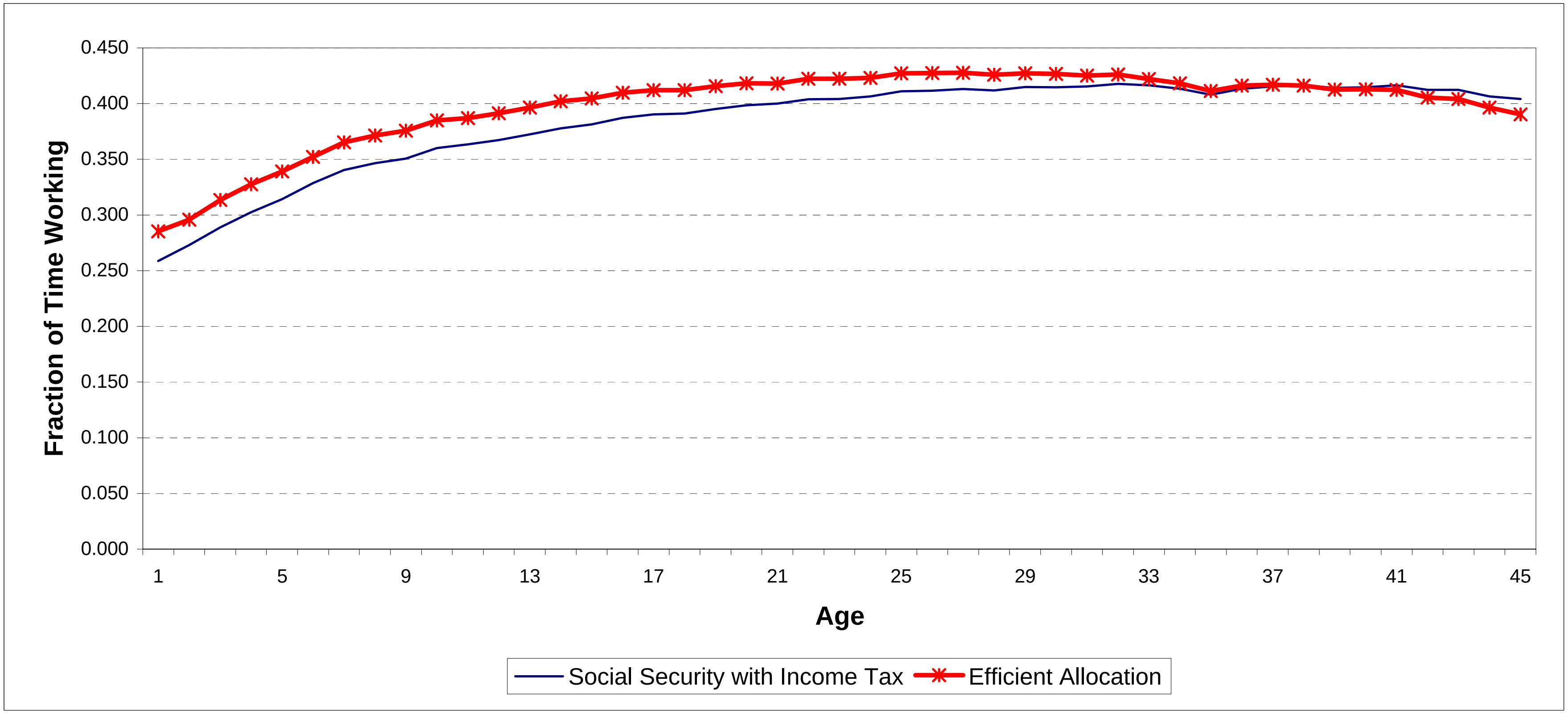

The bold vertical line in Figure 4 highlights the location of the point estimates of the variances described in the text.

## Figure 4: Inefficiency

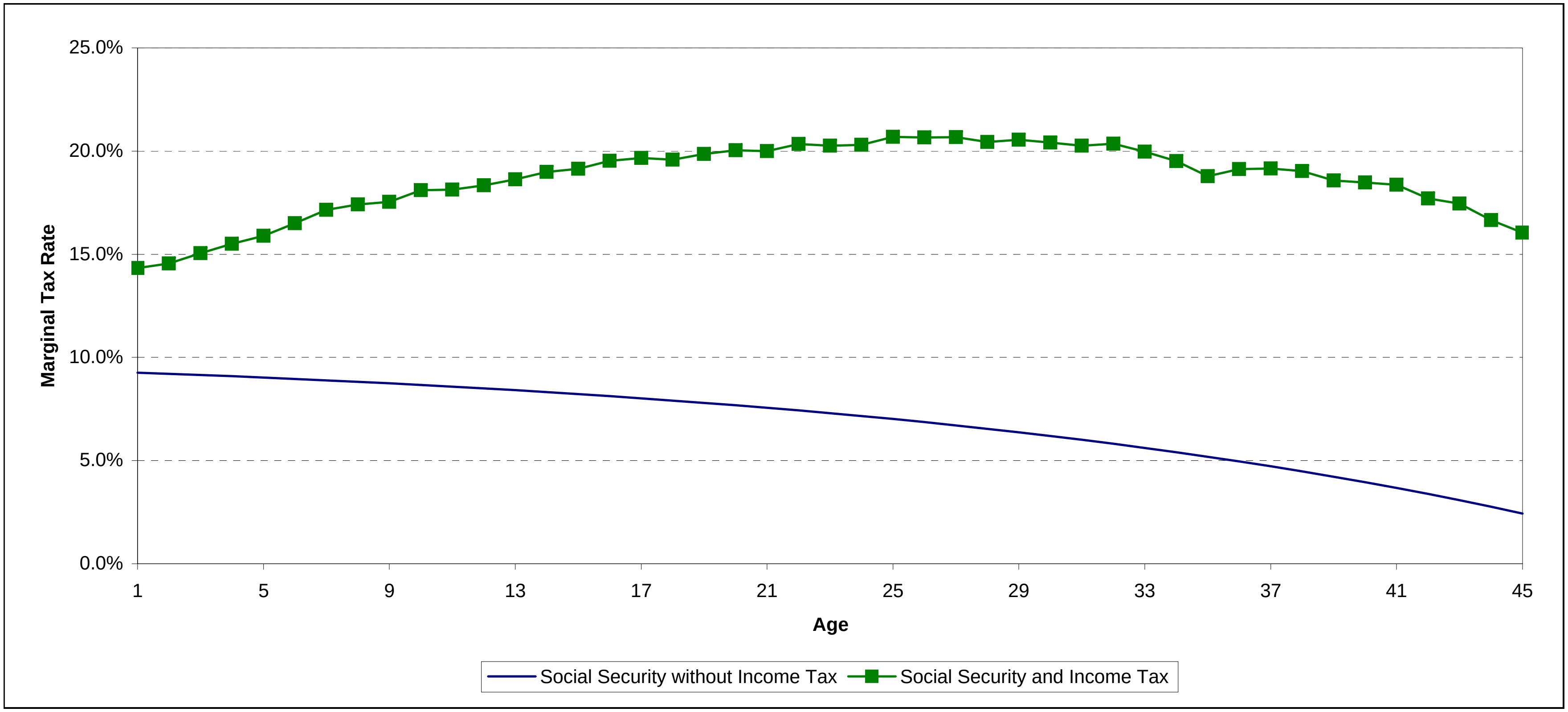



The bold vertical line in Figure 4 highlights the location of the point estimates of the variances described in the text.

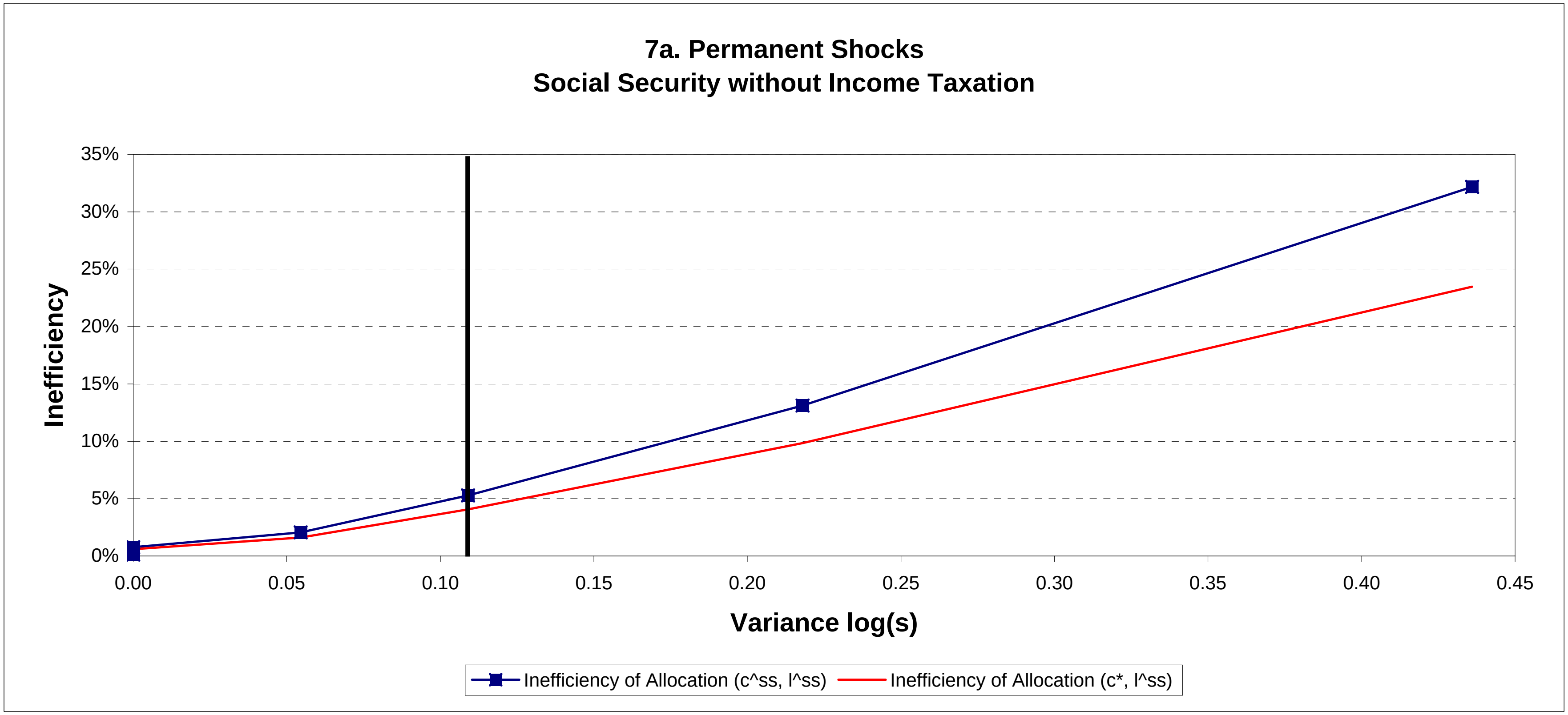
# Figure 5: Labor Profiles Without Idiosyncratic Risk



## Figure 6: Marginal Tax Rate on Earnings

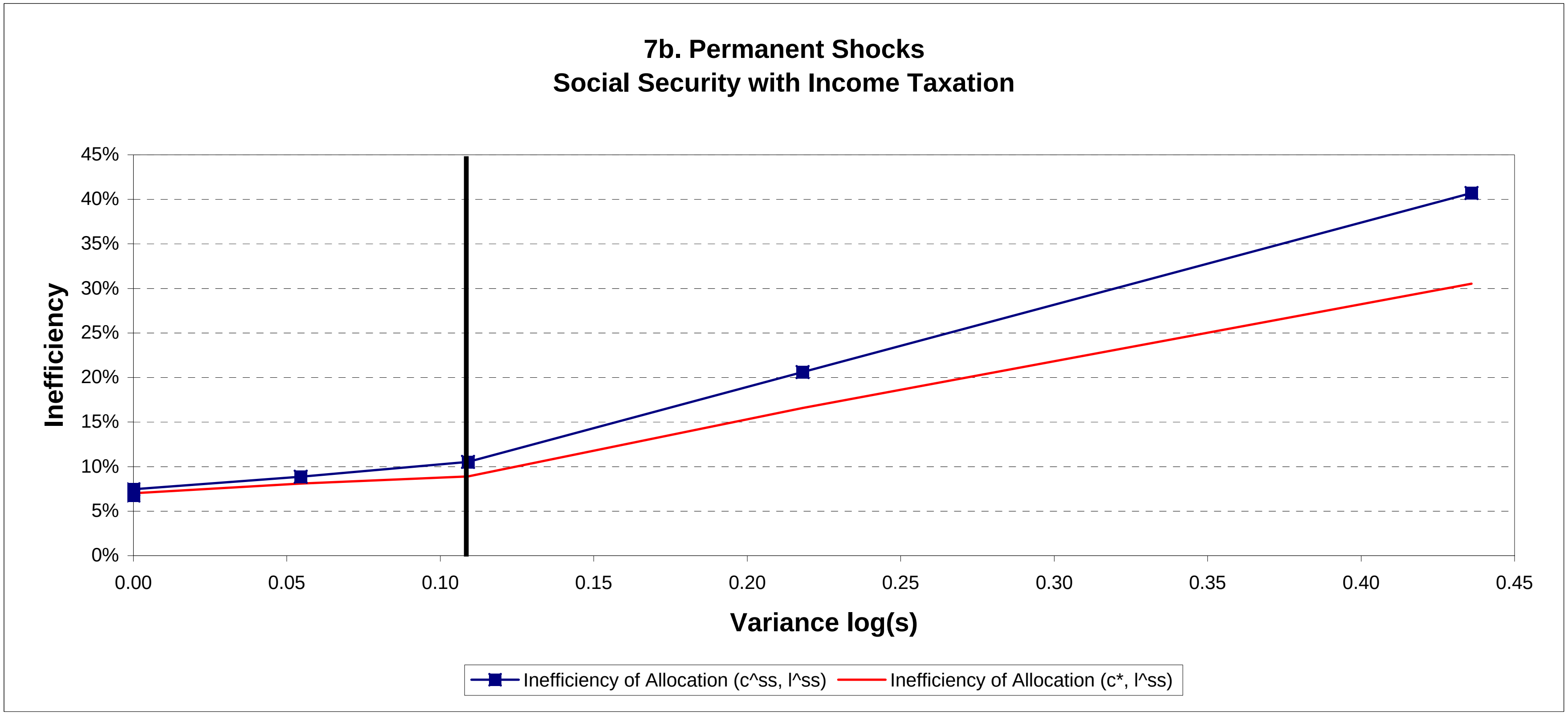


# Figure 7: Decomposition of Inefficiency



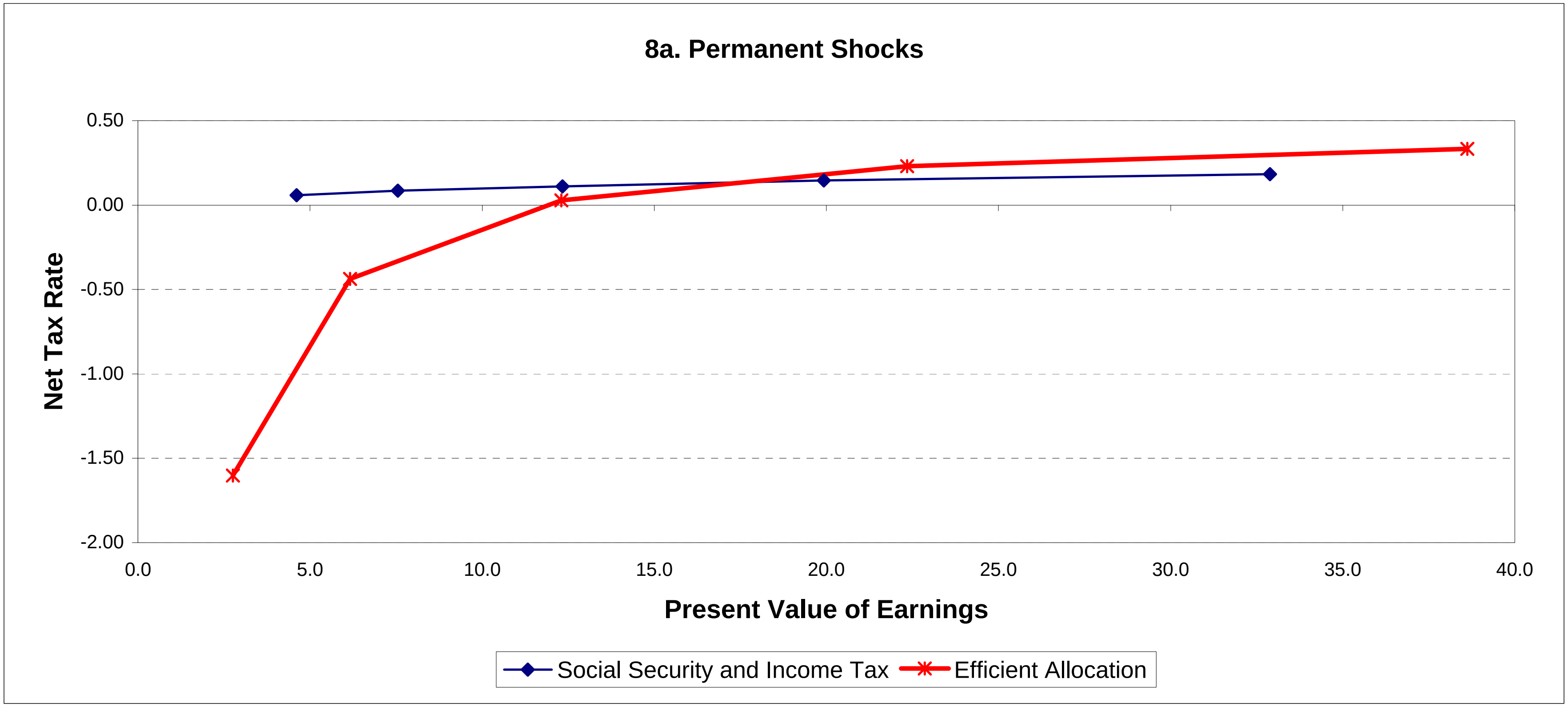

The bold vertical line in Figure 7 highlights the location of the point estimates of the variances described in the text.

## Figure 7: Decomposition of Inefficiency



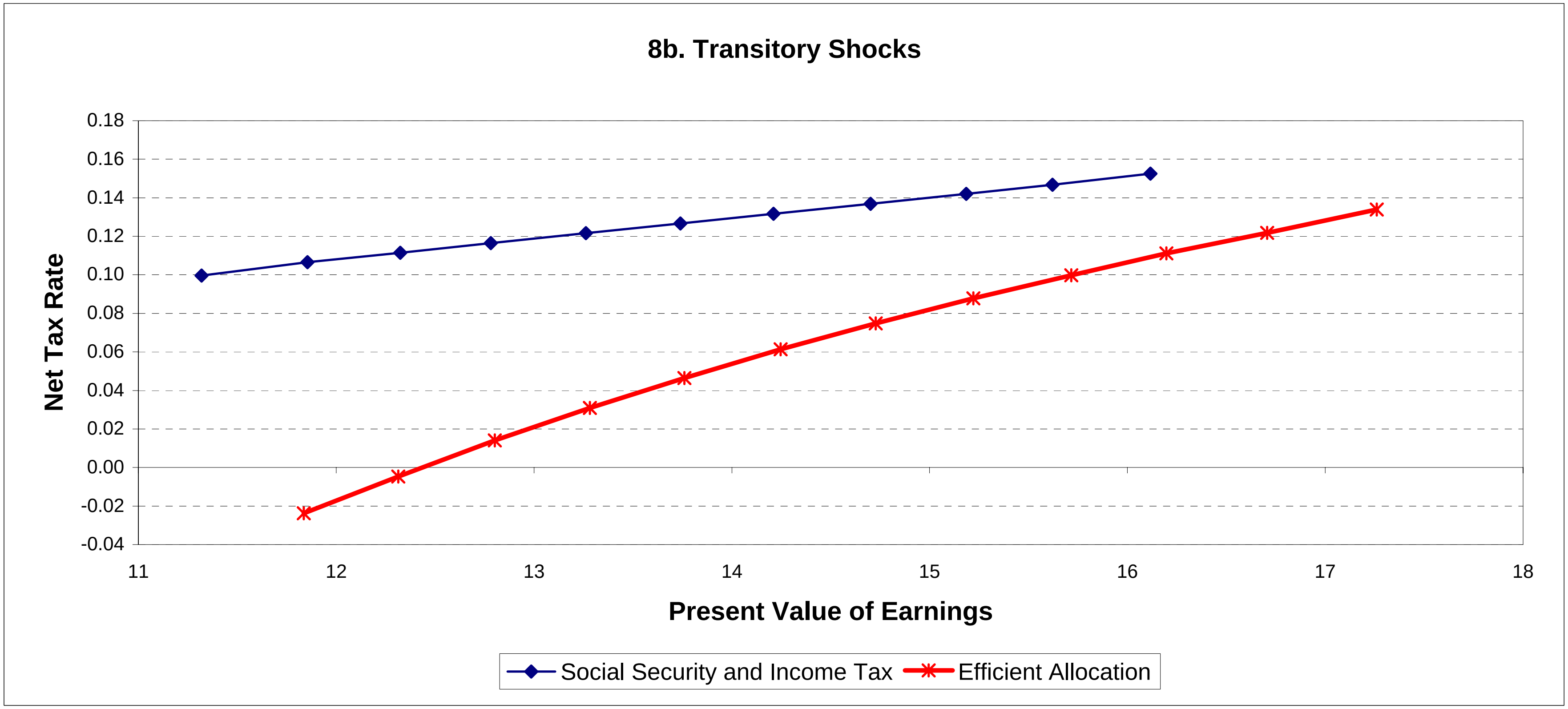

The bold vertical line in Figure 7 highlights the location of the point estimates of the variances described in the text.

## Figure 8: Net Tax Rate




Computations using point estimates of the variances described in the text.

# Figure 8: Net Tax Rate




Computations using point estimates of the variances described in the text.

## Figure 9: Consumption - Labor Wedge

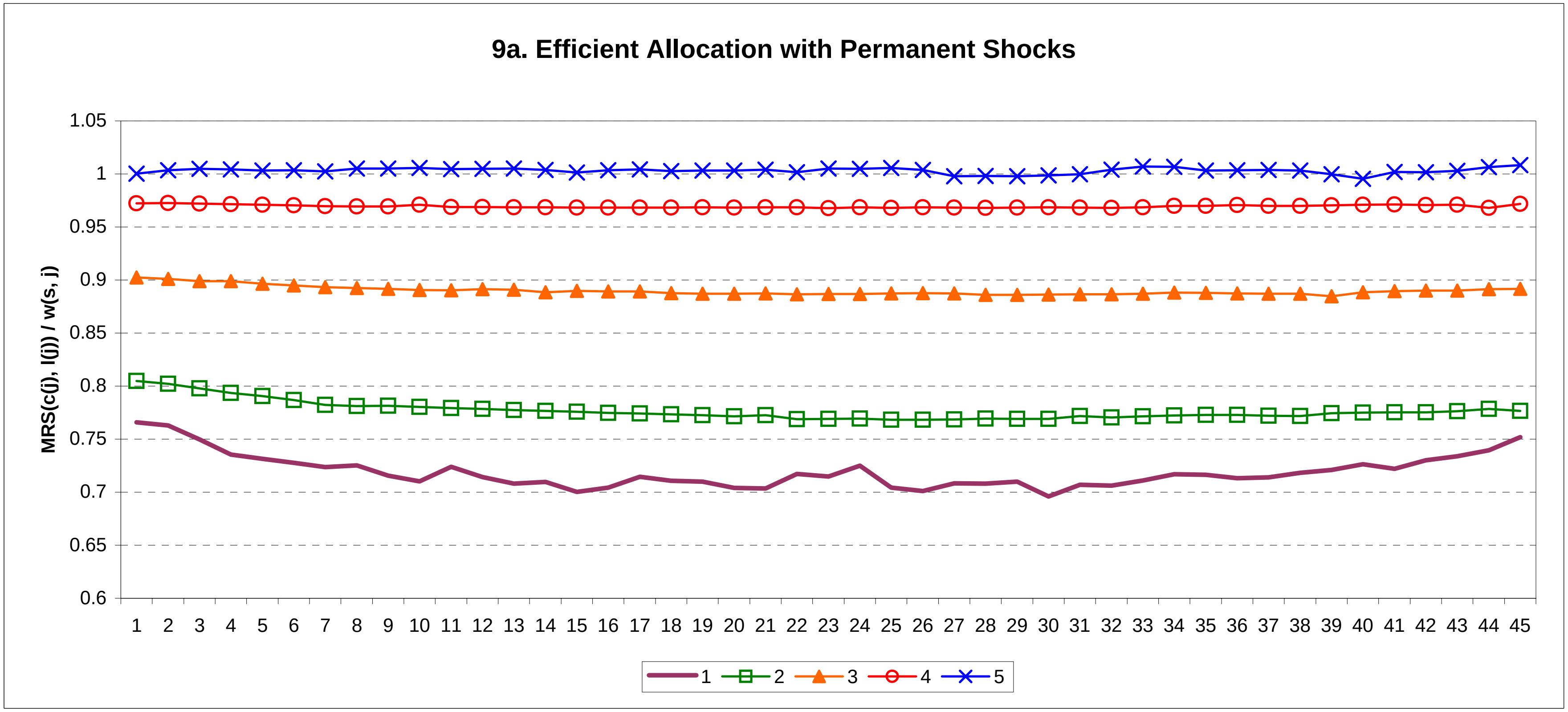


**Figure 9: Consumption - Labor Wedge**