Optimal taxation with work experience as a risky investment

Louis Perrault *

October 27, 2016

Draft

Abstract

This paper characterizes the optimal tax system in a dynamic model where wages depend on stochastic shocks and accumulated work experience. In addition to redistributive and efficiency motives, the taxation of inexperienced workers depends on a second-best requirement to encourage work experience, an implicit social insurance premium and incentive effects. Calibrations using U.S. data yield higher expected optimal marginal income tax rates for experienced workers on much of the domain of the skills distribution of inexperienced workers. They also confirm that the average optimal marginal income tax rate increases (decreases) with age when shocks and the accumulated stock of work experience are substitutes (complements), that is when the elasticity of complementarity is lower (greater) than one. Finally, more variability in experienced workers' earnings prospects leads to increasing pattern of taxes since income taxation implicitly acts as a social insurance mechanism.

JEL-Classification: H21, H24, J24

Keywords: Optimal taxation, Learning-by-doing, Human capital

*Department of Economics, Andrew Young School of Policy Studies, Georgia State University. email: louisperrault@gmail.com. I would like to thank Marcelo Arbex, Robin Boadway, Spencer Bastani, Fabrice Collard, Jean-Denis Garon, Nicolas-Guillaume Martineau, Jean-François Rouillard and Maxim Throshkin for their help and comments.

1 Introduction

Several empirical regularities, such as lower labor supply elasticity for young workers (Blundell & MaCurdy, 1999) and persistence in labor earnings (Storesletten et al., 2004) can be explained with models of work experience accumulation. Recent optimal tax literature has shown that these two facts call for conflicting policy prescriptions. For instance, models featuring persistent productivity shocks prescribe tax rates that increase with age (Farhi & Werning, 2013; Stantcheva, 2014b). Contrastingly, models of learning-by-doing featuring no productivity shocks prescribe the exact opposite (Best & Kleven, 2013; Kapicka, 2014).

Thus, features of the age structure of tax schedules are still an open question. This paper studies an optimal history-dependent tax system when future wage rates, while depending on accumulated work experience, are risky prospects. It adds to the recent strand of literature on age-dependent taxation (Kremer, 2002; Blomquist & Micheletto, 2008; Weinzierl, 2011) and it also enriches the new dynamic optimal tax literature of Golosov et al. (2003); Kocherlakota (2005); Golosov et al. (2007). In particular, I identify new key factors that should help us think through the optimal age-dependent tax structure.

The results I obtain, using data from the 2007 Panel Study of Income Dynamics (PSID), are first that in the optimum a majority of workers (roughly 62% to 77%) will face higher expected marginal labor income tax rates when old.[1] Second, whether the cross-sectional average of marginal labor income tax rates is lower when young then when old depends heavily on the complementarity between the stochastic shock and the accumulated work experience. The information obtained on cross-sectional averages of the labor distortions can yield important insights in light of the results of Farhi & Werning (2013). They show that

[1]Solving the problem of the planner yields the optimal labor and saving distortions. Under a specific implementation of the optimal allocation, the optimal labor distortions can be interpreted as marginal tax rates of a history-dependent tax system.

setting linear age-dependent taxes such that the tax rates at each age are the cross-sectional averages of the fully history-dependent tax system can capture almost all of the welfare gains of the second-best compared to the *laissez-faire* outcome. Using a Constant Elasticity of Substitution (CES) wage function, I illustrate that when the elasticity of complementarity between shock and accumulated work experience is below one, the cross-sectional average marginal income tax rate is lower when young, and when it is above one, the cross-sectional average is lower when old.[2] A possible explanation for this result is that a greater complementarity in the wage function leads to a higher wage elasticity with respect to work experience at the lower end of the shock distribution. In accordance with the results of Best & Kleven (2013) this would push marginal labor income taxes upwards for lower income young workers as their labor effort becomes less elastic to taxation. I show that using the cross-sectional averages to build an age-dependent tax system can also capture much of the welfare gains from the optimal non-linear dynamic tax system.

On theoretical grounds, I provide an analytical characterization of the optimal history-dependent tax system using the characteristics of the optimal second-best allocation. In addition to the redistributive and efficiency motives, the optimal labor distortion (*wedge*) formula reflects a balance between three motives that capture the added effects of having risk and work experience determining the wage when old. The first of these is a second-best rationale that pushes the first period labor wedge downward to encourage work experience accumulation. The second captures the insurance goals of the planner that will either push upward or downward the labor wedge depending on whether the Hicksian complementary coefficient between shock and work experience is above or below one. The third is an incentive motive which takes into account the disincentive on work experience accumulation coming from taxation in the second period. This motive will either push downwards or upwards depending on whether the second period consumption and marginal benefit of work

[2] Note that the elasticity of complementarity is the inverse of the elasticity of substitution. For the CES function used in the simulations, the elasticity of complementarity is also equal to the Hicksian complementarity coefficient.

experience is positively correlated or negatively correlated.

The optimal labor wedge formula when old, although similar to the formula found in the Mirrleesian optimal tax literature, is supplemented by the wage elasticity with respect to the second period shock. The more elastic the wage is with respect to the shock the higher the second period wedge will be. As shown in Golosov et al. (2015), the shape of the optimal tax schedule is heavily influenced by the hazard ratio of the stochastic shock. I show that the behavior of the wage elasticity with respect to the shock, as the second period shock tends to infinity, either reinforces or drastically diminishes the impact of the hazard ratio on the labor wedge formula when old. This new result also depends crucially on the complementarity between the shock and accumulated work experience. I further show that the riskier the second period is, the higher the second period labor wedge should be at the right tail of the distribution of shocks. The impact of risk on the whole tax system is investigated using numerical simulations. Increasing the volatility of stochastic shocks in the second period increases labor wedges in both period but much more drastically when old. In fact, when reducing the volatility of the second period shock, it is possible to see the increasing pattern of taxation with age result be reversed for the cases where the elasticity of complementarity is below one. This pattern can also be reversed in the case where the planner would put more weight on the workers who receive low wages in the first period.

This project is closely related to the new public finance literature that looks at characteristics of the optimal tax system when skill shocks are persistent, such as Farhi & Werning (2013) and Golosov et al. (2015). It builds on these contributions as it follows the first-order approach to solve the planner's problem but models the persistence of productivity directly by incorporating work experience.[3] By endogenizing the wage formation process, the optimal

[3]As we will explain in more detail below, the planner is able to observe the accumulation of work experience. This assumption allows us to keep *common knowledge of preference*, and thus our analysis is also similar to Albanesi & Sleet (2006).

policies take into consideration the effect of taxes and how they effect persistence of earnings. It is also tied to recent work on optimal taxation with human capital and risky environments where human capital is acquired through either schooling or on-the-job training programs (Bohacek & Kapicka, 2008; Kapicka, 2014; Kapicka & Neira, 2014; Stantcheva, 2014b).

Finally, the paper contributes to the literature on optimal taxation with learning-by-doing. Krause (2009) considers optimal taxation in a two-type model similar to Stiglitz (1982), where the planner can commit or not to a two-period income tax schedule. The author finds that the no-distortion-at-the-top result no longer applies and that there are some cases where it is justified to tax the high skilled workers even if it depresses both labor supply and future wages. Best & Kleven (2013) also considers a Mirleesian economy with a continuum of types. They study both age-independent and age-dependent taxation. Their numerical simulations make a strong case for higher age-dependent income tax rates for the young. This is due to endogenous wage rates and a negative correlation between age and innate (first period) ability (conditional on earnings). In contrast to Krause (2009) and Best & Kleven (2013), I incorporate uncertainty in the second period wage rate and the ability for workers to save.

Stantcheva (2014a) considers a more general framework where "training" is a form of human capital investment. In the author's paper, workers face a risky environment where their choice of training effort can be either a substitute or a complement to labor effort. It is shown that one of the limit case of the general model is the learning-by-doing model. However, most of the analysis undertaken in the paper focuses on the impact of the substitutability or complementarity of training effort with respect to labor effort on the optimal tax schedule. This present paper analyses the case where labor supply decisions necessarily involves learning through the accumulation of work experience. In that respect, it is more in line with Krause (2009) and Best & Kleven (2013).

The rest of the paper is organized as follows. Section 2 presents the two-period model and writes the planner's problem recursively. Section 3 considers the characteristics of the optimal income tax systems derived from analytical results. Section 4 presents different numerical simulations to highlight further properties of the optimal allocation. Finally Section 5 is dedicated to concluding comments.

2 The Model

The economy is populated by workers who live for two periods $t = 1, 2$, each period they consume c_t and provide labor effort l_t from which they acquire work experience. Workers obtain wage w_t and and earn gross income $y_t = w_t l_t$. The wage $w_t(\theta_t, e_t)$ is a function of the time-varying shock θ_t and of the stock of effective work experience e_t. It is assumed that:

$$\frac{\partial w_t}{\partial \theta_t} > 0; \quad \frac{\partial w_t}{\partial e_t} > 0 \quad \frac{\partial^2 w_t}{\partial \theta_t \partial e_t} \geq 0; \quad \frac{\partial^2 w_t}{\partial^2 x} \leq 0 \text{ for } x = \{\theta_t, e_t\}, \tag{2.1}$$

where the stock of worker experience evolves according to

$$e_{t+1} = \phi(y_t) + e_t, \tag{2.2}$$

where $\phi_y > 0$ and the starting level of work experience, e_1, is identical for all workers.[4][5] In each period, worker's shock θ_t is distributed according to the density $f^t(\theta_t)$ with support $[\underline{\theta}, \bar{\theta}]$. The focus is restricted to shocks that are independent across periods and endogenize the

[4]There are no assumption made on the second derivative of the function $\phi(y_t)$. The problem of the worker is assumed to be concave due to the properties of the wage function and the disutility function of labor effort.

[5]Making work experience a function of both income and present period shock θ, e.g. $\phi(\theta_t, y_t)$, would remove the *common knowledge of preference* assumption as work experience would no longer be observable to the planner.

persistence of wages through accumulated work experience. Let the per period preferences be represented by the following utility function

$$u(c_t) - h(l_t), \tag{2.3}$$

with u being strictly concave, h strictly convex, and also define θ^t as the history of shocks up to period t. For a given allocation $\{c(\theta^2), y(\theta^2)\}$ that specifies consumption and income for each history of shocks the worker's expected lifetime utility is

$$U(\{c,y\}) \equiv \sum_{t=1}^{2} \beta^{t-1} \int \left[u(c(\theta^t)) - h\left(\frac{y(\theta^t)}{w_t(\theta_t, e_t(\theta^{t-1}))} \right) \right] f^2(\theta_2) f^1(\theta_1) d\theta_2 d\theta_1, \tag{2.4}$$

where β is the discount factor.

Incentive Compatibility and Planner's Problem: The particularity of Mirrleesian optimal taxation is that worker's shocks are private observation and that the planner is constrained by the information he has, i.e. he can only observe labor income and savings.[6] By the revelation principle, it is possible to focus on direct mechanisms where workers will report their type every period. A reporting strategy $\sigma = \{\sigma_1(\theta^1), \sigma_2(\theta^2)\}$ implies a history of per period reports $\sigma^t(\theta^t)$, for which the planner allocates consumption $c(\sigma^t)$, income $y(\sigma^t)$ and by extension accumulation of work experience effective in the second period, i.e.

$$e_2(\sigma_1(\theta^1)) = \phi(y(\sigma_1(\theta^1)) + e_1.^7 \tag{2.5}$$

For an allocation $\{c, y\}$, let $\omega(\theta^t)$ denote the equilibrium continuation utility after history

[6] However, the distribution of shocks is known to the planner.

[7] Work experience $e(\sigma_1(\theta^1))$ is also a function of e_1 but since it is assumed that all workers start with the same level of work experience, for simplicity of exposition, we will omit to include this level of starting work experience for the rest of the paper.

θ^t, defined as the unique solution to

$$\omega(\theta^t) = u\big(c(\theta^t)\big) - h\left(\frac{y(\theta^t)}{w_t(\theta_t, e_t)}\right) + \beta\int \omega(\theta^t, \theta_{t+1}) f^{t+1}(\theta_{t+1}) d\theta_{t+1} \tag{2.6}$$

for $t = 1, 2$ with $\omega(\theta^3) \equiv 0$.[8] For any reporting strategy σ, let the continuation value $\omega^\sigma(\theta^t)$ be the unique solution to

$$\omega^\sigma(\theta^t) = u\big(c(\sigma^t(\theta^t))\big) - h\left(\frac{y(\sigma^t(\theta^t))}{w_t(\theta_t, e_t)}\right) + \beta\int \omega^\sigma(\theta^t, \theta_{t+1}) f^{t+1}(\theta_{t+1}) d\theta_{t+1}. \tag{2.7}$$

An allocation $\{c, y\}$ is said to be incentive compatible if and only if

$$\omega(\theta^t) \geq \omega^\sigma(\theta^t) \quad \forall\theta^t, \forall\sigma. \tag{2.8}$$

Therefore an allocation is incentive compatible if truth telling, i.e. $\sigma^* = \{\sigma^*(\theta^t)\}$ with $\sigma^*_t(\theta_t) = \theta_t$ yields a weakly higher continuation utility.

As in Golosov et al. (2015), the planner evaluates welfare of workers using Pareto weights $a(\theta_1) \geq 0$. Weight $a(\theta_1)$ is given to workers who received θ_1 in the first period. The weights also have the characteristic that $\int_{\underline{\theta}}^{\bar{\theta}} a(\theta_1) f^1(\theta_1) d\theta_1 = 1$. Hence social welfare for a given allocation is

$$SW(\{c, y\}) \equiv \int_{\underline{\theta}}^{\bar{\theta}} a(\theta_1) \left\{ E_1\left(\sum_{t=1}^{2} \beta^{t-1} \left[u(c(\theta^t)) - h\left(\frac{y(\theta^t)}{w_t(\theta_t, e_t(\theta^{t-1}))} \right) \right] \middle| \theta_1 \right) \right\} f^1(\theta_1) d\theta_1,$$

where E_1 is the conditional expectation of the worker that received shock θ_1 in the first period. Let v_0 be the value of social welfare that the planner credibly promises to deliver to society. Also suppose that the economy has a linear technology that transform effective labor into consumption and that the planner can transfer resources across periods at a gross

[8] Note that $\omega(\theta^t, \theta_{t+1})$ stands in for $\omega(\theta^{t+1})$ for clarity of exposition.

interest rate of R. The planner's problem is then to minimize the cost of providing allocation $\{c, y\}$ subject to the allocation being incentive compatible and offering v_0, i.e.

$$K(v_0, e_1) = \min_{\{c,y\}} \left[\sum_{t=1}^{2} \left(\frac{1}{R} \right)^{t-1} \int \left\{ c(\theta^t) - y(\theta^t) \right\} f^2(\theta_2) f^1(\theta_1) d\theta_2 d\theta_1 \right] \tag{2.9}$$

$$\begin{aligned} s.t. \quad & v_0 = SW(\{c, y\}), \\ & \omega(\theta^t) \geq \omega^\sigma(\theta^t) \quad \forall \theta^t, \forall \sigma, \\ & e_2(\theta^1) = \phi(y(\theta^1)) + e_1. \end{aligned}$$

The relaxed problem

Following Farhi & Werning (2013), I use the first-order approach to write a relaxed problem of the planner's problem. The approach relies on changing the incentive constraint (2.8) to a "temporal" incentive constraint which only considers one-shot deviations every period. They show that the set of allocations that satisfy these new incentive constraints from the relaxed problem is a subset of allocations that satisfy (2.8).[9]

I derive the temporal incentive constraints as in Farhi & Werning (2013). Working backwards and starting from period 2, let the continuation value under truthful revelation, $\omega(\theta^2)$, be:

$$\omega(\theta^2) = u(c(\theta^2)) - h\left(\frac{y(\theta^2)}{w_2(\theta_2, e_2(\theta^1))} \right). \tag{2.10}$$

Consider the deviation strategy r where the workers reports truthfully until t but not during period t, i.e. $\sigma^{t-1}(\theta^{t-1}) = \theta^{t-1}$ and $\sigma^t(\theta^t) = r$,where $r \neq \theta_t$. The continuation utility under

[9]However, much like in the static nonlinear optimal tax case, the validity of the first-order approach in a dynamic setting is not guaranteed. In the case of a static optimal tax exercise, it can be shown that when the worker's utility function satisfies the Spence-Mirrlees single-crossing property and the allocation satisfies a monotonicity condition, which corresponds to the second-order condition, the optimal allocation is incentive compatible. In this dynamic setting the envelope condition is a necessary condition, but there are no "simple" conditions like in the static case that guarantee incentive compatibility. Therefore, in the numerical simulations below, as in Farhi & Werning (2013) or Golosov et al. (2015), the incentive compatibility of the allocations from the solutions of the relaxed problems are verified ex-post numerically.

deviation strategy r in the second period is:

$$\omega^{\sigma^r}(\theta^2) = u(c(\theta^1, r)) - h\bigg(\frac{y(\theta^1, r)}{w_2(\theta_2, e_2(\theta^1))}\bigg). \tag{2.11}$$

An allocation is temporary incentive compatible in period 2, if for all histories θ^2,

$$\omega(\theta^2) = \max_r \ \omega^{\sigma^r}(\theta^2). \tag{2.12}$$

For period 1, let the continuation value under truthful revelation, $\omega(\theta^1)$, be the unique solution to:

$$\omega(\theta^1) = u(c(\theta^1)) - h\bigg(\frac{y(\theta^1)}{w_1(\theta_1, e_1)}\bigg) + \beta \int \omega(\theta^1, \theta_2) f^2(\theta_2) d\theta_2. \tag{2.13}$$

Again, considering a one shot-deviation strategy r, let the continuation utility be the unique solution to:

$$\omega^{\sigma^r}(\theta^1) = u(c(r)) - h\bigg(\frac{y(r)}{w_1(\theta_1, e_1)}\bigg) + \beta \int \omega^{\sigma^r}(e_2(\theta_1, r), \theta_2) f^2(\theta_2) d\theta_2, \tag{2.14}$$

where $e(r) = \phi(y(r)) + e_1$.[10] An allocation is temporary incentive compatible in period 1, if for all shock θ^1,

$$\omega(\theta^1) = \max_r \ \omega^{\sigma^r}(\theta^1).^{11} \tag{2.15}$$

An allocation is incentive compatible, i.e. constraints (2.8) hold, if for all histories of shocks both

$$\omega(\theta^1) = \max_r \ \omega^{\sigma^r}(\theta^1) \ \text{ and } \ \omega(\theta^2) = \max_r \ \omega^{\sigma^r}(\theta^2) \tag{2.16}$$

are true. Following the first-order approach, these temporary incentive constraints are re-

[10] $\omega^{\sigma^r}(r, e(r), \theta_2)$ can be any future strategy and has no link to the period 2 one-shot deviation utility $\omega^{\sigma^r}(\theta^2)$.

[11] Equivalently it can be written

$$\omega(\theta^1) = \max_r \bigg\{ u(c(r)) - h\bigg(\frac{y(r)}{w_1(\theta_1, e_1)}\bigg) + \beta \int \omega^{\sigma^r}(r, e_2(r), \theta_2) f^2(\theta_2) d\theta_2 \bigg\}.$$

placed by the following envelope conditions applied to (2.16):

$$\dot{\omega}(\theta^1) := \frac{\partial \omega(\theta^1)}{\partial \theta_1} = \frac{h'\Big(\dfrac{y(\theta^1)}{w_1(\theta_1, e_1)}\Big) y(\theta^1)}{[w_1(\theta_1, e_1)]^2} \frac{\partial w_1(\theta_1, e_1)}{\partial \theta_1}, \tag{2.17}$$

$$\dot{\omega}(\theta^2) := \frac{\partial \omega(\theta^2)}{\partial \theta_2} = \frac{h'\Big(\dfrac{y(\theta^2)}{w_2\big(\theta_2, e_2(y(\theta^1))\big)}\Big) y(\theta^2)}{[w_2\big(\theta_2, e_2(y(\theta^1))\big)]^2} \frac{\partial w_2\big(\theta_2, e_2(y(\theta^1))\big)}{\partial \theta_2}. \tag{2.18}$$

The problem is written recursively starting with the second period problem. Let $\upsilon(\theta^1) = \int \omega(\theta^1, \theta_2) f^2(\theta_2) d\theta_2$ be the expected continuation utility. The second period problem of the planner is to minimize the second period expected costs taking as given $\upsilon(\theta^1)$ and $e(\theta^1)$ (expressed as υ, and e_2 respectively) subject to the envelope condition,

$$\begin{aligned} K(\upsilon, e_2, 2) = & \min_{\{c(\theta), y(\theta)\}} \int [c(\theta) - y(\theta)] f^2(\theta) d\theta \\ s.t. \quad & \omega(\theta) = u\big(c(\theta)\big) - h\Big(\frac{y(\theta)}{w_2(\theta, e_2)}\Big) \\ & \upsilon = \int \omega(\theta) f^2(\theta) d\theta \\ & \text{and (2.18).} \end{aligned} \tag{2.19}$$

The first period problem taking as given υ_0 and starting work experience e_1 is

$$K(\upsilon_0, e_1, 1) = \min_{\{c(\theta), y(\theta), \omega(\theta), \upsilon(\theta)\}} \int \Big[c(\theta) - y(\theta) + \frac{1}{R} K\big(\upsilon(\theta), e(\theta), 2\big)\Big] f^1(\theta) d\theta$$

$$\begin{aligned} s.t. \quad & \omega(\theta) = u\big(c(\theta)\big) - h\Big(\frac{y(\theta)}{w_1(\theta, e_1)}\Big) + \beta \upsilon(\theta) \\ & \upsilon_0 = \int a(\theta) \omega(\theta) f^1(\theta) d\theta \\ & e(\theta) = \phi(y(\theta)) + e_1 \\ & \text{and (2.17).} \end{aligned} \tag{2.20}$$

The value of υ_0 is chosen to be the highest value such that the expected ressource costs of providing this level of social welfare is zero, i.e. $K(\upsilon_0, e_1, 1) = 0$.

3 The Optimal Allocation: Optimal Wedges

3.1 Definitions

To write and interpret the optimal wedge formulas I require the use of several terms which I define below.

Wedges

In the static non-linear optimal taxation literature differences in marginal utility of consumption and marginal disutility of labor are interpreted as marginal taxes. However, in a dynamic setting, these differences can no longer readily be interpret in such a way. It is then convenient to define these differences as *wedges* which help in getting intuition from the solution of the optimal allocation problem. One of the issues with interpreting wedges as taxes is that there can be many different combinations of tax instruments to implement the optimal allocation. Furthermore, as argued by Golosov et al. (2007) each wedge corresponds to a particular choice of the worker taking all other choices fixed at a specific level. Since choices are made jointly in a decentralized economy setting, a particular tax rate equal to the wedge without further restrictions on the tax instruments may lead workers to deviate and the optimal allocation may not be implemented. Nonetheless, some implementations, like the one elaborated in Appendix B, can equate the labor wedges to marginal tax rates. For this reason, this paper sometimes uses those terms interchangeably.

For any allocation $\{c, y\}$ after any history θ^t, let the intertemporal wedges $\tau_K(\theta^1)$ and the labor wedges $\tau_L(\theta^1), \tau_L(\theta^2)$ respectively be:

$$\tau_K(\theta^1) = 1 - \frac{1}{R\beta}\frac{u'(c(\theta^1))}{\mathrm{E}\big[u'(c(\theta^2))\big]} \ \forall\theta^1, \tag{3.1}$$

$$\tau_L(\theta^1) = 1 - \frac{h'\left(\dfrac{y(\theta^1)}{w_1(\theta_1, e_1)}\right)}{u'(c(\theta^1))w_1(\theta_1, e_1)} + \beta\frac{\phi_y(y(\theta^1))}{u'(c(\theta^1))}\mathrm{E}\big[\mathcal{MB}(\theta^2)\big] \ \ \forall\theta^1, \tag{3.2}$$

$$\tau_L(\theta^2) = 1 - \frac{h'\left(\dfrac{y(\theta^2)}{w_2(\theta_2, e(\theta^1))}\right)}{u'(c(\theta^2))w_2(\theta_2, e(\theta^1))} \ \forall\theta^2, \tag{3.3}$$

where

$$\mathcal{MB}(\theta^2) \equiv \frac{h'\left(\dfrac{y(\theta^2)}{w_2(\theta_2, e_2(\theta^1))}\right)y(\theta^2)}{[w_2(\theta_2, e(\theta^1))]^2}\frac{\partial w_2(\theta_2, e(\theta^1))}{\partial e(\theta^1)}$$

is the realized marginal benefit of work experience and $\mathrm{E}[\mathcal{MB}(\theta^2)]$ is the expected marginal benefit of work experience from the point of view of the first period.[12]

Hicksian Complementarity

Let the Hicksian complementarity coefficient between the shock θ_2 and work experience e_2 be defined as

$$\rho(\theta^2, e_2) \equiv \frac{\dfrac{\partial^2 w_2(\theta_2, e_2)}{\partial\theta\partial e_2}w_2(\theta_2, e_2)}{\dfrac{\partial w_2(\theta_2, e_2)}{\partial\theta}\dfrac{\partial w_2(\theta_2, e_2)}{\partial e_2}}. \tag{3.4}$$

This coefficient measures how complementary the shock and work experience are in the production of the second period wage. For any wage function that is additive in shock and

[12]The marginal benefit can also be written using labor instead of income:

$$\mathcal{MB}(\theta^2) \equiv \frac{h'\big(l(\theta^2)\big)\, l(\theta^2)}{w_2(\theta_2, e_2)}\frac{\partial w_2(\theta_2, e_2)}{\partial e}.$$

work experience, this term is zero as they are perfect substitutes. In any situation where $\frac{\partial^2 w_2}{\partial\theta\partial e} > 0$, the Hicksian complementarity coefficient is positive and work experience will increase exposure to risk. As Stantcheva (2014b) has demonstrated it is the relation of the Hicksian coefficient with respect to one that is more relevant to the planner for questions of risk and insurance .[13] More precisely, if the Hicksian coefficient is below 1, it implies that the wage elasticity with respect to the shock is decreasing in work experience, or alternatively the wage elasticity with respect to work experience is decreasing with respect to the shock. If the Hicksian coefficient is greater than 1, the wage elasticity with respect to the shock is increasing in work experience and whenever the coefficient equals 1 the wage elasticity is constant. Note that a CES wage function where $w_2 = (\theta_2^{1-\rho} + e_2^{1-\rho})^{\frac{1}{1-\rho}}$ will have $\rho(\theta^2, e_2) = \rho$.

Elasticities

Optimal tax formulas are generally written with elasticity parameters which capture the efficiency cost of taxation. Let

$$\alpha_t(\theta^t) \equiv \frac{h''(l(\theta^t)l(\theta^t)}{h'(l(\theta^t))} \text{ and } \eta_t(\theta^t) \equiv -\frac{u''(c(\theta^t))c(\theta^t)}{u'(c(\theta^t))},$$

be two elasticity measures.[14] The elasticity measure $\alpha_t(\theta^t)$ is akin to the inverse of the Frisch elasticity of labor supply.[15] The elasticity measure $\eta_t(\theta^t)$ is the inverse of the elasticity of intertemporal substitution or alternatively the measure of relative risk aversion of the preferences.

[13]This fact is demonstrated in the optimal tax literature in both Bovenberg and Jacobs (2011) and Stantcheva (2014b).

[14]These elasticity measures are used in Golosov et al. (2015).

[15]In this setting, in the first period it captures the elasticity of labor supply keeping future wage, consumption in both period and future labor supply constant. In the second period it also captures the elasticity of labor supply keeping consumption in both period and labor in the first period constant.

3.2 Optimal Wedges

The subsection describes and analyzes the solution to the planner's problem. The labor choice distortion in both periods is considered first and the distortion to the saving decisions is considered later. To obtain further insights on the labor wedges, I make the following assumption:

Assumption 1. *The optimal allocation satisfies*

$$\dot{c}(\theta^t) \geq 0 \ \forall \theta^t \ \textit{for} \ t = 1, 2.^{16}$$

Proposition 1. *The optimal labor wedges in periods t=1 and t=2 are:*

$$\begin{aligned}\tau_L^*(\theta^1) = & \frac{\mu(\theta^1)}{f^1(\theta_1)} \frac{h'(l_1)}{w_1} \frac{\epsilon_{\theta 1}}{\theta_1} \big(1 + \alpha_1(\theta^1)\big) - \frac{\phi_y}{R} E\Big[\tau_L(\theta^2) w_{e2} l_2\Big] \\ & - \frac{\phi_y}{R} \left\{ E\Big[\mathcal{MB}(\theta^2) \frac{\mu(\theta^2)}{f^2(\theta_2)} \frac{\epsilon_{\theta 2}}{\theta_2} (1 - \rho_{\theta e2})\Big] + \mathrm{Cov}\Big(\frac{1}{u'(c_2)}, \mathcal{MB}(\theta^2)\Big) \right\} \end{aligned} \tag{3.5}$$

$$\frac{\tau_L^*(\theta^2)}{1 - \tau_L^*(\theta^2)} = u'(c(\theta^2)) \frac{\mu(\theta^2)}{f^2(\theta_2)} \frac{\epsilon_{\theta 2}}{\theta_2} \big(1 + \alpha_2(\theta^2)\big) \geq 0, \tag{3.6}$$

[16] It can be shown that incentive compatibility requires that

$$\frac{\partial [u(c(\theta)) + \beta v(\theta)]}{\partial \theta} \geq 0.$$

I find that Assumption 1 is satisfied in all of our numerical simulations.

where[17]

$$\mu(\theta^1) = \int_{\underline{\theta}}^{\bar{\theta}} \big(1 - g(\theta^1)\big) \frac{1}{u'(c(\theta^1))} f^1(\theta_1) d\theta_1, \text{ with } g(\theta^1) = a(\theta_1) u'(c(\theta^1)) \lambda_1,$$

$$\text{and} \qquad \lambda_1 = \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta^1))} f^1(\theta_1) d\theta_1,$$

$$\mu(\theta^2) = \int_{\underline{\theta}}^{\bar{\theta}} \big(1 - g(\theta^1, \theta_2)\big) \frac{1}{u'(c(\theta^1, \theta_2))} f^2(\theta_2) d\theta_2,$$

$$\text{with } g(\theta^1, \theta_2) = u'(c(\theta^1, \theta_2)) \lambda_2, \text{ and } \lambda_2 = \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta^1, \theta_2))} f^2(\theta_2) d\theta_2.$$

Proof: See Appendix A.1. Also note, that for ease of notation let $l(\theta^t) = l_t,\ c(\theta^t) = c_t,\ e(\theta^1) = e_2,\ w_t(\theta_t, e_t) = w_t,\ \partial w_t / \partial e = w_{et},\ \partial w_t / \partial \theta = w_{\theta t},\ \epsilon_{\theta t} = w_{\theta t} * (\theta / w_t),\ \rho(\theta^2, e_2) = \rho_{\theta e2}$.

The multipliers associated with the envelope conditions in period 1 and period 2 are $\mu(\theta^1)$ and $\mu(\theta^2)$ respectively. The first period's multiplier captures the insurance motive or the redistributive goals of the planner who promised υ_0. And, $\mu(\theta^2)$ is the insurance motive of the planner who promised continuation utility $\upsilon(\theta^1)$ to a worker. Parameter $g(\theta^t)$ is the value to the planner of giving one more dollar to an individual with history θ^t. Notice that the Pareto weight $a(\theta_1)$ only appear in $g(\theta^1)$. Also, λ_1 is the marginal resource cost of providing a marginal increase in promised utility υ_0 and λ_2 is the same for the promised utility υ for period 2 from period 1.

Starting with the second period labor wedge (3.6), notice that it is possible to rewrite it

[17] I use the same definitions found in Stantcheva(2014b).

in the ABC form found in the static optimal tax literature, Diamond (2003):

$$\frac{\tau_L(\theta^2)}{1-\tau_L(\theta^2)} = A(\theta^2)B(\theta^2)C(\theta^2), \tag{3.7}$$

where

$$\begin{aligned}
A(\theta^2) &= (1+\alpha_2(\theta^2))\epsilon_{\theta 2}(\theta^2),\\
B(\theta^2) &= \frac{1-F^2(\theta_2)}{\theta_2 f^2(\theta_2)},\\
C(\theta^2) &= \int_{\theta_2}^{\bar{\theta}} \exp\left(\int_{\theta_2}^{x} \eta(\tilde{x})\frac{\dot{c}(\tilde{x})}{c(\tilde{x})}d\tilde{x}\right)[1-g(x)]\frac{f^2(x)}{1-F^2(\theta_2)}dx.
\end{aligned}$$

The intuition behind this optimal wedge formula goes as follows. Distorting labor supply at shock level θ_2 increases the marginal deadweight burden at this shock level which depends on the labor supply elasticity $\alpha_2(\theta^2)$ and is amplified by the wage elasticity with respect to the shock $\epsilon_{\theta 2}(\theta^2)$. This effect is captured by $A(\theta^2)$. Note that $\epsilon_{\theta 2}(\theta^2)$ is not present in most optimal income tax exercise since a shock (or type) represents the wage rate for a particular worker type.[18] Since wage in this paper is composed of both shock and work experience, the impact of a change in type on the wage will be present in the envelope condition and thus in the optimal labor wedge formula.

Increasing the distortion at shock level θ_2 transfers resources from workers with higher shock levels to the planner who's value for these resources is reflected by $C(\theta^2)$.[19] It is the insurance motive of the planner. And finally, the hazard ratio $B(\theta^2)$ reflects the tradeoff between $A(\theta^2)$ and $C(\theta^2)$ as it captures the number of workers related to $C(\theta^2)$ and the number of workers at $A(\theta^2)$. The fact that the wage elasticity, labor elasticity and the density are

[18] The partial derivative $\partial w/\partial\theta$ in the standard envelope condition is technically present but it is equal to one.

[19] As pointed out in Golosov et al. (2015), the term $\exp\left(\int_{\theta_2}^{x}\eta(\tilde{x})\frac{\dot{c}(\tilde{x})}{c(\tilde{x})}d\tilde{x}\right)$ represents the income effect on labor supply of transferring resources from workers who receive shocks above θ_2.

all positive combined with Assumption 1 leads to the result that the optimal second period wedge is non-negative.[20] From Assumption 1, the monotonicity assumption on consumption, it is possible to show that $\mu(\theta^2) \geq 0$ and thus the result on the optimal wedge follows.[21]

The formula for the first period optimal wedge (3.5) is composed of four parts.[22] The first part is more familiar and the three others are linked to the accumulation of work experience through labor. This first part is composed of the redistributive motive, the shock density, the wage elasticity, and the labor choice elasticity parameter. By the same argument as above, this first part will also be non-negative as it can be shown that $\mu(\theta^1) \geq 0$.

The remaining three novel parts of the optimal first period labor wedge formula are all linked to the different motives the planner has for work experience accumulation. These

[20]From the boundary conditions, it is possible to obtain the classic no distortion at the top and bottom.

[21]See Golosov et al. (2011) and Stantcheva (2014b). The demonstration is simpler than the one in these papers as it is not necessary to take into consideration the impact of persistence of shocks in the present setting.

[22]The first period labor wedge can also be partly written in the ABC form, but its interpretation is much less straightforward. Equation (3.5) can be rewritten in the following way:

$$\begin{aligned}\tau_L(\theta^1) = {} & A(\theta^1)B(\theta^1)C(\theta^1)\left[\frac{h'(l_1)}{u'(c_1)w_1}\right] \\ & - \frac{\phi_y}{R}\left\{\mathrm{E}\Big[\tau_L(\theta^2)w_{e2}l_2\Big] + \mathrm{E}\Big[\mathcal{MB}(\theta^2)\frac{\mu(\theta^2)}{f^2(\theta_2)}\frac{\epsilon_{\theta 2}}{\theta_2}(1-\rho_{\theta e2})\Big] + \mathrm{Cov}\Big(\frac{1}{u'(c_2)}, \mathcal{MB}(\theta^2)\Big)\right\}\end{aligned}$$

where

$$\begin{aligned}A(\theta^1) &= (1+\alpha_1(\theta^1))\epsilon_{\theta 1}(\theta^1), \\ B(\theta^1) &= \frac{1-F^1(\theta_1)}{\theta_1 f^1(\theta_1)}, \\ C(\theta^1) &= \int_{\theta_1}^{\bar{\theta}} \exp\left(\int_{\theta_1}^{x} \eta(\tilde{x})\frac{\dot{c}(\tilde{x})}{c(\tilde{x})}d\tilde{x}\right)[1-g(x)]\frac{f^1(x)}{1-F^1(\theta_1)}dx.\end{aligned}$$

motives are:

$$\underbrace{\mathrm{E}\Big[\tau_L(\theta^2)w_{e2}l_2\Big]}_{\text{Second-Best Motive}} + \underbrace{\mathrm{E}\Big[\mathcal{MB}(\theta^2)\frac{\mu(\theta^2)}{f^2(\theta_2)}\frac{\epsilon_{\theta 2}}{\theta_2}\big(1-\rho_{\theta e2}\big)\Big]}_{t=2\text{'s Social Insurance Motive}} + \underbrace{\mathrm{Cov}\Big(\frac{1}{u'(c_2)}, \mathcal{MB}(\theta^2)\Big)}_{\text{Incentive Motive}}.^{23} \quad (3.8)$$

The first part of (3.8) is referred to as the second-best motive which captures the expected marginal revenues the planner gains from increasing the worker's work experience. In fact, terms like this one can be found in most optimal tax literature where the planner distorts more than one market. The planner takes into account the impact of a change in a tax instrument in one market on the other markets that also face distortions by other tax instruments. Since $\tau_L(\theta^2) \geq 0$, the fact that $w_{e2}l_2 \geq 0$, and that for any given shock θ_1 there is at least one subsequent shock θ_2 where the planner requires $l(\theta^2) > 0$ the first part must be positive. This implies that the second-best motive pushes for a lower optimal first period wedge which encourages work experience accumulation.

The second part of (3.8) captures the social insurance motive of the planner and how imperfect information in the second period affects pre and after-tax insurance. The second part of (3.8) is composed of the marginal benefit of work experience, the insurance motive in the second period $\mu(\theta^2)$, the density $f^2(\theta^2)$, the wage elasticity with respect the shock $\epsilon_{\theta 2}$, and the Hicksian complimentary coefficient $\rho_{\theta e2}$. The sign of this part depends entirely on whether $\rho_{\theta e2}$ is smaller, equal or greater than one. The planner will want to encourage work experience when the elasticity of wage with respect to shocks will be decreasing in work experience, i.e. $\rho_{\theta e2} < 1$, and discourage it when the elasticity of wage with respects to shocks is increasing in work experience, i.e. $\rho_{\theta e2} > 1$. The rationale being that the planner is able to diminish exposure to risk by increasing work experience whenever $\rho_{\theta e2} < 1$. The opposite is true when $\rho_{\theta e2} > 1$ making him want to discourage work experience. This result is similar to the optimal net wedge on human capital expenses in Stantcheva (2014b).

[23] These three parts are multiplied by $-\phi_y/R$, which is removed for expositional simplicity.

However in this model. it is the expected value of similar terms in the second period that matters as the effect of work experience is felt in the second period and not in the period it was accumulated as in Stantcheva (2014b).[24]

The third part can be seen as providing incentive in a second best framework. In contrast with the second part, it is linked to the marginal benefit of work experience in terms of labor disutility and not just benefits in terms of wage. Suppose that the planner was able to perfectly observe types in the second period, he could perfectly insure the worker, i.e. $c(\theta^2) = \hat{c}$ for all θ^2, which would imply that the covariance term would be zero. As the planner is incapable to do this in an asymmetric information set up, he must offer incentives for truthful reporting. Assumption 1 states that the optimal allocation features non-decreasing consumption in period two with respect to shock θ_2, thus the covariance term will be positive if the marginal benefit of work experience is increasing in type. This will depend on two factors, one behavioral i.e. whether labor effort is generally increasing with type and the other technological, whether the Hicksian complementarity coefficient is greater or lesser than one as is shown in the following equations:

$$\frac{\partial \mathcal{MB}(\theta_1, \theta_2)}{\partial \theta_2} = \big(1 + \alpha_2(l_2)\big)\frac{\epsilon_{e2}}{e_2}\dot{l}(\theta^2) - \frac{h'(l_2) l_2 w_{e2}}{w_2}\frac{\epsilon_{\theta 2}}{\theta_2}(1 - \rho_{\theta e2}). \tag{3.9}$$

[24]Note that the first and second terms of (3.8) can be combined. Using the definition of the optimal second period labor wedge it is possible to insert it in the insurance motive term. Rearranging and combining the first two terms of (3.8) it is possible to write the optimal first period labor wedge in the following way:

$$\begin{aligned}\tau_L^*(\theta^1) = &\frac{\mu(\theta^1)}{f^1(\theta_1)}\frac{h'(l_1)}{w_1}\frac{\epsilon_{\theta 1}}{\theta_1}(1 + \alpha_1(\theta^1)) \\ &- \frac{\phi_y}{R}\left\{\mathrm{E}\left[\tau_L^*(\theta^2) w_{e2} l_2 \left(\frac{2 + \alpha_2(\theta^2) - \rho_{\theta e2}}{1 + \alpha_2(\theta^2)}\right)\right] + \mathrm{Cov}\left(\frac{1}{u'(c_2)}, \mathcal{MB}(\theta^2)\right)\right\}.\end{aligned}$$

From this formulation two things come to light. First, the Hicksian complementarity parameter $\rho(\theta^2, e_2)$ must be at least above 2 to push the value of the combined terms to be negative and thus push the first period labor wedge upwards. Second, it highlights the importance of the second period labor elasticity apart from the level of $\tau_L^*(\theta^2)$ in determining the first period labor wedge.

The more a worker works in the second period the more he will gain from an increase in his wage coming from an increase in the shock. Thus, if labor is increasing in type, the marginal benefit of work experience should, in part, also be increasing in type. But as it was argued above, whenever the Hicksian complementarity coefficient is below 1 the wage elasticity with respect to work experience is decreasing in type. Thus in such a situation the marginal benefit of work experience will, in part, be decreasing in shock. So, in that case the effect can push in different directions. If for example labor effort is increasing in shock and the Hicksian complementarity coefficient is above 1 then the marginal benefit of work experience must be increasing in shock. If both second period consumption and the marginal benefit of work experience are increasing in shock, redistribution and information asymmetry would reduce the incentive to acquire work experience. Hence the planner will want to decrease the optimal first period labor wedge to incentivize workers to acquire work experience to counteract the disincentive effect of redistribution in the second period. This logic is reversed if the covariance term becomes negative, in this case the planner will wish to increase the optimal first period labor wedge to discourage work experience accumulation.

Intertemporal Wedge:

As utility is separable between consumption and labor effort in this model, the optimal condition (3.1) has the “inverse Euler equation” feature:

$$\frac{1}{u'(c(\theta^1))} = \frac{1}{R\beta} \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta^2))} f^2(\theta_2) d\theta_2.$$

Proposition 2. *Suppose that the relaxed problem solves the original problem. Then the* *optimal intertemporal wedge is positive, $\tau_K^*(\theta^1) > 0$, and satisfies:*

$$\tau_K^*(\theta^1) = 1 - \frac{\left[\int_{\underline{\theta}}^{\bar{\theta}} [u'(c(\theta^2))]^{-1} f^2(\theta_2) d\theta_2\right]^{-1}}{\int_{\underline{\theta}}^{\bar{\theta}} u'(c(\theta^2)) f^2(\theta_2) d\theta_2} > 0.$$

The proof can be found in Appendix A.2. The result that the intertemporal wedge is positive is obtained by applying Jensen's inequality to the definition of the optimal wedge. This result is found in several NDPF papers such as Kocherlakota (2005); Farhi & Werning (2013); Stantcheva (2014a).[25] The intuition for this distortion is that the planner seeks to discourage savings as savings make separating workers who received different shocks more difficult. For the rest of the paper I will sometimes refer to the intertemporal wedge as the savings wedge.

3.3 Uncertainty for older workers: the CES wage and log-normal distribution of shocks case

To further investigate the impact of introducing uncertainty in the later period of life on the age structure of the tax system, this subsection derives additional characteristics on labor wedges by specifying the functional form of the wage function and the probability distribution of the shocks.[26] Let the functional form of the wage function be

$$w_t(\theta_t, e_t) = (\theta_t^{1-\rho} + e_t^{1-\rho})^{\frac{1}{1-\rho}}. \tag{3.10}$$

[25]See Golosov et al. (2015) for a more general result featuring non-separable preferences between consumption and labor effort.

[26]Many of the results in this subsection hold for more general wage functions as shown in the Appendix.

For this wage function the Hicksian complementarity coefficient, the wage elasticity with respect to shock, and to work experience are

$$\rho(\theta_2, e_2) = \rho, \quad \epsilon_{\theta t} = \left(\frac{\theta_t}{w_t}\right)^{1-\rho} \text{ and } \quad \epsilon_{et} = \left(\frac{e_t}{w_t}\right)^{1-\rho}. \tag{3.11}$$

The probability distribution of the shocks considered in this subsection is the log-normal distribution $\ln \mathcal{N}(\mu, \sigma^2)$ where μ and σ are, respectively, the mean and standard deviation of the random variables natural logarithm and the probability density function of the log-normal distribution is

$$f^t(\theta) = \frac{1}{\theta\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln\theta - \mu)^2}{2\sigma^2}\right). \tag{3.12}$$

3.3.1 No uncertainty in the second period

As a benchmark, consider the characteristics of the optimal allocation when there is no uncertainty in the second period, i.e. $\Pr\{\theta_2 = \tilde{\theta}\} = 1$. This implies that the wage of each worker is his accumulated work experience, i.e. $w_2(\theta^1) = (\tilde{\theta}^{1-\rho} + e_2(\theta^1)^{1-\rho})^{\frac{1}{1-\rho}}$.

Corollary 1. *Supposing that the relaxed problem solves the original problem, when* $Pr\{\theta_2 = \tilde{\theta}\} = 1$ *and the second period wage of workers is* $w_2(\theta^1)$ *the optimal intertemporal wedge is zero,* $\tau_K^*(\theta^1) = 0$*, the second period optimal labor wedge is zero,* $\tau_L^{2*}(\theta^1) = 0$ *and the first period labor wedge is :*

$$\tau_L^{1*}(\theta^1) = \frac{\mu(\theta^1)}{f^1(\theta_1)} \frac{h'(l_1)}{w_1} \frac{\epsilon_{\theta 1}}{\theta_1} \big(1 + \alpha_1(\theta^1)\big).$$

In this special case there is no longer an information problem in the second period as the planner is able to observe work experience accumulation. In this situation the planner does not need to distort the second period labor decision. Because there is no need to distort the second period, there is also no reason for the planner to distort the savings decision as it will not relax the incentive problem in the second period (or in the first period).

This implies that for each worker consumption in both periods is smoothed and follows $u'(c_1(\theta^1)) = \beta R u'(c_2(\theta^1))$. However, since there is still an information problem in the first period, consumption cannot be equalized in the first period for all workers and by extension second period consumption is also not equalized. This result is similar to the one obtained in Atkinson & Stiglitz (1976) where there is no incentive to tax savings when preferences are separable.

For the first period labor distortion, comparing (3.5) with the optimal first period wedge in Corollary 1, the difference between those formulas are the three terms of (3.8). As there is no distortion of the second period labor decisions nor any uncertainty, encouraging or discouraging work experience accumulation by distorting the first period labor decision is no longer necessary. Only the redistribution motive and the efficiency motive of the first period remain.

Note that in this no uncertainty scenario, the results on the second period labor wedge are different then those of Krause (2009) and Best & Kleven (2013). The reason for this is twofold. The first is that information available to the planner is different. In this model, if the worker deviates in the first period, the planner will know his second period wage as he can observe the worker's income. This is not the case in both Krause (2009) and Best & Kleven (2013) as work experience is only a function of labor effort. The planner seeks to reduce the incentive problem by distorting the second period labor choice. Furthermore in Best & Kleven (2013), the planner must use age-dependent taxes, this restricts the tools he has to use to tackle the information problem. The informational structure of this paper offers the most contrasting results in the case of no uncertainty as the distortions when young are necessarily greater than the ones when old.

3.3.2 Impact of complementarity and variance on the second period labor wedge

In opposition to the riskless environment just highlighted above, uncertainty in the second period forces the planner to distort the second period labor market as shown by (3.6). The assumptions made on the wage function and the probability distribution allows to characterize in more detail the labor distortion at the right tail of the distribution. This is done by applying the methodology of Golosov et al. (2015) which consists of investigating the asymptotic behavior of (3.6) as θ_2 goes to infinity. For this, the assumption of a bounded distribution needs to relaxed and assume that the shocks are non-negative, i.e. $\theta_2 \in \Theta = \mathbb{R}_+$ in $t = 2$.

Assumption 2. *$\alpha_2(\theta)$, $\eta_2(\theta)$ have finite, non-zero limits $\bar{\alpha}$, $\bar{\eta}$; $\frac{c_2(\theta)}{y_2(\theta)}$ has a finite non-zero limit; $\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}$ has a finite limit; $\frac{\dot{c}_2(\theta)/c_2(\theta)}{\dot{y}_2(\theta)/y_2(\theta)}$ has a limit as $\theta \to \infty$.*[27]

Assumption 2 is made to guarantee well-behaved cases. In addition, note that as $\theta_2 \to \infty$ the CES wage function (3.10) has a limit $\bar{\epsilon}_{\theta 2}$. For parameter values of $\rho \in [0, 1]$ the wage elasticity with respect to shock goes to 1 as the shock goes to infinity, i.e. $\epsilon_{\theta 2}(\theta, e_2) \to 1$ $(\theta \to \infty)$. For a parameter value above 1 the wage elasticity with respect to shock goes to 0 as the shock goes to infinity, i.e. $\epsilon_{\theta 2}(\theta, e_2) \to 0$ $(\theta \to \infty)$. These two facts lead to the following Corollaries.

Corollary 2.a. *Under Assumption 2, a CES wage function with $0 \leq \rho \leq 1$ and $f^2(\theta)$ distributed $\ln \mathcal{N}(\mu, \sigma^2)$, as $\theta \to \infty$ the second period labor distortion is asymptotically equivalent to*

$$\frac{\tau_L^{2*}(\theta)}{1 - \tau_L^{2*}(\theta)} \sim A(\theta)B(\theta)C(\theta) \sim (1 + \bar{\alpha})\ \bar{\epsilon}_{\theta 2} \left(\frac{\sigma^2}{\ln \theta - \mu} \right),$$

where

$$\bar{\epsilon}_{\theta 2} = 1.^{28}$$

[27] For ease of exposition, we replace $c(\theta^2), y(\theta^2)$ by $c_2(\theta), y_2(\theta)$ and $\tau_L(\theta^2)$ by $\tau_L^2(\theta)$.

[28] Note that if the CES has a scaling factor κ as in the numerical simulations below the limit of the wage

As can be seen from Corollary 2.a, the optimal second period wedge as θ goes to infinity is shaped by the labor elasticity parameter $\bar{\alpha}$, the wage elasticity parameter with respect to second period shock $\bar{\epsilon}_{\theta 2}$ and the variance parameter of the log-normal distribution σ^2. Thus the greater the variance parameter, capturing uncertainty to the worker, the higher the labor distortion when old. The intuition of this result can be obtained by taking each part of (3.7) individually. First, consider $A(\theta)$ which measures the cost of distorting the labor decision. This part converges to a finite and positive limit as θ goes to infinity. The redistributive part $C(\theta)$ can also be shown to have a finite limit of 1.[29] It can be shown numerically that for much of the domain of the shock distribution of young workers, $C(\theta^2)$ climbs very quickly to its asymptotic value. Finally, the hazard ratio of the log-normal distribution can be shown to go to zero as θ goes to infinity, i.e. $B(\theta) \to 0$ $(\theta \to \infty)$. Even though $B(\theta) \to 0$ as θ goes to infinity this rate of convergence is rather slow in fact the tail behavior can be characterized by $B(\theta) \sim \frac{\sigma^2}{\ln \theta - \mu}$. All of these results combined give us Corollary 2.a.

In the case of $\rho > 1$ the result of Corollary 2.a no longer hold has the limit of the wage elasticity with respect to shock is zero and thus $\lim_{\theta \to \infty} A(\theta) = 0$. The following result considers the behavior of the limit of this elasticity.

Corollary 2.b. *Under Assumption 2, a CES wage function with $\rho > 1$ and $f^2(\theta)$ distributed*

elasticity with respect to shock would be $\kappa^{1-\rho}$.

[29] $C(\theta)$ can be written in the following way:

$$\lim_{\theta \to \infty} C(\theta) = \lim_{\theta \to \infty} [1 - \lambda_2 u'(c_2(\theta))] + \lim_{\theta \to \infty} \frac{\eta_2(\theta)}{\eta_2(\theta) + \alpha_2(\theta)} \lim_{\theta \to \infty} \frac{\tau_L^{2*}(\theta)}{1 - \tau_L^{2*}(\theta)}.$$

The limit of $C(\theta)$ can be into two distinct parts. The first part captures the value of extracting a dollar from a worker who received shock θ and as θ goes to infinity, so does $c_2(\theta)$ and thus the value of giving (or leaving) a dollar to shock type θ, measured by $g(\theta) \equiv \lambda_2 u'(c_2(\theta))$, goes to zero. The second part measures the income effect on labor effort from distorting labor on type θ. This income effect is of course influenced by the size of the limit of the second period labor distortion.

$\ln \mathcal{N}(\mu, \sigma^2)$ *, as* $\theta \to \infty$ *the second period labor distortion is asymptotically equivalent to*

$$\frac{\tau_L^{2*}(\theta)}{1-\tau_L^{2*}(\theta)} \sim A(\theta)B(\theta)C(\theta) \sim (1+\bar{\alpha})\left(\frac{e_2}{\theta}\right)^{\rho-1}\left(\frac{\sigma^2}{\ln\theta-\mu}\right).^{30}$$

Note that the change in the wage elasticity does not affect our characterization of $\lim_{\theta\to\infty} C(\theta)$ nor its interpretation. In this particular context the asymptotic behavior of $A(\theta)$ as θ goes to infinity is $A(\theta) \sim (1+\bar{\alpha})\left(\frac{e_2}{\theta}\right)^{\rho-1}$. Importantly, $A(\theta)$, contrary to $B(\theta)$, can go quickly to zero. In fact, the higher the value of ρ above 1 the faster $A(\theta)$ converges to 0. Comparing the results of Corollary 2.a and 2.b, for the right tail of the distribution of shocks, the optimal labor distortion when old, for a given realization of θ_1 in the first period, should be greater whenever $0 \leq \rho \leq 1$ compared to when $\rho > 1$. The intuition is that whenever the complementarity coefficient ρ is greater than one, a higher shock does not translate as much to a higher wage for a given level of work experience. Therefore the government does not gain anything by distorting a wider proportion of the distribution since the change in information rent gets smaller and smaller.[31]

4 Numerical Simulations

This section investigates the properties of the optimal wedges and how they are influenced by different key parameters. To compare the results derived with those of the literature, I proceed with a calibration exercise to match certain empirical moments of the wage distribution at each age and the wage elasticity with respect to work experience found in the meta-analysis of Best & Kleven (2013).

[30] We also obtain

$$\lim_{\theta\to\infty} C(\theta) = 1 + \frac{\bar{\eta}}{\bar{\eta}+\bar{\alpha}} \lim_{\theta\to\infty} \frac{\tau_L^{2*}(\theta)}{1-\tau_L^{2*}(\theta)}$$

[31] Note that the result from Corollary 2.b should extend to distributions other than log-normal that would give a positive top marginal tax rate like the pareto-lognormal distribution.

4.1 Functional Forms, Calibration and Computational Strategy

Functional Forms

The functional form for the per period utility function used is

$$\ln c - \frac{l^{1+\alpha}}{1+\alpha}, \tag{4.1}$$

where $\alpha > 0$. The function that transform income into effective work experience is

$$\phi(y) = y^{\delta}, \tag{4.2}$$

where $\delta \in (0, 2)$.[32] The wage has a the following CES functional form

$$w_t(\theta_t, e_t) = \kappa_t * (\theta_t^{1-\rho} + \xi e_t^{1-\rho})^{\frac{1}{1-\rho}}, \tag{4.3}$$

where κ_t and ξ are scaling parameters with ρ being the Hicksian complementarity parameter. For simplicity, I assume that all workers start with no accumulated work experience, i.e. $e_1 = 0$, and that $\kappa_1 = 1$ in the first period. This implies that in the first period

$$w_1(\theta_1) = \theta_1 \quad \text{and} \quad e_2(\theta_1) = [y(\theta_1)]^{\delta}. \tag{4.4}$$

Using both the wage function and the transformation function, the wage elasticity with

[32] I have no priors related to the value of δ. The range is chosen to insure that the worker's problem of choosing labor effort in the first period is concave.

respect to work experience and the wage elasticity with respect to shock, respectively, are

$$\epsilon_{e2} = \kappa_2^{1-\rho} * \xi * \left(\frac{e_2}{w_2}\right)^{1-\rho}, \quad \epsilon_{\theta 2} = \kappa_2^{1-\rho} * \left(\frac{\theta_2}{w_2}\right)^{1-\rho}.^{33} \tag{4.5}$$

For the calibrations the wage elasticity with respect to first period labor effort is used. This elasticity also coincides with the wage elasticity with respect to first period income

$$\gamma(\theta_2, e_2) \equiv \frac{\partial w_2}{\partial l_1}\frac{l_1}{w_2} = \frac{\partial w_2}{\partial y_1}\frac{y_1}{w_2} = \delta * \kappa_2^{1-\rho} * \xi * \left(\frac{e_2}{w_2}\right)^{1-\rho}. \tag{4.6}$$

Calibration

The different calibrations sets δ, κ_2 and ξ to match three target moments. The first target moment is the mean of the estimated wage elasticity with respect to experience in Best & Kleven (2013), i.e. $\hat{\gamma} = 0.29$.[34] The second and third target moments are the mean and standard deviation of the wages of head of household workers age 41 and over in the 2007 round of the Panel Study of Income Dynamics (PSID). These moments were obtained by taking the hourly wages and age from the 2007 PSID and splitting the sample in two where the cut off is the median age as in Best & Kleven (2013).[35] I then approximated the distribution $F^{young}(w)$ of the first period of life by a log-normal distribution with $(\mu_{young}, \sigma_{young}) = (2.805, 0.672)$ and the second period of life $F^{old}(w)$ with $(\mu_{old}, \sigma_{old}) = (3.165, 0.814)$.[36] For the calibration and a majority of the numerical simulations I imposed that $F^1(\theta) = F^2(\theta) = F^{young}(w)$. The estimated distribution $F^{old}(w)$ is then used to obtain the mean and standard deviation of wages in the second period of life.

[33] Thus for $0 \leq \rho \leq 1$ this implies that $\epsilon_{\theta 2} \to \kappa_2^{1-\rho}$ as $\theta_2 \to \infty$ and for $\rho > 1$ it implies that $\epsilon_{\theta 2} \to 0$ as $\theta_2 \to \infty$.

[34] In their numerical simulations this elasticity is then taken to be the wage elasticity with respect to first period labor effort. They use three scenarios where $\gamma = 0$, $\gamma = 0.2$ and $\gamma = 0.4$. A similar logic is followed in taking the wage elasticity with respect to experience to be the wage elasticity with respect to first period labor.

[35] To make sure the sample is representative I only keep observations that are related to the original SRC(Survey Research Center) sample.

[36] To approximate the distribution the command *fitdistr* from the MASS package in R was used. See `http://cran.r-project.org/web/packages/MASS/index.html`.

I then created a *model economy* where heterogeneous workers, who will face risk in the second period of their life, decide how much to work and consume in both periods of their life, and how much to save at gross interest rate R.[37] The following key parameters were set to be $\alpha = 2$, $\beta = 0.6$ and $R = 1/\beta$. The assumption on β and R implies that savings are for insurance purposes. The 2007 US tax system is approximated by a linear labor tax following the methodology of Jacquet et al. (2013). Due to limited empirical evidence on the complementarity between the shock and work experience, several calibrations were undertaken where ρ takes on different values.[38][39][40][41]

Computational Strategy

The computational strategy closely follows the strategy used in Golosov et al. (2015). The recursive formulation of the planner's problem is used to solve a two period discrete-time

[37]The debt limit was set to 0, i.e. savings must be non-negative.

[38]As pointed out in Stantcheva (2014b), there is some evidence of complementarity with respect to skill and on-the-job training. See and OECD (2004) and Huggett et al. (2011).

[39]In every calibration, the parameters are chosen as to minimize the following loss function

$$LOSS = \left|\frac{\tilde{\gamma}}{0.29} - 1\right| + \left|\frac{\tilde{w_2}}{\bar{w}_2^{PSID}} - 1\right| + \left|\frac{\tilde{sd_2}}{\bar{sd}_2^{PSID}} - 1\right|,$$

where $\tilde{\gamma}$, $\tilde{w_2}$, $\tilde{sd_2}$ are, respectively, the mean of the wage elasticity with respect to first period income, the mean second period wage and the standard deviation of the second period wage from the model economy. And $\bar{w}_2^{PSID}, \bar{sd}_2^{PSID}$ are the first two moments of the wage distribution $F^{old}(w)$ obtained from the PSID. Note that $\tilde{\gamma}$ is the discrete version of

$$\tilde{\gamma} = \int\int \delta\kappa_2^{1-\rho}\xi\left(\frac{e(\theta_1)}{w_2(\theta_2, e(\theta_1))}\right)^{1-\rho} f^1(\theta_1)f^2(\theta_2)d\theta_1 d\theta_2.$$

[40]The third calibration, Case 3, considers when ρ tends to 1. From the CES functional form of the wage function assumed above, the wage function obtained as ρ tends to 1 is

$$w_t(\theta_t, e_t) = \kappa_t \theta_t^{\frac{1}{1+\xi}} e_t^{\frac{\xi}{1+\xi}}$$

with

$$\epsilon_{e2} = \frac{\xi}{1+\xi},\ \epsilon_{\theta 2} = \frac{1}{1+\xi} \text{ and } \gamma(\theta_2, e_2) = \delta\frac{\xi}{1+\xi}.$$

[41]The values of the three moments under each calibration can be found in Appendix XXXX in table XXXX.

Table 1: Calibration Parameters

Parameter	Case 1	Case 2	Case 3	Case 4	Case 5	Type
ρ	0.2	0.5	$\to 1$	1.2	1.5	Exogenous
α	2	2	2	2	2	Exogenous
β	0.6	0.6	0.6	0.6	0.6	Exogenous
R	$1/\beta$	$1/\beta$	$1/\beta$	$1/\beta$	$1/\beta$	Exogenous
δ	0.975	1.15	1.325	1.4	1.3	Endogenous
ξ	0.4	0.26	0.275	0.325	0.4	Endogenous
κ_2	1.1	0.95	1.4	5.5	3	Endogenous

dynamic program with a two-dimensional continuous state space. It is solved by backward induction. First, the second period value function is approximated by tensor products of Chebyshev orthogonal polynomials evaluated at root nodes. Each node problem is solved using an interior-point algorithm which consists of replacing the nonlinear programming problem by a series of barrier subproblems controlled by a barrier parameter.[42]

For each second period node problem, a discrete-type version of (2.19) is solved where only the downward incentive constraints linking two adjacent workers and a monotonicity condition on consumption is assumed.[43] As was shown by Hellwig (2007) both the continuous type model and discrete type model are in a sense mathematically equivalent. Bastani (2014) has demonstrated that simulations using either continuous models or discrete type models produced similar results as long as the number of types used to represent the skill distribution is large enough. These results lend confidence that the numerical simulations of this paper are similar to the continuous type problem of (2.19) while taking advantage of solving a well-behaved convex programming problem of the discrete type model.

[42]We use algorithm 1 in KNITRO. See `http://www.artelys.com/tools/knitro_doc/` .

[43]I have also solved each problem without imposing the monotonicity condition and found the same solutions. This condition is imposed as it speeds up computation time required to solve each problem and it imposes the second order condition explicitly as a set of constraints.

Once the second period value function is approximated, it is incorporated in the first period problem (2.20). I imposed $e_1 = 0$ and looked for the υ_0 such that $K(\upsilon_0, 0, 1) = 0$. Assuming that the first period value function is continuous, a value of υ_0 big enough is chosen so that $K < 0$ and a discrete version of the first period problem is solved using the chosen υ_0.[44] Following this step, another value of υ_0 small enough is chosen such that $K > 0$. I then proceed by bisection and solved a new first period problem with an updated υ_0, and this is done until the value where $K(\upsilon_0) = 0$ is found. With this solution, the optimal allocation is obtained by forward induction. The solution is then verified to be incentive compatible. Note that for all five cases, the planner is Utilitarian, i.e. $a(\theta_1) = 1 \; \forall \; \theta_1$.

4.2 Results

I start by analyzing the results of the benchmark case of $\rho = 0.5$, i.e. Case 2 as it features some clear results. The upper left graphic of Figure 1 illustrates many of the important results found in this paper. It features the first period labor wedge and the expected value of the second period labor wedge a worker of wage θ_1 will face.[45] Note that except for the very low wages and high wages in the first period, workers usually face a lower first period wedge compared with the expected value of the second period wedge.[46] In fact, in this simulation, 77.3% of the workers in the population are in that situation. That is true since a significant mass of workers have shocks below the average value of the first period shock, i.e. 20.73$.

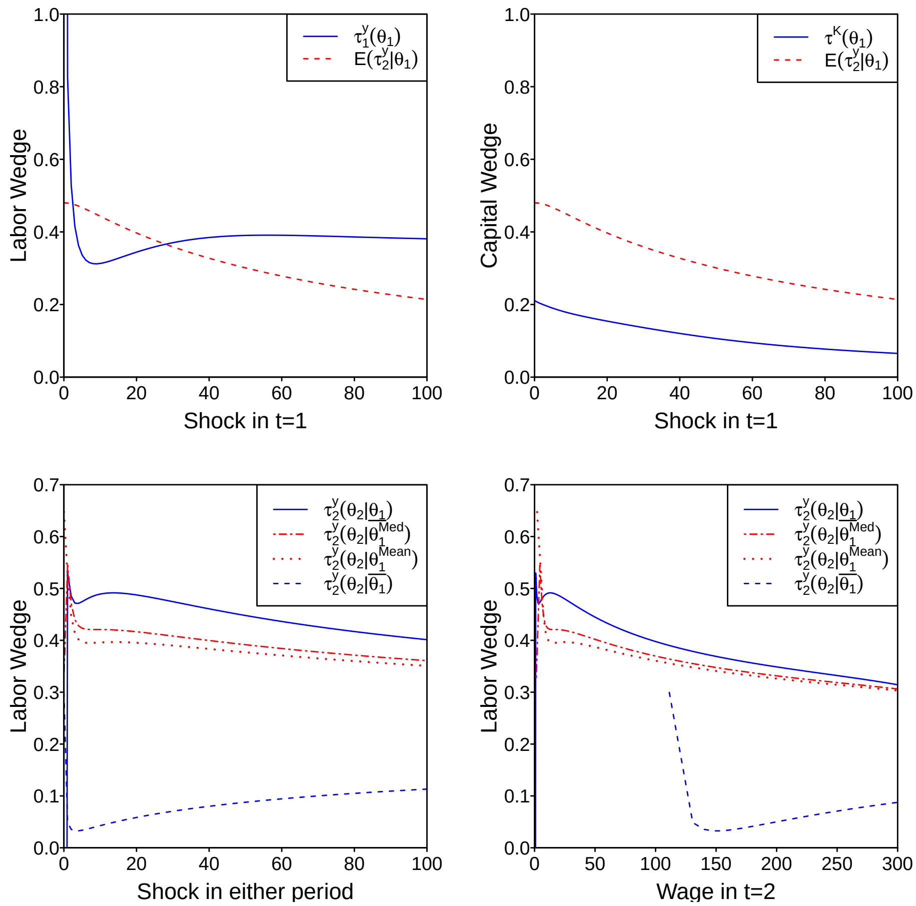
It is possible to see this more clearly by looking at the lower left graphic where the whole second period optimal labor wedge schedule is graphed for workers that received the lowest,

[44] In the first period, no monotonicity condition is imposed, but the downward incentive constraints are still used. The monotonicity and incentive compatibility is verified ex-post.

[45] In the simulations shocks go from 0.01 to 499.01. I restricted the x-axis since it allows a better look at what happens to the average skilled individuals and lower. Note that the share of the population that has a first period shock lower or equal to 100 is 99.6%.

[46] Very low skill workers face optimal first period wedges above 100% as it requires a high distortion to discentivize labor effort as work experience is valuable in the second period.

Figure 1: Optimal Wedges from Calibration $\rho = 0.5$

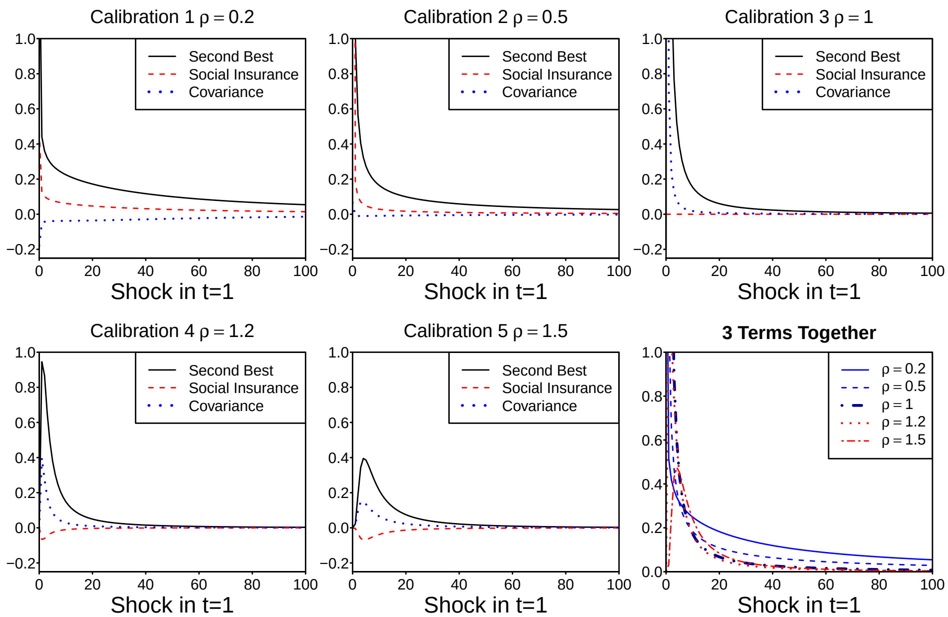


the median, the mean shock and the highest shock, respectively $\underline{\theta_1}$, θ_1^{Med}, θ_1^{Mean} and $\bar{\theta_1}$.[47] This gives the upper and lower bound of the optimal wedge schedule in the second period for workers of different history θ^2. The workers who received the median and mean shock in the first period should have second period labor wedges above the labor wedge they face in the first period. In fact they should face a higher wedge then most workers in the first period except for those at the very bottom of the distribution in period 1. In the case of the highest shock in the first period, they will face a greater labor wedge in the second period unless they also receive the highest shock in the second period. The upper right figure plots the optimal savings wedge and the expected value of the second period labor wedge. These two curves feature similar patterns as the savings wedge is used to help in separating workers of different types in the second period.

To further illustrate the forces shaping the first period labor wedge, Figure 2 shows the values of the three motives for each first period shock θ_1 for all calibrations. The first thing to notice is that the second-best motive appears to be dominant force out of all the three motives. As was shown in the section above the other two motives are influenced by whether the complementarity parameter is below or above one. The social insurance motive is positive for both calibrations where ρ is below one and negative for the calibrations where ρ is above one. The incentive motive captured by the covariance of consumption and the marginal benefit of work experience appears to also be prominently determined by the complementarity coefficient. In Case 1, the covariance is slightly negative since ρ is smaller than one. On the contrary, in Case 4 and 5, the covariance is positive as the complementarity coefficient above one ensures that the marginal benefit of work experience is increasing. One of the interesting results comes from looking at the bottom right graph. It shows that when ρ is above one the combined value of three terms appear to be the highest at the bottom of the distribution. This would imply that it is the calibrations where you would expect the optimal first period labor wedge to be the lowest. This turns out not to be the case, as shown

[47] In the simulations these values are $\underline{\theta_1} = 0.01\$$, $\theta_1^{\text{Med}} = 17.01\$$, $\theta_1^{\text{Mean}} = 20.73\$$ and $\bar{\theta_1} = 499.01\$$.

Figure 2: Optimal Labor Wedge in t=1: Three Motives



in Table 3, in fact it is much higher than the first period labor wedges in the calibration where ρ is below one. One plausible explanation for this result could be that the higher ρ increases the wage elasticity with respect to labor effort and therefore renders the first period labor supply much less elastic, at least at very low levels of income, since in this situation the payoffs to work are much greater. This explanation would be in the spirit of the results found in Best & Kleven (2013). In fact, as illustrated by Table 2, parameter ρ appears to be a determining factor in whether the average labor distortion is increasing with age are not in the calibrations.

Table 2: Average Wedges in Each Period: All Cases

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\rho = 0.2$	0.342	0.385	65.2%	0.151
$\rho = 0.5$	0.343	0.402	77.3%	0.156
$\rho \to 1$	0.418	0.407	72.2%	0.159
$\rho = 1.2$	0.502	0.401	56.6%	0.164
$\rho = 1.5$	0.518	0.391	61.5%	0.16
No WE	0.294	0.493	97.9%	0.196

Source: Authors' calculations. Numbers rounded.

The pattern of the average wedges in each period of life obtain from the optimal allocation has important implications for tax policy as demonstrated in Farhi & Werning (2013). In a similar manner as these authors and Stantcheva (2014b), I consider a simpler set of age-dependent tax policies. For each case, I evaluate the welfare gains to an economy where the age-dependent linear taxes are set at the cross-sectional average of the labor wedges and saving wedges compared to a *laissez-faire* economy. The tax revenues generated in a given period by the linear taxes are used to finance a demogrant that is given back to workers in the same period. Also note that the linear tax on saving incomes is collected in the second period of life. The welfare gains are measured by the constant percentage increase in consumption (Δ) offered in all periods and all state to the *laissez-faire* economy required to obtain the same expected lifetime utility given by a specific allocation, i.e.

$$\sum_{t=1}^{2}\beta^{t-1}\int\left[u\big((1+\Delta)c^{LF}(\theta^t)\big)-h\left(\frac{y^{LF}(\theta^t)}{w_t(\theta_t,e_t^{LF}(\theta^{t-1}))}\right)\right]f^2(\theta_2)f^1(\theta_1)d\theta_2 d\theta_1$$
$$=\sum_{t=1}^{2}\beta^{t-1}\int\left[u\big(c^*(\theta^t)\big)-h\left(\frac{y^*(\theta^t)}{w_t(\theta_t,e_t^*(\theta^{t-1}))}\right)\right]f^2(\theta_2)f^1(\theta_1)d\theta_2 d\theta_1, \tag{4.7}$$

where $\{c^{LF}, y^{LF}\}$ and $\{c^*, y^*\}$ are respectively the *laissez-faire* allocation and the considered allocation. The *laissez-faire* economy used in this analysis features no constraint to saving or borrowing, and workers can do so at gross interest rate R.

Table 3: Welfare Gains: All Cases

	Case 1	Case 2	Case 3	Case 4	Case 5
ρ	0.2	0.5	$\rightarrow 1$	1.2	1.5
% increase for Second Best	10.17%	10.35%	10.87%	12.64%	4.85%
% increase for age-dependent linear taxes	8.87%	9.32%	9,84%	10.33%	3.89%
% of simple policy in terms of Second Best	87.26%	89.99%	90.46%	81.72%	80.16%

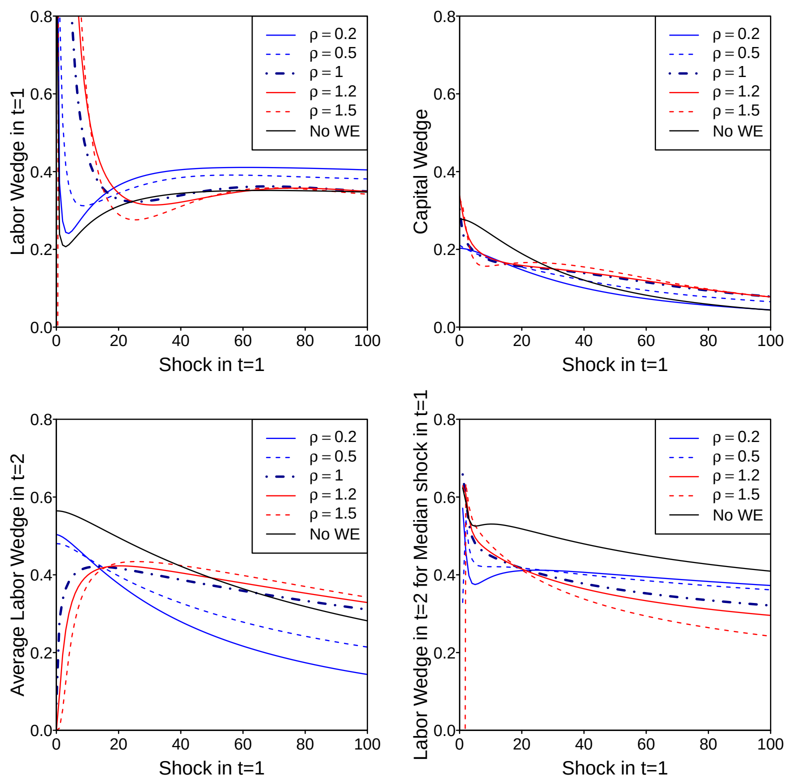
Source: Authors' calculations. Numbers rounded.

As it can be seen from Table 3, the age-dependent linear tax policies capture a large portion of the welfare gains obtained from the fully history-dependent non-linear tax policy. However, the simpler tax policies seem to be less effective when ρ is greater than one. As can be seen from Figure 3, in both periods the labor wedges appear to be less linear than the ones from the cases where ρ is lower than one. Also note that the difference en welfare gain for the simulation are lower for Case 5. This maybe due to the fact that this calibration features the highest average second period wage as can be seen in Table XX shown in Appendix C.1.

4.2.1 Changes in ρ

To explore in greater details the effects of the value of the Hicksian complementarity parameter two sets of simulations are presented. One set of simulation has the values of the

Figure 3: All Cases



coefficient ρ below one and the other above one. This is done because the behavior of the wage function changes drastically as ρ goes from values below one to above one.[48] The first set of simulations take the calibrated parameters δ and ξ from Case 1, i.e. $\rho = 0.2$, and increase ρ to 0.5 and 0.8. For each new change in ρ, κ_2 is adjusted to match the mean of the estimated mean of the second period wage.[49] The second set of simulations takes Case 4, i.e. $\rho = 1.2$ as its basis and ρ is increased to 1.5 and 1.8. For the later simulations, the parameter δ and ξ from Case 4 are kept and κ_2 is adjusted to match the first moment of the second period wage. In addition, the results obtained when there is no work experience (No WE) in the model are reported. The second period wage in this scenario is $w_2 = \kappa_2 * \theta_2$. For this model, κ_2 was calibrated again to match the estimated mean second period wage.

Looking at Figure 4 several things jump out as ρ is increased in the first set of simulations, i.e. where $\rho < 1$. The first is how the first period wedge increase for the very low skilled workers but decreases for workers who have skill levels greater than 10$. This shift pushes down the average labor wedge when young by about 2%. As shown in Table 4, the mean first period labor wedge goes from 34.2% to 32.1%.

The change in ρ also has an impact on the second period labor wedge. In the bottom left figure of Figure 4, the average second period labor wedge schedule flattens with an increase in ρ. Furthermore, the lowest skilled worker in the first period's average second period labor wedge goes from the highest in the population to one of the lowest of most workers in the economy. However, this flattening does not really impact the cross-sectional average second period labor wedge. A similar pattern emerges for the savings wedge schedule.

[48]This fact extends to the optimal labor wedge from both our analytical results of the asymptotic behavior of the second period labor wedge and the numerical results found in Figure 2.

[49]As ρ increases from 0.2 to 0.8 the calibrated κ_2 is decreased from 1.1 to 0.325. In the end, this slightly reduces both the wage elasticity with respect to labor and the standard variation of the second period wage.

Figure 4: Optimal Wedges: $\rho = 0.2$ with Changes in ρ

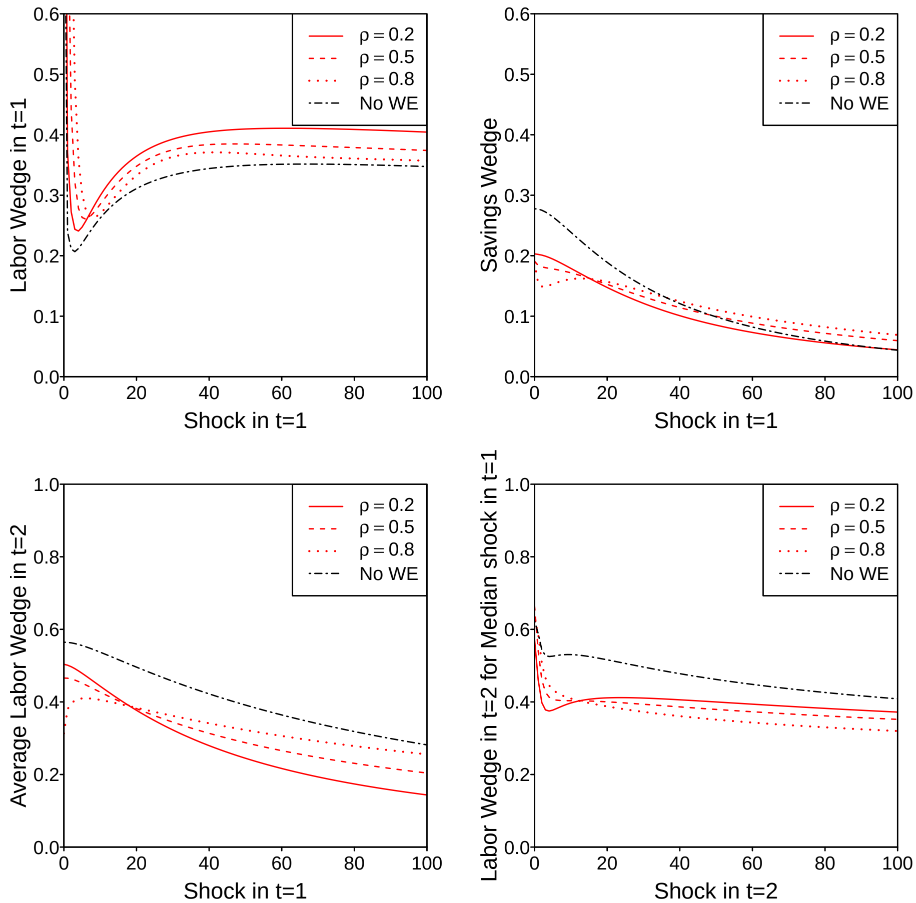
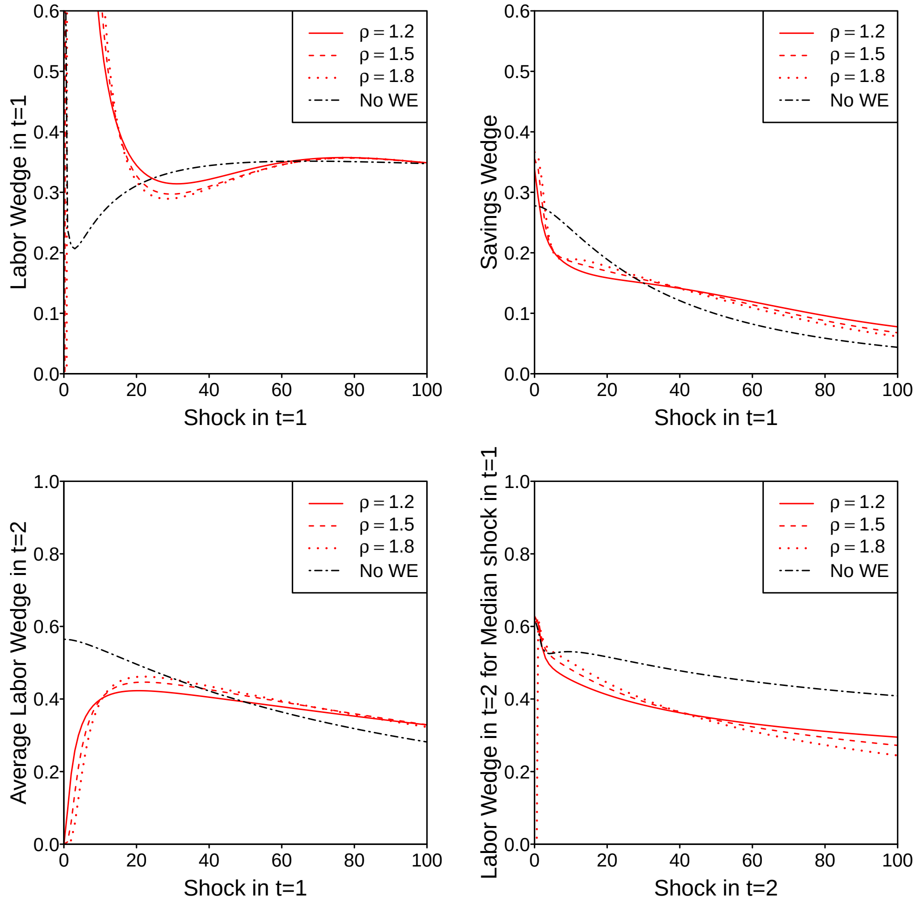


Table 4: Average Wedges in Each Period: Using Calibration of $\rho = 0.2$ with Changes in ρ

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\rho = 0.2$	0.342	0.385	65.2%	0.151
$\rho = 0.5$	0.329	0.386	72.1%	0.152
$\rho = 0.8$	0.321	0.386	79.5%	0.15
No WE	0.294	0.493	97.9%	0.196

Source: Authors' calculations. Numbers rounded.

Figure 5: Optimal Wedges: $\rho = 1.2$ with Changes in ρ



The second set of simulations, i.e. $\rho > 1$, gives somewhat similar results. The top left figure shows the results of increasing ρ from 1.2 to 1.5. It increases the optimal first period labor wedge for workers with skills lower than the median skill shock but also lowers it for a section of mid to high skill range. This change increases the cross-sectional average labor wedge of the young workers as can be seen in Table 5.

The second period labor wedge follows a similar pattern as the first set of simulations with $\rho < 1$. Increasing ρ lowers the average second period labor wedge the first period low skill shock levels will face in the second period. But at around first period shock of 10$, the average labor wedge increases for much of the rest of the distribution. But as mentioned above, a higher value of ρ increases the wage elasticity with respect to labor effort at the bottom of the distribution which may explain the reason why the cross-sectional average appear to be higher in the first period of life compared to the one in the second period. It is also possible to see that the increase in ρ can be seen to lower the tail of the second period labor wedge of the median shock worker. This result is inline with the analytical result derived above.

One thing to notice is that in both sets of simulations the relationship between the cross-sectional averages of each life period is not altered. For simulations with $\rho < 1$, the cross-sectional average of the labor wedge is smaller when young than when old. And for those with $\rho > 1$, the cross-sectional average of the labor wedge is greater when young than when old.

Table 5: Average Wedges in Each Period: Using Calibration of $\rho = 1.2$ with Changes in ρ

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\rho = 1.2$	0.502	0.401	56.6%	0.164
$\rho = 1.5$	0.569	0.407	60.9%	0.171
$\rho = 1.8$	0.637	0.406	60.8%	0.175
No WE	0.294	0.493	97.9%	0.196

Source: Authors' calculations. Numbers rounded.

4.2.2 Changes in δ

The next set of simulations investigates the role of the parameter that influences how income is transformed into work experience. It is the parameter that more directly influences the elasticity of future wages with respect to labor effort. Contrary to changes in ρ its effect is felt by all workers not more keenly felt by specific parts of the distribution. Taking Case 2 and Case 4 as the basis for the simulations, the value of δ is lowered for each set of simulations.[50] For each reduction of δ an extra calibration is done where κ_2 is adjusted such that the mean second period wage matches the estimated mean of the second period wage.[51]In Case 2 the reduction in δ decreases radically the average wage elasticity with respect to labor effort and slightly increases the variation of the second period wage. Whereas in Case 4 the reduction in δ decreases slightly the average wage elasticity with respect to labor effort but slightly decreases the variation of the second period wage.

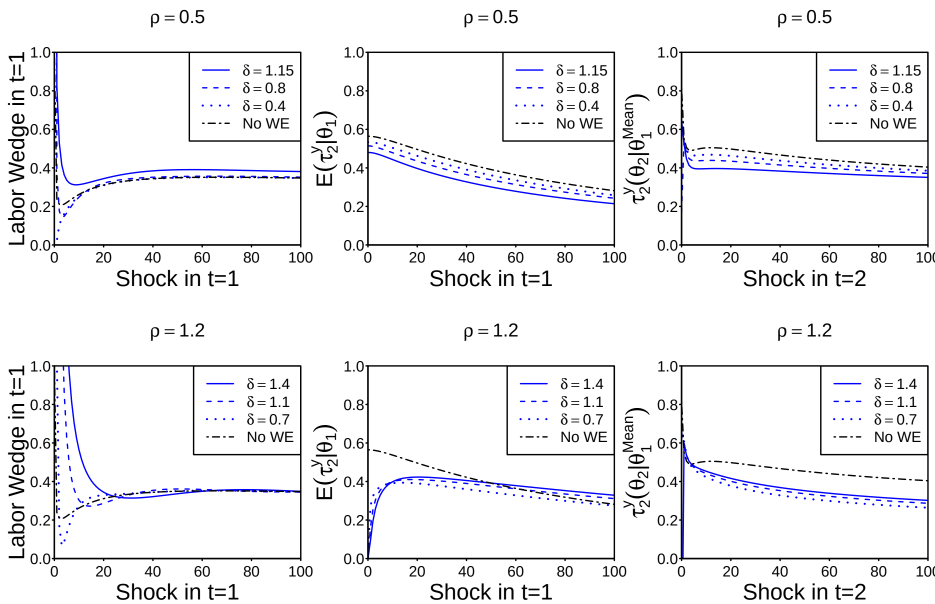
Results from decreasing δ can be observed in Figure 6 and Table 6. I find that the lower δ is, and lower the elasticity of the second period wage is with respect to labor, the lower the optimal wedge schedule in the first period in both sets of simulations. From the top and bottom left figures, it can be seen that the effect is more pronounced for the low skilled. The effect is even more drastic for simulations based on Case 4. It is such that the result of higher distortions for the young compared to the old is reversed. The results here are in line with the numerical results of Best & Kleven (2013) who finds that a decrease in γ usually leads to lower labor distortions for the young. However in contrast with their result, and probably due to the risky nature of the environment considered in this paper it is possible for young workers to be taxed less than older workers.

What happens in the second period of life differs in both sets of simulation. As can be seen in both Figure 6 and Table 6, labor wedges are increased in the case where $\rho = 0.5$ but decrease in case where $\rho = 1.2$. This is most likely due to the change in riskiness of the second period wage brought about by the change in parameter δ. As mentioned above, in the simulations based on Case 2, the variation of the second period wage is increased and hence the planner is required to offer more insurance and thus needs to distort labor more. However, this is not the case for simulations with $\rho = 1.2$ and the decrease in δ is also

[50]Recall that $\delta = 1.15$ is used in Case 2 and $\delta = 1.4$ in Case 4.

[51]The impact of those changes on a model economy can be seen in Table XX1 found in Appendix C.1.

Figure 6: Optimal Wedges: $\rho = 0.5$ and $\rho = 1.2$ with Changes in δ



accompanied by a decrease in variation of the second period wage and thus less reasons to distort labor. Both of these effects are minor compared to the strong effect the change in δ has on the first period labor wedge.

Table 6: Average Wedges in Each Period: $\rho = 0.5$ and $\rho = 1.2$ with Changes in δ

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\rho = 0.5$				
$\delta = 1.15$	0.343	0.402	77.3%	0.156
$\delta = 0.8$	0.29	0.438	93%	0.175
$\delta = 0.4$	0.286	0.464	96.2%	0.186
$\rho = 1.2$				
$\delta = 1.4$	0.502	0.401	56.6%	0.164
$\delta = 1.1$	0.347	0.395	85.7%	0.151
$\delta = 0.7$	0.297	0.381	93.8%	0.158
No WE	0.294	0.493	97.9%	0.196

Source: Authors' calculations. Numbers rounded.

4.2.3 Changes in σ

This set of simulations focuses on the change in the *riskiness* of the second period.[52] This is done by changing the value of the scale parameter σ of the log-normal distribution of the second period distribution of shocks. The distribution of shocks is no longer identical in both period. The first period distribution is the same as in all other simulations above, i.e. $F^1(\theta) = F^{young}(\theta)$. The second period distribution will be a modification of $F^{young}(\theta)$ as the new scale parameter σ is changed to obtain $F^2(\theta)$. An increase in σ results in an increase in the variance of the shock and hence the variance of the second period wage for a given level of work experience. These changes will have an impact on the mean of the shock but not on the median shock. Again, Case 2 is used for most of the parameters but κ_2 is changed to keep the mean of the second period wage matching the same as the above simulations.[53][54]

Figure 9 and Table 7 illustrates the changes for the first period labor wedge as σ, i.e.

[52] In this paper a change in σ has an impact on the variance and the kurtosis of the distribution. See Golosov et al. (2015) for a discussion on the effects of a change of the highest σ of a mixture of lognormals.

[53] Note that $\sigma = 0.67$ simulation is the same found in Figure 1.

[54] Simulations using Case 4 were also run and the results are very similar to the ones found with Case 2. I report these simulations in Figure XX2 and Table XX2 in Appendix C.1.

the riskiness, is increased. The labor wedge slightly increases and that is reflected by the increase of 5 percentage points of the average first period labor wedge. The most important change happens in the second period of life. This is in line with the analytical results that showed that an increase in σ, when the distribution of shocks is log-normal, results in an increase in the second period labor wedge. As shown in Table 7, the mean second period labor wedge increases by roughly 16 percentage points as σ is taken from 0.47 to 0.77. As risk increases the difference between the average labor distortion when young and when old changes as well.[55] In the cases where σ is below 67, the result obtained that the average labor wedges are lower when young is completely reversed. In fact even the result where a majority of workers faces lower labor distortion when young also disappears as shown in the simulation with $\sigma = 0.47$. Therefore uncertainty is an important explanation for the the age structure of taxation. Furthermore, as an increase in σ implies a greater labor distortions in the second period, the savings wedge is also increased to facilitate the separation of types.

Table 7: Average Wedges in Each Period: $\rho = 0.5$ with Changes in t=2's σ

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\sigma = 0.47$	0.381	0.282	0%	0.115
$\sigma = 0.57$	0.361	0.346	48.9%	0.138
$\sigma = 0.67$	0.343	0.402	77.3%	0.156
$\sigma = 0.77$	0.331	0.442	89.6%	0.167

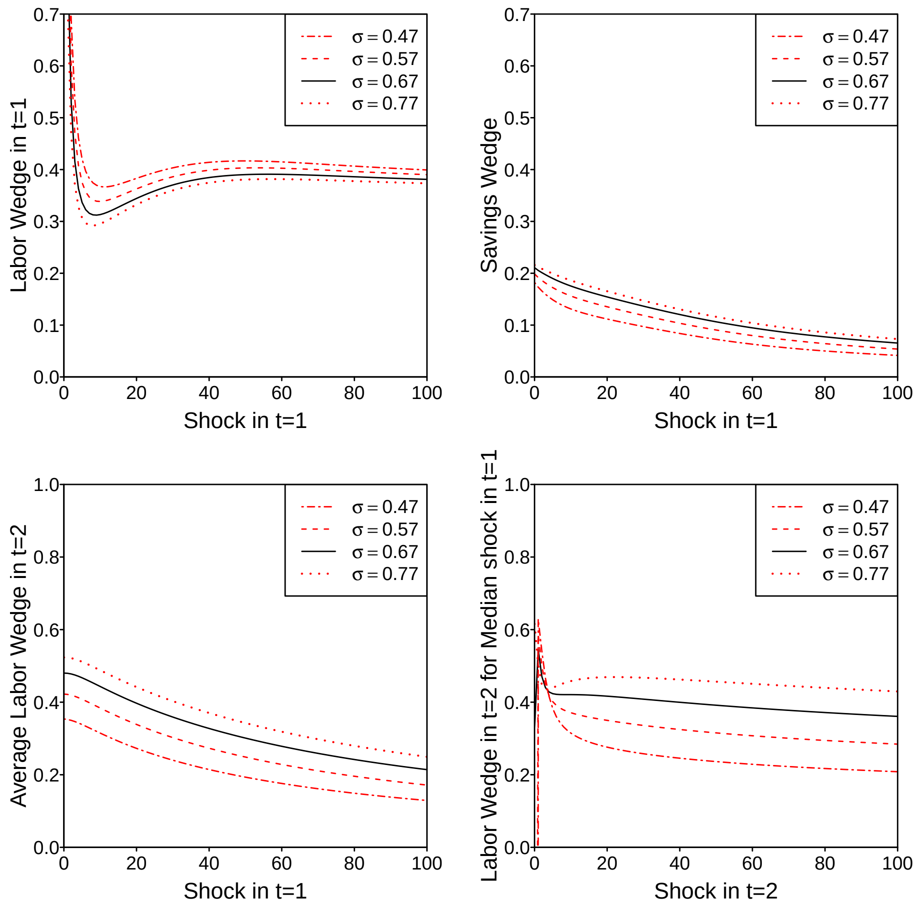
Source: Authors' calculations. Numbers rounded.

4.2.4 Changes in Pareto weights

So far all simulations have been made with the assumption that the planner's preferences where utilitarian. This final set of simulations considers changes to the planner's preferences

[55] It is quite possible that the log-normal distribution used in the numerical simulations of this paper overestimates the risk workers face in the second period of life.

Figure 7: Optimal Wedges: $\rho = 0.5$ with Changes in t=2's σ



using different Pareto weights. Let the functional form for the Pareto weights be

$$a(\theta_1) = \frac{\theta_1^{-\zeta}}{\int\limits_{\underline{\theta}}^{\bar{\theta}} \theta_1^{-\zeta} f^1(\theta_1) d\theta_1},$$

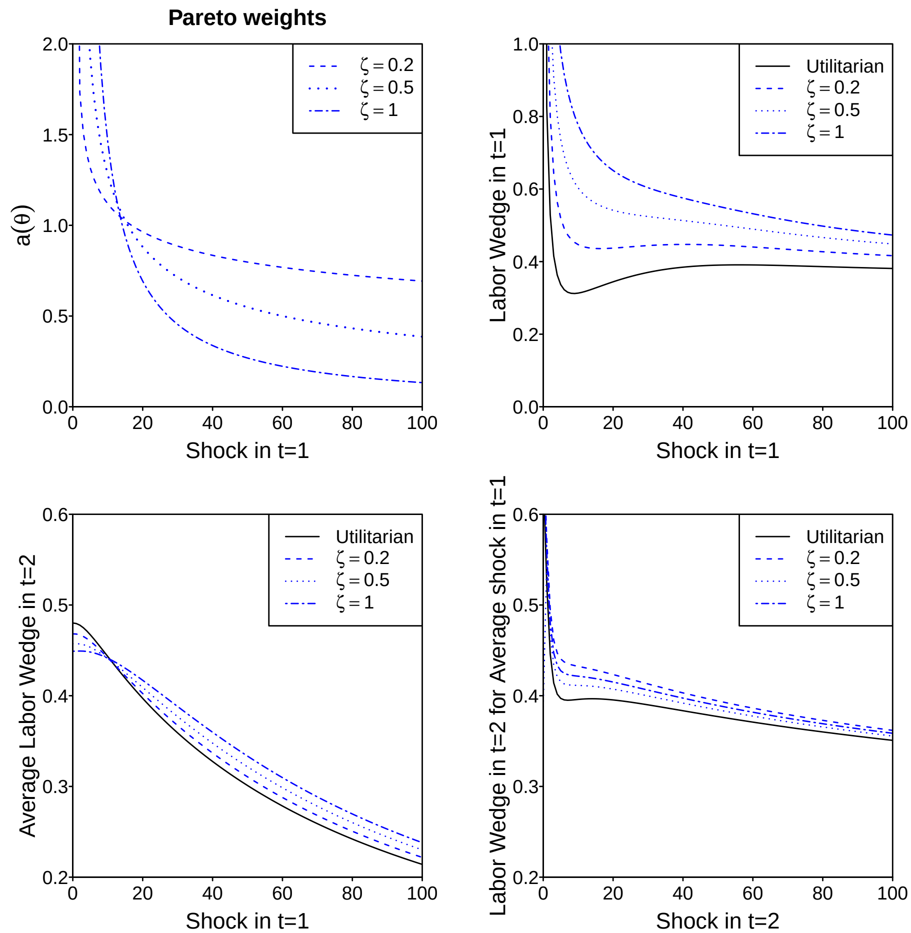
where $\zeta \geq 0$. Here the greater the value of ζ the more weight the lower skilled workers have. One could interpret this parameter as the planner's redistributive taste. This social welfare preference violates the anonymity axiom, nonetheless it is useful in obtaining some information on labor and savings distortions under different redistributive taste of the planner.

Again, Case 2 is used for the simulations. There is no need to change any parameters of the wage function as only the preferences of the planner are changed.[56] The simulations considers an increase in ζ from 0.2 to 1. The shape of the Pareto weight function under different values of ζ can be seen in the top left graphic in Figure 8. From both Figure 8 and Table 8, it is possible to see the effect the change in ζ has on the first period labor wedge. This result can be understood by looking at 3.5 from Proposition 1. Notice that $a(\theta_1)$ only appears in the first period labor wedge formula. By increasing ζ, the value of $a(\theta_1)$ decreases, except for low values of θ_1, and thus it pushes upward the value of the optimal labor wedge as giving a dollar to higher skilled workers is valued less and less by the planner. Also, due to the value of acquiring work experience in the future, the labor supply elasticity is also reduced in the first period of life. Therefore "taxing" more heavily in the first period to finance an increasing need to redistribute to the lower skilled worker lowers the cost of this redistribution.

The effect on the second period labor wedge is marginal and it only slightly increases with greater values of ζ. The planner seeks to redistribute in the first period and only distort

[56]Simulations were also made for Case 4 and are not reported here as they are similar in qualitative information. They can be found in Appendix C.1.

Figure 8: Optimal Wedges: $\rho = 0.5$ with Changes in ζ



labor in the second period to provide insurance. Since risk hasn't changed there aren't any more motivation for the planner to distort the second period labor market even more than in the utilitarian case. Because the second period labor wedge isn't affected and the first period labor wedge is pushed up drastically, the result of the increasing pattern of distortion with age no longer holds. In fact, as the planner's preference change the amount of workers that would face higher distortions in the future go from 77.3% to 0%.

Table 8: Average Wedges in Each Period: $\rho = 0.5$ with Changes in ζ

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
Utilitarian	0.343	0.402	77.3%	0.156
$\zeta = 0.2$	0.451	0.404	0%	0.157
$\zeta = 0.5$	0.574	0.408	0%	0.16
$\zeta = 1$	0.706	0.413	0%	0.163

Source: Authors' calculations. Numbers rounded.

5 Conclusion

This paper studies the design of an optimal dynamic tax system when wages depend on both accumulated work experience and a stochastic shock. The analysis is made under the assumption that work experience is observable to the planner. Using a first-order approach to solve the planner's problem, I find that in addition to standard considerations found in the standard optimal taxation, the first period labor tax is a balance of three motives that captures the added effects of having wages be a function of risk and work experience.

The main contribution of the paper is to highlight the importance of the parameter determining the complementarity between the stochastic shock and the accumulated work experience on the age structure of the optimal tax system. I find that a majority of workers, in numerical simulations calibrated to US data, would expect to face a higher marginal labor income tax when old compared to the one they faced when young. Nevertheless, whether the cross-sectional average of the marginal labor income tax increases or decreases with age depends on the complementarity parameter. A parameter value smaller than one leads to increasing average labor taxes with age whereas a parameter value above one leads to declining average labor taxes with age. The above results illustrate the importance of obtaining estimates of the value of the complementarity parameter if policy makers are considering tax reforms towards an age-dependent tax system.

In future work, I plan on relaxing the assumption of observable work experience accumulation and consider the case where work experience is a function of first period shock and labor effort. This would allow for different types of job opportunities and rewards to be part of the model. However, this comes at the cost of losing the *common knowledge of preference* assumption as work experience becomes unobservable. In the model above, work experience in one profession is in a way perfectly transferable to another profession. Considering the impact of labor income taxation on job-specific work experience may relax the incentive constraints on workers that decide to change occupation in response to labor income taxes.

References

Albanesi, S. & Sleet, C. (2006). Dynamic optimal taxation with private information. *Review of Economic Studies*, *73(1)*, 1–30.

Atkinson, A. B. & Stiglitz, J. (1976). The design of tax structure: Direct vs. indirect taxation. *Journal of Public Economics*, *6*, 55–75.

Bastani, S. (May 2014). Using the discrete model to derive optimal income tax rates. *Working Paper*.

Best, M. C. & Kleven, H. J. (February 2013). Optimal income taxation with career effects of work effort. *Working Paper*.

Blomquist, S. & Micheletto, L. (2008). Age-related optimal income taxation. *The Scandinavian Journal of Economics*, *110(1)*, 45–71.

Blundell, R. & MaCurdy, T. (1999). Labor supply: A review of alternative approaches. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics*, volume 3 chapter 27, (pp. 1559–1695). North Holland: Elsevier.

Bohacek, R. & Kapicka, M. (2008). Optimal human capital policies. *Journal of Monetary Economics*, *55*, 1–16.

Diamond, P. A. (2003). *Taxation, Incomplete Markets, And Social Security*. Cambride, MA: MIT Press.

Farhi, E. & Werning, I. (2013). Insurance and taxation over the life cycle. *Review of Economic Studies*, *80(2)*, 596–635.

Golosov, M., Kocherlakota, N., & Tsyvinsky, A. (2003). Optimal indirect and capital taxation. *Review of Economic Studies*, *70*, 569–587.

Golosov, M., Troshkin, M., & Tsyvinski, A. (2011). Optimal dynamic taxes. *Working Paper*.

Golosov, M., Troshkin, M., & Tsyvinski, A. (2015). Redistribution and social insurance. *Working Paper.*

Golosov, M., Tsyvinsky, A., & Werning, I. (2007). New dynamic public finance: A userís guide. *NBER Macroannual 2006*, 317–387.

Hellwig, M. F. (2007). A contribution to the theory of optimal utilitarian income taxation. *Journal of Public Economics*, *91*, 1449–1477.

Huggett, M., Ventura, G., & Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review*, *101(7)*, 2923–54.

Jacquet, L., Lehmann, E., & Van der Linden, B. (2013). The optimal marginal tax rates with both extensive and intensive responses. *Journal of Economic Theory*, *148(5)*, 1770–1805.

Kapicka, M. (2014). Pareto efficient taxation with learning by doing. *Working Paper.*

Kapicka, M. & Neira, J. (October 2014). Optimal taxation in a life-cycle economy with endogenous human capital formation. *Working Paper.*

Kocherlakota, N. (2005). Zero expected wealth taxes: A mirrlees approach to dynamic optimal taxation. *Econometrica*, *73*, 1587ñ1621.

Krause, A. (2009). Optimal nonlinear income taxation with learning-by-doing. *Journal of Public Economics*, *93*, 1098ñ1110.

Kremer, M. (2002). Should taxes be independent of age. *Mimeo.*

OECD (2004). Improving skills for more and better jobs: Does training make a difference? *Discussion Paper, OECD Employment Outlook.*

Stantcheva, S. (2014b). Optimal taxation and human capital policies over the life cycle. *Working Paper.*

Stantcheva, S. (December 2014a). Learning and (or) doing: Human capital investments and optimal taxation. *Working Paper.*

Stiglitz, J. (1982). Self-selection and pareto efficient taxation. *Journal of Public Economics*, *17*, 213–240.

Storesletten, K., Telmer, C., & Yaron, A. (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics*, *51(3)*, 609ñ633.

Weinzierl, M. (2011). The surprising power of age-dependent taxes. *Review of Economic Studies*, *78*, 1490–1518.

Werning, I. (September 2011). Nonlinear capital taxation. *Working Paper*.

Appendix

A The Hamiltonians

First, use the first constraint in the first period problem and manipulate it such that

$$c(\theta) = u^{-1}\left(\omega(\theta) + h\left(\frac{y(\theta)}{w_1(\theta, e_1)}\right) - \beta\upsilon(\theta)\right).$$

The Hamiltonian for period $t = 1$ is:

$$\left[u^{-1}\left(\omega(\theta) + h\left(\frac{y(\theta)}{w_1(\theta, e_1)}\right) - \beta\upsilon(\theta)\right) - y(\theta) + \frac{1}{R}\ K(\upsilon(\theta), e(\theta), 2)\right] f^1(\theta)$$
$$+ \lambda_1[\upsilon_0 - a(\theta)\omega(\theta)f^1(\theta)] + \mu(\theta)\left[\frac{h'\left(\frac{y(\theta)}{w_1(\theta,e_1)}\right)y(\theta)}{[w_1(\theta, e_1)]^2}\frac{\partial w_1(\theta, e_1)}{\partial\theta_1}\right],$$

where

$$e(\theta) = \phi(y(\theta)) + e_1,$$

and the Hamiltonian for period $t = 2$ is

$$\left[u^{-1}\left(\omega(\theta) + h\left(\frac{y(\theta)}{w_2(\theta, e_2)}\right)\right) - y(\theta)\right] f^2(\theta)$$
$$+ \lambda_2[\upsilon - \omega(\theta)f^2(\theta)] + \mu(\theta)\left[\frac{h'\left(\frac{y(\theta)}{w_2(\theta,e_2)}\right)y(\theta)}{[w_2(\theta, e_2)]^2}\frac{\partial w_2(\theta, e_2)}{\partial\theta_2}\right].$$

Furthermore using the envelope theorem we get:

$$K_\upsilon(\upsilon, e_2, 2) = \lambda_2,$$

and so from the point of view of the first period, we have:

$$K_{\upsilon}(\upsilon(\theta), e(\theta), 2) = \lambda(\theta).$$

The FOCs for the second period problem are:

$y(\theta)$:

$$\left[-1+\frac{h'\left(\frac{y(\theta)}{w_2(\theta,e_2)}\right)}{u'(c(\theta))w_2(\theta,e_2)}\right]f^2(\theta)+\mu(\theta)\left[\frac{h''\left(\frac{y(\theta)}{w_2(\theta,e_2)}\right)y(\theta)}{[w_2(\theta,e_2)]^3}+\frac{h'\left(\frac{y(\theta)}{w_2(\theta,e_2)}\right)}{[w_2(\theta,e_2)]^2}\right]\frac{\partial w_2(\theta,e_2)}{\partial\theta}=0. \tag{A.1}$$

The law of motion for the co-state variable $\omega(\theta)$:

$$-\dot{\mu}(\theta)=\frac{f^2(\theta)}{u'(c(\theta))}-\lambda_2 f^2(\theta). \tag{A.2}$$

The FOCS for the first period are:

$y(\theta)$:

$$\left[-1+\frac{h'(l(\theta))}{u'(c(\theta))w_1(\theta,e_1)}+\frac{1}{R}\phi'(y(\theta))\frac{\partial K}{\partial e(\theta)}\right]f^1(\theta)+\mu(\theta)\left[\frac{h''(l(\theta))\,l(\theta)}{[w_1(\theta,e_1)]^2}+\frac{h'(l(\theta))}{[w_1(\theta,e_1)]^2}\right]\frac{\partial w_1(\theta,e_1)}{\partial\theta}=0, \tag{A.3}$$

where

$$\begin{aligned}\frac{\partial K}{\partial e(\theta)} = & - \int \frac{h'\left(\frac{y(\theta)}{w_2(\theta,e_2)}\right) y(\theta)}{u'(c(\theta))[w_2(\theta,e_2)]^2} \frac{\partial w_2(\theta,e_2)}{\partial e} f^2(\theta) d\theta \\ & - \int \frac{\mu(\theta)}{f^2(\theta)} \left[\frac{h''\left(\frac{y(\theta)}{w_2(\theta,e_2)}\right) y(\theta)}{[w_2(\theta,e_2)]^3} + \frac{h'\left(\frac{y(\theta)}{w_2(\theta,e_2)}\right)}{[w_2(\theta,e_2)]^2} \right] \frac{\partial w_2(\theta,e_2)}{\partial \theta} \frac{\partial w_2(\theta,e_2)}{\partial e(\theta)} \frac{y(\theta)}{w_2(\theta,e_2)} f^2(\theta) d\theta \\ & - \int \frac{\mu(\theta)}{f^2(\theta)} \frac{h'\left(\frac{y(\theta)}{w_2(\theta,e_2)}\right) y(\theta)}{[w_2(\theta,e_2)]^3} \frac{\partial w_2(\theta,e_2)}{\partial \theta} \frac{\partial w_2(\theta,e_2)}{\partial e(\theta)} \left[1 - \frac{\frac{\partial^2 w_2(\theta,e_2)}{\partial\theta \partial e(\theta)}}{\frac{\partial w_2(\theta,e_2)}{\partial \theta} \frac{\partial w_2(\theta,e_2)}{\partial e(\theta)}} \right] f^2(\theta) d\theta. \end{aligned} \quad \text{(A.4)}$$

$\upsilon(\theta)$:

$$-\frac{\beta}{u'(c(\theta))} + \frac{1}{R} \frac{\partial K}{\partial \upsilon(\theta)} = 0, \quad \text{(A.5)}$$

The law of motion for the co-state variable $\omega(\theta)$:

$$-\dot{\mu}(\theta) = \frac{f^1(\theta)}{u'(c(\theta))} - \lambda_1 a(\theta) f^1(\theta). \quad \text{(A.6)}$$

A.1 Labor wedge

A.1.1 Labor wedge in Period 2

Starting for t=2, using the FOC for $y(\theta^2)$ and rearranging, we get:

$$1 - \frac{h'(l(\theta^2))}{u'(c(\theta^2)) w_2(\theta_2, e_2)} = \frac{\mu(\theta^2)}{f^2(\theta_2)} \frac{h'(l(\theta^2))}{[w_2(\theta_2,e_2)]^2} \left[1 + \frac{h''\big(l(\theta^2)\big)\, l(\theta^1)}{h'(l(\theta^2))} \right] \frac{\partial w_2(\theta_2, e_2)}{\partial \theta}. \quad \text{(A.7)}$$

Using the definition our elasticity measure and

$$\epsilon_{\theta t} \equiv \frac{\partial w_t(\theta_t, e(\theta^{t-1}))}{\partial \theta_t} \frac{\theta_t}{w_t(\theta_t, e(\theta^{t-1}))},$$

we obtain the formula (3.6) from Proposition 1.

To write the optimal second period wedge formula (3.6) into the ABC format use the definition from above, and see that

$$\mu(\theta^2)u'(c(\theta_2)) = \int_{\theta_2}^{\bar{\theta}} \left(\frac{u'(c(\theta_2))}{u'(c(x))} \right) [1 - \lambda_2 u'(c(x))] f^2(x) dx,$$

and

$$\begin{aligned} \frac{u'(c(\theta))}{u'(c(x))} &= \exp\left(\ln \frac{u'(c(\theta))}{u'(c(x))} \right) = \exp\left(- \int_\theta^x \frac{du'(c(\tilde{x}))}{u'(c(\tilde{x}))} \right) \\ &= \exp\left(- \int_\theta^x \frac{u''(c(\tilde{x}))dc(\tilde{x})}{u'(c(\tilde{x}))} \right) = \exp\left(- \int_\theta^x \frac{u''(c(\tilde{x}))c(\tilde{x})}{u'(c(\tilde{x}))} \frac{\dot{c}(\tilde{x})}{c(\tilde{x})} d\tilde{x} \right) \\ &= \exp\left(\int_\theta^x \eta(\tilde{x}) \frac{\dot{c}(\tilde{x})}{c(\tilde{x})} d\tilde{x} \right). \end{aligned}$$

Multiplying on both sides by $(1 - F^2(\theta))/(1 - F^2(\theta))$ (3.6) and using the new definitions, I obtain

$$\frac{\tau_L^*(\theta^2)}{1 - \tau_L^*(\theta^2)} = (1 + \alpha(\theta^2))\epsilon_{\theta 2} \left[\frac{1 - F^2(\theta_2)}{\theta_2 f^2(\theta_2)} \right] \frac{\int_{\theta_2}^{\bar{\theta}} \exp\left(\int_\theta^x \eta(\tilde{x}) \frac{\dot{c}(\tilde{x})}{c(\tilde{x})} d\tilde{x} \right) [1 - \lambda_2 u'(c(\tilde{x}))] f^2(x) dx}{1 - F^2(\theta_2)}.$$

A.1.2 Labor wedge in Period 1

Using the FOC for $y(\theta^1)$ and rearranging, we get:

$$1 - \frac{h'(l(\theta^1)}{u'(c(\theta^1))w_1(\theta_1, e_1)} - \frac{1}{R}\phi'(y(\theta^1))\frac{\partial K}{\partial e(\theta^1)} = \frac{\mu(\theta^1)}{f^1(\theta_1)}\frac{h'(l(\theta^1))}{[w_1(\theta_1, e_1)]^2}\left[1 + \frac{h''\big(l(\theta^1)\big)\, l(\theta^1)}{h'(l(\theta^1))}\right]\frac{\partial w_1(\theta_1, e_1)}{\partial \theta_1}. \tag{A.8}$$

Notice that

$$\begin{aligned}\int \frac{h'\big(l(\theta^2)\, l(\theta^2)}{u'(c(\theta^2))w_2(\theta_2, e_2)}\frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2)d\theta_2 = &\int \frac{1}{u'(c(\theta^2))} f^2(\theta_2)d\theta_2 \\ &\times \int \frac{h'\big(l(\theta^2)\, l(\theta^2)}{w_2(\theta_2, e_2)}\frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2)d\theta_2 \\ &+ \mathrm{Cov}\!\left(\frac{1}{u'(c(\theta^2))}, \frac{h'\big(l(\theta^2)\, l(\theta^2)}{w_2(\theta_2, e_2)}\frac{\partial w_2(\theta_2, e_2)}{\partial e}\right),\end{aligned}$$

and using the result from the intertemporal wedge, we have:

$$\begin{aligned}\int \frac{h'\big(l(\theta^2)\, l(\theta^2)}{u'(c(\theta^2))w_2(\theta_2, e_2)}\frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2)d\theta_2 = &\frac{R\beta}{u'(c(\theta^1))}\int \frac{h'\big(l(\theta^2)\, l(\theta^2)}{w_2(\theta_2, e_2)}\frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2)d\theta_2 \\ &+ \mathrm{Cov}\!\left(\frac{1}{u'(c(\theta^2))}, \frac{h'\big(l(\theta^2)\, l(\theta^2)}{w_2(\theta_2, e_2)}\frac{\partial w_2(\theta_2, e_2)}{\partial e}\right).\end{aligned}$$

Reinserting this in $\partial K/\partial e$, and using the rewritten FOC, we have:

$$
\begin{aligned}
&1-\frac{h'(l(\theta^1)}{u'(c(\theta^1))w_1(\theta_1,e_1)}+\beta\frac{\phi'(y(\theta^1)}{u'(c(\theta^1))}\int\frac{h'\big(l(\theta^2)\,l(\theta^2)}{w_2(\theta_2,e_2)}\frac{\partial w_2(\theta_2,e_2)}{\partial e}f^2(\theta_2)d\theta_2=\\
&\quad\frac{\mu(\theta^1)}{f^1(\theta_1)}\frac{h'(l(\theta^1))}{[w_1(\theta_1,e_1)]^2}\left[1+\frac{h''\big(l(\theta^1)\big)\,l(\theta^1)}{h'(l(\theta^1))}\right]\frac{\partial w_1(\theta_1,e_1)}{\partial\theta_1}\\
&\quad-\frac{\phi'(y(\theta^1))}{R}\left\{\mathrm{Cov}\left(\frac{1}{u'(c(\theta^2))},\frac{h'\big(l(\theta^2)\,l(\theta^2)}{w_2(\theta_2,e_2)}\frac{\partial w_2(\theta_2,e_2)}{\partial e}\right)+\int\tau_L(\theta^2)\frac{\partial w_2(\theta_2,e_2)}{\partial e}l(\theta^2)f^2(\theta_2)d\theta_2\right.\\
&\quad\left.+\int\frac{\mu(\theta^2)}{f^2(\theta_2)}\frac{h'\big(l(\theta^2)\big)\,l(\theta^2)}{[w_2(\theta_2,e_2)]^2}\frac{\partial w_2(\theta_2,e_2)}{\partial\theta}\frac{\partial w_2(\theta_2,e_2)}{\partial e(\theta^1)}\left[1-\frac{\dfrac{\partial^2 w_2(\theta_2,e_2)}{\partial\theta\partial e(\theta^1)}}{\dfrac{\partial w_2(\theta_2,e_2)}{\partial\theta_2}\dfrac{\partial w_2(\theta_2,e_2)}{\partial e(\theta^1)}}\right]f^2(\theta_2)d\theta_2\right\}.
\end{aligned}
\tag{A.9}
$$

Using the definitions of the Hicksian complementarity, the labor elasticity and the wage elasticity, we obtain the results of Proposition 1.

A.2 Intertemporal Wedge

Using the FOC on $\omega(\theta)$ for period t=2 and the following boundary conditions:

$$\lim_{\theta\to\underline{\theta}}\mu(\theta)=0\quad\text{and}\quad\lim_{\theta\to\bar{\theta}}\mu(\theta)=0,$$

we get

$$\mu(\theta)=\int_{\theta}^{\bar{\theta}}\left[\frac{1}{u'(c(\theta))}-\lambda_2\right]f^2(\theta)d\theta,$$

and

$$\int_{\underline{\theta}}^{\bar{\theta}}\frac{1}{u'(c(\theta))}f^2(\theta)d\theta=\lambda_2.$$

Using the FOC on $v(\theta)$ in t=1 and the envelope result from above, we have

$$\lambda(\theta^1) = \frac{\beta R}{u'(c(\theta^1))}.$$

Combining both we obtain the inverse Euler equation:

$$\frac{1}{u'(c(\theta^1))} = \frac{1}{R\beta} \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta^2))} f^2(\theta_2) d\theta_2.$$

A.3 Proof of Corollary 2.a and 2.b

The proof of Corollary 2.a and 2.b follows the proof of Corollary 1 in Golosov et al. (2015).[57] Properties of the log-normal distribution: Use the fact that

$$\frac{\theta f^{2'}(\theta)}{f^2(\theta)} = 1 + \frac{\ln\theta - \mu}{\sigma^2},$$

and by L'Hôpital's rule, it is possible to obtain

$$\lim_{\theta\to\infty} \frac{1 - F^2(\theta)}{\theta f^2(\theta)} = \lim_{\theta\to\infty} \frac{-1}{\theta f^{2'}(\theta)/f^2(\theta) + 1} = 0,$$
$$\lim_{\theta\to\infty} \frac{(1 - F^2(\theta))(\ln\theta - \mu)/\sigma^2}{\theta f^2(\theta)} = \lim_{\theta\to\infty} \frac{(\ln\theta - \mu)/\sigma^2 + (1 - F^2(\theta))/(\theta f^2(\theta)\sigma^2)}{\theta f^{2'}(\theta)/f^2(\theta) + 1} = 1.$$

From this the following result follows: $\frac{1-F^2(\theta)}{\theta f^2(\theta)} \sim \frac{\sigma^2}{\ln\theta - \mu}$.

Proof of Corollary 2.a

Using Assumption 2, and also specifically the fact that both $\frac{c_2(\theta)}{y_2(\theta)}$ and $\frac{\dot{c}_2(\theta)/c_2(\theta)}{\dot{y}_2(\theta)/y_2(\theta)}$ have finite

[57]However, I do not prove every lemma as I do not require all of them for the proof of Corollary 2.a and 2.b.

limits, it follows that there exist a limit for $\lim_{\theta\to\infty}\frac{\dot{c}_2(\theta)}{\dot{y}_2(\theta)}$ since

$$\frac{\dot{c}_2(\theta)}{\dot{y}_2(\theta)}=\frac{\dot{c}_2(\theta)/c_2(\theta)}{\dot{y}_2(\theta)/y_2(\theta)}\frac{c_2(\theta)}{y_2(\theta)}. \tag{A.10}$$

Also note that $c_2(\theta), y_2(\theta)\to\infty$ as $\theta\to\infty$ must hold. If this was not the case, it would imply that $1-\tau_L^2(\theta)=\frac{h'(y_2(\theta)/w_2(\theta,e))}{u'(c(\theta))w_2(\theta,e)}\to 0$ as θ goes to infinity which would contradict the assumption that $\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}$ has a finite limit.

Using L'Hôpital's rule

$$\lim_{\theta\to\infty}\frac{c_2(\theta)}{y_2(\theta)}=\lim_{\theta\to\infty}\frac{\dot{c}_2(\theta)}{\dot{y}_2(\theta)},$$

and using (A.10) I obtain

$$1=\frac{\lim\limits_{\theta\to\infty}\dfrac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}{\lim\limits_{\theta\to\infty}\dfrac{\dot{y}_2(\theta)\theta}{y_2(\theta)}}=\frac{\lim\limits_{\theta\to\infty}\dfrac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}{\lim\limits_{\theta\to\infty}\epsilon_{\theta 2}(\theta,e_2)+\dfrac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}. \tag{A.11}$$

Using the fact that $\lim_{\theta\to\infty}\tau_L^2(\theta)=\bar{\tau}_L^2<1$, and using $(1-\bar{\tau}_L^2)=\frac{h'(l_2(\theta)}{u'(c(\theta))w_2(\theta,e)}$, and by L'Hôpital's rule it is possible to get the following result

$$\begin{aligned}1&=\frac{\lim\limits_{\theta\to\infty}\dfrac{h'(l_2(\theta)}{u'(c(\theta))w_2(\theta,e)}}{1-\bar{\tau}_L^2}=\frac{1}{1-\bar{\tau}_L^2}\lim_{\theta\to\infty}\frac{h''(l_2(\theta)\dot{l}(\theta)}{u''(c(\theta))\dot{c}(\theta)w_2(\theta,e)+u'(c(\theta))\frac{\partial w_2}{\partial\theta}},\\&=\frac{1}{1-\bar{\tau}_L^2}\lim_{\theta\to\infty}\frac{h'(l_2(\theta)}{u'(c(\theta))w_2(\theta,e)}\frac{\frac{h''(l_2(\theta)\dot{l}(\theta)}{h'(l_2(\theta))}}{\frac{u''(c(\theta))\dot{c}(\theta)}{u'(c_2(\theta))}+\frac{\partial w_2}{\partial\theta}\frac{1}{w_2(\theta,2)}}\\&=\frac{1}{1-\bar{\tau}_L^2}\lim_{\theta\to\infty}\frac{h'(l_2(\theta)}{u'(c(\theta))w_2(\theta,e)}\lim_{\theta\to\infty}\frac{\alpha(\theta)\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}{\epsilon_{\theta 2}(\theta,e_2)-\eta(\theta)\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}\\&=\lim_{\theta\to\infty}\frac{\alpha(\theta)\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}{\epsilon_{\theta 2}(\theta,e_2)-\eta(\theta)\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}.\end{aligned} \tag{A.12}$$

From the above assumptions let $\alpha(\theta)\to\bar{\alpha},\ \eta(\theta)\to\bar{\eta},$ and $\epsilon_{\theta 2}(\theta,e_2)\to\bar{\epsilon_{\theta 2}}$ as $\theta\to\infty$, then if $\lim_{\theta\to\infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}$ and $\lim_{\theta\to\infty} \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}$ are finite (A.11) and (A.12) can be used to obtain

$$\lim_{\theta\to\infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)} = \lim_{\theta\to\infty} \left(\frac{1+\alpha(\theta)}{\eta(\theta)+\alpha(\theta)}\right) \epsilon_{\theta 2}(\theta, e_2) = \left(\frac{1+\bar{\alpha}}{\bar{\eta}+\bar{\alpha}}\right) \bar{\epsilon}_{\theta 2}, \tag{A.13}$$

$$\lim_{\theta\to\infty} \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)} = \lim_{\theta\to\infty} \left(\frac{1-\eta(\theta)}{\eta(\theta)+\alpha(\theta)}\right) \epsilon_{\theta 2}(\theta, e_2) = \left(\frac{1+\bar{\eta}}{\bar{\eta}+\bar{\alpha}}\right) \bar{\epsilon}_{\theta 2}. \tag{A.14}$$

Using assumption 2 it can be shown that these limits are finite. If either $\lim_{\theta\to\infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}$ or $\lim_{\theta\to\infty} \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}$ are infinite, (A.11) would imply that

$$\lim_{\theta\to\infty} \frac{\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}{\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}} = 1.$$

Suppose that $\lim_{\theta\to\infty} \left|\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}\right| = \infty$, then by (A.12) this result follows

$$1 = \lim_{\theta\to\infty} \frac{\alpha(\theta)\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}{\epsilon_{\theta 2}(\theta, e_2) - \eta(\theta)\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}} = \lim_{\theta\to\infty} \frac{\alpha(\theta)}{\frac{\epsilon_{\theta 2}(\theta,e_2)}{\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}} - \eta(\theta)} \frac{\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}{\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}} = -\frac{\bar{\alpha}}{\bar{\eta}} < 0$$

which is a contradiction. This result also relies on the fact that $\epsilon_{\theta 2}(\theta, e_2)$ was assumed to have a finite limit.

Using the result on the limit of $\lim_{\theta\to\infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}$ the behavior of $C(\theta)$ can be characterized as $\theta \to \infty$. First define $q(\tilde{x}) \equiv \eta(\tilde{x})\frac{\dot{c}_2(\tilde{x})\tilde{x}}{c_2(\tilde{x})}$, and rewrite $C(\theta)$ with the assumption that the distribution is unbounded:

$$C(\theta) = \frac{\exp\left(-\int_0^\theta \frac{q(\tilde{x})}{\tilde{x}} d\tilde{x}\right) \int_\theta^\infty \exp\left(\int_0^x \frac{q(\tilde{x})}{\tilde{x}} d\tilde{x}\right) [1 - \lambda_2 u'(c_2(x))] f^2(x) dx}{1 - F^2(\theta)}.$$

By L'Hôpital's rule

$$
\begin{aligned}
\lim_{\theta\to\infty} C(\theta) = \lim_{\theta\to\infty} & \frac{-\exp\left(-\int_0^\theta \frac{q(\tilde{x})}{\tilde{x}}d\tilde{x}\right)\frac{q(\theta)}{\theta}\int_\theta^\infty \exp\left(\int_0^x \frac{q(\tilde{x})}{\tilde{x}}d\tilde{x}\right)[1-\lambda_2 u'(c_2(x))]f^2(x)dx}{-f^2(\theta)} \\
+ & \frac{-\exp\left(-\int_0^\theta \frac{q(\tilde{x})}{\tilde{x}}d\tilde{x}\right)\exp\left(-\int_0^\theta \frac{q(\tilde{x})}{\tilde{x}}d\tilde{x}\right)[1-\lambda_2 u'(c(\theta))]f^2(\theta)}{-f^2(\theta)} \\
= & \lim_{\theta\to\infty}[1-\lambda_2 u'(c(\theta))] + \lim_{\theta\to\infty}\frac{q(\theta)C(\theta)[1-F^2(\theta)]}{\theta f^2(\theta)}.
\end{aligned}
$$

And since, $c_2(\theta)\to\infty$, $u'(c_2(\theta))\to 0$ and thus

$$
\lim_{\theta\to\infty} C(\theta) = 1 + \lim_{\theta\to\infty} q(\theta)\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}\frac{1}{A(\theta)}.
$$

The above assumptions, (A.13) and the above result are used to obtain

$$
\lim_{\theta\to\infty} C(\theta) = 1 + \frac{\bar{\eta}}{(1+\bar{\alpha})\bar{\epsilon}_{\theta 2}}\left(\frac{1+\bar{\alpha}}{\bar{\eta}+\bar{\alpha}}\right)\bar{\epsilon}_{\theta 2}\lim_{\theta\to\infty}\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}.
$$

Because $\lim_{\theta\to\infty}\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}$ was assumed to have a finite limit, $C(\theta)$ must have a finite limit. From assumption 2, $A(\theta)$ also has a finite limit, using the above result on $C(\theta)$ and the property of the log-normal distribution such that $\lim_{\theta\to\infty}\frac{1-F^2(\theta)}{\theta f^2(\theta)} = 0$, which implies that $B(\theta)\to 0$ as $\theta\to\infty$, then this results follows

$$
\lim_{\theta\to\infty}\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)} = \lim_{\theta\to\infty} A(\theta)B(\theta)C(\theta) = 0.
$$

The last result implies that with the log-normal distribution $\lim_{\theta\to\infty} C(\theta) = 1$. Now consider $\lim_{\theta\to\infty}\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}\frac{\ln\theta-\mu}{\sigma^2}$,

$$
\begin{aligned}
\lim_{\theta\to\infty}\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}\frac{\ln\theta-\mu}{\sigma^2} = & \lim_{\theta\to\infty}(1+\alpha(\theta))\epsilon_{\theta 2}(\theta,e_2)\lim_{\theta\to\infty}\frac{1-F^2(\theta)}{\theta f^2(\theta)}\frac{\ln\theta-\mu}{\sigma^2}\lim_{\theta\to\infty} C(\theta), \\
= & (1+\bar{\alpha})\bar{\epsilon}_{\theta 2} \qquad \text{(A.15)}
\end{aligned}
$$

From the results above I get

$$\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)} \sim (1+\bar{\alpha})\bar{\epsilon}_{\theta 2}\left(\frac{\sigma^2}{\ln\theta-\mu}\right) \text{ as } \theta\to\infty.$$

Proof of Corollary 2.b

The proof of Corollary 2.b is almost identical to the one of Corollary 2.a. Using Assumption 2 and the wage function

$$w_2(\theta,e_2)=(\theta^{1-\rho}+e_2^{1-\rho})^{\frac{1}{1-\rho}},$$

I get

$$\lim_{\theta\to\infty}\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}=\lim_{\theta\to\infty}\left(\frac{1+\alpha(\theta)}{\eta(\theta)+\alpha(\theta)}\right)\epsilon_{\theta 2}(\theta,e_2)=\lim_{\theta\to\infty}\left(\frac{1+\alpha(\theta)}{\eta(\theta)+\alpha(\theta)}\right)\left(\frac{w_2(\theta,e_2)}{\theta}\right)^{\rho-1}, \tag{A.16}$$

$$\lim_{\theta\to\infty}\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}=\lim_{\theta\to\infty}\left(\frac{1-\eta(\theta)}{\eta(\theta)+\alpha(\theta)}\right)\epsilon_{\theta 2}(\theta,e_2)=\lim_{\theta\to\infty}\left(\frac{1-\eta(\theta)}{\eta(\theta)+\alpha(\theta)}\right)\left(\frac{w_2(\theta,e_2)}{\theta}\right)^{\rho-1}, \tag{A.17}$$

since with the above CES function the wage elasticity with respect to shock is $\epsilon_{\theta 2}=\left(\frac{w_2(\theta,e_2)}{\theta}\right)^{\rho-1}$. A particularity of this function is that whenever $\rho>1$ the limit of this elasticity is zero. This implies that both $\lim_{\theta\to\infty}\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}$ and $\lim_{\theta\to\infty}\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}$ have a finite limit of zero.

In this situation, the limit of $C(\theta)$ can still be written in the following way:

$$\lim_{\theta\to\infty}C(\theta)=1+\left(\frac{\bar{\eta}}{\bar{\eta}+\bar{\alpha}}\right)\lim_{\theta\to\infty}\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)},$$

since

$$\lim_{\theta\to\infty}\frac{\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}{A(\theta)}=\lim_{\theta\to\infty}\frac{1}{\eta(\theta)+\alpha(\theta)}.$$

The limit of $C(\theta)$ is still finite by the assumption on the finite limit of $\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}$. However, the asymptotic properties of $A(\theta)$ have changed since the wage elasticity goes to zero. This also does not change the following result

$$\lim_{\theta\to\infty} \frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)} = \lim_{\theta\to\infty} A(\theta)B(\theta)C(\theta) = 0,$$

but it changes the result on $\lim_{\theta\to\infty} \frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)} \frac{\ln\theta-\mu}{\sigma^2}$,

$$\begin{aligned}\lim_{\theta\to\infty} \frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)} \frac{\ln\theta-\mu}{\sigma^2} &= \lim_{\theta\to\infty} (1+\alpha(\theta))\epsilon_{\theta 2}(\theta, e_2) \lim_{\theta\to\infty} \frac{1-F^2(\theta)}{\theta f^2(\theta)} \frac{\ln\theta-\mu}{\sigma^2} \lim_{\theta\to\infty} C(\theta),\\ &= 0\times 1\times 1 = 0.\end{aligned}$$

Notice that with the CES function with $\rho > 1$, this holds

$$\lim_{\theta\to\infty} \epsilon_{\theta 2}(\theta, e_2)\theta^{\rho-1} = e_2^{\rho-1}.$$

Now consider

$$\begin{aligned}\lim_{\theta\to\infty} \frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)} \frac{\theta^{\rho-1}(\ln\theta-\mu)}{\sigma^2} &= \lim_{\theta\to\infty} (1+\alpha(\theta))\epsilon_{\theta 2}(\theta, e_2)\theta^{\rho-1} \lim_{\theta\to\infty} \frac{1-F^2(\theta)}{\theta f^2(\theta)} \frac{\ln\theta-\mu}{\sigma^2} \lim_{\theta\to\infty} C(\theta),\\ &= (1+\bar{\alpha})e_2^{\rho-1}. \end{aligned} \tag{A.18}$$

Using this result

$$\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)} \sim (1+\bar{\alpha}) \left(\frac{e_2}{\theta}\right)^{\rho-1} \left(\frac{\sigma^2}{\ln\theta-\mu}\right) \text{ as } \theta\to\infty.$$

B Implementation: No savings in equilibrium

In this section, I consider the decentralization of the optimal allocation through a tax system. We follow Werning (2011) methodology which augments any given mechanism and allows a choice over savings subject to a nonlinear savings tax. I also follow Kapicka & Neira (2014) where the the design of the tax on savings is first considered and than proceed to incorporate this tax in the full tax system.

Recall that the equilibrium values $\omega(\theta^1)$ and $\omega(\theta^2)$ from an incentive compatible allocation $\{c, y\}$ satisfy

$$\begin{aligned}
\omega(\theta^1) &= u(c(\theta^1)) - h\left(\frac{y(\theta^1)}{w_1(\theta_1, e_1)}\right) + \beta \int \omega(\theta^1, e(\theta^1), \theta_2) f^2(\theta_2) d\theta_2, \\
\omega(\theta^2) &= u(c(\theta^2)) - h\left(\frac{y(\theta^2)}{w_2(\theta_2, e(\theta^1))}\right).
\end{aligned}$$

And as above, let the worker's optimization problem be

$$\begin{aligned}
\omega^r(\theta_1) &= \max_{r^1} \left\{ u(c(r^1)) - h\left(\frac{y(r^1)}{w_1(\theta_1, e_1)}\right) + \beta \int \omega^{\sigma^r}(r^1, e(r^1), \theta_2) f^2(\theta_2) d\theta_2 \right\}, \\
\omega^r(r^1, \theta_2) &= \max_{r^2} \left\{ u(c(r^2)) - h\left(\frac{y(r^2)}{w_2(\theta_2, e(r^1))}\right) \right\},
\end{aligned}$$

where r^1 and r^2 are the reports of their types given in each period to the planner. Therefore incentive compatibility implies

$$\begin{aligned}
\omega^r(\theta_1) &= \omega(\theta^1), \\
\omega^r(r^1, \theta_2) &= \omega(r^1, \theta_2) = \omega(\theta^2).
\end{aligned}$$

So, for any incentive compatible allocation $\{c, y\}$, consider these two budget constraints

$$\begin{aligned}
\tilde{c_1} + M(x_2, r^1) &\leq c(r^1), \\
\tilde{c_2} &\leq x_2 + c(r^2),
\end{aligned}$$

where x_2 is after-interest savings and $M(x_2, \sigma^1)$ is referred to as a tax function based on after-interest savings and the worker's report of his type r^1 in period 1. Again, the goal is to augment the direct mechanism by allowing the worker to save. If I suppose that the net interest rate is i then $M(x_2, r^1) - x_2/(1+i)$ is a nonlinear tax on savings in period 1. An important part is that $M(0, r^1) = 0$ so that a worker that decides not to save pays no taxes.

Under the augmented mechanism, the worker's problem in the first and second period is:

$$V(x_1, e_1, \theta_1) = \max_{r^1, x_2} \left\{ u\big(c(r^1) - M(x_2, r^1)\big) - h\left(\frac{y(r^1)}{w_1(\theta_1, e_1)}\right) + \beta \int V(x_2, r^1, e(r^1), \theta_2) f^2(\theta_2) d\theta_2 \right\},$$
$$V(x_2, r^1, e(r^1), \theta_2) = \max_{r^2} \left\{ u\big(c(r^2) + x_2\big) - h\left(\frac{y(r^2)}{w_2(\theta_2, e(r^1))}\right) \right\},$$

where $x_1 = 0$. Now define $W(x_2, r^1, \theta_1)$ has the right hand side of the first equation above, i.e.:

$$W(x_2, r^1, \theta_1) \equiv u\big(c(r^1) - M(x_2, r^1)\big) - h\left(\frac{y(r^1)}{w_1(\theta_1, e_1)}\right) + \beta \int V(x_2, r^1, e(r^1), \theta_2) f^2(\theta_2) d\theta_2.$$

Now I impose that

$$W(x_2, r^1, \theta_1) \leq \omega(\theta_1) \ \forall x_2, r^1, \theta_1$$
$$W(0, \theta_1, \theta_1) = \omega(\theta_1) \ \forall \theta_1.$$

The inequality is there to ensure that in period 1, it is optimal for the individual to report truthfully, $r^1 = \theta_1$ and not save, $x_2 = 0$. The equality is there to make sure that this mechanism delivers the same utility as the original mechanism. Imposing these inequalities is equivalent to imposing

$$M(x_2, r^1) \geq M^*(x_2, r^1),$$

where

$$M^*(x_2, r^1) \equiv \max_{\theta_1} \tilde{M}^*(x_2, r^1, \theta_1),$$

and

$$\tilde{M}^*(x_2, r^1, \theta_1) \equiv c(r^1) - u^{-1}\left(\omega(\theta^1) + h\left(\frac{y(r^1)}{w_1(\theta_1, e_1)}\right) - \beta \int V(x_2, r^1, e(r^1), \theta_2) f^2(\theta_2) d\theta_2\right). \tag{B.1}$$

The last definition, $\tilde{M}^*$, is a hypothetical tax function that ensures that individuals with shock θ_1 are indifferent to any savings or report, i.e. $W(x_2, r^1, \theta_1) = \omega(\theta_1)\ \forall x_2, r^1, \theta_1$. The innovation of Werning (2011) is to show that since this tax function is implausible since it would need to depend on both θ_1 and report r^1, the upper envelope over true types θ_1 can be used which gives the function M^* which is only conditioned on report r^1. This then rules out misreporting in period 1 and 2.[58]

Assuming M^* is differentiable at $x_2 = 0$, the first-order condition of the worker augmented-mechanism problem at zero savings and thruth-telling, when R is the gross rate of return is

$$u'(c(\theta^1)) M^*_x(0, \theta_1) = \beta R \int \frac{\partial V(0, \theta_1, \theta_2)}{\partial x_2} f^2(\theta_2) d\theta_2,$$

and this implies that

$$M^*_x(0, \theta_1) = \beta R \frac{E(u'(c(\theta^2))}{u'(c(\theta^1))} = \frac{1}{1 - \tau_K(\theta^1)},$$

where $\tau_K(\theta^1)$ is the intertemporal wedge as defined above.

Now, consider the history-dependent tax system with labor income tax, $T = \Big(T_1(y_1), T_2(y_1, y_2)\Big)$,

[58] Tax function M^* is the lowest possible tax that prevents a deviation.

and savings tax , $M(x_2, y_1)$, where y_t is labor income in a decentralized economy. The worker has then the following budget constraints

$$c_1 + M(x_2, y_1) \leq y_1 - T_1(y_1), \tag{B.2}$$

$$c_2 \leq x_2 + y_2 - T_2(y_1, y_2), \tag{B.3}$$

The worker's problem is then to maximize his lifetime utility subject to (B.2) and (B.3). The market allocation of this decentralized economy is given by vectors $\{\tilde{c}, \tilde{y}, \tilde{x}\}$. Following Kapicka & Neira (2014) it is possible to prove a version of the *taxation principle*

***Modified Taxation Principle*:** If an allocation, $\{c, y\}$ is incentive compatible, i.e. (2.8) is satisfied, then there exist a tax system such that $M(0, y_1) = 0$ for all y_1 and and allocation $\{c, y, 0\}$ solves the worker's decentralized problem. Reciprocally, consider a tax system, composed of T and M, that is such that $\{c, y, x_2\}$ solves the worker's decentralized problem, then the allocation is also incentive compatible.

The proof of this goes as follows, let

$$\begin{aligned} T_1(y_1(\theta^1)) &= c_1(\theta^1) - y_1(\theta^1) \\ T_2(y_1(\theta^1), y_2(\theta^2)) &= c_2(\theta^2) - y_2(\theta^2) \\ M(x_2, y_1(\theta^1)) &= M^*(x_2, \theta_1), \end{aligned} \tag{B.4}$$

with T_1 and T_2 set very high for values of income not considered in the optimal plan such that no worker would ever choose them. Defining M^* has before and thus by construction I have

$$\omega(\theta^1) = W(0, \theta_1, \theta_1) \geq \max_{x_2} W(x_2, r^1, \theta_1) \geq W(0, r^1, \theta_1) \ \forall r^1 \in [\underline{\theta}, \bar{\theta}],$$

and

$$\omega(\theta^2) = V(0, \theta_2, \theta_2).$$

So choosing $\{c, y, 0\}$ yields lifetime utility $\omega(\theta^1)$ for an individual of type θ^1, and therefore any other choice would lead to lower or equal lifetime utility. So this implies that $\{c, y, 0\}$ is the solution to the worker's problem.

In the opposite way, if you take a tax system (T, M), and let $\{c, y, x_2\}$ be the solution to the worker's decentralized problem. Then it must be that a worker with history θ^2 will prefer allocation $\{c(\theta^2), y(\theta^2), x_2(\theta^1)\}$ over any other allocation $\{c(\tilde{\theta^2}), y(\tilde{\theta^2}), x_2(\tilde{\theta^2})\}$. This implies that allocation $\{c, y\}$ is incentive compatible.

From there it is straightforward to show that the optimal marginal tax rate on savings will be $\frac{1}{1-\tau_K(\theta^1)}$ and that the marginal labor tax rates T_1' and T_2' evaluated at the optimal allocation must coincide with the optimal labor wedges $\tau_L(\theta^1)$ and $\tau_L(\theta^2)$.

C Computational Information

C.1 Further Numerical Simulations

Figure 9: Optimal Wedges: $\rho = 1.2$ with Changes in t=2's σ

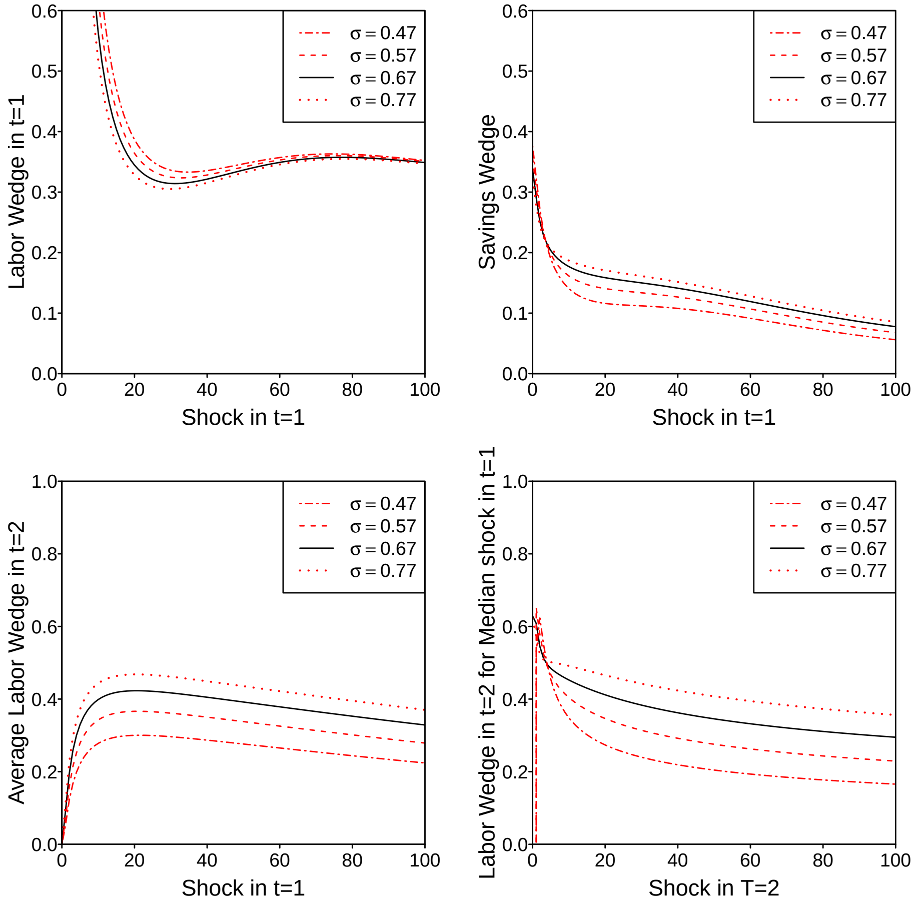


Figure 10: Optimal Wedges: $\rho = 1.2$ with Changes in ζ