

## ADEMU WORKING PAPER SERIES

# Quantifying the Welfare Gains from History Dependent Income Taxation

Marek Kapička<sup>†</sup>

August 2017

WP 2017/071

[www.ademu-project.eu/publications/working-papers](http://www.ademu-project.eu/publications/working-papers)

### Abstract

I quantify the welfare gains from introducing history dependent income tax in an incomplete markets overlapping generations framework where individuals face uninsurable idiosyncratic shocks. I assume that the income tax paid is a function of a geometrical weighted average of past incomes, and solve for the optimal weights. I find that the two main factors that determine the nature of history dependence are the degree to which the government discounts future generations and the degree of mean reversion in the productivity process. The welfare gains from history dependence are large, about 1.76 percent of consumption. I decompose the total effect into an efficiency effect that increases labour supply, and an insurance effect that reduces volatility of consumption and find that, quantitatively, the insurance effect dominates the efficiency effect. The optimal tax increases consumption insurance by trading higher tax progressivity with respect to past incomes for a reduced tax progressivity with respect to the current income.

**Keywords:** risk sharing, income taxation, incomplete markets

**Jel codes:** E6, H2

---

<sup>†</sup> CERGE-EI, Prague, [mkapicka@gmail.com](mailto:mkapicka@gmail.com)

## Acknowledgments

I would like to thank seminar participants at IIES Stockholm, conference participants at 2016 WAMSLAEF conference, 2016 National Tax Association conference, 2017 Swiss Macro Workshop, 2017 Bristol Macro Workshop, 2017 SED in Edinburgh and 2017 NBER Macro-Public Finance Workshop. I am also grateful for financial support from the ADEMU project "A Dynamic Economic and Monetary Union", funded by the European Union's Horizon 2020 Program under grant agreement No 649396. All errors are mine.

---

The ADEMU Working Paper Series is being supported by the European Commission Horizon 2020 European Union funding for Research & Innovation, grant agreement No 649396.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.



# 1 Introduction

Recent research on dynamic optimal taxation, starting with [Goloso et al. \(2003\)](#), has shown one robust feature of optimal tax systems in economies with idiosyncratic shocks: they should in general depend on the full history of individual's incomes.<sup>1</sup> Yet, our understanding of how specifically the income taxes should depend on past incomes is limited. How important is one's income last year relative to one's income ten years ago? What are the welfare gains from history dependence, and where do the welfare gains come from? What parameters of the environment are key for determining the gains from history dependence? How much is lost by restricting history dependence to a limited number of periods? Robust answers to questions like these have not yet been provided.

This paper answers those questions in an analytically tractable framework similar to the one used by [Heathcote et al. \(2014\)](#). Individuals face uninsurable random walk productivity shocks, and the government uses a nonlinear income tax of the form  $T_t(y^t) = y_t - \lambda \bar{y}_t^{1-\tau}$ , where  $y_t$  is a current income, and  $\bar{y}_t$  is an aggregator of past incomes:  $\bar{y}_t = \prod_{j=0}^t (y_{t-j})^{\theta_j}$ . The tax is thus history dependent, with the relative importance of past incomes embodied in the coefficients  $\{\theta_j\}$ . The framework is flexible enough to incorporate history independent income taxes, income taxes that depend on a finite number of lags, and income taxes that depend on the full history of incomes. It also retains tractability that characterizes this environment with history independent taxes.

Why is history dependence useful? In principal, history dependence could either help reduce distortions of labor supply (*incentive effect*), or provide a more efficient consumption insurance (*insurance effect*). One can write the tax system in a way that the incentive effect is determined only by the parameter  $\tau$ , called the progressivity wedge: higher  $\tau$  drives a wedge between wage and marginal rate of substitution, and reduces labor supply. The extent of consumption insurance is, on the other hand, driven by both the progressivity wedge  $\tau$  and the history dependence coefficients  $\theta$ . I show that the coefficients  $\theta$  depend only on a small number of underlying parameters. Notably, they are independent of the progressivity parameter  $\tau$ .

I show that the character of history dependence depends manly on two things: how the government discounts future generations, and how mean reverting the underlying

---

<sup>1</sup>Exceptions to this rule involve some economies where all uncertainty is resolved in the initial period and, to some extent, economies with IID shocks, where history dependence can be replaced by dependence on assets ([Albanesi and Sleet \(2006\)](#)).

productivity process is. Government's discounting of future generations drives a wedge between how the government effectively discounts future periods, and how the agent discount future periods. When the government is more patient than the agents, it can exploit the differences in discounting to shift consumption insurance across ages, again without changing labor supply incentives. In particular, since the government is more patient, it prefers to shift consumption insurance towards later ages, and away from earlier ages. This is achieved by choosing a tax system that is regressive with respect to the current income, and mildly progressive with respect to the past incomes. The second factor is the degree of mean reversion in the labor productivity process. Mean reversion means that there are gains from spreading the impact of a productivity shock over many periods. This is done by making the tax system more progressive with respect to the current income, and regressive with respect to past incomes. That way, current consumption responds to the current shock positively but very little, while at the same time responding positively but very little to past shocks. This increases consumption insurance but, since individuals take into account the response of future consumption to current income, their incentives to work are not reduced. Both factors thus give a very different policy prescription when it comes to the nature of history dependence, and I show conditions under which they cancel each other out, and history independence is optimal.

While the history dependence parameters  $\theta$  are chosen independently of the progressivity wedge  $\tau$ , the reverse is not true. The progressivity wedge is chosen to balance the distortions of labor supply and, again, a reduction in consumption dispersion. Since history dependence already reduces consumption dispersion, the government responds by reducing the progressivity wedge  $\tau$  relative to the case with history independent taxation. Thus, history dependence in the end reduces both the progressivity of the tax system, and the dispersion of consumption.

I calibrate the model to the U.S. economy where the productivity shocks follow a random walk, thus leaving the differences in discounting as the key factor that determines the importance of history dependence. I follow [Heathcote et al. \(2014\)](#) who show that the U.S. tax system can be well approximated by a (history independent) tax system with a progressivity parameter  $\tau = 0.161$ . I show that the welfare gains from history dependence are large: they are about 1.76 percent in consumption equivalents. Quantitatively, almost all of the benefits of history dependent income taxation, about 92 percent of it, come from the fact that it reduces the cross-sectional variance of log consumption,

which is reduced by about a third. The remaining 8 percent comes from efficiency gains, i.e. from the reduction of the progressivity wedge  $\tau$ . In the optimum,  $\tau$  is increased modestly from 0.161 to 0.234. In contrast, if the tax system is history independent, then the current U.S. tax system should be much more progressive, since the parameter  $\tau$  should be increased to 0.296. I also compute welfare gains from a limited history dependence. There are large welfare gains even from relatively short history dependence: Taxes that depend only on the past 10 annual incomes generate about 50 percent of the potential welfare gains from an unlimited history dependence, while taxes that depend on the past 16 incomes generate about 75 percent of the potential welfare gains.

Finally, I compare the welfare gains from a history dependent taxes to taxes that depend only on age. Age dependent taxation achieves a similar reduction of consumption dispersion, by increasing tax progressivity with age. This, however, introduces unnecessary variations in the progressivity wedge over time - something that history dependent taxation is able to evade. As a result, the welfare gains from age dependent taxation, while still large, are only about 48 percent of the overall gains.

The paper connects two strands of the existing literature. On one hand, it uses insights from the recent dynamic public finance literature (Golosov et al. (2003), Kocherlakota (2005)), Albanesi and Sleet (2006), Battaglini and Coate (2008), Farhi and Werning (2005), Werning (2007) Golosov et al. (2016) and Farhi and Werning (2012) and many others) that shows that history dependence in income taxation is optimal. On the other hand, in order to achieve tractability, the paper does not use a standard mechanism design approach to gain insights about the optimal policies. Instead, it follows the tractable analytical framework pioneered by Constantinides and Duffie (1996) and further extended by Heathcote et al. (2014) and Heathcote et al. (2016), who include insurable transitory shocks as well as labor supply decision.<sup>2</sup> As in Benabou (2002), who studies educational decisions in a related framework, I assume that the agents cannot borrow and save to self-insure to gain tractability. Each of the last three papers assumes that income tax function is a power function but, importantly, none of them allows for history dependence. My framework includes history dependence, but retains enough tractability to quantitatively study problems with multiple sources of heterogeneity, with overlapping generations, and in general equilibrium, none of which has been a focus of the dynamic optimal taxation literature.

---

<sup>2</sup>Other functional forms used in the literature can be found in (Conesa and Krueger (2006), Kindermann and Krueger (2017), and others.

## 2 Setup

The model is a “perpetual youth” overlapping generation model. There is a measure one of agents in the population. Each period, a fraction  $1 - \delta$  of each generation dies and is replaced by a new generation of size  $1 - \delta$  as well. The measure of age- $j$  agents is thus  $(1 - \delta)\delta^j$ . The utility function takes the form

$$w_0 = \mathbb{E}_0 \sum_{j=0}^{\infty} (1 - \beta\delta)(\beta\delta)^j \left( \ln c_j - \frac{\phi}{1 + \eta} h_j^{1+\eta} \right), \quad 0 \leq \beta < 1. \quad (1)$$

The parameter  $\eta \geq 0$  is inverse of the Frisch elasticity of labor supply, while  $\phi$  determines the relative weight of disutility from working. The parameter  $\eta$  is identical for all agents, while  $\phi$  is a random variable, drawn once at age zero, with a cumulative distribution function  $F_\phi$ . At each age, agents received labor productivity  $z_j \in Z \equiv (0, \infty)$ . Their productivity, together with hours worked  $h_j$  determines the output  $y_j = z_j h_j$ . Log productivity  $z_j$  follows an autoregressive process,

$$\ln z_j = \rho \ln z_{j-1} + \omega_j,$$

where  $\omega_j$  is an iid random variable that has a cumulative distribution function  $F_\omega$  with density  $f_\omega$ , and  $\rho$  is autocorrelation of productivity shocks. The distribution is normalized so that  $\mathbb{E}(e^\omega) = 1$ . Both the initial “seed”  $z_{-1}$  and the initial shock  $\omega_0$  are equal to zero, implying that the initial productivity  $z_0$  is equal to one for everyone.<sup>3</sup>

**Market Structure.** I assume that there is *no* insurance against the permanent shocks  $z$ . That includes self-insurance: the agents are not allowed to save to hedge against the permanent shocks. This is a strong assumption, but it allows me to get closed-form solution for the equilibrium allocations even for the tax systems that are history dependent. [Benabou \(2002\)](#) makes the same assumption for the same reasons. I will revisit this assumption later.

---

<sup>3</sup>The model can be easily generalized to allow for initial heterogeneity in productivity, without changing the main results.

## 2.1 A Tax with History Dependence

The government taxes individual incomes by using an income tax that is history dependent: the tax paid depends on individual's history of earnings. The tax function has the following functional form: an individual of age  $j$  with a history of incomes  $y_0, y_1, \dots, y_j$  pays taxes

$$T_j(y_0, y_1, \dots, y_j) = y_j - \lambda_j (\bar{y}_j)^{1-\tau}, \quad (2)$$

where  $\bar{y}_t$  is a weighted geometric average of current and past incomes,

$$\bar{y}_j = \prod_{k=0}^j (y_{j-k})^{\theta_k}.$$

The *progressivity wedge*  $\tau$  determines the overall progressivity of the tax system. The *history dependence* parameters  $\theta = (\theta_0, \theta_1, \dots)$  represent how the current tax paid depends on income realizations in the past. Specifically, the parameter  $\theta_k$  represents the weight on income of lag  $k$ . A history independent tax is a special case with  $\theta_0 = 1$  and  $\theta_k = 0$  otherwise. The tax function  $T_j$  depends on age directly through the level tax parameters  $\lambda_j$ . An age specific level tax parameter allows the government to choose a trend in average consumption independently of the remaining parameters. On the other hand, both the history dependence parameters and the progressivity wedge are time and age invariant, and that is one of the key restrictions in the paper.<sup>4</sup> The parameters  $\tau$  and  $\theta$  jointly determine the progressivity of the tax system. The average income-weighted marginal tax with respect to the current income is  $1 - (1 - \tau)\theta_0$ , and so if  $\theta_0$  is sufficiently large, the tax system may be regressive with respect to the current income. The average income-weighted marginal tax with respect to the income of lag  $k$  is  $-(1 - \tau)\theta_k$ . If  $\theta_k$  is positive then the tax system is regressive with respect to past incomes, while if  $\theta_k$  is negative, the tax system is progressive.

**Incentive keeping constraint.** The tax function leaves one degree of freedom in the tax parameters  $\tau$  and  $\theta$ . This allows us to simplify the problem by choosing a convenient normalization. The first-order conditions yield the following expression for equilibrium

---

<sup>4</sup>Taxes paid also depends indirectly on age, because the length of individual histories depends on age. However, if one assumes that incomes before being born are all equal to one, then the tax function can be written as a time and age invariant function of an infinite history of incomes.

hours worked:

$$\ln h_j^* = \frac{1}{1 + \eta} \left[ \ln(1 - \tau) \sum_{k=0}^{\infty} (\beta\delta)^k \theta_k - \ln \phi \right]. \quad (3)$$

This suggests that a useful way to normalize the tax system is to impose the following *incentive keeping constraint*:

$$\sum_{k=0}^{\infty} (\beta\delta)^k \theta_k = 1. \quad (4)$$

This normalization implies that labor supply decisions of the agents are independent of the history dependent coefficients  $\theta$ , and depend only on the progressivity wedge  $\tau$ . The logic is as follows. Each individual, when choosing hours worked, takes into account the incentive effects of all future taxes paid from current income. If the normalization (4) holds, reducing marginal tax rates in the current period by reducing  $\theta_0$  by one unit must be exactly offset by an increase in the marginal tax rate in some period  $k$  by  $(\beta\delta)^{-k}$  units. Since future taxes paid are effectively discounted at rate  $\beta\delta$ , this trade-off does not change work incentives. The incentive keeping constraint then implies that the incentives to work are determined exclusively by the progressivity wedge  $\tau$ , and are independent of the parameters  $\theta$ .

## 2.2 Allocations

Equation (5) and the incentive keeping constraint (4) imply that the optimal hours worked are

$$\ln h_j^* = \frac{1}{1 + \eta} [\ln(1 - \tau) - \ln \phi]. \quad (5)$$

Hours worked are independent of the productivity shocks, because with log utility the income and substitution effects cancel out. Heterogeneity in the preference parameter  $\phi$  is the only source of heterogeneity in hours worked in this model. Due to the incentive keeping constraint (4), hours worked are independent of the history dependence parameters  $\theta$ , and tax policies only affect hours worked by the progressivity wedge  $\tau$ .

Individual consumption can be obtained by using the budget constraint  $c_j = \lambda_j (\bar{y}_j)^{1-\tau}$ ,

and by substituting in the equilibrium hours worked:

$$\ln c_j = \ln \lambda_j + \frac{1 - \tau}{1 + \eta} \sum_{k=0}^j \theta_k [\ln(1 - \tau) - \ln \phi] + (1 - \tau) \sum_{k=0}^j \theta_k \ln z_{j-k}. \quad (6)$$

Consumption depends on all past productivities only because taxes paid depend on past incomes. The key in determining the nature of history dependence are, of course, the history dependence parameters  $\theta$ . Note that consumption in general can move predictably with age, first because  $\lambda$  depends on age and, second, because the expected value of the weighted average of past incomes  $\bar{y}_j$  is changing with age. Specifically, if  $\sum_{k=0}^j \theta_k$  is decreasing in  $j$ , then the expected consumption will be increasing with age.

### 3 Partial Equilibrium

I will start by analyzing a government's problem in a partial equilibrium, where prices are exogenous, and there is only one generation, born at time zero. The government maximizes the following social welfare function:

$$\mathcal{W} \equiv (1 - \alpha\delta) \sum_{j=0}^{\infty} (\alpha\delta)^j \mathbb{E}_0(u_j). \quad (7)$$

where  $u_j$  is period utility, and  $\alpha \in [\beta, 1)$  is the government's discount factor. The government is therefore discounting future at a rate that may be different from the agent's discount rate. This assumption can be taken as a primitive assumption about the social welfare function, as in [Farhi and Werning \(2005\)](#). In that case, the social welfare function is non utilitarian. But, as we shall see later, the assumption can be also justified in an overlapping generation setting, where the government has a utilitarian social welfare function for each generation, but cares more about future generations than about the current generation (in which case  $\alpha > \beta$ ). In any case, the additional flexibility of allowing for a different discount factor for the government will prove useful for clarifying what matters for history dependence.

The tax system has aggregate cost equal to the present value of aggregate consump-

tion minus the present value of aggregate earnings:

$$\mathcal{P} \equiv (1 - q\delta) \sum_{j=0}^{\infty} (q\delta)^j \mathbb{E}_0(c_j - y_j), \quad (8)$$

where the interest rate  $q \in (0, 1)$  is exogenously given. The government chooses the tax parameters  $\lambda$ ,  $\tau$  and  $\theta$  to maximize the social welfare function (7) subject to the resource constraint  $\mathcal{P} = 0$  and the incentive keeping constraint (4), taking the policy functions (5) and (6) as given.<sup>5</sup>

Substituting the policy functions (5) and (6) into the social welfare function (7) and the resource constraint and optimizing with respect to the level parameters  $\lambda$  yields the following optimality condition: the parameters  $\lambda$  should be chosen in a way that the aggregate consumption is growing at a rate equal to  $\alpha/q$ :

$$\mathbb{E}_0(c_j) = \mathbb{E}_0(c_0) \left( \frac{\alpha}{q} \right)^j. \quad (9)$$

Aggregate consumption is thus constant if  $\alpha = q$ , and the government's discount factor is exactly offset by the rate of return. Note that individual discount factor  $\beta$  does not play any role in determining the trend in aggregate consumption because an individual cannot shift resources over time on his own. This is reminiscent of [Calvo and Obstfeld \(1988\)](#) finding that aggregate consumption in the Pareto optimum of an overlapping generations economy is determined by the social planner's rate of time preference, and individual's time preference does not play any role. It is shown in Appendix A that if the parameters  $\lambda$  are set optimally, the aggregate welfare can be written as a function of  $\tau$  and  $\theta$  only, and is given by

$$\begin{aligned} \mathcal{W}(\tau, \theta) = & \bar{u}(\tau) + \ln \left( \frac{1 - \alpha\delta}{1 - q\delta} \right) + \frac{\alpha\delta}{1 - \alpha\delta} (\ln \alpha - \ln q) \\ & - \sum_{j=0}^{\infty} (\alpha\delta)^j \left[ \alpha\delta \ln B_{\omega} \left( (1 - \tau) \sum_{k=0}^j \rho^{j-k} \theta_k \right) + (1 - \alpha\delta) \ln B_{\phi} \left( - (1 - \tau) \sum_{k=0}^j \frac{\theta_k}{1 + \eta} \right) \right] \\ & + (1 - \tau) \left( \frac{\alpha\delta}{1 - \alpha\delta\rho} \mathbb{E}\omega - \frac{1}{1 + \eta} \mathbb{E} \ln \phi \right) \sum_{j=0}^{\infty} (\alpha\delta)^j \theta_j, \end{aligned} \quad (10)$$

<sup>5</sup>It is easy to allow for exogenous government spending  $G$ , in which case the resource constraint becomes  $\mathcal{P} + G = 0$ .

where  $B_\omega(a) = \mathbb{E}(e^{a\omega})$  and  $B_\phi(a) = \mathbb{E}(e^{a \ln \phi})$  are moments of the underlying distributions, and  $\bar{u}(\tau) = \frac{1}{1+\eta} [\ln(1-\tau) - 1 + \tau]$  is the representative agent's period utility.<sup>6</sup> The government's problem is now to maximize the objective function (10) by choosing the tax parameters  $\tau$  and  $\theta$  subject to the incentive keeping constraint (4).

### 3.1 Lognormal Distribution

The government's problem will be greatly simplified if both the productivity shocks and the preference shocks are lognormally distributed. To that end, assume that

$$\begin{aligned}\omega &\sim N\left(-\frac{\sigma_\omega^2}{2}, \sigma_\omega^2\right) \\ \ln \phi &\sim N\left((1+\eta)\frac{\sigma_\phi^2}{2}, (1+\eta)^2\sigma_\phi^2\right).\end{aligned}$$

The mean and variance of the taste shock distribution are normalized so as to simplify algebra later. Then  $B_\omega(a) = e^{-\frac{1}{2}a(1-a)\sigma_\omega^2}$  and  $B_\phi(a) = e^{\frac{1}{2}(1+\eta)a(1+(1+\eta)a)\sigma_\phi^2}$  and the government's objective function (10) reduces to

$$\mathcal{W}(\tau, \theta) = \bar{u}(\tau) - \frac{1}{2}(1-\tau)^2 \left[ P_{\alpha, \rho}(\theta) \kappa \sigma_\omega^2 + P_{\alpha, 1}(\theta) \sigma_\phi^2 \right], \quad (11)$$

where  $\kappa = \frac{\alpha\delta}{1-\alpha\delta\rho^2}$  is a parameter that scales the productivity shocks, and  $P_{\alpha, \rho}(\theta)$  is a quadratic function of only of the history dependence parameters, given by

$$P_{\alpha, \rho}(\theta) = \sum_{j=0}^{\infty} (\alpha\delta)^j \left( \theta_j^2 + 2 \sum_{k=0}^{j-1} \rho^{j-k} \theta_k \theta_j \right).$$

The functional form of the social welfare function (11) shows that the optimal choice of the history dependence parameters is determined by minimizing the weighted average of  $P_{\alpha, \rho}$ , which represents the contribution of the productivity shocks, and  $P_{\alpha, 1}$ , which represents the contribution of the preference shocks. The weights are given by the relative variances of both shocks, with the productivity shock being scaled by  $\kappa$  to correct for the fact that the shocks are persistent and discounted. The welfare function (11) also shows that the choice of the history dependence parameters  $\theta$  is independent of the choice of

---

<sup>6</sup>In principle, there is an additional constant term reflecting the aggregate production gains from the dispersion in  $\phi$ . This term is zero due to the normalization of the taste shock distribution.

the progressivity wedge  $\tau$ , although the reverse is clearly not true. This simplifies the problem, both technically, and substantially. Furthermore, it is easy to see that, under history independence, the value of  $P_{\alpha,\rho}$  equals one for all values of  $\alpha$  and  $\rho$ . The welfare gains from a history dependent tax policy are determined by the reduction of the weighted average of  $P_{\alpha,\rho}$  and  $P_{\alpha,1}$ . The size of the welfare gains is also proportional to the square of the progressivity wedge  $\tau$ .

Turning to the progressivity wedge  $\tau$ , an increase in  $\tau$  weighs the costs in terms of labor supply distortions against the benefits of reduction in the dispersion of consumption (see 11). Since history dependence reduces consumption dispersion by reducing  $P$  below one, the benefits from higher  $\tau$  are reduced, and the balance shifts in favor of lower labor supply distortion:

**Proposition 1.** *The optimal progressivity wedge  $\tau^*$  decreases when history dependence is allowed.*

Before solving for the optimum history dependence coefficients, it is worth considering a simple example that sheds light on why, and under what conditions, is history dependence in income taxation welfare improving.

### 3.2 Optimal History Dependence

The social welfare function shows that the optimal history dependence parameters are a solution to the following simple minimization problem:

$$\theta^* = \arg \min_{\theta} \left\{ P_{\alpha,\rho}(\theta) \kappa \sigma_{\omega}^2 + P_{\alpha,1}(\theta) \sigma_{\phi}^2 \right\} \quad \text{s.t. (4)}.$$

This minimization problem makes it clear that the optimal history dependence coefficients depend only on the discount factors  $\beta$  and  $\alpha$ , on the survival rate  $\delta$ , on the autocorrelation of shocks  $\rho$ , and on the ratio of variances  $\sigma_{\omega}^2 / \sigma_{\phi}^2$ . Notably, they are independent of the progressivity wedge  $\tau$ , and the level of the two shock variances. The first-order condition in  $\theta_k$  is

$$\sum_{l=0}^k \left(1 + \kappa \rho^{k-l}\right) \theta_l + \sum_{l=k+1}^{\infty} (\alpha \delta)^{l-k} \left(1 + \kappa \rho^{l-k}\right) \theta_l = \left(\frac{\beta}{\alpha}\right)^k \zeta,$$

where  $\zeta$  is an (appropriately normalized) Lagrange multiplier on (4). This is a difference equation that, together with (4), can be solved for the coefficients  $\theta$ , and for the Lagrange multiplier  $\zeta$ .

In general, the difference equation above has to be solved numerically. In special cases, there is a closed form solution. We will first investigate two examples with a closed form solution. The examples will highlight two of the main sources for welfare gains: differences in discounting between government and the agent, and mean reversion in the productivity process.

### 3.3 Example 1: The role of differences in discounting

If  $\rho = 1$ , it is easy to see that the above minimization problem is equivalent to the minimization of  $P_{\alpha,1}$  subject to (4). It also follows, that the history dependence coefficients are independent of both  $\sigma_\omega^2$  and  $\sigma_\phi^2$ , and that the welfare now depends only on the total variance of shocks  $\kappa\sigma_\omega^2 + \sigma_\phi^2$ .

**Why History Dependence?** To understand how differences in discounting shape history dependence, consider the following perturbation to a history independent tax with  $\theta_0 = 1$  and  $\theta_1 = 0$ . Suppose that  $\theta_0$  increases by  $d\theta_0 > 0$ . An increase in  $\theta_0$  has two effects. First, it makes the tax system less progressive at all ages, and directly changes welfare by  $\psi \times d\theta_0 < 0$ . Second, a change in  $\theta_0$  requires an adjustment in the tax parameters  $\lambda_0$  and  $\lambda_1$  to keep mean consumption unchanged, and the government budget constraint holds. This changes welfare by  $-(\psi + \sigma_\phi^2) \times d\theta_0 < 0$ . Taken together, the level effects cancel each other out and an increase in  $\theta_0$  on net decreases welfare by  $\sigma_\phi^2 \times d\theta_0$ . The effect is negative, because an increase in  $\theta_0$  reduces consumption insurance at every age.

An increase in  $\theta_0$  is, however, compensated by a decrease in  $\theta_1$ . The incentive keeping constraint implies that  $\theta_1$  must change by  $d\theta_1 = -(\beta\delta)^{-1}d\theta_0 < 0$ . The effects now have the opposite sign: a decrease in  $\theta_1$  changes utility directly by  $\alpha\delta\psi \times d\theta_0 > 0$  and indirectly through adjustment in  $\lambda_1$  by  $\alpha\delta(\psi + \sigma_\phi^2) \times d\theta_1$ , with a positive net effect  $-\alpha\delta\sigma_\phi^2 \times d\theta_1 > 0$ . The positive net effect now comes from an increased consumption insurance for agents with age one and more; that's why its size is  $\alpha\delta$  times smaller than what a perturbation

in  $\theta_0$  produces. The overall change in welfare is

$$d\mathcal{W} = -\sigma_\phi^2 \times d\theta_0 - \delta\alpha\sigma_\phi^2 \times d\theta_1 = \left(\frac{\alpha}{\beta} - 1\right) \sigma_\phi^2 \times d\theta_0.$$

If  $\alpha > \beta$  then the total effect is positive, and it is optimal to increase  $\theta_0$  and decrease  $\theta_1$ . To understand this result, note that, by changing history dependence, the government changes consumption insurance across ages without changing labor supply incentives (as long as the incentive keeping constraint holds). The net effect depends on two things: at what rate can consumption insurance be traded across ages, and how the government values consumption insurance at different ages. An increase in  $\theta_0$  decreases consumption insurance at all ages, while a decrease in  $\theta_1$  increases consumption insurance in the future (at ages one and on). The trade-off is determined by the individual discount factor  $\beta\delta$ . Since  $\beta\delta < 1$ , an increase in  $\theta_0$  allows for a proportionally larger reduction in  $\theta_1$ , and the above perturbation decreases consumption insurance at age zero while increasing consumption insurance in the future. Now, how important this trade-off is for the government is determined by the government's discount factor  $\alpha\delta$ . If the government is more patient than the agent, an increase in future consumption insurance dominates, and the trade-off is worthwhile.

If  $\beta = \alpha$ , then a perturbation in  $\theta_0$  or  $\theta_1$  each has nonzero first-order effect on welfare, and the trade-off between consumption insurance across ages still exists. But the effects exactly cancel each other in the social welfare function, and a history independent tax policy is an optimal one. If, on the other hand,  $\sigma_\phi = 0$  then there is no trade-off. any change in  $\theta$  is exactly compensated by changes in  $\lambda_0$  and  $\lambda_1$ , and any perturbation in  $\theta_0$  or  $\theta_1$  has zero welfare effect. It is again optimal to have history independent tax because it smooths consumption for the representative agent.

**Main result.** I will now show that an analogous intuition extends to the more general case with infinite history dependence. It also carries over to the case when there is idiosyncratic uncertainty about productivity shocks, rather than preference shocks.

**Proposition 2.** *If  $\rho = 1$  then the welfare maximizing coefficients  $\theta^*$  are*

$$\theta_0^* = \frac{1 - \frac{\beta^2\delta}{\alpha}}{1 - \beta\delta}, \quad \theta_k^* = - \left(1 - \frac{\beta}{\alpha}\right) \left(\frac{\beta}{\alpha}\right)^{k-1} \theta_0^*, \quad k > 0,$$

and the value of  $P_{\alpha,1}(\theta^*)$  is

$$P_{\alpha,1}(\theta^*) = \frac{(1 - \delta\alpha) \left(1 - \frac{\beta^2\delta}{\alpha}\right)}{(1 - \beta\delta)^2}.$$

Proposition 2 offers a very simple and practical characterization of the optimal progressivity coefficients. It immediately implies that the optimal coefficient on the current income is positive and greater than one, while the coefficients on past incomes are negative. Relatively to a history independent tax, the tax system thus becomes less progressive with respect to the current income and more progressive with respect to past incomes. The results, and the intuition, are both analogous to the example above.

The coefficients on past incomes are geometrically increasing, and converge to zero at a rate equal to the ratio of both discount factors,  $\beta/\alpha$ . They will typically be significantly smaller than the coefficient on the current income: for example, the coefficient on the previous period income is only  $1 - \beta/\alpha$  the coefficient on the current income. Even for  $\alpha = 1$ , it will be only around 2-4 percent of what the coefficient on the current income is.

**Reinterpreting History Dependent Tax.** The history dependent tax function (2) can, for the optimum coefficients  $\theta^*$ , be reinterpreted in two useful ways. First, by rearranging the terms, one can write the geometric average of past incomes as

$$T_t = y_t - \left[ \prod_{k=0}^t \left( \frac{y_{t-k}}{y_{t-k-1}} \right) \left( \frac{\beta}{\alpha} \right)^k \right]^{(1-\tau)\theta_0}.$$

That is one the optimal tax can be written as a function of a weighted average of past income *growth rates*, as opposed to a weighted average of the levels of income. The result follows intuitively from the fact that the productivity shocks are essentially shocks to the growth rate of productivity, as opposed to shocks to the level of productivity. Alternatively, and perhaps more usefully from a practical perspective, one can define an average past income as  $x_t$  by  $x_t = \prod_{k=0}^t y_{t-k}^{(1-\beta/\alpha)(\beta/\alpha)^k}$ . Then the tax paid is

$$T_t = y_t - \lambda \left( \frac{y_t}{x_{t-1}} \right)^{(1-\tau)\theta_0}.$$

The past income average is simply updated recursively according to  $x_t = y_t^{1-\beta/\alpha} x_{t-1}^{\beta/\alpha}$ , and every agent starts with  $x_0 = 1$ . This formulation is useful since one needs only one state variable, the average of past incomes  $x_{t-1}$ , to characterize the current tax function.

**Limited History Dependence** What if the government is only allowed to use income history of length  $K$ ? Next proposition shows the closed form solution to the truncated case:

**Proposition 3.** *Assume that  $\theta_k = 0$  for  $k > K$ , for some  $K > 0$ . Then the welfare maximizing coefficients are*

$$\begin{aligned}\theta_0^* &= \frac{1 - \frac{\beta^2 \delta}{\alpha}}{1 - \beta \delta} \left[ 1 - \frac{\left(1 - \frac{\beta}{\alpha}\right)^2 \delta \alpha}{(1 - \beta \delta)^2} \left(\frac{\beta^2 \delta}{\alpha}\right)^K \right]^{-1} \\ \theta_k^* &= - \left(1 - \frac{\beta}{\alpha}\right) \left(\frac{\beta}{\alpha}\right)^{k-1} \theta_0^*, \quad k = 1, \dots, K-1 \\ \theta_K^* &= - \frac{1}{1 - \beta \delta} \left(1 - \frac{\beta}{\alpha}\right) \left(\frac{\beta}{\alpha}\right)^{K-1} \theta_0^*.\end{aligned}$$

and the value of  $P_{\alpha, \rho}(\theta^*)$  is

$$P_{\alpha, \rho}(\theta^*) = \frac{(1 - \delta \alpha) \left(1 - \frac{\beta^2 \delta}{\alpha}\right)}{(1 - \beta \delta)^2} \left[ 1 - \frac{\left(1 - \frac{\beta}{\alpha}\right)^2 \delta \alpha}{(1 - \beta \delta)^2} \left(\frac{\beta^2 \delta}{\alpha}\right)^K \right]^{-1}.$$

The coefficients in the full history dependence problem are just a limiting case with  $K$  going to infinity. On the other hand, there is a substantial difference between a tax that includes only current and past income ( $K = 1$ ), and a tax under full history dependence. For example, when  $\tau = 0.2$ ,  $\alpha = 1$ ,  $\beta = 0.96$  and  $\delta = 0.97$  then  $1 - (1 - \tau)\theta_0^* = -0.75$ . The average marginal income tax rate with respect to the current income is thus -75 percent. In contrast, the current income is taxed only at a rate of -23 percent with full history dependence. There is also a substantial difference between a tax with no history dependence ( $K = 0$ ) and a tax that includes only current and past income. However, including one period lag reduces variance of consumption by only about 5 percent, significantly less than under full history dependence. Although the value of  $P_{\alpha, \rho}(\theta^*)$  is decreasing in  $K$ , one needs more than one lag to capture most of the potential benefits

from history dependence.

### 3.4 Example 2: The role of mean reversion in productivity

Differences in discounting between the government and the agent are one source of welfare gains from history dependent income taxation. A second source of welfare gains comes from the fact that productivity shocks are mean reverting. To investigate this source of welfare gains, assume now that  $\alpha = \beta$ , so that there is no difference in discounting, but allow for any  $\rho \in [0, 1]$ . If, in addition, there is no heterogeneity in the preference parameter  $\phi$ , the problem again has a closed form solution. We get

**Proposition 4.** *If  $\alpha = \beta$  and  $\sigma_\phi^2 = 0$  then the welfare maximizing coefficients  $\theta^*$  are*

$$\theta_0^* = \frac{1 - \beta\delta}{1 - \beta\delta\rho}, \quad \theta_k^* = (1 - \rho)\theta_0, \quad k > 0,$$

and the value of  $P_{\beta,\rho}(\theta^*)$  is

$$P_{\beta,\rho}(\theta^*) = \frac{(1 - \beta\delta)(1 - \beta\delta\rho^2)}{(1 - \beta\delta\rho)^2}.$$

The current coefficient is always smaller than one, while the coefficients on past incomes are positive. That is, the tax on past incomes is negative, and the optimal tax system increases progressivity with respect to the current income, while decreasing progressivity with respect to the past incomes. Interestingly, the coefficients on past incomes are constant, and do not change with the length of the history.

The autocorrelation of shocks is critical in determining the gains from history dependence, with lower  $\rho$  delivering higher welfare gains. If  $\rho = 1$  and the shocks follow a random walk, then there are *no* gains from history dependence:<sup>7</sup>

**Proposition 5.** *If  $\alpha = \beta$  and  $\rho = 1$  then history independent tax system is optimal.*

At the other extreme, if  $\rho = 0$  and the shocks are iid, then  $\theta_0 = \theta_k = 1 - \beta\delta$  for all  $k > 0$ , and all the shocks have the same weight. It is therefore optimal to simply take an unweighted geometric average of all the past incomes.

To illustrate the welfare gains, I calibrate the stochastic process for productivity and preference shocks according to [Kaplan \(2012\)](#). The autocorrelation of shocks is  $\rho = 0.958$ ,

---

<sup>7</sup>The proposition applies even if  $\sigma_\phi^2 > 0$ .

the variance of productivity innovations is  $\sigma_\omega^2 = 0.017$ , and the variance of preference shocks is  $\sigma_\phi^2 = 0.065$ . We also set  $\beta = 0.96$ ,  $\delta = 1$  (no mortality shock) and  $\eta = 2$ . If one starts with the best available history independent tax system, the welfare gain is about 0.66 % in consumption equivalents. Almost all of the welfare gains come from the reduction of consumption variance. Only a small part comes from the reduction in the progressivity wedge  $\tau$  which decreases from 0.299 under the history independent tax system to 0.277. The welfare gains are somewhat larger if one starts with the current U.S. tax code, which is represented by a history independent tax function with  $\tau = 0.161$ . The welfare gains from history dependence only are 0.89 %, with additional welfare gains coming from increasing the progressivity wedge to 0.277.

**Limited History Dependence** If the income history is exogenously restricted to be of length  $K$ , the solution has to be modified. Next Proposition shows the solution:

**Proposition 6.** *Assume that  $\theta_k = 0$  for  $k > K$ , for some  $K > 0$ . If  $\alpha = \beta$  and  $\sigma_\phi^2 = 0$  then the welfare maximizing coefficients are*

$$\begin{aligned}\theta_0^* &= \frac{1 - \beta\delta}{1 - \beta\delta\rho} \left[ 1 - (\beta\delta)^{K+1} \left( \frac{1 - \rho}{1 - \beta\delta\rho} \right)^2 \right]^{-1}, \\ \theta_k^* &= (1 - \rho)\theta_0, \quad k = 1, \dots, K - 1 \\ \theta_K^* &= \frac{1 - \rho}{1 - \beta\delta\rho} \theta_0.\end{aligned}$$

and the value of  $P_{\beta,\rho}(\theta^*)$  is

$$P_{\beta,\rho}(\theta^*) = \frac{(1 - \beta\delta)(1 - \beta\delta\rho^2)}{(1 - \beta\delta\rho)^2} \left[ 1 - (\beta\delta)^{K+1} \left( \frac{1 - \rho}{1 - \beta\delta\rho} \right)^2 \right]^{-1}.$$

The initial history dependence coefficient decreases with the length of permissible history  $K$ . Interestingly, the ration of the coefficient on the past income to the coefficient on the current income is still  $1 - \rho$ . The exception is the coefficient on the last permissible income  $\theta_K$ , that now increases to (imperfectly) pick up for the missing coefficients on higher incomes. The coefficients in the full history dependence problem are again a limit when  $K$  goes to infinity.

Figure 1 illustrates the welfare gains from limited history dependence, and the cor-

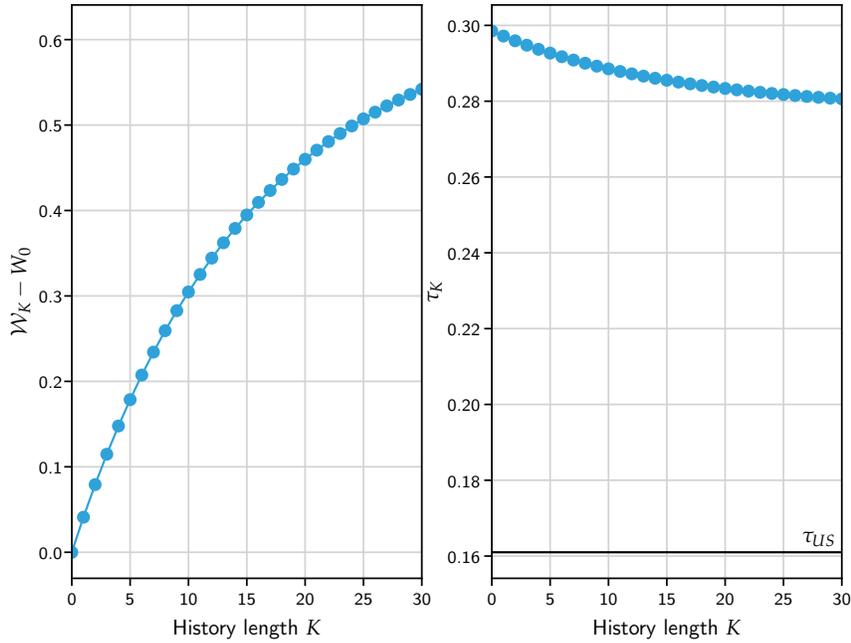


Figure 1: Welfare gains from partial history dependence in pct consumption equivalents (left panel) and the progressivity wedge (right panel).

responding progressivity wedge. The welfare gains from a short permissible income history are rather small: including 5 past incomes lead to a welfare gain of about 0.18%, less than a third of the potential welfare gain under full history dependence. Including 15 past incomes produces a welfare gain of 0.4%, a little less than two thirds of the potential welfare gains. A relatively long history is thus needed to capture most of the welfare gains.

### 3.5 When is history independence optimal?

Previous section has identified one case when it is optimal to have history independence: when the shocks follow a random walk, and there is no difference in discounting. It is, however, optimal to have history independent taxation in one additional "knife-edge" case: when the pressure for more progressivity with respect to past incomes coming from the fact that  $\alpha > \beta$  exactly cancels out with the pressure for less progressivity with respect to past income coming from the fact that  $\rho < 1$ . This will happen under the

following conditions:

**Proposition 7.** *If  $\rho = \beta$ ,  $\alpha = 1$  and  $\sigma_\phi^2 = 0$  then history independent tax system is optimal.*

Previous results show that there are two opposing forces in play. First, if  $\alpha > \beta$ , it is optimal to shift consumption insurance forward. If  $\rho < 1$ , it is optimal to shift it backward. Under the conditions of the proposition, both forces cancel out, and history independence is optimal.

### 3.6 Age Independent Tax Function

Up until now, the government was allowed to use a tax function that depended directly on age through the parameters  $\lambda_j$ . What happens if the government does not have the opportunity to choose age specific  $\lambda$ ? I derive a full solution in Appendix A. I show several results there. First, the inability to transfer resources across age through variations in  $\lambda$  means that the parameters  $\tau$  and  $\theta$  will, at least in part, be chosen so as to substitute for age varying  $\lambda$ . The optimal history dependence parameters  $\theta$  now also depend on the intertemporal price of consumption  $q$ . The exact form of the objective function can be found in the Appendix. Here I only mention that an exercise analogous to the perturbation exercise in section 3.3 yields that a perturbation of  $\theta_0 = 1$  by  $d\theta_0$  and a corresponding perturbation of  $\theta_1 = 0$  by  $d\theta_1 = -(\beta\delta)^{-1}d\theta_0$  will yield a welfare change

$$d\mathcal{W} = \left( \frac{q}{\beta} - 1 + \frac{\alpha - q}{2\beta} \right) \sigma_\phi^2 \times d\theta_0.$$

Hence, it will be welfare improving if, first,  $q > \beta$  and, second, if  $\alpha > q$ . The first requirement is that the government is more patient than the agent, but it is now the intertemporal price of consumption, rather than the government discount factor, that is being compared to the agent's discount factor  $\alpha$ . The second condition requires that the government is patient relatively to its discount factor. We know from (9) that, in that case, the government wants to implement an increasing pattern of consumption over time. It cannot do it directly now. However, a progressive history dependent tax acts as an imperfect substitute for that, because the progressivity parameters accumulate and increase consumption as individual ages.

It is easy to see that if  $\alpha = q$  then the condition collapses to the one in section 3.2. In fact, I show more than that. Whenever  $\alpha = q$  and the distribution is lognormal, the

welfare function (11) is *approximately* correct, and so are the optimal welfare coefficients. More specifically, a first-order approximation to the time invariant parameter  $\lambda$  yields the welfare function identical to (11). As it turns out, for empirically relevant parameter values, namely standard deviations of the idiosyncratic shocks, the approximation is very close to the unrestricted solution. In those cases, an age independent tax function is a very good approximation of the optimal one.

## 4 General Equilibrium

I will now analyze a problem for a full overlapping generation economy, and show how it can be malled to the partial equilibrium framework. The social welfare function aggregates lifetime utilities of generations that are not yet born, and the remaining lifetime utilities of generations that are currently alive. The government weights the welfare of a generation born in period  $t$  by  $\pi^t$  for some Pareto weight parameter  $\pi \in (0, 1)$ . This includes generations not yet born, but also generations that are currently alive, i.e. born before time  $t = 0$ . Their Pareto weights are then whose weight is then greater than one. The social welfare function is then

$$\mathcal{W} \equiv \frac{(\pi - \beta\delta)(1 - \pi)}{\pi(1 - \beta\delta)} \left[ \sum_{t=0}^{\infty} \pi^t \mathbb{E}_0(w_t) + \sum_{t=-\infty}^{-1} \pi^t \mathbb{E}_0(\hat{w}_t) \right], \quad (12)$$

where  $\mathbb{E}_0(w_t)$  is the expected lifetime utility of someone who will be born at time  $t \geq 0$ ,

$$\mathbb{E}_0(w_t) = (1 - \beta\delta) \sum_{j=0}^{\infty} (\beta\delta)^j \mathbb{E}_0(u_j) \quad t \geq 0, \quad (13)$$

and  $\mathbb{E}_0(\hat{w}_t)$  is the remaining lifetime utility of someone who is currently alive and was born at time  $t < 0$ . Importantly, the remaining lifetime utility is discounted back to their birthdate:

$$\mathbb{E}_0(\hat{w}_t) = (1 - \beta\delta) \sum_{j=-t}^{\infty} (\beta\delta)^j \mathbb{E}_0(u_j) \quad t < 0. \quad (14)$$

Discounting everyone's period utilities back to their birthdates and discounting every generation's lifetime utilities at rate  $\pi$  ensures that everyone is treated symmetrically, and implies that the resulting social welfare function is time consistent (Calvo and Ob-

stfeld (1988)).

The expected period utilities  $\mathbb{E}_0(u_j)$  depend on age, but not explicitly on time, or one's cohort. That is because the tax function is not explicitly dependent on time (although it can depend on age or individual history). Then all individuals of age  $j$  have the same utility  $u_j$ . Substituting the expressions (13) and (14) into the welfare function (12) yields that the welfare is a weighted average of period utilities, with future utilities discounted at a rate determined jointly by the discount factor, the survival rate, and the Pareto parameter:

$$\mathcal{W} = \left(1 - \frac{\beta\delta}{\pi}\right) \sum_{j=0}^{\infty} \left(\frac{\beta\delta}{\pi}\right)^j \mathbb{E}_0(u_j). \quad (15)$$

The general equilibrium case can thus be mapped into the partial equilibrium model with  $\alpha = \beta/\pi$ . If the benchmark case when the government discounts at the same rate as the agents,  $\pi = \beta$  and the welfare is simply the average of period utilities weighted by the survival rate  $\delta$ .<sup>8</sup>

Similar calculations can be made regarding the resource costs. The annualized present value of the aggregate costs is the sum of the discounted expected cost of the generations not yet born, and the expected costs of the currently living generations for the remainder of their life:

$$\mathcal{P} \equiv \frac{(1-q)(1-\delta)}{1-q\delta} \left[ \sum_{t=0}^{\infty} q^t \mathbb{E}_0(p_t) + \sum_{t=-1}^{-\infty} \mathbb{E}_0(\hat{p}_t) \right],$$

where  $\mathbb{E}_0(p_t)$  is the expected present value of costs of a generation born in period  $t$  discounted back to the present value at rate  $q$ ,

$$\mathbb{E}_0(p_t) = (1-q\delta) \sum_{j=0}^{\infty} (q\delta)^j \mathbb{E}_0(c_j - y_j) \quad t \geq 0, \quad (16)$$

and  $\mathbb{E}_0(\hat{p}_t)$  for  $t < 0$  is the expected present value of of costs for the currently living generation for the remainder of their lifetimes,

$$\mathbb{E}_0(\hat{p}_t) = (1-q\delta)\delta^{-t} \sum_{j=0}^{\infty} (q\delta)^j \mathbb{E}_0(c_{j-t} - y_{j-t}) \quad t < 0. \quad (17)$$

---

<sup>8</sup>One interpretation is that, conditional on being alive, all period utilities are weighted equally.

If the period cost again do not depend on time, then the aggregate cost are independent of the intertemporal price of consumption, and are simply equal to the period costs, where the cost of older generations is discounted at rate  $\delta$  because of a decreasing generation size:

$$\mathcal{P} = (1 - \delta) \sum_{j=0}^{\infty} \delta^j \mathbb{E}_0(c_j - y_j). \quad (18)$$

The general equilibrium case can again be mapped into the partial equilibrium model with  $q = 1$ . Overall, the relationship between the results for the partial equilibrium economy and for the general equilibrium economy are summarized as follows:

**Remark.** The solution to the general equilibrium model is identical to the solution to the partial equilibrium model with  $\alpha = \beta/\pi$  and  $q = 1$ .

#### 4.1 A Case with $\pi = \beta$

I will now study a benchmark case, where the government weights the generations at the discount rate  $\pi = \beta$ .<sup>9</sup> The welfare function (15) is now

$$\mathcal{W}(\tau, \theta) = \bar{u}(\tau) - \frac{1}{2}(1 - \tau)^2 P_{1,1}(\theta)(\sigma_z^2 + \sigma_\phi^2), \quad (19)$$

where  $\sigma_z^2 = \frac{\delta}{1-\delta}\sigma_\omega^2$  is a cross-sectional variance of productivity shocks. One implication of the fact that welfare can be represented by (19) is that the preference shocks  $\ln \phi$  and the productivity shocks  $z$  enter symmetrically; in fact, only the sum of both variances matters for aggregate welfare. This is intuitive, since both shocks affect the agent's ability to produce output similarly, and each innovation in the agent's productivity  $\omega$  is permanent, just like the preference shock. Since the cross-sectional variance of log consumption is

$$\text{Var}(\ln c) = (1 - \tau)^2 P_{1,1}(\sigma_\phi^2 + \sigma_z^2),$$

the welfare function in the benchmark case is simply a welfare from a representative agent's allocation  $\bar{u}(\tau)$ , minus one half of the cross-sectional variance of log consumption  $\text{Var}(\ln c)$ . The optimal history dependence parameters can be characterized as follows:

---

<sup>9</sup>Note that, since  $\alpha = q = 1$ , an age independent tax function will be approximately correct in this case, as discussed previously.

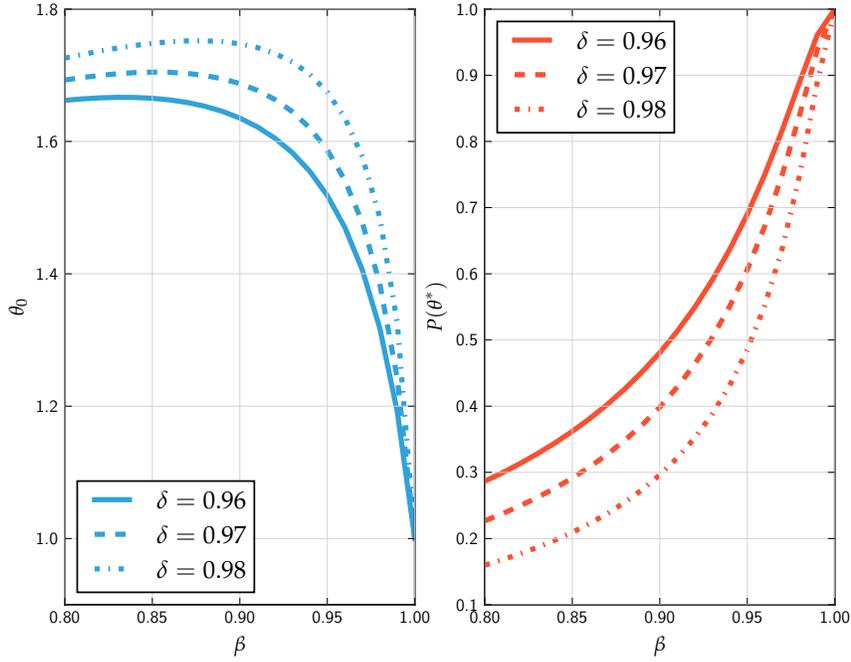


Figure 2: The value of the history dependence parameter  $\theta_0$  (left panel) and the reduction in the cross-sectional variance of consumption  $P_{\beta/\pi,1}(\theta^*)$  (right panel) for various values of  $\beta$  and  $\delta$ .

**Proposition 8.** *The coefficients  $\theta^*$  maximize social welfare if and only they minimize the cross-sectional variance of log consumption subject to the incentive keeping constraint (4).*

The optimal history dependence parameters are given in Proposition 6, with  $\alpha$  set equal to one. The coefficients on past incomes are thus geometrically increasing at rate  $\beta$ . The left panel of Figure 2 plots the value of the coefficient  $\theta_0$  for a range of values of  $\beta$  and for three values of  $\delta$ . The value of  $\theta_0$  is significantly greater than one. In fact, for plausible value of the progressivity wedge  $\tau$ , the value of “nominal” progressivity with respect to the current income  $1 - (1 - \tau)\theta_0$  is negative. Since the marginal tax rate with respect to the past incomes is positive, the tax system “subsidizes” current income, and then taxes it for the rest of the agent’s life.

The value of  $P(\theta^*; 1)$  converges to one as  $\beta$  approaches unity (more precisely, when  $\beta$  converges to  $\alpha = 1$ ), and so for extremely low discount rates the reduction in the cross-sectional variance of consumption will not be very high. For more plausible values the reduction can be significant: for  $\beta$  around 0.96 the variance of log consumption will be reduced by about a third, as the right panel of Figure 2 illustrates.

**Consumption Dynamics.** The optimal income tax changes dramatically the dynamics of individual consumption. While consumption follows a random walk when taxes are history independent, it is a moving average process under the history dependent tax system:

$$\ln c_j = \ln \bar{c}_j - \frac{1 - \tau}{1 + \eta} \beta^j \ln \phi + (1 - \tau) \theta_0 \sum_{k=0}^{j-1} \beta^k \omega_{j-k},$$

where  $\ln \bar{c}_j$  is a deterministic component of consumption. The left panel of Figure 3 shows a typical response to a unit increase in the innovation to the permanent component  $\omega$  for a given progressivity wedge. Under history independent taxation, the effect on consumption is permanent. Under full history dependence, consumption responds more in the short run, but then the effect of the shocks disappears. Clearly, in order to maintain the same progressivity wedge, there needs to be a trade-off, and consumption cannot always respond less under history dependence. But the government can convert the long-run consumption risk (under history independence) into a short-run risk. Figure 3 also shows the consumption impulse response function and the variance of log consumption under partial history dependence for  $K = 5$ . The result combines features of both extremes: Not all long-run risk is eliminated, although it is significantly reduced. The variance of log consumption initially increases faster than under history independence, and then grows without bounds, although at a smaller rate than under history independence.

This has important implications for the nature of consumption dispersion. The right panel of Figure 3 shows the variance of log consumption by age. Under history independent taxation, the variance of individual consumption  $\text{Var}(\ln c_j)$  grows over time linearly without bounds.<sup>10</sup> Under the optimal history dependent scheme, on the other hand, the variance of log consumption is bounded away from infinity,

$$\text{Var}(\ln c_j) = (1 - \tau)^2 \theta_0^2 \frac{1 - \beta^{2j}}{1 - \beta^2} \sigma_\omega^2, \quad (20)$$

and converges over time to a finite limit  $\theta_0^2 / (1 - \beta^2)(1 - \tau)^2 \sigma_\omega^2$ . Since  $\theta_0 > 1$ , the optimal policy increases the immediate impact of the shock and, in the short run, the variance of log consumption is higher than under history independence. But then the shock's impact decays at a rate equal to the discount rate.

---

<sup>10</sup>What keeps the cross-sectional consumption from growing without bounds is that individuals die at rate  $\delta$ .

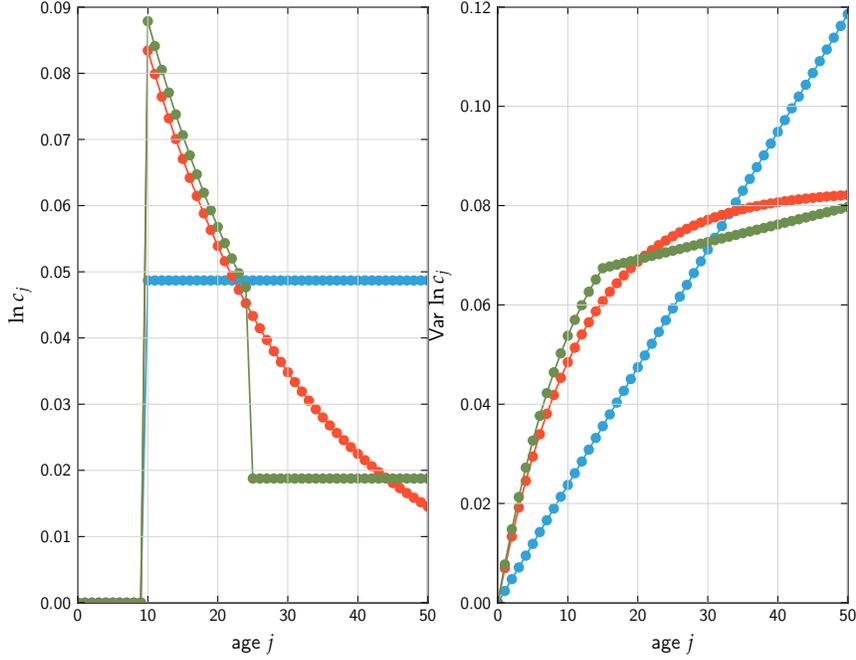


Figure 3: Response of log consumption to a unit increase in  $\omega$  (left panel) and variance of log consumption by age (right panel). Blue lines represent taxes dependent only on current income. Red lines represent taxes with full history dependence. Green lines represent taxes with limited history dependence ( $K = 5$ ).

**Welfare gains.** What are the welfare gains from optimal history dependent tax? Previous discussion shows that history dependence increases welfare in two ways: first, it reduces consumption dispersion for any given progressivity wedge  $\tau$  and, second, it reduces the progressivity wedge itself. The contribution of each of those components will be examined quantitatively later. The welfare gain from the reduction of consumption dispersion is, however, easy to characterize. Let  $\hat{\mathcal{W}}_0(\tau)$  be welfare under a history dependent tax with progressivity wedge  $\tau$ , and  $\hat{\mathcal{W}}(\tau)$  be welfare under a history dependent tax where the coefficients  $\theta$  are chosen optimally, but the progressivity wedge is still  $\tau$ . Using the fact that with log utility the welfare gains are approximately equal to the difference between both values of welfare, we get

$$\hat{\mathcal{W}}(\tau) - \hat{\mathcal{W}}_0(\tau) = \frac{1}{2}(1 - \tau)^2 [1 - P_{1,1}(\theta^*)] (\sigma_\phi^2 + \sigma_z^2).$$

If  $\tau = \tau_0^*$ , the optimal progressivity wedge under the history independent tax system, then  $\hat{\mathcal{W}} - \hat{\mathcal{W}}_0$  represents a lower bound on the welfare gain from history dependence

since there is an additional welfare gain from replacing  $\tau_0^*$  by  $\tau^*$ . If  $\tau$  equal to a value that approximate U.S. tax code, then  $\hat{\mathcal{W}} - \hat{\mathcal{W}}_0$  is the lower bound on the welfare gains from reforming the current U.S. tax code. The latter calculation can in fact be done very easily, since  $(1 - \tau_{US})^2(\sigma_\phi^2 + \sigma_z^2)$  is the measured variance of consumption under the U.S. tax code and is approximately 0.14. Thus, the welfare gains are at least  $0.07 [1 - P(\theta^*)]$ . If the variance of consumption is reduced by a third, then the welfare gains are at least 2.3 percent in consumption equivalents. As we shall see, this is a pretty accurate estimate of the welfare gains.

## 4.2 Parameterization

The model is parameterized to match U.S. experience. The parameterization is similar to [Heathcote et al. \(2016\)](#), although differences are necessary given that the model of this paper does not include the education sector. The U.S. tax code is history independent and its progressivity parameter is estimated by [Heathcote et al. \(2016\)](#) to be  $\tau = 0.161$ . I set  $\eta = 2$ , implying a Frisch elasticity of 0.5,  $\beta = 0.957$  to match an interest rate of 4 percent in the calibrated economy, and  $\pi = \beta$ . The survival rate is set to  $\delta = 0.971$ , to match an expected working life of 35 years.

Under the U.S. tax code, the measured cross-sectional variance of consumption is given by

$$\widehat{\text{Var}}(\ln c) = (1 - \tau)^2 (\sigma_z^2 + \sigma_\phi^2) + \xi_c, \quad (21)$$

where  $\xi_c$  is the measurement error in consumption. To set the cross-sectional variance of the permanent productivity shock  $\sigma_z^2$  and the variance of the taste shock  $\sigma_\phi^2$ , I use equation (21) as a restriction on feasible parameter values. Based on the Panel Study of Income Dynamics and Consumer Expenditure Survey data from 2000 to 2006, [Heathcote et al. \(2016\)](#) report empirical moments  $\widehat{\text{Var}}(\ln c) = 0.18$  and its measurement error  $\xi_c = 0.04$ . Equation (21) then provides a restriction on plausible values of  $\sigma_z^2$  and  $\sigma_\phi^2$ .<sup>11</sup> [Heathcote et al. \(2016\)](#) find  $\sigma_\omega^2 = 0.003$  in their baseline calculations. Other estimates reported in the literature are larger: [Heathcote et al. \(2010\)](#) find  $\sigma_\omega^2 = 0.00625$  by using a longer time period 1967-2005, although in a framework that does not include taste

<sup>11</sup>There are corresponding expressions for the variance of hours worked and earnings. Unlike equation (21) they potentially depend on the variance of insurable productivity shocks (which are not modeled in this paper). For this reason I do not use them to parameterize the model.

shocks.<sup>12</sup> I will set the benchmark values to  $\sigma_\omega^2 = 0.005$ , but I will consider other values satisfying (21) as well. The remaining benchmark values are  $\sigma_z^2 = 0.167$  and  $\sigma_\phi^2 = 0.039$ . All the benchmark parameters are summarized in Table 1.

Table 1: Benchmark Parameters

$\beta$	$\delta$	$\eta$	$\sigma_\omega^2$	$\sigma_\varepsilon^2$	$\sigma_\phi^2$
0.957	0.971	2.000	0.005	0.234	0.039

Figure 4 plots the aggregate welfare as a function of the progressivity wedge  $\tau$  for both the tax system that depends only on the current income (blue line), and for a tax system with optimally chosen full history dependence (red line). The US tax system, being history independent, is represented by a point on the blue line. Since the utility is logarithmic, the welfare gains in consumption equivalents are well approximated by the differences in the welfare. The welfare maximizing progressivity wedge in a tax system based only on current income  $\tau^{\text{HI}}$  is 0.296, a significant increase in progressivity relative to the current U.S. value of 0.161. Introduction of history dependence decreases the progressivity wedge  $\tau^{\text{FH}}$  substantially, down to 0.234, although still more than the current U.S. value.

The welfare gains from replacing the best tax system based only on the current income by the optimal fully history dependent tax system are large: they are equal to 1.77 percent of consumption equivalents. The welfare gain from replacing the current U.S. tax system with the optimal fully history dependent tax system is naturally even larger, 2.48 percent of consumption. In contrast, the welfare gain from replacing the current U.S. tax system with the best tax system based on the current income only is only 0.72 percent of consumption.

To understand the source of the welfare gains, recall that one can decompose the optimal tax reform into two parts: a choice of the history dependence parameters for any given progressivity wedge  $\tau$ , and a choice of the welfare maximizing progressivity wedge  $\tau$ . The first part is represented in Figure 4 by a vertical move from the blue line to the red line, and the second part is represented by a move along the red line. Starting with the best history independent tax system, the figure shows that most of the gains

<sup>12</sup>Older estimates are provided by [Heathcote et al. \(2005\)](#) who estimate  $\sigma_\omega^2 = 0.0095$  and [Storesletten et al. \(2004\)](#) who estimate  $\sigma_\omega^2 = 0.0161$ , both by using data up to 1996.

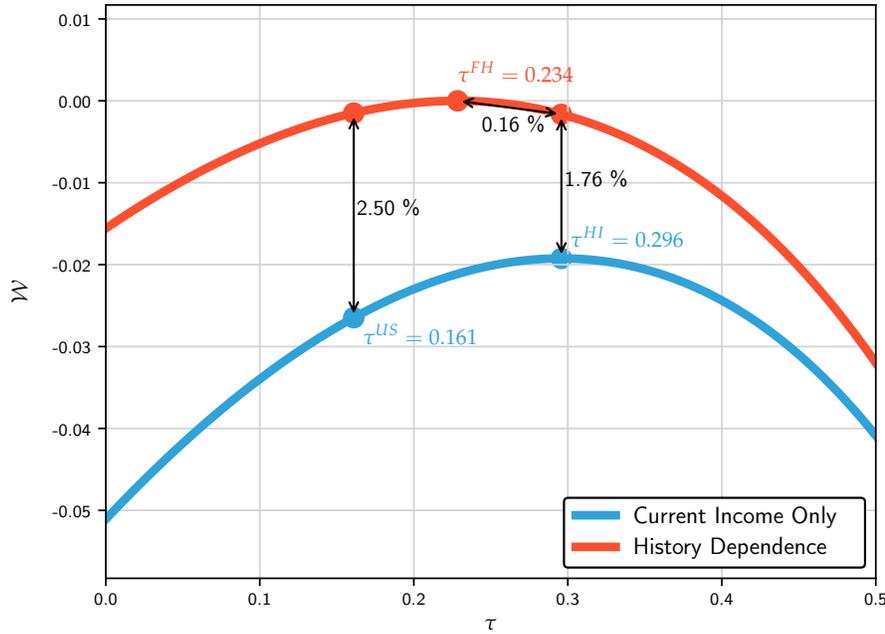


Figure 4: Welfare gains from history dependence

are realized from the introduction of history dependence: 1.63 percent of consumption, which represents is 92.1 percent of all gains. The remaining 7.9 percent of the gains, that is 0.14 percent of consumption equivalents, is then realized by reducing the progressivity wedge from  $\tau^{HI}$  to  $\tau^{FH}$ .

Introducing a limited history dependence reduces the welfare gains from the optimal tax reform. The left panel of Figure 5 plots the welfare gains as a function of the length of history dependence, ranging from 0 to  $K = 30$ . About 16 percent of the overall welfare gains can be obtained when the tax is allowed to depend only on the current and previous income. To reach 50 percent of the overall welfare gains, one needs to include 10 past incomes; including 16 past incomes allows the government to capture 75 percent of the gains. Given the size of the overall welfare gains, conditioning income taxes on even a few past incomes has large benefits, although longer lags are needed to capture the majority of benefits.

The progressivity wedge  $\tau$  gradually decreases as the length of history dependence increases, as the right panel of Figure 5 shows. The average marginal tax rate out of the current income, on the other hand, is not monotone. It equals the progressivity

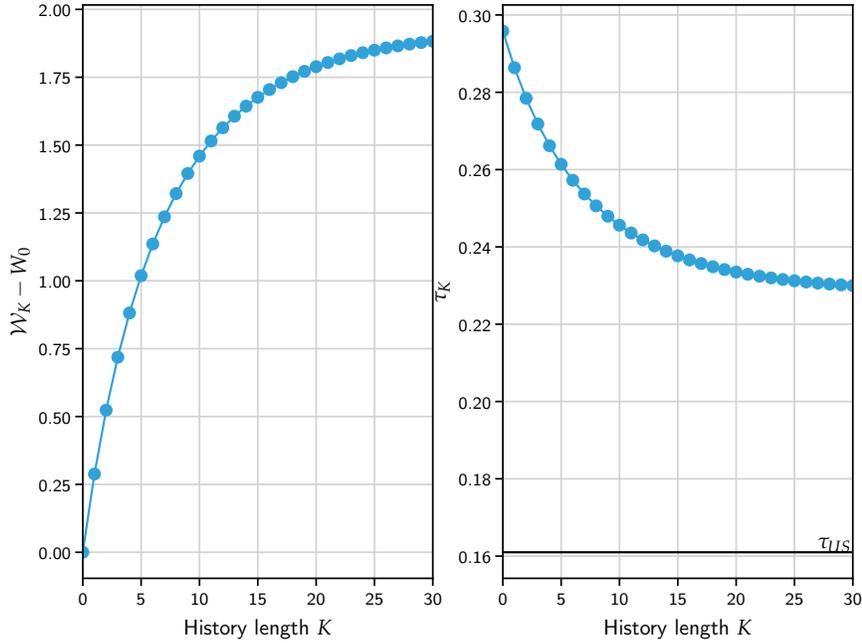


Figure 5: Welfare gains from limited history dependence in percent consumption equivalents (left panel) and the progressivity wedge (right panel).

wedge of 29.5% when  $K = 0$ , then increases to 57.9% for  $K = 3$  as the preference shock considerations dominate, then decreases before increasing again.

## 5 Age Dependent Progressivity

Consider now a tax function that is history independent, but its remaining parameters are allowed to depend on age,  $T_j(y) = y - \lambda_j y^{1-\tau_j}$ . The parameters  $\tau_j$  are again the progressivity wedges, and are now age dependent. The government's problem is to choose the sequences  $(\lambda_j, \tau_j)$  subject to the resource constraint.

The details of the government's problem are relegated to Appendix C, but the nature of the optimum is easy to describe. The parameters  $\lambda_j$  are again chosen to equate the average consumption over time. The progressivity wedge is increasing with age, because the variance of consumption increases with age. As a result, the government sacrifices more progressivity in favor of higher redistribution. The progressivity wedge is thus very low at the beginning, and converges to one as age goes to infinity.

Both history dependent and age dependent taxation share the same objective: to reduce the variance of consumption. Age dependent taxation, however, comes with important negative side-effects: age varying progressivity wedge introduces deterministic variations in labor supply that are not motivated by efficiency considerations. If  $\bar{\tau} = (1 - \delta) \sum_{j=0}^{\infty} \delta^j \tau_j$  is the average progressivity wedge, then the aggregate welfare is

$$\mathcal{W} = \bar{u}(\bar{\tau}) - \frac{1}{2} \text{Var}(\ln c) - \Delta,$$

where  $\bar{u}(\bar{\tau})$  is the representative agent's utility from the average progressivity wedge  $\bar{\tau}$ ,  $\gamma_{\varepsilon}(\bar{\tau})$  is the gain from insurable shocks given the average tax, and  $\Delta$  is the welfare loss from age varying progressivity wedge,

$$\Delta = -\log \left[ (1 - \delta) \sum_{t=0}^{\infty} \delta^t \left( \frac{1 - \tau_t}{1 - \bar{\tau}} \right)^{\frac{1}{1+\eta}} \right].$$

One can then decompose the effects of age dependent taxes into a reduction of variance of log consumption, change in the representative agent's utility plus gains from insurable transitory shocks, and distortions of the deterministic profile  $\Delta$  (see Table 2). The variance of log consumption is reduced by 14.3 percent to 0.084. This is comparable to the reduction in the variance of consumption from history dependent income taxation (20.8 percent), and produces a welfare gain of about 0.707 percent of consumption. The gain in the representative agent's utility  $\bar{u}(\bar{\tau})$  is also significant, about 0.536 percent of consumption, due to the fact that  $\bar{\tau}$  is lower than the optimal tax under the age-independent system. However, age variations in the progressivity wedge produce a cost of about 0.468 percent of consumption, approximately offsetting the gain in the representative agent's utility. Overall, the gain from age dependent taxation is about 0.775 percent of consumption.

Relative to the history dependent tax system, the age dependent system is somewhat less efficient in increasing the representative agent's utility and in reducing variance of consumption. The main difference is, however, that age dependent taxes produce unwelcome variations in the progressivity wedge, while history dependent tax system does not. As a result, the welfare gain from history dependent tax system is 1.765 percent, and age dependency can only recover about 43.9 percent of the gains.

Interestingly, if one shuts down the preference shocks, age dependent tax system can

capture 57.6 percent of the overall welfare gains. Preference shocks introduce additional source of variation in consumption, and limit the gains from age varying progressivity wedge, especially from setting it very low initially. The gains from time varying progressivity wedge is thus limited. On the other hand, if one shuts down productivity shocks then the gains from age dependent taxation are exactly zero, because the cross-sectional variance of consumption does not naturally increase with age. These findings suggests that the relative attractiveness of age dependent taxation may be sensitive to the number of underlying shocks.

Table 2: Contributions to Welfare Gains

	History Dependence	Age Dependence
Representative agent's utility	0.741	0.536
Variance of log consumption	1.025	0.707
Variation of progressivity wedge	–	–0.468
Welfare gains	1.765	0.775

Note: Welfare gains relative to U.S. tax system. Contribution of RA's utility is  $\bar{u}(\tau^*) - \bar{u}(\tau^{US})$ . Contribution of variance of log consumption is  $-\frac{1}{2}[\text{Var}(\ln c) - \text{Var}(\ln c^{US})]$ . Contribution of the variation in the progressivity wedge is  $-\Delta$ .

## 6 Conclusions

This paper studies the nature of history dependent taxation, and the welfare gains from it, in a parametric framework that is easy to analyze. There are two main factors that determine the importance of history dependent income taxation. The first one is mean reversion of productivity shocks. The second one are differences in discounting between government and an individual, in particularly if the government discounts future utilities at a lower rate, which comes naturally in an overlapping generations economy. Both factors are significant, but each of them produces a different nature of history dependence. If productivity shocks are mean reverting, the tax system should be more progressive with respect to the current income and regressive with respect to past incomes, while the opposite is true if the government discounts at a lower rate.

I show that in a general equilibrium overlapping generations economy where the productivity shocks follow a random walk (so that the effect of mean reversion is absent),

the history dependence parameters should be chosen with a simple goal in mind: to minimize the cross-sectional variance of log-consumption while keeping the efficiency wedge constant. I show that this is achieved by discounting past incomes at a rate equal to the discount rate. The welfare gains from history dependent taxation itself are large, about 1.76 percent of consumption. About 50 percent of the gains is captured by having taxes depend on past 10 incomes. Taxes that depend on age, but not on individual history, are able to capture about 44 percent of the potential gains.

The results of this paper were derived under one strong assumption: that the agents are not allowed to borrow and save. This assumption is not restrictive under history independent taxation when consumption follows a random walk. However, under history dependent taxes consumption regresses back to its mean, and individuals have incentives to borrow or save to further smooth their consumption. History dependence thus elicits a demand for individual borrowing and saving. The welfare implications are not clear: higher insurance may be beneficial given that the tax function taxes a particular parametric form that is not optimal, but saving will reduce incentives to work. The assumption of no saving can also be relaxed by assuming that the government chooses a savings tax in a way to prevent individuals from savings. The properties of such a tax function are easy to derive. Such a savings tax may not be optimal, but this means that the welfare gains in this paper are form a lower bound on the overall welfare gains that can be achieved by choosing the savings tax optimally.

The question that future research needs to answer is whether the principles found in this paper are a good and reliable guide in more complex environments. Mirrlees economies, where there are no restrictions on the shape of the tax function, or economies where agents can borrow and save, are especially interesting environments to answer these questions.

## References

- Albanesi, S. and C. Sleet (2006). Dynamic optimal taxation with private information. *The Review of Economic Studies* 73(1), 1–30. [2](#), [4](#)
- Battaglini, M. and S. Coate (2008). Pareto efficient income taxation with stochastic abilities. *Journal of Public Economics* 92(3-4), 844–868. [4](#)
- Benabou, R. (2002). Tax and education policy in a heterogeneous-agent economy: What

- levels of redistribution maximize growth and efficiency? *Econometrica* 70(2), 96–129. 4, 5
- Calvo, Guillermo, A. and M. Obstfeld (1988). Optimal time-consistent fiscal policy with finite lifetimes. *Econometrica* 56(2), 411–432. 9, 20
- Conesa, J. C. and D. Krueger (2006). On the optimal progressivity of the tax code. *Journal of Monetary Economics* 53(7), 1425–1450. 4
- Constantinides, G. M. and D. M. Duffie (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy* 104(2), 219–240. 4
- Farhi, E. and I. Werning (2005). Inequality and social discounting. *Journal of Political Economy* 115(1), 365–402. 4, 8
- Farhi, E. and I. Werning (2012). Insurance and taxation over the life-cycle. *Review of Economic Studies* 80(2), 596–635. 4
- Golosov, M., N. R. Kocherlakota, and A. Tsyvinski (2003). Optimal indirect and capital taxation. *The Review of Economic Studies* 70(3), 569–587. 2, 4
- Golosov, M., A. Tsyvinski, and M. Troshkin (2016). Redistribution and social insurance. *American Economic Review* 106, 359–386. 4
- Heathcote, J., F. Perri, and G. Violante (2010). Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006. *Review of Economic Dynamics* 13, 15–51. 26
- Heathcote, J., K. Storesletten, and G. Violante (2005). Two views of inequality over the life-cycle. *Journal of European Economic Association* 3, 543–552. 27
- Heathcote, J., K. Storesletten, and G. Violante (2014). Consumption and labor supply with partial insurance: An analytical framework. *American Economic Review* 104(7), 2075–2126. 2, 3, 4
- Heathcote, J., K. Storesletten, and G. Violante (2016). Optimal tax progressivity: An analytical framework. Working paper, Federal Reserve Bank of Minneapolis. 4, 26
- Kaplan, G. (2012). Inequality and the life cycle. *Quantitative Economics* 3(3), 471–525. 16

- Kindermann, F. and D. Krueger (2017). High marginal tax rates on the top 1%? Technical report, University of Pennsylvania. 4
- Kocherlakota, N. R. (2005). Zero expected wealth taxes: A Mirrless approach to dynamic optimal taxation. *Econometrica* 73(5), 1587–1621. 4
- Storesletten, K., C. Telmer, and A. Yaron (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics* 51(3), 609–633. 27
- Werning, I. (2007). Optimal fiscal policy with redistribution. *Quarterly Journal of Economics* 122(3), 925–967. 4

## Appendix A: Deriving the Welfare Function

The government chooses the tax parameters  $\{\lambda_j\}$ ,  $\tau$  and  $\{\theta_j\}$  to maximize the social welfare function (7) subject to  $\mathcal{P} = 0$  and (4), taking (5) and (6) as given. The expected values of period utility and period costs are

$$\begin{aligned}\mathbb{E}_0(u_j) &= \ln \lambda_j + \frac{1-\tau}{1+\eta} [\ln(1-\tau) - \mathbb{E} \ln \phi] \sum_{k=0}^j \theta_k + (1-\tau) \sum_{k=0}^j \frac{1-\rho^{j-k}}{1-\rho} \theta_k \mathbb{E} \omega - \frac{1-\tau}{1+\eta} \\ \mathbb{E}_0(c_j) &= \lambda_j B_\phi \left[ -(1-\tau) \sum_{k=0}^j \frac{\theta_k}{1+\eta} \right] \prod_{k=0}^{j-1} B_\omega \left[ (1-\tau) \sum_{l=0}^k \rho^{k-l} \theta_l \right] e^{\frac{1-\tau}{1+\eta} \ln(1-\tau) \sum_{k=0}^j \theta_k} \\ \mathbb{E}_0(z_j h) &= (1-\tau)^{\frac{1}{1+\eta}},\end{aligned}$$

where the term  $(1-\rho^{j-k})/(1-\rho)$  equals  $j-k$  when  $\rho = 1$ . Substitute  $\mathbb{E}_0(u_j)$  into (7),  $\mathbb{E}_0(c_j)$  and  $\mathbb{E}_0(z_j h)$  into (8), and let  $\zeta$  be the Lagrange multiplier on the resource constraint. The first-order conditions in  $\lambda_j$  are

$$\frac{1}{\lambda_j} = \zeta \frac{1-q\delta}{1-\alpha\delta} \left(\frac{q}{\alpha}\right)^j B_\phi \left[ -\sum_{k=0}^j \frac{1-\tau}{1+\eta} \theta_k \right] \prod_{k=0}^{j-1} B_\omega \left[ (1-\tau) \sum_{l=0}^k \rho^{k-l} \theta_l \right] e^{\frac{1-\tau}{1+\eta} \ln(1-\tau) \sum_{k=0}^j \theta_k}$$

One can then rewrite the aggregate consumption as

$$\mathbb{E}_0(c_j) = \frac{1}{\zeta} \frac{1-\alpha\delta}{1-q\delta} \left(\frac{\alpha}{q}\right)^j,$$

which implies (9). The resource constraint then yields  $\zeta = (1-\tau)^{-\frac{1}{1+\eta}}$ . After eliminating  $\lambda$  and  $\zeta$  from the objective function, one can write it as (10).

**Lognormal distribution.** With productivity shocks being lognormally distributed,

$$\begin{aligned}\ln B_{\omega j} &= -\sum_{k=0}^j \left[ (1-\tau) \rho^{j-k} \theta_k - (1-\tau)^2 \rho^{2(j-k)} \theta_k^2 - 2(1-\tau)^2 \sum_{l=0}^{k-1} \rho^{2j-k-l} \theta_l \theta_k \right] \frac{\sigma_\omega^2}{2} \\ \ln B_{\phi j} &= -\sum_{k=0}^j \left[ (1-\tau) \theta_k - (1-\tau)^2 \theta_k^2 - 2(1-\tau)^2 \sum_{l=0}^{k-1} \theta_l \theta_k \right] \frac{\sigma_\phi^2}{2}.\end{aligned}$$

Simplifying and cancelling terms yields (11).

## Appendix B: Welfare Function when $\lambda$ is age independent.

Assume that  $\alpha = q$ . By substituting  $\mathbb{E}_0(u_j)$  into (7) and using  $(1 - \alpha\delta) \sum_{j=0}^{\infty} (\alpha\delta)^j \sum_{k=0}^j \theta_k = \sum_{j=0}^{\infty} (\alpha\delta)^j \theta_j$  and that  $(1 - \alpha\delta) \sum_{j=0}^{\infty} (\alpha\delta)^j \sum_{k=0}^j (j - k)\theta_k = \alpha\delta / (1 - \alpha\delta) \sum_{j=0}^{\infty} (\alpha\delta)^j \theta_j$ , one can write the welfare function as

$$\mathcal{W} = \ln \lambda - \frac{1 - \tau}{1 + \eta} + (1 - \tau) \left[ \frac{\ln(1 - \tau)}{1 + \eta} - \frac{1}{1 + \eta} \mathbb{E} \ln \phi + \frac{\alpha\delta}{1 - \alpha\delta} \mathbb{E} \omega \right] \sum_{k=0}^{\infty} (\alpha\delta)^k \theta_k.$$

The (age independent) tax parameter  $\lambda$  is obtained from the resource constraint  $\mathcal{P} = 0$ . Using  $\mathbb{E}_0(c_j)$  and  $\mathbb{E}_0(z_j h)$  yields

$$\lambda^*(\tau, \theta) = \frac{(1 - \tau)^{\frac{1}{1+\eta}}}{(1 - \delta q) \sum_{j=0}^{\infty} (\delta q)^j B_{\phi} \left[ - \sum_{k=0}^j \frac{1-\tau}{1+\eta} \theta_k \right] \prod_{k=0}^{j-1} B_{\omega} \left[ (1 - \tau) \sum_{l=0}^k \theta_l \right] e^{\frac{1-\tau}{1+\eta} \ln(1-\tau) \sum_{k=0}^j \theta_k}}.$$

**Lognormal distribution.** With productivity shocks being lognormally distributed, the function  $\lambda^*$  is approximated as follows:

**Lemma 9.** *If the shocks are lognormally distributed then*

$$\begin{aligned} \ln \lambda^*(\tau, \theta) &\approx \sum_{k=0}^{\infty} (\delta q)^k \left( (1 - \tau)\theta_k - (1 - \tau)^2 \theta_k^2 - 2(1 - \tau)^2 \sum_{l=0}^{k-1} \theta_l \theta_k \right) \frac{\frac{\delta q}{1 - \delta q} \sigma_{\omega}^2 + \sigma_{\phi}^2}{2} \\ &\quad - \sum_{k=0}^{\infty} (\delta q)^k (1 - \tau) \frac{\ln(1 - \tau)}{1 + \eta} \theta_k + \frac{\ln(1 - \tau)}{1 + \eta}. \end{aligned}$$

*Proof.* Let  $\Gamma = (1 - \tau)^{-\frac{1}{1+\eta}} \lambda^*$ . I will approximate the function  $\Gamma$ . Moreover, since  $q\delta$  always appears together, I will set  $q = 1$  throughout, to reduce notation. Then the function  $\Gamma$  can be compactly written as

$$\Gamma(\tau, \theta)^{-1} = (1 - \delta) \sum_{j=0}^{\infty} \delta^j \bar{B}_j B_{\phi j} \prod_{k=0}^{j-1} B_{\omega k},$$

where

$$\begin{aligned}
B_{\omega j} &\equiv B_{\omega} \left[ (1 - \tau) \sum_{k=0}^j \theta_k \right] = e^{-\sum_{k=0}^j [(1-\tau)\theta_k - (1-\tau)^2\theta_k^2 - 2(1-\tau)^2 \sum_{l=0}^{k-1} \theta_l \theta_k] \frac{\sigma_{\omega}^2}{2}} = e^{-\frac{1}{2} m_j (1 - m_j) \sigma_{\omega}^2} \\
B_{\phi j} &\equiv B_{\phi} \left[ -\frac{1 - \tau}{1 + \eta} \sum_{k=0}^j \theta_k \right] = e^{-\sum_{k=0}^j [(1-\tau)\theta_k - (1-\tau)^2\theta_k^2 - 2(1-\tau)^2 \sum_{l=0}^{k-1} \theta_l \theta_k] \frac{\sigma_{\phi}^2}{2}} = e^{-\sum_{k=0}^j n_k \frac{\sigma_{\phi}^2}{2}} \\
\bar{B}_j &\equiv e^{\sum_{k=0}^j \frac{1-\tau}{1+\eta} \ln(1-\tau)\theta_k} = e^{\sum_{k=0}^j o_k},
\end{aligned}$$

and

$$\begin{aligned}
m_k &= (1 - \tau) \sum_{l=0}^k \theta_l \\
n_k &= (1 - \tau)\theta_k - (1 - \tau)^2\theta_k^2 - 2(1 - \tau)m_{k-1}\theta_k \\
o_k &= \frac{1 - \tau}{1 + \eta} \ln(1 - \tau)\theta_k.
\end{aligned}$$

Note, for future reference, that the relationship between  $m$  and  $n$  can be written as  $m_k - m_k^2 = m_{k-1} - m_{k-1}^2 + n_k$ , and that  $m_0 - m_0^2 = n_0$ .

I first truncate the history at an arbitrary length  $K$ , approximate the resulting function  $\Gamma$ , and then take the limit as  $K$  goes to infinity. Assume that  $\theta_k = 0$  for  $k > K$  for some  $K > 0$ . Truncation implies that  $\bar{B}_j = \bar{B}_K$ ,  $B_{\phi j} = B_{\phi K}$  and  $B_{\omega j} = B_{\omega K}$  for  $j > K$ . Then

$$\begin{aligned}
\Gamma(\tau, \theta)^{-1} &= (1 - \delta) \left[ \sum_{j=0}^K \delta^j \bar{B}_j B_{\phi j} \prod_{k=0}^{j-1} B_{\omega k} + \frac{\delta^{K+1}}{1 - \delta B_{\omega K}} \bar{B}_K B_{\phi K} \prod_{k=0}^K B_{\omega k} \right] \\
&= (1 - \delta) e^{o_0 - \frac{1}{2} n_0 \sigma_{\phi}^2} \left[ \sum_{j=0}^K \delta^j \prod_{k=0}^{j-1} e^{o_{k+1} - \frac{1}{2} n_{k+1} \sigma_{\phi}^2} B_{\omega k} + \frac{\delta^{K+1}}{1 - \delta B_{\omega K}} \prod_{k=0}^K e^{o_{k+1} - \frac{1}{2} n_{k+1} \sigma_{\phi}^2} B_{\omega k} \right].
\end{aligned}$$

This can be expressed recursively by means of the following relationship:

$$\begin{aligned}
\Gamma(\tau, \theta)^{-1} &= e^{o_0 - \frac{1}{2} n_0 \sigma_{\phi}^2} A_0 \\
A_k &= 1 - \delta + \delta e^{o_{k+1} - \frac{1}{2} n_{k+1} \sigma_{\phi}^2 - \frac{1}{2} m_k (1 - m_k) \sigma_{\omega}^2} A_{k+1}, \quad k = 0, \dots, K - 1 \\
A_K &= (1 - \delta) \left( 1 + \delta \frac{B_{\omega K}}{1 - \delta B_{\omega K}} \right) = \frac{1}{\frac{1}{1 - \delta} - \frac{\delta}{1 - \delta} e^{-\frac{1}{2} m_K (1 - m_K) \sigma_{\omega}^2}}.
\end{aligned}$$

The terminal term  $A_K$  can be approximated as

$$A_K \approx \frac{1}{1 + \frac{\delta}{1-\delta} m_K (1 - m_K) \frac{1}{2} \sigma_\omega^2} \approx e^{-\frac{\delta}{1-\delta} (m_K - m_K^2) \frac{1}{2} \sigma_\omega^2},$$

The first approximation uses the fact that  $e^{-a} \approx 1 - a$  for small  $a$ , while the second one uses the same fact in the form  $\ln(1 + a) \approx a$  for small  $a$ . Now suppose that

$$A_k \approx e^{-\frac{\delta}{1-\delta} (m_k - m_k^2 + \sum_{l=1}^{K-k} \delta^l n_{k+l}) \frac{\sigma_\omega^2}{2} - \sum_{l=1}^{K-k} \delta^l n_{k+l} \frac{\sigma_\phi^2}{2} + \sum_{l=1}^{K-k} \delta^l o_{k+l}}. \quad (22)$$

Clearly,  $A_K$  takes this form. Write

$$\begin{aligned} A_{k-1} &= 1 - \delta + \delta e^{o_k - n_k \frac{1}{2} \sigma_\phi^2 - m_{k-1} (1 - m_{k-1}) \frac{1}{2} \sigma_\omega^2} A_k \\ &= 1 - \delta + \delta e^{-(m_{k-1} - m_{k-1}^2) \frac{1}{2} \sigma_\omega^2 - n_k (1 - n_k) \frac{1}{2} \sigma_\phi^2 + o_k} A_k \\ &= 1 - \delta + \delta e^{-[m_{k-1} - m_{k-1}^2 + \frac{\delta}{1-\delta} (m_k - m_k^2 + \sum_{l=1}^{K-k} \delta^l n_{k+l})] \frac{1}{2} \sigma_\omega^2 - \sum_{l=0}^{K-k} \delta^l n_{k+l} \frac{1}{2} \sigma_\phi^2 + \sum_{l=0}^{K-k} \delta^l o_{k+l}} \\ &= 1 - \delta + \delta e^{-[\frac{1}{1-\delta} (m_{k-1} - m_{k-1}^2) + \frac{\delta}{1-\delta} n_k + \frac{\delta}{1-\delta} \sum_{l=1}^{K-k} \delta^l n_{k+l}] \frac{1}{2} \sigma_\omega^2 - \sum_{l=0}^{K-k} \delta^l n_{k+l} \frac{1}{2} \sigma_\phi^2 + \sum_{l=0}^{K-k} \delta^l o_{k+l}} \\ &\approx e^{-\frac{\delta}{1-\delta} (m_{k-1} - m_{k-1}^2 + \sum_{l=1}^{K-k+1} \delta^l n_{k-1+l}) \frac{1}{2} \sigma_\omega^2 - \sum_{l=1}^{K-k+1} \delta^l n_{k-1+l} \frac{1}{2} \sigma_\phi^2 + \sum_{l=1}^{K-k+1} \delta^l o_{k-1+l}}. \end{aligned}$$

Thus,  $A_{k-1}$  takes the form in (22) as well. The third equality uses the fact that  $m_k - m_k^2 = m_{k-1} - m_{k-1}^2 + n_k$ , and the last line uses the same approximations as in the initial step. Continuing by induction and noting that  $m_0 - m_0^2 = n_0$  and that  $\Gamma(\tau, \theta)^{-1} = e^{o_0 - \frac{1}{2} n_0 \sigma_\phi^2} A_0$ , one obtains that

$$\ln \Gamma(\tau, \theta) \approx - \sum_{k=0}^K \delta^k \left( o_k - n_k \frac{\sigma_\phi^2}{2} - \frac{\delta}{1-\delta} n_k \frac{\sigma_\omega^2}{2} \right).$$

Letting  $K$  go to infinity, replacing  $\delta$  with  $q\delta$  and substituting in for  $n_k$  and  $o_k$  yields the desired result.  $\square$

Substituting  $\lambda^*$  into the welfare function  $\mathcal{W}$  yields

$$\mathcal{W} = \frac{1}{1 + \eta} [\ln(1 - \tau) - (1 - \tau)] - (1 - \tau)^2 \sum_{k=0}^{\infty} (\delta q)^k \left( \theta_k^2 + 2 \sum_{l=0}^{k-1} \theta_l \theta_k \right) \left( \frac{\delta q}{1 - \delta q} \frac{\sigma_\omega^2}{2} + \frac{\sigma_\phi^2}{2} \right).$$

## Accuracy of the Approximation

The approximate solution, as characterized by the welfare function (19) and by Proposition 2 is only as good as the approximation that underlies it. It is easy to see that the approximation abstracts from some potentially important factors. Most prominently, it suppresses the importance of consumption smoothing. The welfare function (19) implies that the history dependence coefficients are independent of the variance parameters, and thus hold even in case of  $\sigma_\phi^2 = \sigma_z^2 = 0$ . But that is clearly incorrect in the true solution. In the absence of idiosyncratic shocks it is optimal to have constant consumption over time, which is achieved by a history independent tax system. Thus, we know that  $\theta_0 = 1$  and  $\theta_k = 0$  for  $k > 0$  is optimal, and the approximate solution is far away from the true one. But how good is the approximation for realistic parameter values? Figure 6 shows the approximate history dependence coefficients for history length  $K = 15$ , and compares them to the true history dependence coefficients. In computing the true history dependence parameters I take a benchmark value for the overall variance of shocks to be  $\sigma^2 = \sigma_\phi^2 + \sigma_z^2 = 0.198$  and , and the benchmark optimal progressivity wedge  $\tau = 0.238$  and show the results for various alternative values of the overall variance of shocks  $\hat{\sigma}^2$ .<sup>13</sup>

If the standard deviation of shocks is zero,  $\hat{\sigma} = 0$ , then the true coefficients are zero, and the approximate solution is obviously inaccurate. This is also true when the standard deviation is only 10 percent of the benchmark standard deviation of shocks, and the consumption smoothing factor still dominates. However, if the standard deviation is one half of the benchmark deviation then the true solution is already quite close to the approximate solution. If the standard deviation is 90 percent of the benchmark value or equal to the benchmark value then the approximate solution is almost identical to the true solution. For realistic values of the

---

<sup>13</sup>To simplify exposition I only plot coefficients  $\theta_k$  for  $k = 1, \dots, K - 1$  and do not show  $\theta_0$  and  $\theta_K$  that have different magnitudes.

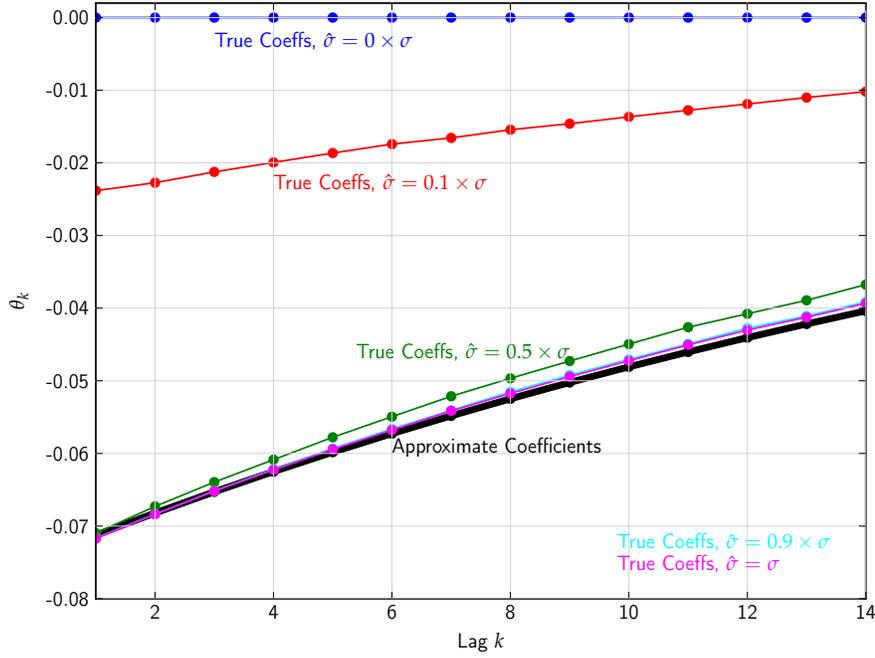


Figure 6: Accuracy of the Approximate Solution for  $K = 15$ ,  $\tau = 0.238$  and  $\sigma^2 = \sigma_\phi^2 + \sigma_z^2 = 0.198$ .

## Appendix C: Age dependent progressivity

Suppose that the tax function parameters depend on age,  $T_j(y) = y - \lambda_j y^{1-\tau_j}$  and that  $\pi = \beta$ . Then the optimal policy functions (with only permanent shocks) are

$$\ln h_j(z) = \frac{1}{1+\eta} (\ln(1-\tau_j) - \ln \phi)$$

$$\ln c_j(z^j) = \ln \lambda_j + \frac{1-\tau_j}{1+\eta} [\ln(1-\tau_j) - \ln \phi] + (1-\tau_j) \ln z_j.$$

$$u_j(z^j) = \ln \lambda_j + \frac{1-\tau_j}{1+\eta} [\ln(1-\tau_j) - \ln \phi - 1] + (1-\tau_j) \ln z_j$$

Then the government maximizes

$$\mathcal{W} = (1-\delta) \sum_{j=0}^{\infty} \delta^j \mathbb{E}_0(u_j)$$

where

$$\mathbb{E}_0(u_j) = \ln \lambda_j + \frac{1 - \tau_j}{1 + \eta} [\ln(1 - \tau_j) - 1] - (1 - \tau_j) \frac{\sigma_\phi^2}{2} - (1 - \tau_j) j \frac{\sigma_\omega^2}{2},$$

subject to the resource constraint

$$\mathcal{P} = (1 - \delta) \sum_{j=0}^{\infty} \delta^j \mathbb{E}_0(r_j) = 0$$

where the periods costs are

$$\mathbb{E}_0(r_j) = \lambda_j b_{1,\phi}(\tau_j) (1 - \tau_j)^{\frac{1-\tau_j}{1+\eta}} B_\omega (1 - \tau_j)^j - (1 - \tau_j)^{\frac{1}{1+\eta}},$$

where  $b_{1,\phi}(\tau) = \mathbb{E} \left( e^{-\frac{1-\tau}{1+\eta}\phi} \right) = e^{-\frac{1}{2}(1-\tau)\tau\sigma_\phi^2}$ , and the right-hand side uses  $\mathbb{E} \left( e^{-\frac{1}{1+\eta}\phi} \right) = 1$ . Let  $\zeta$  be the Lagrange multiplier on the resource constraint. The first-order conditions in  $\lambda_j$  are

$$\frac{1}{\lambda_j} = \zeta b_{1,\phi}(\tau_j) B (1 - \tau_j)^j (1 - \tau_j)^{\frac{1-\tau_j}{1+\eta}}. \quad (23)$$

The first condition simply requires that average consumption is equalized across periods and equals  $\zeta^{-1}$ . Using (23), write the period utility as

$$\mathbb{E}_0(u_j) = -\ln \zeta - \frac{1 - \tau_j}{1 + \eta} - (1 - \tau_j)^2 \frac{\sigma_\phi^2}{2} - j(1 - \tau_j)^2 \frac{\sigma_\omega^2}{2},$$

and the government's problem as

$$\mathcal{W} = \ln \hat{\zeta} - (1 - \delta) \sum_{j=0}^{\infty} \delta^j \left[ \frac{1 - \tau_j}{1 + \eta} + (1 - \tau_j)^2 \frac{\sigma_\phi^2}{2} + j(1 - \tau_j)^2 \frac{\sigma_\omega^2}{2} \right],$$

where

$$\hat{\zeta} = (1 - \delta) \sum_{j=0}^{\infty} \delta^j (1 - \tau_j)^{\frac{1}{1+\eta}}.$$

where  $\hat{\zeta} = 1/\zeta$ .

The first-order condition in  $1 - \tau_j$  is

$$\frac{1}{1 + \eta} + (1 - \tau_j)\sigma_\phi^2 + j(1 - \tau_j)\sigma_\omega^2 = \zeta (1 - \tau_j)^{\frac{1}{1+\eta}} \frac{1}{1 + \eta} \frac{1}{1 - \tau_j},$$

where the last term comes from simplifying the derivative of the exponential term.

Note that if  $\tau$  is constant then welfare reduces to the welfare in the history independent case. This also shows that the optimal choice of  $\lambda$  only has a second-order effect on welfare, since the welfare is the same in both cases.