# Quantifying the Welfare Gains from History Dependent Income Taxation


Marek Kapička*

CERGE-EI, Prague


May 13, 2022


**Abstract**

I quantify the welfare gains from introducing history dependent income tax in an incomplete markets framework where individuals face uninsurable idiosyncratic shocks. I assume that taxes paid are a function of a geometrically weighted average of past incomes, and solve for the optimal weights. I find that the three main factors that determine the nature of history dependence are the degree of mean reversion in the productivity process, the discount factor, and relative weights of the underlying shocks.

The welfare gains from history dependence itself are between 1.72 and 2.98 percent of consumption, depending on whether one starts with the best history independent system or the current U.S. tax system. I decompose the total effect into an efficiency effect that reduces distortion of labor supply, and an insurance effect that reduces volatility of consumption. Quantitatively, the insurance effect strongly dominates the efficiency effect.



J.E.L Codes: E6, H2

Keywords: risk sharing, income taxation, incomplete markets



*I would like to thank seminar participants at IIES Stockholm, University of Munich, University of Manchester, Sabanci University, conference participants at 2016 WAMS-LAEF 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. The research was financially supported by the Czech Science Foundation grant P402/18-13933S. All errors are mine.

# 1 Introduction

The idea that labor income taxes should depend not only on the current income, but also on past incomes, has a long history in economics. It goes back at least to Vickrey (1939, 1947), who proposed averaging of labor income taxes with the objective of eliminating the impact of temporary fluctuations in earnings on the taxes paid.[1] More recently, research on dynamic optimal taxation, starting with Golosov et al. (2003), has shown one general feature of optimal tax systems in economies with unobservable idiosyncratic productivity shocks: they should in general depend on the full history of individual's incomes. Yet, our understanding of how specifically the income taxes should depend on past incomes is limited. How important is last year's income relative to 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 welfare 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 Heathcote et al. (2014). Individuals face uninsurable productivity and preference shocks, and the government uses a nonlinear income tax that exhibits a he relative importance of past incomes embodied in their geometric average weights. The framework is flexible enough to incorporate history independent income taxes, income taxes that depend on a small number of past incomes, and income taxes that depend on the full history of incomes. It also retains tractability even under realistic assumptions about the underlying stochastic process for wages that includes a permanent component, a persistent component, and a transitory component. It is standard in the macro-labor literature to decompose the stochastic process for productivity into those three components, but the literature on dynamic optimal taxation typically studies only one shock at a time.

Why is history dependence useful? History dependence can either help reduce distortions of labor supply (*incentive effect*), or provide a more efficient consumption insurance (*insurance effect*). I provide remarkably simple formulas for the welfare gains and show that one can restrict attention to tax systems where the incentive effect is de-

[1] Although history dependent income taxes has not been used frequently, examples can be found. Most notably, U.S. income tax allowed, between 1964 and 1986, for tax averaging over the current income and incomes in the past 4 years, subject to certain conditions. See e.g. Schmalbeck (1984).

termined only by the overall progressivity wedge of the tax function and not by the geometric weights on past incomes, called history dependence coefficients. The extent of consumption insurance is, on the other hand, driven by both the progressivity wedge and the history dependence coefficients. I show that the optimal history dependence coefficients depend only on a small number of underlying parameters: the relative variances of the underlying shocks, the autocorrelation the persistent component, and the discount factor. Notably, they are independent of the progressivity wedge itself, but also of the level of government spending, the Frisch elasticity of labor, and of the interest rate.

I show that it is optimal to have a history dependent tax system that is more progressive with respect to the current income than a history independent tax system, but regressive with respect to past incomes. This improves consumption smoothing: a temporary increase in the current income translates into a smaller increase in the current after-tax income, but also increases future after-tax incomes. It also does not reduce incentives to supply labor, since individuals know that the return to working is spread over all future periods. I show that the weights on past incomes are geometrically decreasing and converge to zero. The speed of convergence is driven by the autocorrelation of the persistent component and is, for realistic parameter values, very slow. The short-run dynamics is driven mainly by the transitory component. History independence emerges as a special case when the transitory component is not present and the persistent component follows a random walk.

While the history dependence parameters are chosen independently of the progressivity wedge, 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 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. Remarkably, the progressivity wedge is the only tax parameter that affects the redistribution of welfare across permanent types. Thus, the optimal tax system provides more insurance against transitory and persistent shocks, while reducing redistribution across permanent types.

I calibrate the model to the U.S. economy. 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 $\tau = 0.181$. The stochastic process for productivity is parametrized according to Kaplan (2012). I find that the welfare gains from history dependence are large. If one starts with the best history independent tax system, then history dependence itself adds about 1.72 percent in consumption equivalents. This is the insurance effect, and is almost as large as the gains from optimally changing only progressivity. Additional 0.15 percent comes from the incentive effect, i.e. from the reduction of the progressivity wedge. Quantitatively, almost all of the benefits of history dependent income taxation thus come from the insurance effect. If, on the other hand, one starts with the current US tax system and introduces history dependence without changing the overall progressivity, the welfare gains are a whooping 2.98 % in consumption equivalents. Turning to tax systems with limited history dependence, I find that even a short history dependence produces relatively large welfare gains. Taxes that depend only on the current and previous income generate 43 percent of the potential welfare gains, while adding 6 past incomes to the tax function generate about 75 percent of the potential welfare gains.

The optimal history dependent tax system assumes that the level of taxes depends directly on age; this allows the government to separate the problem of transferring resources across ages from the problem of transferring resources across different histories. But how important is age dependence? I show that age independent tax system is approximately optimal when the discount factor equals to intertemporal price of consumption.[2] The reason is that age dependence is useful in implementing a decreasing or increasing path of aggregate consumption. If the discount factor equals to intertemporal price of consumption, it is optimal to have aggregate consumption constant over time. Age independence is optimal only approximately, because history dependence itself introduces intertemporal variations in aggregate consumption. In principle, age dependence is needed in order to "undo" those variations. However, the gains from doing so are of second order.

The theoretical framework of this paper is flexible enough to incorporate various alternative assumptions about idiosyncratic productivity shocks. As an extension, I incorporate heterogeneity in wage profiles, as in Guvenen (2007). Heterogeneity in income profiles by itself prescribes a dramatically different pattern of the optimal history dependence. It is optimal to include only the current and previous income, and the tax function is progressive with respect to the previous income, not regressive. The optimal

[2]It should be noted that I do not allow for age dependence in the degree of overall progressivity, as in Heathcote et al. (2020), which is a source of additional potential welfare gains.

tax system is equivalent to a tax system then depends only on the growth rate of one's incomes, and not on its level. This is intuitive: such a tax system translates differences in the growth rate of incomes to differences in the level of consumption, and is superior to a history independent tax system that translates differences in the growth rate of income to differences in the growth rate of consumption. If heterogeneous wage profiles are considered in isolation, the welfare gains from the optimal tax system are large. When, however, the heterogeneity in wage profiles is weighted agains other components of the wage process, their role is diminished and the resulting welfare gains are comparable to the welfare gains in the benchmark case.

## 1.1 Related Literature

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 (2012b) and many others) that shows that history dependence in income taxation is optimal.[3] 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 of Benabou (2002), and further extended by Heathcote et al. (2014) and Heathcote et al. (2016), who include insurable transitory shocks as well as labor supply decision.[4] 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. 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. I will return to this important point in the concluding remarks.

There is a growing literature that studies how income taxes should depend on age, for

[3]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)).

[4]Other functional forms used in the literature can be found in (Conesa and Krueger (2006), Kindermann and Krueger (2017), and others.

example Kremer (2002), Blomquist and Micheletto (2003), or Weinzierl (2011). Weinzierl (2011) considers optimal age dependence in a calibrated Mirrleesean life-cycle economy and argues that optimal age dependence captures more than 60 percent of the welfare gains from the full reform (that features both age and history dependence). Heathcote et al. (2020) is the closest to the approach in this paper. They consider an environment where the degree of tax progressivity, as well as the level of taxes, are allowed to depend on age. They do not consider history dependence; on the other hand, this paper does not consider age dependent degree of progressivity. I discuss the relationship between age dependence and history dependence in another paper (Kapicka (2019)). Here it suffices to say that the relative benefits of age dependent taxation relative to history dependent taxation depend on a number of underlying assumptions, especially on the sources of heterogeneity. It is not in general true that one system dominates the other.

# 2 Setup

There is a measure one of infinitely lived agents. Their preferences are represented by

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

where $\eta$ is the inverse of the Frisch elasticity of labor, and $\phi$ is a parameter that determines the relative weight of disutility from working. It is a random variable, drawn once at age zero, and the distribution is normalized by assuming that $\mathbb{E}(\phi^{-(1+\eta)^{-1}}) = 1$. The agents receive hourly wages $w_j \in W \equiv (0, \infty)$ at age $j$ that are exogenously determined by the following stochastic process:

$$\ln w_j = \kappa + z_j + \varepsilon_j \tag{2}$$

$$z_j = \rho z_{j-1} + \omega_j, \tag{3}$$

where $\kappa$ is a permanent component drawn once at the beginning of period zero, $z_j$ is a persistent component with autocorrelation $\rho \in [0, 1)$ and innovation $\omega_j$, and $\varepsilon_j$ is an iid component. The initial value of the persistent component $z_{-1}$ is equal to zero.[5] The distributions of all components are normalized so that $\mathbb{E}(e^{\kappa}) = \mathbb{E}(e^{\omega}) = \mathbb{E}(e^{\varepsilon}) = 1$.

[5] The model can be easily generalized to allow for initial heterogeneity in productivity, without changing the main results.

Wages, together with hours worked $h_j$ determines the agent's output $y_j = w_j h_j$.

**Market Structure.** I assume that there is *no* insurance against the wage shocks. That includes self-insurance: the agents are not allowed to save to hedge against the 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) or Sleet and Yazici (2017) make the same assumption for the same reason. I will revisit this assumption later.

## 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

$$T_j(y_0, y_1, \dots, y_j) = y_j - \lambda_j \left(\bar{y}_j\right)^{1-\tau}, \tag{4}$$

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

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

There are three sets of parameters that define the tax policy $T = \{T_j\}$. The *progressivity wedge* $\tau$ determines the overall progressivity of the tax system. The *history dependence* parameters $\theta = \{\theta_i\}$ represent how the current tax payment depends on income realizations in the past, with $\theta_k$ representing the weight on income with lag $k$. A history independent tax is a special case with $\theta_0 = 1$ and $\theta_k = 0$ otherwise. Both $\tau$ and $\theta$ are age invariant; that is one of the key restrictions in the paper. On the other hand, the level tax parameters $\lambda = \{\lambda_j\}$ depend directly on age, in order to allow the government to choose a trend in average consumption independently of the remaining parameters.[6] The parameters $\tau$ and $\theta$ jointly determine the progressivity of the tax system.

[6]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, its expected value, then the tax function can be taken as a time and age invariant function of an infinite history of incomes.

If $\theta_k$ is positive for $k > 0$ then the tax system is regressive with respect to past incomes. To see this, note that

$$\frac{dT_{j+k}}{d \log y_j} = -\theta_k(1-\tau)\left(y_{j+k} - T_{j+k}\right). \tag{5}$$

If $\theta_k$ is positive, future tax liabilities decrease with current income. In other words, the marginal income tax with respect to past incomes is negative. If $\theta_k$ is negative then the opposite is true, and the tax system is progressive with respect to past incomes. It can also be shown that the average income-weighted marginal tax with respect to the income of lag $k$ is $-\theta_k(1-\tau)$. The average income-weighted marginal tax with respect to the current income is $1 - \theta_0(1-\tau)$, and it can be negative even if $\tau < 1$, if $\theta_0$ is high enough.

**Incentive keeping constraint.** Since the tax function raises past incomes to the power of $\theta_k(1-\tau)$, it leaves one degree of freedom in the tax parameters $\tau$ and $\theta$. This allows us to simplify the problem by choosing a convenient normalization. There are many possibilities for normalizing the parameters: for example, one could set $\tau$ equal to zero. However, it turns out to be more convenient to normalize the tax parameters in a way that eliminates any relationship between the history dependent parameters and hours worked. To construct such a normalization, use the assumption that the agents are not allowed to save and solve for the optimal hours worked:

$$\ln h_j^* = \frac{1}{1+\eta}\left[\ln(1-\tau)\sum_{k=0}^{\infty}\beta^k\theta_k - \ln\phi\right]. \tag{6}$$

Labor supply decisions of the agents are independent of the history parameters $\theta$ if the history dependence parameters $\theta$ satisfy the following restriction, called an *incentive keeping constraint*:

$$\sum_{k=0}^{\infty}\beta^k\theta_k = 1. \tag{7}$$

If the incentive keeping constraint holds, hours worked depend only on the progressivity wedge $\tau$, and any variation in the history dependence parameters keeps the incentives to work unchanged, which gives the constraint its name. The reason why any variation of the history dependence parameters that satisfies (7) has no effect on hours worked is the following. Each individual, when choosing hours worked, takes into account the incentive effects of all future taxes paid from current income. If the incentive keeping constraint holds, reducing marginal tax rates in period $i$ by reducing $\theta_i$ by one unit must be exactly offset by an increase in the marginal tax rate in some other period $j$ by increasing $\theta_j$ by $\beta^{i-j}$ units. Since taxes paid are effectively discounted by a discount factor $\beta$ as well, this trade-off does not change work incentives. The normalization allows us to clearly separate the incentive aspect of the problem, represented by the choice of $\tau$, and the insurance aspect of the problem, represented by the history dependence coefficients $\theta$. It is worth stressing that (7) is not a constraint on the tax system and only makes it very transparent that the parametric tax system (4) keeps the incentives to work constant across time and states, which is one of its key, if implicit, assumptions.

## 2.2 Allocations

Substituting the incentive keeping constraint (7) into (6) implies that the optimal hours worked are

$$\ln h_j^* = \frac{1}{1+\eta}\left[\ln(1-\tau) - \ln\phi\right]. \tag{8}$$

Hours worked are independent of the productivity shocks, because with log utility the income and substitution effects cancel out. The preference parameter $\phi$ is the only source of heterogeneity in hours worked. By substituting in the optimal hours worked into the budget constraint, one obtains individual consumption:

$$\ln c_j = \ln\lambda_j + \frac{1-\tau}{1+\eta}\sum_{k=0}^{j}\theta_k\left[\ln(1-\tau) - \ln\phi\right] + (1-\tau)\sum_{k=0}^{j}\theta_k \ln w_{j-k}. \tag{9}$$

Consumption depends on past wages 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 can, in general, move predictably with age, first because $\lambda$ depends on age and, second, because the expected value of the weighted average of past incomes $\bar{y}_j$ changes with age. For example, if $\theta_k$ is nonzero for all $k$ then consumption will move deterministically with age even if $\lambda$ is constant, because the second term in (9) changes with age. Moreover, the deterministic component will be increasing for some and decreasing for others, depending on the sign of $\ln(1-\tau) - \ln\phi$.

## 3 Government's problem

The tax system has aggregate cost equal to the present value of aggregate consumption minus the present value of aggregate earnings:

$$(1-q)\sum_{j=0}^{\infty} q^j\,\mathbb{E}_0(c_j - w_j h_j) + G = 0, \tag{10}$$

where both the interest rate $q \in (0,1)$ and the annual government consumption $G$ are given exogenously. The government chooses the tax parameters $\lambda$, $\tau$ and $\theta$ to maximize the agent's expected utility $\mathcal{W}$ subject to the resource constraint (10) and the incentive keeping constraint (7), taking the policy functions (8) and (9) as given.

I will first characterize the optimal choice of the level parameters $\lambda$, and then proceed to the full characterization of the optimal tax system. Substituting the policy functions (8) and (9) into the utility function $\mathcal{W}$ and the resource constraint (10) and optimizing with respect to the level parameters $\lambda$ yields the following optimal values in terms of the primitives of the model and of the statistical moments of the underlying shock distribiutions:

**Proposition 1.** *The optimal value of the level tax parameters $\{\lambda\}$ is*

$$\lambda_j = \frac{1-\beta}{1-q}\left(\frac{\beta}{q}\right)^j \frac{e^{-\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}}{\prod_{k=0}^{j}\left(B_{\omega k}B_{\varepsilon k}\right)B_{\phi j}B_{\kappa j}}\left[\nu\,(1-\tau)^{\frac{1}{1+\eta}} - G\right], \tag{11}$$

*where $\nu = \exp(-\frac{(1-\rho)q\rho}{(1-q\rho)(1-q\rho^2)}\frac{\sigma_\omega^2}{2})$ is the present value of the persistent component, and the moments $B_{\omega j}$, $B_{\varepsilon j}$, $B_{\kappa j}$ and $B_{\phi j}$ are given by*

$$B_{\omega j} = \mathbb{E}e^{(1-\tau)\sum_{k=0}^{j}\rho^{j-k}\theta_k\omega}, \quad B_{\varepsilon j} = \mathbb{E}e^{(1-\tau)\theta_j\varepsilon}, \quad B_{\kappa j} = \mathbb{E}e^{(1-\tau)\sum_{k=0}^{j}\theta_k\kappa}, \quad B_{\phi j} = \mathbb{E}e^{-\frac{1-\tau}{1+\eta}\sum_{k=0}^{j}\theta_k\ln\phi}.$$

Proposition 1 is proven in Appendix A. While the expression for the optimal tax parameters $\lambda$ is rather complex, its implication for the optimal consumption profile are simple. By substituting the optimal $\lambda$ back to the consumption process (9), we see that they are chosen in a way that the aggregate consumption grows at a rate equal to $\beta/q$:

$$\mathbb{E}_0(c_j) = \mathbb{E}_0(c_0)\left(\frac{\beta}{q}\right)^j. \tag{12}$$

Aggregate consumption is thus constant if $\beta = q$, and the agent's discount factor is exactly offset by the rate of return. It is, however, not true that if $\beta = q$ then $\lambda$ is constant over time. If $\lambda$ is constant, aggregate consumption will be changing over age, first because of the "fanning out" of the persistent productivity component and, second, because history dependence itself introduces deterministic age variations in aggregate consumption. As one can see from equation (11), time variations in $\lambda$ are therefore needed to "undo" those deterministic changes. I will, however, return to this point later, and show that if $\beta = q$, time variations in $\lambda$ are negligible.

If the parameters $\lambda$ are set optimally then the aggregate welfare can be expressed as a function of the remaining tax parameters $\tau$ and $\theta$, and of the statistical moments of the underlying shock distributions $B_{\omega j}$, $B_{\varepsilon j}$, $B_{\kappa j}$ and $B_{\phi j}$ that are themselves a function of both $\tau$ and $\theta$. The expression for the social welfare is:

$$\begin{aligned}\mathcal{W}(\tau,\theta) = \bar{u}(\tau) + (1-\tau)\left(\frac{\mathbb{E}\omega}{1-\beta\rho} + \mathbb{E}\varepsilon + \mathbb{E}\kappa - \frac{\mathbb{E}\ln\phi}{1+\eta}\right) \\ - \sum_{j=0}^{\infty}\beta^j\left[\ln B_{\omega j} + \ln B_{\varepsilon j} + (1-\beta)(\ln B_{\kappa j} + \ln B_{\phi j})\right].\end{aligned} \tag{13}$$

The first term on the right-hand side is the lifetime utility of a representative agent that faces a tax function with a progressivity wedge $\tau$, when the level parameters $\lambda$ are chosen optimally according to Proposition 1:[7]

$$\overline{u}(\tau) = \ln\left[\nu(1-\tau)^{\frac{1}{1+\eta}} - G\right] - \frac{1-\tau}{1+\eta} + \ln\left(\frac{1-\beta}{1-q}\right) + \frac{\beta}{1-\beta}\ln\left(\frac{\beta}{q}\right).$$

The last two terms on the right-hand side of (13) represent two ways in which the idiosyncratic shocks affect welfare: directly by its contribution to individual consumption, and indirectly through the level tax parameters $\lambda$. The direct effect is represented by the second term on the right-hand side of (13). The indirect effect is represented by the moments $B_{\omega j}$, $B_{\varepsilon j}$, $B_{\kappa j}$ and $B_{\phi j}$, which show up in the third term on the right-hand side of (13).

The welfare formula (13) allows us to separate model parameters that don't have any connection to history dependence from those that do. Government consumption $G$ and the utility parameter $\eta$ show only in the representative utility agent's utility $\overline{u}$.

[7] 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.

Since $\overline{u}$ is independent of the history dependence parameters $\theta$, $G$ and $\eta$ affect optimal history dependence at best indirectly, through the choice of the progressivity wedge $\tau$. The intertemporal price $q$ also affects only the representative agent's utility. In this case, however, its irrelevance for history dependence is even stronger: it has no effect at all on either the progressivity wedge or the history dependence parameters, and will only determine the level parameters $\lambda$ so as to ensure that (12) holds.

**Welfare redistribution across permanent types.** How does the tax system redistribute welfare across agents with different permanent characteristics $\kappa$ and $\phi$? It is easy to show that the expected lifetime utility conditional on the permanent type, $\mathcal{W}(\tau, \theta|\kappa, \phi)$ is given by

$$\mathcal{W}(\tau, \theta|\kappa, \phi) = \mathcal{W}(\tau, \theta) + (1 - \tau) \left( \kappa - \mathbb{E}\kappa - \frac{\ln \phi - \mathbb{E} \ln \phi}{1 + \eta} \right).$$

The second term depends on the progressivity wedge $\tau$, but is independent of the history dependence parameters $\theta$. History dependence thus does not affect redistribution of welfare across types. This does not mean that the optimal history dependence parameters will be independent of the distribution of the permanent components. To the contrary, the presence of the terms $B_{\kappa j}$ and $B_{\phi j}$ in (11) shows that this is not the case. The tax level parameters $\{\lambda_j\}$ and the average welfare $\mathcal{W}(\tau, \theta)$ will depend on the interaction of both. But the dispersion around the average welfare will not. This observation is important for easily assessing how different tax systems evaluate redistribution across permanent types relative to insurance against transitory and persistent shocks. If $\tau$ decreases then redistribution of welfare across types will decrease, regardless of the history dependence parameters.

The remaining part of the government's problem is to maximize the objective function (13) by choosing the tax parameters $\tau$ and $\theta$ subject to the incentive keeping constraint (7). I will now simplify the structure by assuming that all the shocks have lognormal distribution.

# 4 Lognormally Distributed Shocks

The government's problem will be substantially simplified if both the productivity shocks and the preference shocks are lognormally distributed, in which case both the expected values and the coefficients $B_{\omega j}$, $B_{\varepsilon j}$, $B_{\kappa j}$ and $B_{\phi j}$ can be written as simple functions of the parameters. To that end, assume that all the shocks are lognormally distributed:

**Assumption 1.** The idiosyncratic shocks $\omega$, $\varepsilon$, $\kappa$ and $\phi$ are distributed according to

$$\omega \sim N\left(-\frac{\sigma_\omega^2}{2}, \sigma_\omega^2\right), \quad \varepsilon \sim N\left(-\frac{\sigma_\varepsilon^2}{2}, \sigma_\varepsilon^2\right), \quad \kappa \sim N\left(-\frac{\sigma_\kappa^2}{2}, \sigma_\kappa^2\right), \quad \frac{\ln \phi}{1+\eta} \sim N\left(\frac{\sigma_\phi^2}{2}, \sigma_\phi^2\right).$$

As will soon become clear, it will be useful to define the total effective variance of the idiosyncratic shocks to be the sum of present discounted values of the variances of all shocks, with the variance of the innovation of the persistent wage component adjusted to account for its persistence:

$$\sigma^2 = \frac{\sigma_\omega^2}{1-\beta\rho^2} + \sigma_\varepsilon^2 + \sigma_\kappa^2 + \sigma_\phi^2.$$

The first term on the right-hand side is derived as follows. The conditional variance of the persistent component $z_j$ is $\sigma_{z_j}^2 = \sum_{k=0}^{j} \rho^{2k} \sigma_\omega^2$, and its present discounted value is $\sum_{j=0}^{\infty} \beta^j \sigma_{z_j}^2$. Rearranging and simplifying gives the first term in the expression for $\sigma$. The shares of each shock's variance in the total effective variance is defined as

$$s_\omega = \frac{1}{1-\beta\rho^2}\frac{\sigma_\omega^2}{\sigma^2}, \quad s_\varepsilon = \frac{\sigma_\varepsilon^2}{\sigma^2}, \quad s_\kappa = \frac{\sigma_\kappa^2}{\sigma^2}, \quad s_\phi = \frac{\sigma_\phi^2}{\sigma^2}.$$

The share of the persistent component is "leveraged" relative to other shocks because of its persistency, especially for high degrees of persistence. In the extreme case of random walk, the share of the persistent component will be its variance multiplied 25 times if $\beta = 0.96$.

The government's objective function (13) can now be compactly written as follows:

**Proposition 2.** *Suppose that Assumption 1 holds. Then the social welfare function is*

$$\mathcal{W}(\tau, \theta) = \overline{u}(\tau) - \frac{1}{2}(1-\tau)^2 p(\theta)\sigma^2, \tag{14}$$

*where*

$$p(\theta) = s_\omega P_\rho(\theta) + s_\varepsilon P_0(\theta) + (s_\kappa + s_\phi)P_1(\theta), \tag{15}$$

*and the quadratic form $P_\rho(\theta)$ is given by*

$$P_\rho(\theta) = \sum_{j=0}^{\infty} \beta^j \left( \theta_j^2 + 2 \sum_{k=0}^{j-1} \rho^{j-k} \theta_k \theta_j \right). \tag{16}$$

With lognormal distribution, welfare is a function of several simple terms: the representative agent utility $\overline{u}$, the square of the progressivity wedge $(1-\tau)^2$, the total effective variance of the shocks $\sigma^2$, and a function $p(\theta)$ that is quadratic in the history dependence parameters. Since the history dependence parameters enter only through the function $p$, their choice is now, unlike the general formulation in (13), independent of the progressivity wedge $\tau$. This further simplifies the problem, because the optimal history dependence parameters will also be independent of the utility parameter $\eta$ and government spending $G$. Fluctuations in government spending will then show up in fluctuations in the tax parameters $\tau$ and $\lambda$, but not in $\theta$. The reverse implication is, however, clearly not true: the optimal choice of the progressivity wedge $\tau$ will depend on the history dependence parameters $\theta$.

The function $p(\theta)$ is a weighted average of three quadratic forms in $\theta$, one representing the persistent shock, one representing the transitory shock, and one representing both permanent shocks. The weight of each of the three quadratic forms is given by the respective variance shares of each shock. One can rewrite the expression for $p(\theta)\sigma^2$ as

$$p(\theta)\sigma^2 = P_\rho(\theta)\frac{\sigma_\omega^2}{1-\beta\rho^2} + P_0(\theta)\sigma_\varepsilon^2 + P_1(\theta)(\sigma_\kappa^2 + \sigma_\phi^2).$$

The values of $P_0$, $P_1$ and $P_\rho$ can now be interpreted as "risk loadings" for their respective shocks, and the function $p$ as the average risk loading, where weighted by the variance shares. The role of the history dependence parameters is that it determines the risk loadings of each shock.

One can further reinterpret the objective function (14) as follows. It can be shown that the second term in (14) is equal to one half times the present value of the variance of log consumption,

$$(1-\tau)^2 P(\theta)\sigma^2 = (1-\beta)\sum_{j=0}^{\infty} \beta^j \text{Var} \ln c_j,$$

where $\text{Var} \ln c_j = \mathbb{E}[(c_j - \mathbb{E}c_j)^2]$. Then the welfare function equals the representative agent's utility minus one half of the present value of the variance of log consumption. Since the history dependence parameters only enter the second term, they will be chosen so as to satisfy a very simple rule:

**Proposition 3.** *The coefficients $\theta$ maximize welfare if and only they minimize the present value of variance of log consumption subject to the incentive keeping constraint (7).*

The social welfare function (14) and the functional form (15) suggest an easy characterization of what is the role of history dependence in income taxation, and how it should be chosen. History dependence coefficients determine how the shocks impact the dispersion of consumption. The relationship between both depends on the persistency of the shock, and is summarized by the "risk loading" factors $P_0$, $P_\rho$ and $P_1$. Each type of shock would, by itself, dictate a different pattern of history dependence parameters, and I will investigate those patterns in the next section. But since there is only one set of history dependence parameters, the optimum will minimize a weighted average of the shock specific risk loading factors, with weights being their variance shares. Note that both permanent shocks $\kappa$ and $\phi$ enter $p$ symmetrically, although one of them is a preference shock, and one of them is a productivity shock. This distinction is unimportant: both affect earnings (either through hours worked or through productivity) and none of them affects period utility.[8] The symmetry extends to the persistent component if it follows the random walk, i.e. if $\rho = 1$. In that case, only the total weight of the three shocks $s_\omega + s_\kappa + s_\phi$ matters.

The welfare function (14) implies that the optimal choice of the progressivity wedge $\tau$ weighs the costs in terms of labor supply distortions against the benefits of reduction in the dispersion of consumption. In the absence of history dependence, the value of $p$ equals one. Optimally chosen history dependence also reduces consumption dispersion by decreasing $p$ below one. The benefits from higher $\tau$ are thus reduced, and the balance shifts in favor of lower labor supply distortion:

**Proposition 4.** *The optimal progressivity wedge $\tau^*$ decreases when history dependence is allowed.*

It follows from the previous discussion that Proposition 4 also implies that allowing for history dependence decreases redistribution of welfare across permanent types $\kappa$ and $\phi$.

[8] In case of preference shocks, this is because the direct effect of $\phi$ on utility is exactly compensated by lower hours worked, as can be readily verified.

Suppose, in addition, that one considers a limited history dependence, where only the most recent $K$ past incomes are included. Increasing the number of past incomes will, in the optimum, only decrease the value of $p$. By the logic of Proposition 4, allowing for longer history dependence will reduce the optimal progressivity wedge $\tau$ more.

### 4.1 Optimal History Dependence

The welfare function (14) shows that the optimal history dependence parameters are a solution to the following simple minimization problem:

$$\theta^* = \arg\min_{\theta} \; p(\theta) \quad \text{s.t.} \quad (7). \tag{17}$$

This minimization problem makes it clear that the optimal history dependence coefficients are independent of the progressivity wedge $\tau$, government consumption $G$, utility parameter $\eta$, and also of the effective variance of shocks $\sigma^2$. Let $\zeta$ be the Lagrange multiplier on the incentive keeping constraint (7). The first-order condition in $\theta_k$ is

$$\sum_{k=0}^{j} \left( s_\phi + s_\kappa + s_\omega \rho^{j-k} \right) \theta_k + \sum_{k=j+1}^{\infty} \beta^{k-j} \left( s_\phi + s_\kappa + s_\omega \rho^{k-j} \right) \theta_k + s_\varepsilon \theta_j = \zeta, \quad j \geq 0. \tag{18}$$

Equations (18) constitute a fourth-order linear difference equation in the history dependence coefficients that, together with the incentive keeping constraint (7), can be solved for the coefficients $\theta$, and for the Lagrange multiplier $\zeta$. The problem has a closed form solution, given in the next proposition. The proposition restricts attention to cases where all three types of shocks have a strictly positive variance share. Limiting cases where one or more of the shocks are not present are qualitatively different, and are studied below separately.

**Proposition 5.** *Suppose that $s_\omega > 0$, $s_\varepsilon > 0$, and $s_\kappa + s_\phi > 0$. The welfare maximizing coefficients $\theta^*$ are*

$$\theta_0^* = \frac{(1-\beta\mu_1)(1-\beta\mu_2)}{1-\beta\rho}, \quad \theta_j^* = \left[ \alpha \mu_1^j + (1-\alpha)\mu_2^j \right] \theta_0^*, \quad j > 1,$$

*where $\mu_1 \in (0, \rho)$ and $\mu_2 \in (\rho, 1)$ are solutions to the characteristic equation*

$$\frac{1-\beta\rho^2}{(\mu-\rho)(1-\beta\mu\rho)} s_\omega + \frac{1}{\mu} s_\varepsilon = \frac{1-\beta}{(1-\mu)(1-\beta\mu)}(s_\phi + s_\kappa), \tag{19}$$

*and $\alpha = \frac{\rho-\mu_1}{\mu_2-\mu_1}$ and $1-\alpha$ are their weights.*

Proposition (5) completely characterizes the optimal history dependence parameters as a function of the discount factor $\beta$, persistence of the wage component $\rho$, and the variance shares $s_\omega$, $s_\varepsilon$ and $s_\phi$. In fact, the history dependence can be completely characterized by the two roots $\mu_1$ and $\mu_2$, and its relative weight $\alpha$, who are themselves functions of the underlying parameters.

Proposition (5) has two main implications for the optimal history dependence. First, $\theta_0$ must be strictly less than one because the coefficients on past incomes are all strictly positive. That is, the optimal history dependent tax system is more progressive with respect to the current income than a history independent tax system, with the effective progressivity with respect to the current income $\hat{\tau} = 1 - (1-\tau)\theta_0 > \tau$.[9] It is, however, regressive with respect to all past incomes, since the marginal tax with respect to the past income is negative by (5). An individual with temporarily higher earnings will experience a high marginal tax rate today, but a negative marginal tax on his today's earnings in the future. Due to the incentive keeping constraint (7), future negative marginal tax rates will exactly offset higher current marginal tax rates so as not to change his incentives to supply labor.

Second, the history dependence parameters $\theta$ decrease with lag and converge to zero. How fast they converge depends on the two roots of the characteristic equation $\mu_1$ and $\mu_2$.[10] The smaller of the two roots, $\mu_1 \in (0, \rho)$, is mainly responsible for the dynamics of coefficients on relatively recent incomes. The larger of the two roots, $\mu_2 \in (\rho, 1)$ determines how fast the history dependence coefficients converge to zero. Since $\mu_2 > \rho$, the rate at which the history dependence coefficients ultimately converge to zero is always greater than the autocorrelation of the persistent component. For empirically reasonable values of $\rho$, the history dependence parameters will converge to zero at a very slow rate.

[9] This reasoning does not take in to effect that $\tau$ itself will decrease. We will see that, quantitatively, this effect will be relatively small.

[10] Since (18) is a fourth-order difference equation, there are four roots of the characteristic equation. The remaining two, however, are greater than $1/\beta$. in order to satisfy the incentive keeping constraint, they must have zero weight.

## 4.2 The contribution of each shock

Why do the history dependence parameters take the functional form given in Proposition 5? The resulting functional form is a "compromise" among three forces. Each shock by itself would prescribe a different form of history dependence, and the resulting optimum balances each of the three patterns based on their variance shares. I will now turn to the special cases where some of the shocks are turned off. There are three main results that emerge from this exercise. First, the permanent shocks are responsible in driving the long-run convergence of the history dependence coefficients to zero; in their absence, the coefficients converge to a strictly positive limit (as long as $\rho$ is not zero). Second, it is the persistent component that causes slow convergence to zero. Third, it is the transitory component that is mainly responsible for the short-run dynamics of the history dependence coefficients.

**No permanent shocks.** Consider first the case when the deterministic component is absent. Then $\mu_2$, the large of the two roots of the characteristic equation (19), equals one. That is, it is no longer true that the history dependence coefficients converge to zero. Although the system exhibits a short-run dynamics due to the fact that $\mu_1$ is nonzero, the history dependence coefficients will converge to $(1-\alpha)\theta_0$. Equivalently,

**Corollary 6.** *Suppose that there are no permanent shocks ($\sigma_\phi^2 + \sigma_\kappa^2 = 0$). Then $\mu_2 = 1$ and $\theta_j$ converges to $(1-\rho)/(1-\mu_1)\theta_0$.*

The limiting value will be strictly positive, unless $\rho = 1$ as well, in which case the persistent component follows a random walk. A case with $\rho = 1$ essentially introduces back permanent shocks and, ss we shall see below, a random walk component is treated very similarly to the permanent component. We conclude that the permanent component is fully responsible for driving the history dependence coefficients to zero.

**No transitory shocks.** In the absence of the transitory component, $s_\varepsilon = 0$, it is now the smaller root $\mu_1$ that simplifies the problem, as it is equal to zero. Thus, the optimal tax system exhibits no short-run dynamics and, by eliminating $\alpha$, one can characterize the dynamics of the history dependence coefficients simply as follows:

**Corollary 7.** *Suppose that there are no transitory shocks ($\sigma_\varepsilon^2 = 0$). Then $\mu_1 = 0$ and $\theta_j^* = \rho\mu_2^{j-1}\theta_0$.*

The value of $\mu_2$ is still strictly between $\rho$ and 1, and the system thus exhibits a slow convergence to zero in environments with high persistence. We can conclude that the transitory component is fully responsible for the short-run dynamics of the history dependence coefficients.

**Only persistent shocks.** Suppose now that both transitory and permanent shocks are turned off , and only the persistent component remains. The solution from Proposition 5 has an even simpler form: in addition to $\mu_2 = 1$, we obtain $\mu_1 = 0$ and $\alpha = \rho$. To summarize,

**Corollary 8.** *Suppose that there are no permanent shocks ($\sigma_\phi^2 + \sigma_\kappa^2 = 0$) and no transitory shocks ($\sigma_\varepsilon^2 = 0$). Then $\theta_0^* = (1-\beta)/(1-\beta\rho)$, $\theta_j^* = (1-\rho)\theta_0^*$ for $j > 0$ and $P_\rho^* = (1-\beta)(1-\beta\rho^2)/(1-\beta\rho)^2$.*

The coefficients on past incomes now do not change with the length of the history and are all equal to a fraction $1-\rho$ of the coefficient on the current income. The autocorrelation of the persistent component is critical in determining the magnitude of $P_\rho^*$ and the gains from history dependence. Lower $\rho$ allows for more consumption insurance and produces larger welfare gains. When $\rho$ approaches one, then $P_{\rho^*}$ approaches one as well, and there are no gains from history dependence. We will return to this observation in the next subsection.

As $\rho$ approaches zero, the loading factor $P_{\rho^*}$ approaches $1-\beta$, a substantial reduction from $P_\rho = 1$ under history independence. When $\rho = 0$, the problem is essentially a problem with only a transitory component. In such case, $\theta_0^* = \theta_j^* = 1-\beta$ and it is optimal to simply take an unweighted geometric average of all the past incomes. In his seminal work, Vickrey (1947) has proposed tax systems based on simple arithmetic income averaging. Such a tax system (with the modification that it uses geometric, rather than arithmetic, averaging) is optimal when only transitory shocks are present.

**No persistent shocks.** We have seen above that autocorrelation of the persistent component is a key factor that determines the speed of convergence to zero. What happens when $s_\omega = 0$ and the persistent component is not present? The larger of the two roots $\mu_2$ receives zero weight, because $\alpha = 1$. The dynamics is then fully determined by the first root $\mu_1$. Unlike the larger root, $\mu_1$ is not bounded below by $\rho$, and so it is possible that that the convergence to zero is relatively fast. It is possible to show that $\mu_1$ is increasing in the share of the transitory component $s_\varepsilon$. One would then expect that the convergence to zero will be especially fast in environments where the transitory component has a small share.

### 4.3 When is History Independence Optimal?

History independent tax systems correspond to a solution where $\theta_j = 0$ for all $j > 0$. For it to be optimal, there must be no gains from insurance of idiosyncratic shocks. The following proposition shows conditions under which this is indeed the case:

**Corollary 9.** *Suppose that there are no transitory shocks ($\sigma_\varepsilon^2 = 0$) and, in addition, $\rho = 1$. Then a history independent tax system is optimal.*

To understand Corollary 9, consider first the case when only the deterministic shocks are present. History dependent income tax would in such case create a deterministic variation of consumption over age (see equation 9), but also a time varying cross-sectional dispersion of consumption. History independence, on the other hand, produces a consumption profile with a constant cross-sectional dispersion of consumption, and so dominates any history dependent policy.[11]

As Corollary 9 shows, the intuition extends to the case when there is, in addition, a random walk component. That is because random walk and permanent shocks are very similar: in both cases, it is not possible to improve insurance by reallocating consumption over time and states. They therefore produce identical policy responses, namely history independent tax systems.

To understand why random walk produces history independence, consider the case when there are no permanent shocks, but allow for an arbitrary autocorrelation of the persistent component. Under a history independent tax system, log consumption is an AR(1) process with drift:

$$\ln c_j = \ln(\beta/q) + \rho \ln c_{j-1} + \tilde{\nu}_j, \tag{20}$$

where $\tilde{\nu}_j = (1-\tau)(\omega_j + \sigma_\omega^2/2) - (1-\tau)^2\sigma_\omega^2/2$. The drift of log consumption is therefore $\ln(\beta/q) - (1-\tau)^2\sigma_\omega^2/2$. Under the optimal tax system given in Corollary 8, the

[11]Formally, as $s_\omega$ and $s_\varepsilon$ converge to zero, $\mu_1$ converges to zero and $\mu_2$ converges to $\rho$, keeping the left-hand side of (19) bounded away from zero. As a result, $\alpha$ converges to one.

stochastic process for consumption is different. It is easy to show that log consumption is now a random walk with drift:

$$\ln c_j = \ln(\beta/q) + \ln c_{j-1} + \nu_j, \tag{21}$$

where $\nu_j = (1-\tau)\theta_0(\omega_j + \sigma_\omega^2/2) - (1-\tau)^2\theta_0^2\sigma_\omega^2/2$. The drift of log consumption is $\ln(\beta/q) - (1-\tau)^2\theta_0^2\sigma_\omega^2/2$. Since $\theta_0 < 1$, the innovations in the new random walk process are smaller than the innovation in the original AR(1) process. The optimal tax system thus makes the innovations in the consumption process permanent, but less volatile.

A random walk consumption process is optimal because it eliminates all the gains from consumption insurance. To see the intuition, assume that the optimal tax reform produces another AR(1) process for after-tax incomes, instead of a random walk. Since the tax reform is optimal, it increases welfare. Now, instead of letting the agents consume their after-tax incomes, consider a ”second-round” of income taxation that applies the tax function (4) once again to the after-tax incomes. Since the after-tax incomes follow an AR(1) process, it must further increase welfare. Due to the functional form given in (4), combining both rounds of taxation into one preserves the functional form. Therefore, there exists a tax system that achieves the combined increase in welfare in one step, contradicting the claim that the first-round of taxation was optimal.

### 4.4 Loading Factors and Comparative Statics

It is possible to compute the optimal values of the risk loading functions $P_0^*$, $P_\rho^*$ and $P_1^*$ in a closed form. The formulas are not very informative, however, and are not presented here. It follows from the functional form of $P_\rho$ that, since all the history dependence coefficients are strictly positive, the function $P_\rho$ is increasing in $\rho$. Thus, we always have

$$P_0^* < P_\rho^* < P_1^*. \tag{22}$$

The transitory shock will always reduce the variance of consumption the most (per unit of variance), and the permanent shock will reduce it the least. It is possible to show even stronger result, that $P_1^* \geq 1$, and it is qual to one only under history independence. The permanent shock will always work in the opposite direction against other shocks and its contribution will in fact be negative. In other words, if permanent shocks magically disappear, the welfare gains will be larger.

In addition, the average loading factor is monotone in the variance shares, in that higher variance share of the permanent shocks reduces the average loading function, while higher variance share of the transitory shock increases the average loading function:

**Proposition 10.** *i) Suppose that $s_\phi + s_\kappa$ and $s_\omega / s_\varepsilon$ increase. Then $p(\theta(^*)$ increases. ii) Suppose that $s_\varepsilon$ and $s_\omega / (s_\phi + s_\kappa)$ increase. Then $p(\theta^*)$ decreases.*

The intuition behind the result is fairly straightforward. Since permanent shocks have the largest risk loading factor and transitory shocks have the smallest one (by 22), shifting variance shares away from the permanent shocks and towards the transitory component must decrease the average risk loading factor. Shifting variance shares in the opposite direction increases the average risk loading factor. There is also an "intermediate" case, when variance shares shift towards the persistent component. In that case, however, we cannot sign the change in $p$. The value of $p$ can either increase or decrease: shifting variance shares away from the permanent shocks decreases it, but shifting variance shares away from the transitory component increases it.

## 4.5 A Recursive Formulation

The fact that the optimal history dependence coefficients are exponentially decaying, together with the functional form (4) suggests that it is possible to rewrite the optimal tax policy in a recursive manner, where an appropriately constructed average of past incomes is a sufficient statistics for the current taxes paid. This is indeed the case, although the two roots of (19) imply that one needs to keep track of two averages of past incomes. Define

$$S_{1,j} = \prod_{k=0}^{j-1} \left(y_{j-1-k}\right)^{\mu_1^k}, \quad S_{2,j} = \prod_{k=0}^{j-1} \left(y_{j-1-k}\right)^{\mu_2^k}$$

to be the two corresponding weighted averages of past incomes. Since $\mu_1$ is small and $\mu_2$ is large, $S_1$ will mostly represent the short-term average of recent incomes, and $S_2$ will represent the long-term average income, and we can think of it as an imperfect measure of the permanent income. The two averages update recursively:

$$S_{1,j+1} = y_j \left(S_{1,j}\right)^{\mu_1}, \quad S_{2,j+1} = y_j \left(S_{2,j}\right)^{\mu_2},$$

and one can write the tax function by using $S_1$ and $S_2$ as

$$T_j = y_j - \lambda_j \left[ y_j \left(S_{1,j}\right)^{\alpha} \left(S_{2,j}\right)^{1-\alpha} \right]^{(1-\tau)\theta_0}.$$

The tax function can then be interpreted as follows. The current income is taxed with a larger effective progressivity rate $\hat{\tau} = 1 - (1-\tau)\theta_0 > \tau$ today but, in addition, the agents gets a marginal subsidy on the weighted average of past incomes $S = S_1^{\alpha} S_2^{1-\alpha}$, which itself depends on the long-term and short-term averages of part incomes weighted by $\alpha$ and $1-\alpha$.

If there are no idiosyncratic shock then it follows from Corollary 7 that the short-run average of past incomes is equal to one, and the tax function depends only on the current income, and the permanent income measure $S_2$. If, on the other hand, there are no permanent shocks, then all past incomes will receive the same weight in the permanent measure $S_2$.

# 5 Quantifying Optimal History Dependence

It follows from Proposition 5 that the optimal tax policies, and the associated welfare gains, depend only on a small number of parameters. The history dependence parameters depend only on the discount factor $\beta$, the autocorrelation of persistent shocks $\rho$, and the variance shares $s_\omega$, $s_\varepsilon$ and $s_\kappa + s_\phi$. The progressivity wedge depends, in addition, on the elasticity parameter $\eta$. A time period is one year and so I set $\beta = 0.96$. The intertemporal price of consumption is $q = \beta$ to eliminate all aggregate trends in consumption. I also set $\eta = 2$ in the baseline calibration implying a Frisch elasticity of labor supply equal to one half. I parameterize the stochastic process for wages according to Kaplan (2012), his Table 4. The persistent component has autocorrelation $\rho = 0.958$, and the variance of productivity innovations $\sigma_\omega^2 = 0.017$. The variance of the transitory and permanent component are set to $\sigma_\varepsilon^2 = 0.081$ and $\sigma_\kappa^2 = 0.065$. Unlike productivity shocks, the variance of the preference shocks depends on other aspects of the model. Kaplan (2012) estimates the variance of the preference shock to be 0.107 in his benchmark estimation. However, allowing for unemployment shocks reduces the variance of the preference shock to about 0.005. Heathcote et al. (2016) also estimate the variance of preference shocks, this time allowing for reporting error and for imperfect substitutability among skill levels to be other factors. Their resulting estimate is $\sigma_\phi^2 = 0.036$, which

I take as a baseline value. All the model parameters are summarized in Table 1. When comparing the optimal policies to the U.S. tax system, which is history independent, we take its progressivity wedge $\tau$ to be $\tau^{US} = 0.181$, as estimated by Heathcote et al. (2014).

The persistent component then has the highest variance share $s_\omega = 0.439$. It is followed by the permanent components with a variance share $s_\kappa + s_\phi = 0.311$. About two thirds of its variance share come from the permanent wage component. The variance share of the transitory component is the smallest but not negligible, with $s_\varepsilon = 0.249$. The overall effective variance is $\sigma^2 = 0.325$.

Table 1: Baseline Parameters

| $\beta$ | $q$ | $\eta$ | $\rho$ | $\sigma^2_\omega$ | $\sigma^2_\varepsilon$ | $\sigma^2_\kappa$ | $\sigma^2_\phi$ |
|---|---|---|---|---|---|---|---|
| 0.960 | 0.960 | 2.000 | 0.958 | 0.017 | 0.081 | 0.065 | 0.036 |

Figure 1 plots the first 30 history dependence coefficients for the baseline calibration (blue line). The roots of the optimal tax system are $\mu_1 = 0.354$ and $\mu_2 = 0.988$. The figure confirms that the smaller root $\mu_1 = 0.354$ vanishes fast: it plays very little role only for the first 10 lagged incomes. On the other hand, the larger root $\mu_2 = 0.988$ makes the convergence to zero extremely slow. While the half-life of the smaller root is only 1.7 years, the half-life of the larger root is 56.6 years. As a result, the convergence to zero is almost invisible for the first 30 past incomes. To put an economic interpretation to the magnitudes in Figure 1, if the progressivity wedge is unchanged from the calibrated U.S. value of 0.181, the average marginal tax rate on previous year income is $-13.4\%$. It gets reduced to 1.72% in period 5. Afterwards, it converges very slowly to zero.

The remaining three lines in Figure 1 shows the history dependence coefficients for cases when one of the sources of heterogeneity is shut down. In the absence of permanent shocks (red line), the short-run dynamics of the history dependence coefficients is quite similar to the baseline scenario. The difference is in the long-run dynamics, because the history dependent coefficients now converge to a strictly positive value of 0.0214, which translate to marginal subsidy of 1.75%. In the absence of the persistent component (green line), the history dependence coefficients converge very quickly, in about 10 periods, to zero. This is balanced by significantly larger coefficients in the short run. The coefficient $\theta_1$ equal 16.4, a marginal income subsidy of 13.4%. Finally, in the absence of the transitory component, the short-run dynamics disappears, and the coef-

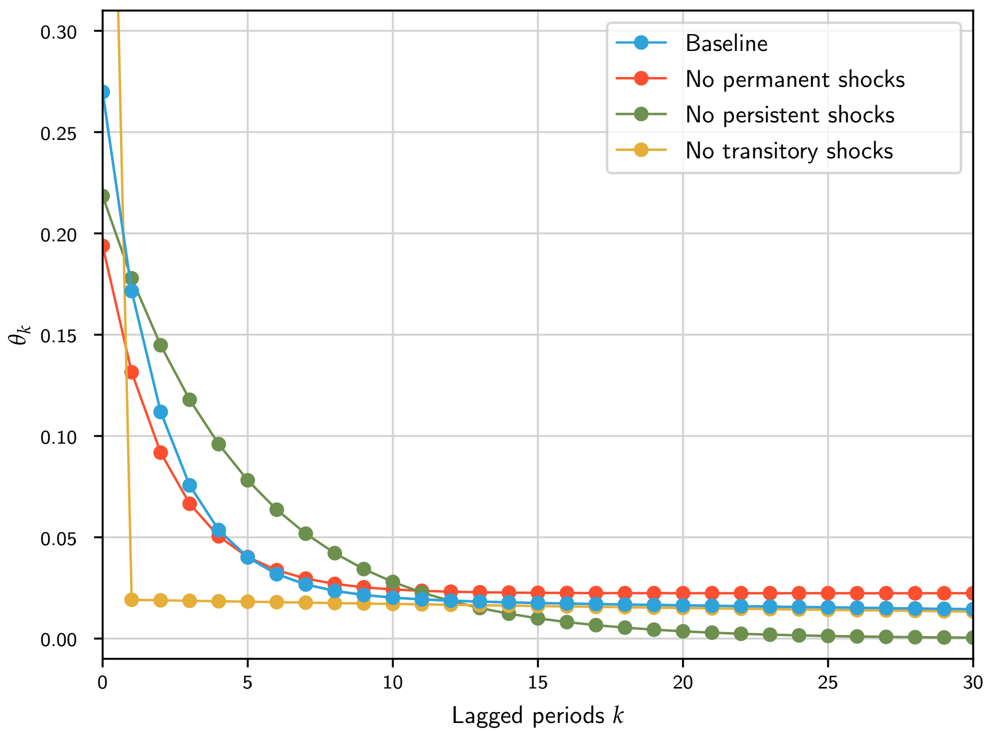


Figure 1: History dependence coefficients for the baseline calibration and when one of the shocks is absent. The current income coefficient $\theta_0$ for the case of no transitory shocks is not shown; its value is 0.64.

ficients converge to zero at a rate that is almost indistinguishable (though not exactly equal) to the baseline scenario.

# 6 Welfare Gains

The welfare formula (14) is remarkably simple. We will now decompose the welfare gains into two components, one coming from the reallocation of hours worked and one coming from the reallocation of consumption, and show an even simpler formula for the latter. Consider an arbitrary tax reform that replaces a history independent tax system with a progressivity wedge $\tau_0$ with a history dependent tax system with progressivity wedge $\tau$ and history dependence parameters $\theta$. The welfare gain from such a tax reform is

$$\Delta = \mathcal{W}(\tau, \theta) - \mathcal{W}(\tau_0, \theta^{HI}), \tag{23}$$

where $\theta^{HI} = \{1, 0, 0, \ldots\}$ reflects history independence. The tax reform increases welfare in two ways: first, it reduces consumption dispersion for any given progressivity wedge $\tau$ by decreasing $p(\theta)$. This represents the *insurance effect*. Second, it changes the progressivity wedge itself. This represents the *incentive effect*. The welfare gain can by accordingly decomposed into two corresponding components:

$$\Delta = \Delta^{\text{ins}} + \Delta^{\text{inc}},$$

where $\Delta^{\text{ins}} = \mathcal{W}(\tau_0, \theta) - \mathcal{W}(\tau_0, \theta^{HI})$ is the insurance effect, and $\Delta^{\text{inc}} = \mathcal{W}(\tau, \theta) - \mathcal{W}(\tau_0, \theta)$ is the incentive effect. If both $\tau$ and $\theta$ are chosen optimally, both components must be positive. Moreover, the welfare gain from increasing insurance is very easy to characterize. Using the fact that with log utility the welfare gains are approximately equal to the difference between levels of welfare, we get that the insurance effect is

$$\Delta^{\text{ins}} \approx \frac{1}{2}(1 - \tau_0)^2 \left[1 - p(\theta)\right] \sigma^2. \tag{24}$$

The welfare gain $\Delta^{\text{ins}}$ provides an easy-to-compute lower bound on the total welfare gain. If the initial progressivity wedge $\tau_0$ is equal to the value that best approximates U.S. tax code $\tau^{US}$, then $\Delta^{\text{ins}}$ is the lower bound on the welfare gains from reforming the current U.S. tax code. If $\tau_0$ is equal to the best progressivity wedge under history independence, then $\Delta^{\text{ins}}$ is the lower bound on the additional welfare gains from introducing history dependence. Note that the welfare gain from history dependence only is negatively correlated with the progressivity wedge. Higher progressivity wedge already provides more consumption insurance by itself, and so history dependence is less valuable.

Figure 2 plots the welfare as a function of the progressivity wedge $\tau$. The welfare gain from introducing history dependence without changing the progressivity wedge $\Delta^{\text{ins}}$ is represented by a vertical movement from the blue line to the red line. The welfare gain from changing only the progressivity wedge $\Delta^{\text{inc}}$ is represented by a movement along the lines. Starting with the current U.S. tax code, introducing history dependence increases welfare by $\Delta^{\text{ins}} = 2.98\%$ in consumption equivalents. Increasing then the progressivity wedge from 0.181 to its optimal value of $\tau^{HD} = 0.324$ further increases the welfare gains

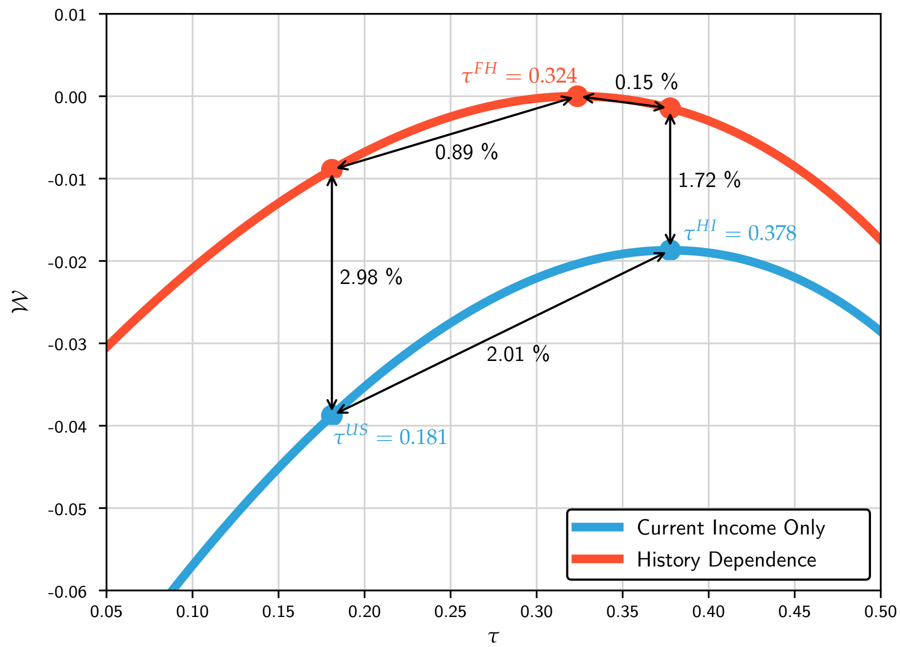


Figure 2: Welfare gains, baseline calibration. Vertical axis shows welfare $\mathcal{W}$ relative to the optimum $\mathcal{W}(\tau^*, \theta^*)$.

by $\Delta^{\text{ins}} = 0.89\%$, bringing the total welfare gain $\Delta = 3.88\%$. About 2.98/3.88 = 77% of the welfare gains are brought about by the insurance channel. In contrast, moving to the best history independent tax system with $\tau^{HI} = 0.378$ represents a static welfare gain of 2.01%.[12] Welfare gains from introducing history dependence are thus substantially larger than the welfare gains from only changing progressivity.

Perhaps a more important exercise is to consider the welfare gains from history dependence after all the gains from optimal progressivity itself are already exhausted. Introducing history dependence only after moving to the best history independent progressivity wedge $\tau^{HI} = 0.378$, reduces the welfare gains from history dependence $\Delta^{\text{ins}}$ to 1.72%. It is to be expected that the insurance effect is now smaller, since formula (24) shows that $\Delta^{\text{ins}}$ is lower if the progressivity wedge is higher. By Proposition 4, the progressivity wedge now needs to be reduced in order to maximize welfare (the incentive effect). The reduction is small, however, from 0.378 to 0.324, and further increases welfare by $\Delta^{\text{inc}} = 0.15\%$, only 8 % of the welfare gains from history dependence it-

[12]The static welfare gain is $\Delta^{\text{s}} = \mathcal{W}(\tau, \theta^{HI}) - \mathcal{W}(\tau_0, \theta^{HI})$.

self. Overall, the welfare loss from restricting oneself to history independent tax systems is 1.72% +0.15% = 1.87%, which represents 1.87/3.88 = 48 % of the maximum welfare gains. The gains from history dependence are thus about as large as the gains from only changing progressivity.

**Decomposing the welfare gains.** Table 2 looks at the contribution of each of the three types of shocks to the overall welfare gains of moving from the current U.S. tax system. The decomposition compares the baseline welfare gains with welfare gains under scenarios when one of the shocks, or two of the shocks, are shut down.

Shutting down the permanent shocks has three effects on the overall welfare gains $\Delta$. First, as shown in Proposition 10, it decreases the value of the average loading factor $P$, which increases the welfare gains. Second, it decreases the value of the effective variance $\sigma^2$. Overall, the first effect dominates, and the insurance gain $\Delta^{\text{ins}}$ increases from 2.98% to 3.67%. But there is also a third effect: lower $\sigma^2$ reduces the optimal progressivity wedge to 0.213, significantly closer to the U.S. value, and the value of $\Delta^{\text{inc}}$ is almost zero. The last effect dominates, and the overall welfare gain decreases to 3.70%. Overall, the absence of the permanent shocks increases the role of the insurance channel, but reduces the role of the incentive channel.

The absence of the persistent component reduces the welfare gain from the insurance channel to 2.12%. It is now driven only by the transitory component, which has the smallest variance share. The persistent component, whose contribution to the welfare gain is in general ambiguous, thus for the baseline calibration increases the welfare gains. The gains from the incentive channel almost zero, just like in the previous case. Overall, the welfare gain is reduced significantly to only 2.17%.

Consider now a complementary decomposition when only one component is active. If the persistent component is the only one present, the welfare gains from the insurance channel are now, 1.26% significantly lower than 2.98% in the baseline scenario. That is, the permanent component, which tends to lower $\Delta^{\text{ins}}$, is dominated by the transitory component, which tends to increase $\Delta^{\text{ins}}$. The transitory component has the smallest variance share, but it generate by far the largest welfare gains "per unit of variance share". Since the welfare gains from the incentive channel are now virtually zero, the overall welfare gain is reduced to 1.27%. If the permanent component is the only one active, the overall welfare gain is almost zero: it is exactly zero from the insurance channel, and close to zero from the incentive channel.

Table 2: Welfare gains, decomposition

| specification | $\Delta$ | $\Delta^{\text{ins}}$ | $\Delta^{\text{inc}}$ | $\tau^{HD}$ |
|---|---|---|---|---|
| baseline | 3.88 | 2.98 | 0.89 | 0.324 |
| no permanent shocks ($\sigma_\kappa^2 + \sigma_\phi^2 = 0$) | 3.70 | 3.67 | 0.03 | 0.213 |
| no persistent shocks ($\sigma_\omega^2 = 0$) | 2.17 | 2.12 | 0.04 | 0.218 |
| no transitory shocks ($\sigma_\varepsilon^2 = 0$) | 1.51 | 0.78 | 0.74 | 0.313 |
| only persistent shocks ($\sigma_\kappa^2 + \sigma_\phi^2 = \sigma_\varepsilon^2 = 0$) | 1.27 | 1.26 | 0.01 | 0.201 |
| only permanent shocks ($\sigma_\omega^2 = \sigma_\varepsilon^2 = 0$) | 0.01 | 0 | 0.01 | 0.196 |
| only transitory shocks ($\sigma_\kappa^2 + \sigma_\phi^2 = \sigma_\omega^2 = 0$) | 3.18 | 2.61 | 0.57 | 0.010 |

Note: Welfare gains in percent consumption equivalents, relative to U.S. tax system.

**The role of shock persistence.** As shown in Proposition 5 and Corollary 8, the autocorrelation of the persistent component is one of the key factors that determine the gains from history dependence. We now inxvestigate the welfare gains as a function of the autocorrelation $\rho$. For each value of $\rho$ (without changing $\sigma_\omega^2$), the optimal tax systems with and without history dependence are computed, and the difference is plotted in Figure 3. As one can see, the relationship between the persistence and welfare gains is not monotone. On one hand, keeping the loading parameters $P_0$, $P_\rho$ and $P_1$ fixed, the gains from a tax reform are increasing in the persistence of shocks, because higher dispersion of the shock increases the gains from reducing volatility of consumption. On the other hand, the ability to reduce volatility of consumption decreases as the persistent component becomes more permanent: in the extreme case of $\rho = 1$, only the transitory component can be a source of welfare gains from history dependence. Those two opposing forces produce a nonmonotone relationship between the persistence of shocks and welfare gains, and the maximum of welfare gains is reached for $\rho_{\max} = 0.832$, well below the U.S. value $\rho_{\text{US}} = 0.958$.

**Sensitivity to parameter values** I now investigate the welfare gains under alternative assumption about the dispersion of preference shocks, discount factor, and the Frisch elasticity of labor. The results are shown in Table 3. The second and third line recalibrates the preference shocks. Equation (8) implies that the variance of log hours worked in the model equals variance of preference shocks. Heathcote et al. (2016) estimate the variance

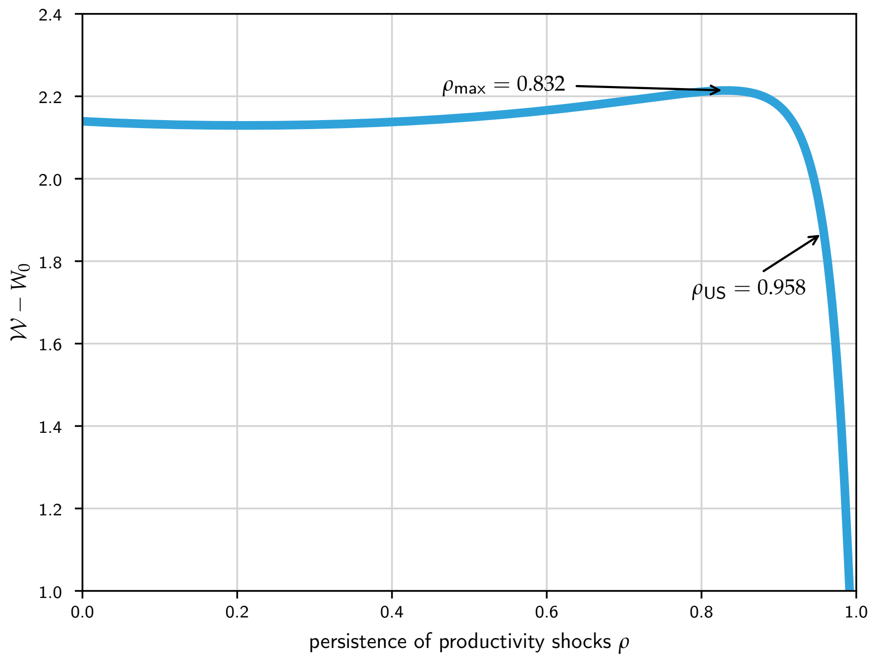


Figure 3: Welfare gains from a tax reform with history dependence, as function of persistence $\rho$. Welfare gains are relative to the best history independent tax system.

of log hours worked to be 0.11. Ignoring all other sources of variation in hours work but correcting for the reporting error with variance of 0.036 yields $\sigma_\phi^2 = 0.074$. On the other hand, Heathcote et al. (2016) in their alternative calibration estimate a lower value $\sigma_\phi^2 = 0.023$. Table 3 shows that increasing the variance of permanent shocks has two effects on the welfare gains. First, since permanent component is not insurable, higher variance reduces the gains from the insurance channel. On the other hand, welfare gains from increasing the progressivity wedge increase. The second channel is more significant, and higher variance of the permanent component increases welfare.

Discount rate is a key parameter in determining the size of the welfare gains, as evidenced from the next two reforms in Table 3. Increasing the patience of agents (as well as the government) from 0.96 to 0.99 increases the welfare gains to a whooping 6.14%, while reducing it to 0.93% reduces the welfare gains to 3.09%. Most of the welfare changes come from the insurance channel. As noted by Farhi and Werning (2012a), one interpretation of a higher discount factor is that the length of the period to which the model is calibrated decreases, but without a corresponding reduction in the variance

Table 3: Welfare gains, alternative specifications

| specification | $\Delta$ | $\Delta^{\text{ins}}$ | $\Delta^{\text{inc}}$ | $\tau^{HD}$ |
|---|---|---|---|---|
| baseline | 3.88 | 2.98 | 0.89 | 0.324 |
| higher preference shocks ($\sigma_\phi^2 = 0.074$) | 4.22 | 2.83 | 1.39 | 0.351 |
| lower preference shocks ($\sigma_\phi^2 = 0.023$) | 3.78 | 3.04 | 0.74 | 0.313 |
| lower discounting ($\beta = 0.99$) | 6.14 | 5.59 | 0.55 | 0.298 |
| higher discounting ($\beta = 0.93$) | 3.09 | 2.24 | 0.84 | 0.320 |
| lower labor elasticity ($\eta = 3$) | 4.38 | 2.98 | 1.40 | 0.372 |
| higher labor elasticity ($\eta = 1$) | 3.32 | 2.98 | 0.33 | 0.259 |

Note: Welfare gains in percent consumption equivalents, relative to U.S. tax system.

of shock innovations. That is, the variance of shocks for a given fixed period of time increases, which in turn increases the gains from history dependence and the insurance channel.

Finally, changes in the Frisch elasticity of labor have no effect on the insurance component, since neither the history dependence parameters nor $\Delta^{\text{ins}}$ depend on it. On the other hand, changes in $\eta$ have a significant effect on the optimal progressivity wedge and on the welfare gains from the incentive channel. Lowering Frisch labor elasticity from 1/2 to 1/3 allows the government to increase the progressivity wedge more, which in turn increases the welfare gains to 4.38%. Increasing Frisch labor elasticity from 1/2 to 1 then reduces the welfare gains to 3.32%.

# 7 Limited History Dependence

Although the welfare gains from history dependence are large, they rely on the assumption that there are no restrictions on the length of income histories. From a practical perspective at least, it would be convenient if very distant histories were relatively unimportant, and most of the welfare gains could be captured by allowing for a limited history dependence. To answer this question, one needs to modify the optimization problem (17) so as to accommodate a restriction that the income history must be at most of length $K$, that is by restricting $\theta_k = 0$ for $k > K$. The analytical solution to the general problem is no longer available. However, one gets a closed form solution for a special case when the persistent component is the only source of heterogeneity, a limited history counterpart to Corollary 8.

**Proposition 11.** *Suppose that $\theta_k = 0$ for $k > K$, for some $K > 0$. Suppose, in addition, that there are no permanent shocks ($\sigma_\phi^2 + \sigma_\kappa^2 = 0$) and no transitory shocks ($\sigma_\varepsilon^2 = 0$). Then the welfare maximizing coefficients are*

$$\theta_0^* = \frac{1-\beta}{1-\beta\rho}\left[1 - \beta^{K+1}\left(\frac{1-\rho}{1-\beta\rho}\right)^2\right]^{-1},$$
$$\theta_k^* = (1-\rho)\theta_0, \quad k = 1, \ldots, K-1$$
$$\theta_K^* = \frac{1-\rho}{1-\beta\rho}\theta_0,$$

*and the value of $P_\rho(\theta^*)$ is*

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

The main difference from the infinite history solution is that the coefficient on the last permissible income $\theta_K$ now increases relatively to other coefficients. It acts as an imperfect proxy for the missing coefficients on more distant incomes. As Proposition 11 shows, the coefficient on the last permissible income is multiplied by $1/(1-\beta\rho)$ times. This factor increases in $\rho$, and so the last coefficient will be relatively different in situations with high persistence. In the other extreme, if $\rho = 0$ and the shocks are transitory, the last coefficient will not differ from the coefficients on less distant incomes. Interestingly, the history dependence coefficients between both endpoints are still $1-\rho$ times the initial coefficient, as in Corollary 8. This will not be a general feature of the solution, however. Finally, note that the coefficients in the full history dependence problem are again a limit when $K$ goes to infinity.

The risk loading factor $P_\rho(\theta^*)$ is clearly decreasing in the length of the permissible history dependence $K$. The rate at which it decreases is driven by the discount factor $\beta$.

Figure 4 shows the optimal history dependence coefficients for the general case when all three components are active, for $K = 10$. The history dependence coefficients in the baseline scenario are now U-shaped. They are increasing for more distant histories for

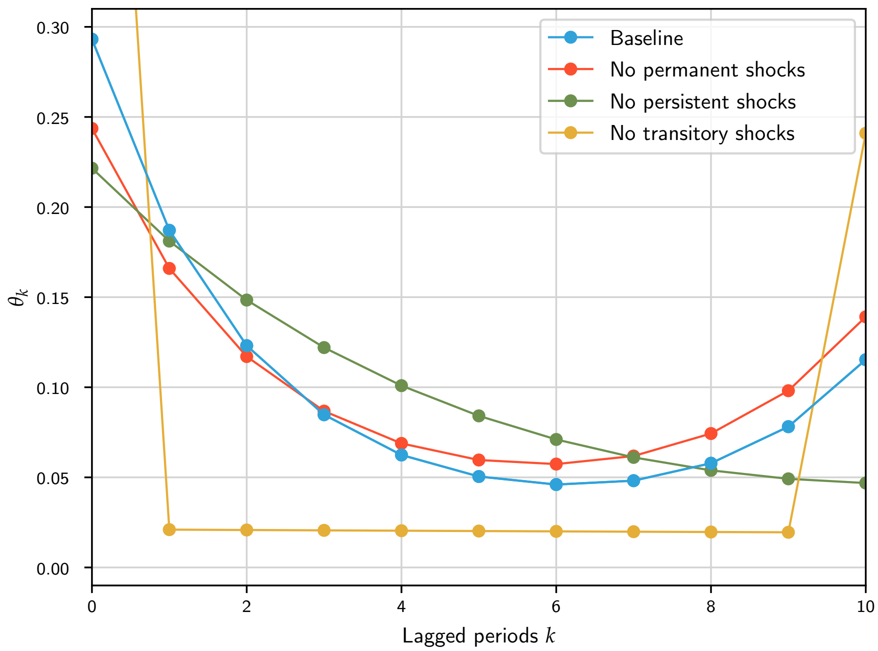


Figure 4: History dependence coefficients for the baseline calibration and when one of the shocks is absent for $K = 10$. The current income coefficient $\theta_0$ for the case of no transitory shocks is not shown; its value is 0.69.

the same reason that the last permissible coefficient in Proposition 11 was higher: they increase in order to replace more distant coefficients that are no longer allowed to be used. The increase is due to the persistent component: in its absence, the coefficients are decreasing with the length of history. The main difference is that the increase is now more gradual than in Proposition 11. This is due to the combination of permanent and transitory shocks. In the absence of the transitory shocks, the history dependence coefficients are constant and only increase at the very end, as the yellow line shows. On the other hand, in the absence

Figure 5 illustrates the welfare gains from limited history dependence, and the corresponding progressivity wedge. The welfare gains from even a short income history are significant: including only current and previous period income captures 43 percent of the potential welfare gains from history dependence. Since the overall welfare gain from history dependence is 1.87 percent relative to the best history independent tax system, adding last period's income to the tax function represents a welfare gain of 0.8 percent.

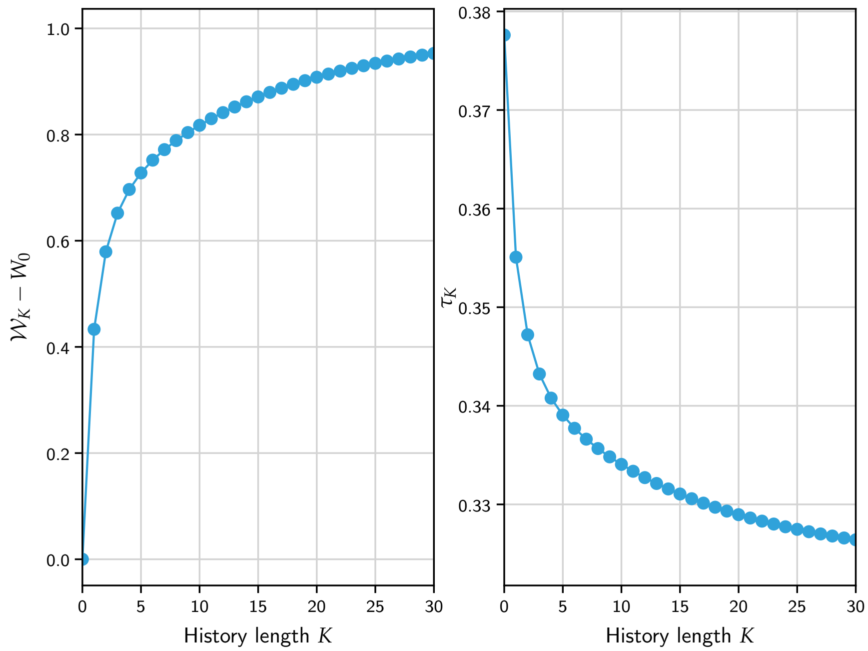


Figure 5: Welfare gains from partial history dependence (left panel) and the progressivity wedge (right panel). Welfare gains are expressed as a fraction of the welfare gains with full history dependence.

6 past incomes lead to a welfare gain capture 75 percent of the potential welfare gain under full history dependence. Even a short history dependence thus generates significant welfare gains.

**Current and past income only.** Since conditioning taxes only on the current and previous income generates more than 40 percent of the overall welfare gains, it is useful to investigate this case in more detail. One can obtain a closed form solution for the history dependence parameters $\theta_0$ and $\theta_1$:

$$\theta_0 = \frac{1 - \beta\tilde{s}}{1 + \beta - 2\beta\tilde{s}}, \quad \theta_1 = \frac{1 - \tilde{s}}{1 + \beta - 2\beta\tilde{s}},$$

where $\tilde{s} = s_\kappa + s_\phi + \rho s_\omega$ is the "persistence adjusted" variance share of the non-transitory shocks. It is a key factor in determining the role of history dependence. It is easy to see that $\theta_0 > \theta_1$ and so previous income always gets a smaller weight. In special cases, the expressions also reduce to the now familiar patterns: if there are only transitory shocks then $\theta_0 = \theta_1$, while if there are only permanent shocks then $\theta_0 = 1$ and $\theta_1 = 0$. The benchmark calibration yields $\theta_0 = 0.536$ and $\theta_1 = 0.483$. The value of the risk loading factor $p$ is

$$p = \frac{1 - \beta\tilde{s}^2}{1 + \beta - 2\beta\tilde{s}}.$$

If only transitory shocks are present, then $\tilde{s} = 0$ and $p = 1/(1+\beta)$, a minimal value one can obtain with only current and previous incomes. For $\beta = 0.96$, it is approximately equal to one half. Differentiating $p$ with respect to $\tilde{s}$ and evaluating at $\tilde{s} = 0$, yields

$$\left.\frac{dp}{d\tilde{s}}\right|_{\tilde{s}=0} = \frac{2\beta}{(1+\beta)^2} \approx 0.5,$$

where the approximation again uses $\beta = 0.96$. An increase in the share of non-transitory shocks increases the risk loading factor by about half of that increase. The relationship is quite linear except for when $\tilde{s}$ is very close to one. The variance share of the "persistence adjusted" transitory shocks $1 - \tilde{s}$ thus appears to be the most important factor in the determination of welfare gains: for a unitary increase in its variance share, $p$ declines by one half, and the welfare gains correspondingly increase, according to (24).

# 8 When is Age Independence Approximately Optimal?

Previous results are based on an assumption that the government can use tax functions that depend directly on age through the level parameters $\lambda$. What happens if the government does not have the ability to choose age specific values of $\lambda$? The inability to transfer resources across age through variations in $\lambda$ means that the remaining parameters $\tau$ and $\theta$ will, at least in part, be chosen so as to substitute for age varying $\lambda$. The choice of the optimal history dependence parameters $\theta$ is now more complex, and depends, for example, on the intertemporal price of consumption $q$.

Let $\overline{\mathcal{W}}(\tau, \theta)$ be the welfare from a tax policy that is restricted to use an age invariant $\lambda$. Then the welfare gain from introducing time varying $\lambda$ is, for any $\tau$ and $\theta$,

$$\Delta^{\text{agg}}(\tau, \theta) = \mathcal{W}(\tau, \theta) - \overline{\mathcal{W}}(\tau, \theta).$$

Starting with an age and history independent tax function, one can now think of a tax reform in two steps. In the first step, the assumption of age independent $\lambda$ is relaxed, yielding aggregate welfare gain $\Delta^{\text{agg}}(\tau, \theta^{HI})$. In the second step, history independence is introduced, and the values of $\lambda$ are re-optimized. The second step yields a welfare gain $\Delta$, as defined previously by (23). This particular sequencing is useful, because $\Delta^{\text{agg}}(\tau, \theta^{HI})$ can be computed easily. It can be shown that it is *approximately* equal to

$$\Delta^{\text{agg}}(\tau, \theta^{HI}) \approx \ln\left(\frac{1-\beta}{1-q}\right) + \frac{\beta}{1-\beta}\ln\left(\frac{\beta}{q}\right) + \left[(1-\tau)\left(\frac{1}{1-\beta\rho} - \frac{1}{1-q\rho}\right) - (1-\tau)^2\left(\frac{1}{1-\beta\rho^2} - \frac{1}{1-q\rho^2}\right)\right]\frac{\sigma_\omega^2}{2}.$$

The approximation is based on the assumption that the shocks are normally distributed, and is derived by repeatedly using a standard approximation $\ln(1+a) \approx a$ for small $a$.[13] The approximate welfare gain $\Delta^{\text{agg}}(\tau, \theta^{HI})$ is nonnegative, and has a minimum of zero when $\beta = q$. In this case, $\mathcal{W}(\tau, \theta^{HI}) \approx \overline{\mathcal{W}}(\tau, \theta^{HI})$, and welfare under a tax function with constant values of $\lambda$ is approximately equal to the welfare under a tax function when $\lambda$ is optimally varying with age.[14] Next proposition shows that this important property extends beyond history independent tax functions. For any $\tau$ and any history dependence parameters, the approximate welfare gain is zero if the discount factor equals the intertemporal price of consumption:

**Proposition 12.** *If $\beta = q$ then $\Delta^{agg}(\tau, \theta) \approx 0$.*

The welfare is thus approximately equal to the expression given in (14) for any history dependence parameters. Since the approximate welfare formula is the same as in equation (14), the optimal coefficients found in Proposition (5) are also approximately correct.

**Accuracy of the Approximation** The approximate solution, as characterized by Proposition 12 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 (14) implies

[13] The aggregate welfare gain corresponds to the aggregate gains in Farhi and Werning (2012a); in fact, the value of $\Delta^{\text{agg}}(\tau, \theta^{HI})$ is an approximation of their welfare gain in a partial equilibrium with linear technology.

[14] This property has in fact been used in the calibration, when the welfare under the U.S. tax code, which is age independent, was plotted on the blue line of Figure 2.

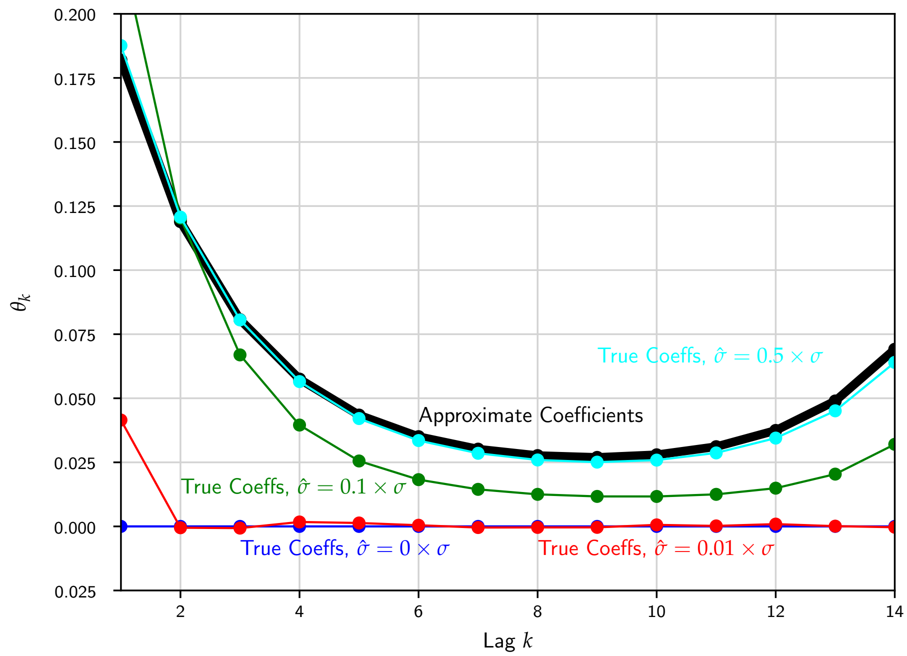


Figure 6: Accuracy of the approximate solution for $K = 15$, $\tau = 0.181$ and $\sigma^2 = 0.325$. For each experiment, all standard deviations are reduced at the same proportion. $\theta_0$ and $\theta_K$ are not shown.

that the history dependence coefficients are independent of the variance parameters, and thus hold even in case of $\sigma^2 = 0$. But that is clearly not a correct solution. In the absence of idiosyncratic shocks it is optimal to have constant consumption over time, which is, for a constant $\lambda$, achieved by a history independent tax system. Thus, we know that if $\sigma^2 = 0$, $\theta_0 = 1$ 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 8 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 $\sigma^2 = 0.325$ , and the U.S. progressivity wedge $\tau = 0.181$ and show the exact coefficients when the standard deviation of all shocks are proportionately reduced to 50 %, 10 %, 1 % and 0 % of its benchmark value.[15]

If the standard deviation of shocks is zero, $\hat{\sigma} = 0$, then the true coefficients are zero,

[15] To simplify exposition I only plot coefficients $\theta_k$ for $k = 1, \ldots K - 1$ and do not show $\theta_0$ and $\theta_K$ that have different magnitudes.

and the approximate solution is obviously inaccurate. This is also true when the standard deviation is only 1 percent of the benchmark standard deviation of shocks, and the consumption smoothing factor still dominates. If the standard deviation is ten percent of the benchmark deviation then the true solution is getting closer to the approximate solution. If the standard deviation is 50 percent of the benchmark value or equal to the benchmark value then the approximate solution is almost identical to the true solution. For even higher values, both solutions are indistinguishable. Thus, for realistic parameter values, the approximate solution provides an excellent approximation to the true coefficients.

# 9 Heterogeneous Income Profiles

Guvenen (2007) proposes and estimates an alternative statistical decomposition of the wage process, where the agents are exposed to less persistent earnings shocks but, in addition, face heterogeneity in the lifecycle wage profiles. Since the approach used in this paper is flexible enough to add additional aspects of wage profiles relatively easily, I now investigate the implications of the heterogeneous wage profiles for the optimal tax design.

Under the heterogeneous income profiles specification, the wages $w_j$ are exogenously determined according to the following stochastic process:

$$\ln w_j = \kappa + \gamma(j+1) + z_j + \varepsilon_j \tag{25}$$

$$z_j = \rho z_{j-1} + \omega_j, \tag{26}$$

where the slope of the lifecycle wage profile is given by an idiosyncratic factor $\gamma$. One can show that the social welfare expression (13) can also be extend relatively easily by adding moments related to the heterogeneous wage profiles. I will, however, assume directly that $\gamma$ is also lognormally distributed,

$$\gamma \sim N\left(-\frac{\sigma_\gamma^2}{2}, \sigma_\gamma^2\right).$$

Under this assumption, a close analogue of Proposition 2 exists. The total effective variance of shocks now includes the present discounted value of the variance of the hetereogeneous wage profile component $\gamma$:

$$\sigma^2 = \frac{\sigma_\omega^2}{1-\beta\rho^2} + \sigma_\varepsilon^2 + \sigma_\kappa^2 + \sigma_\phi^2 + \frac{1+\beta}{(1-\beta)^2}\sigma_\gamma^2,$$

and the share of the new term is

$$s_\gamma = \frac{1+\beta}{(1-\beta)^2}\frac{\sigma_\gamma^2}{\sigma^2}.$$

Compared to the persistent component, the variance of the heterogeneous wage profile component $\sigma_\gamma^2$, is leveraged by an additional order of magnitude: with $\beta = 0.96$, $s_\gamma$ is 1225 times its variance. Even a very small heterogeneity in the wage component will carry a significant weight in the determination of the optimal history dependence. Proposition 2 can now be extended as follows.

**Proposition 13.** *Suppose that the shocks are lognormally distributed. Then the social welfare function $\mathcal{W}(\tau,\theta)$ is given by (14), with $p(\theta)$ now defined as*

$$p(\theta) = s_\omega P_\rho(\theta) + s_\varepsilon P_0(\theta) + (s_\kappa + s_\phi)P_1(\theta) + s_\gamma \hat{P}(\theta), \tag{27}$$

*where $P_\rho(\theta)$ is given by (16) and $\hat{P}(\theta)$ is*

$$\hat{P}(\theta) = \sum_{j=0}^{\infty}\beta^j\left[\theta_j^2 + 2\textstyle\sum_{k=0}^{j-1}\left(1+\dfrac{1-\beta}{1+\beta}(j-k)\right)\theta_k\theta_j\right].$$

In contrast to other shocks, the interaction terms $\theta_k\theta_j$ have larger weight if $k$ is further away from $j$. A positive value of $\theta$ at two distant lags thus increases the risk loading factor $\hat{P}$ relatively more than if the two lags were closer to each other. This will have important consequences for the optimal history dependence coefficients. For example, exponentially decaying coefficients, in the spirit of Proposition 5, are not likely to be optimal, when heterogeneous income profiles are present: one can show that the value of $\hat{P}$ is strictly greater than one, in such a case, except for a case with history independence. But, unlike permanent (or random walk) shocks, history independence is not optimal, if profile heterogeneity is the only source of heterogeneity. One can do better. The solution for this special case is shown in the next proposition.

**Proposition 14.** *Suppose that $\sigma_\omega^2 = \sigma_\varepsilon^2 = \sigma_\phi^2 + \sigma_\kappa^2 = 0$. Then $\theta_0^* = 1/(1-\beta)$, $\theta_1^* =$*

$-1/(1-\beta)$, $\theta_j^* = 0$ *for* $j > 1$, *and* $p = 1/(1+\beta)$.

The proposition thus paints a very different picture about how history dependence should look like. Only a very short history dependence, only the current and previous income, are enough; the tax function is progressive with respect to the previous income, not regressive; and there are significant welfare gains from a short history dependence. To understand the mechanics behind the optimal tax design, note that the optimal tax function can be written as

$$T(y_j, y_{j-1}) = y_j - \lambda_j \left(\frac{y_j}{y_{j-1}}\right)^{\frac{1-\tau}{1-\beta}}.$$

The agents' consumption $c_j = y_t - T(y_j, y_{j-1})$ thus depends on the growth rate of one's incomes. Since the agents differ in the growth rates of incomes, the optimal tax policy translates differences in the *growth rates* of income to differences in the *levels* of consumption. This produces flat consumption profiles for each agent.[16] It is a much better consumption profile than the under a tax policy with history independence, which translates differences in the growth rates of incomes to differences in the growth rates of consumption. The welfare gains from the optimal tax policy are likely to be significant, since the loading factor $p = \hat{P}$ is only a little bit more than one half for realistic values of $\beta$.

It is, of course, an open question how will the forces displayed in Proposition 14 compare to the remaining forces in the model that were displayed in Proposition 5 and that point towards long history dependence and regressivity with respect to past incomes. A closed form solution for the general case is no longer available. However, one can easily solve the model numerically. To recalibrate the model, I have used the fact that, since hours worked are constant conditional on $\phi$, the stochastic process for wages exhibits the same moments as the stochastic process for incomes. I calibrate the moments of the wage process to be the same as the moments of the income process in the benchmark HIP calibration of Guvenen (2007), as reported in his Table 1. The remaining parameters are identical to my benchmark calibration. The parameters are in Table 4.

Relative to the benchmark calibration, the persistent component exhibits less persistence ($\rho$ equals only 0.821 vs. 0.958 in the benchmark calibration) and slightly more variance (0.029 vs 0.017), and transitory shocks exhibit less variance (0.047 vs 0.081).

[16] Assuming that $\lambda_j$ is constant.

Table 4: Heterogeneous Income Profile Parameters

| $\beta$ | $q$ | $\eta$ | $\rho$ | $\sigma^2_\omega$ | $\sigma^2_\varepsilon$ | $\sigma^2_\kappa$ | $\sigma^2_\gamma$ | $\sigma^2_\phi$ |
|---|---|---|---|---|---|---|---|---|
| 0.960 | 0.960 | 2.000 | 0.821 | 0.029 | 0.047 | 0.022 | 0.00038 | 0.036 |

Most importantly, the heterogeneous wage profile component has a variance of 0.00038, which looks small, but carries a large weight, as shown above.

Table 5 shows the welfare gains from the optimal history dependence. The overall welfare gains from the optimal tax reform are huge, at 9.20 percent of consumption, but about two thirds of that gain come from increasing the progressivity parameter $\tau$ from 0.181 to 0.470. The welfare gains from history dependence itself are 3.21 percent. That is slightly larger than 2.98 percent in the benchmark model without heterogeneous income profiles. The welfare gain from history dependence is thus robust to the introduction of heterogeneous income profiles.

Table 5: Welfare gains, decomposition

| specification | $\Delta$ | $\Delta^{\text{ins}}$ | $\Delta^{\text{inc}}$ | $\tau^{HD}$ |
|---|---|---|---|---|
| baseline | 9.20 | 3.21 | 5.98 | 0.470 |
| only heterogeneous profiles ($\sigma^2_\omega = \sigma^2_\varepsilon = \sigma^2_\phi + \sigma^2_\kappa = 0$) | 8.56 | 7.65 | 0.91 | 0.325 |
| no heterogeneous profiles ($\sigma^2_\gamma = 0$) | 2.75 | 2.74 | 0.01 | 0.202 |

Note: Welfare gains in percent consumption equivalents, relative to U.S. tax system.

Figure 7 shows the optimal history dependence coefficients for 60 periods. Proposition 14 has shown that heterogeneity in profiles manifests itself as a force for progressivity with respect to past incomes, and negative history dependence coefficients. This needs to be weighted against the forces that are due to the persistent and transitory component, and favor regressivity with respect to past incomes. The blue line of Figure 7 shows the resulting compromise: the optimal history dependence coefficients are U-shaped, first positive, then negative, and then converging back to zero. In the short run, the "conventional" forces from the benchmark model dominate, and the history dependence coefficients are strictly positive. After about 15 periods, the heterogeneity in profiles become the dominant force, and the history dependence coefficients become

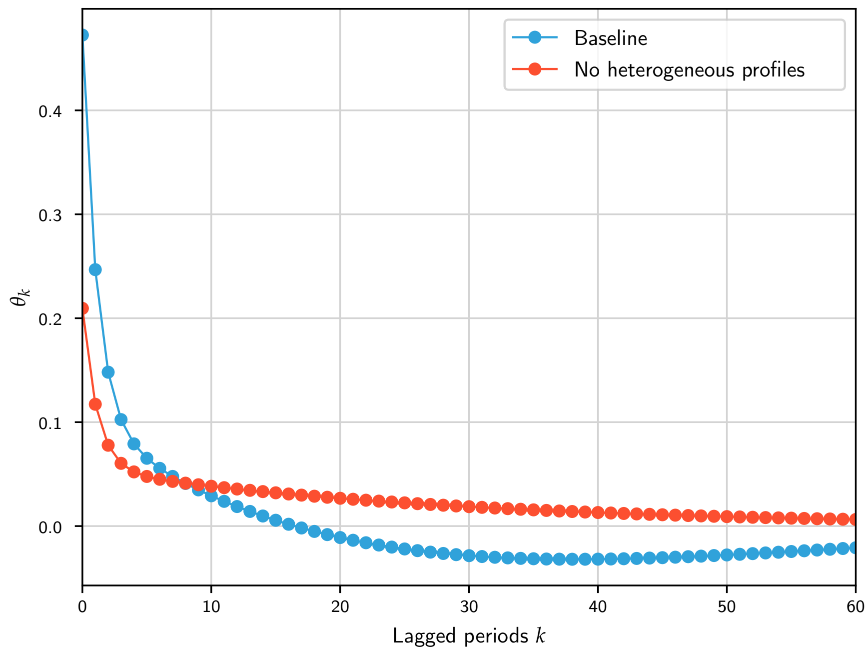


Figure 7: History dependence coefficients for the baseline calibration with and without heterogeneous profiles.

negative. In the long run, the permanent component drives them back to zero.

If heterogeneous profiles are the only source of heterogeneity, then Proposition 14 applies. While the overall welfare gains are only slightly lower, almost all of them comes from the optimal history dependence, which increases welfare by 7.65 percent. The optimal history dependence coefficients are not plotted in Figure 7 because they are of an order of magnitude larger, but they are given theoretically in Proposition 14: $\theta_0 = 25$, $\theta_1 = -25$, and all other coefficients are zero. Finally, if heterogeneity in profiles is absent, then the overall welfare gain is 2.75 percent, with 2.74 percent coming from history dependence. The gains from history dependence are not far off from the gains of 2.98 percent found in the benchmark calibration.[17] This suggests that the results are robust to a reasonable reparameterization of the wage process. The history dependence coefficients are also not far off, as one can see by comparing the red line in Figure 7, and the blue line in Figure 1. Overall, Table 5 shows that introducing heterogeneity in

[17]Both exercises are close, but not identical: while the benchmark wage process was calibrated without heterogeneity in income profiles, it is now calibrated to include heterogeneity in income profiles, and then the heterogeneity is shut down.

the wage profiles has a potential to substantially increase welfare gains, but since its effects need to be weighted against the effects of other shocks, the additional gains end up being rather modest.

# 10 A Bewley-Aiyagari Economy

I will now relax the assumption that agents are not allowed to borrow and save by assuming that agents have access to a risk-free bond. That is, I will solve for an Aiyagari economy. To make the problem tractable, I will further restrict the tax function (4) by assuming that the history dependence parameters take the form

$$\theta_j = \theta_0 \left[ \alpha \mu_1^j + (1-\alpha)\mu_2^j \right], \quad j \geq 1.$$

The promise keeping constraint (7) restricts the four parameters of the tax function $\theta_0$, $\alpha$, $\mu_1$ and $\mu_0$ by

$$\theta_0 \left( \frac{\alpha}{1-\beta\mu_1} + \frac{1-\alpha}{1-\beta\mu_2} \right) = 1.$$

This tax system includes a history independent tax system (for $\theta_0 = 1$ and $\mu_1 = \mu_2 = 0$) and the optimal tax system (for $\mu_1$ and $\mu_2$ chosen according to the proposition xxx).

The above representation allows us to write the tax function recursively by

$$T(y_t) = y_t - \lambda \left( y_t S_{1,t}^{\alpha} S_{2,t}^{1-\alpha} \right)^{(1-\tau)\theta_0}, \quad S_{1,t+1} = y_t S_{1,t}^{\mu_1}, \quad S_{2,t+1} = y_t S_{2,t}^{\mu_2}.$$

The agents choose consumption, $c$ and next period assets $a'$, subject to the budget constraint

$$c + qa' \leq \lambda \left( whs_1^{\alpha} s_2^{1-\alpha} \right)^{(1-\tau)\theta_0} + a, \tag{28}$$

where $q$ is the intertemporal price of consumption. When choosing hours worked, the agents take into account the fact that their choice will affect their tax liabilities in the future via an aggregate ”past income” variable $s$:

$$s_1' = whs_1^{\mu_1}, \quad s_2' = wzs_2^{\mu_2}. \tag{29}$$

We also assume that the agents are subject to the borrowing constraint $\underline{b}(z) \leq 0$:

$$a' \geq \underline{b}(z). \tag{30}$$

The agent's problem is to choose consumption, savings and hours worked to maximize

$$V_{\phi}(a, s_1, s_2, w) = \max_{c,h,a'} \left\{ (1-\beta)\big(\ln c - \frac{\phi}{1+\eta} h^{1+\eta}\big) + \beta \mathbb{E}\left[V_{\phi}(a', s_1', s_2', w') | w\right] \right\}$$

subject to (28), (29) and (30), where wages $w$ follow (2).

We can rewrite the dynamic program as

$$V(a, s_1, s_2, z, \varepsilon, \kappa, \phi) = \max_{c,h,a'} \left\{ (1-\beta)\big(\ln c - \frac{\phi}{1+\eta} h^{1+\eta}\big) + \beta \mathbb{E}\left[V(a', s_1', s_2', z', \varepsilon', \kappa, \phi) | z\right] \right\}$$

subject to

$$\begin{aligned} c + qa' &\leq \lambda \left( e^{z\varepsilon\kappa} h s_1^{\alpha} s_2^{1-\alpha} \right)^{(1-\tau)\theta_0} + a \\ s_1' &= e^{z\varepsilon\kappa} h s_1^{\mu_1} \\ s_2' &= e^{z\varepsilon\kappa} z s_2^{\mu_2} \\ z' &= \rho z + \omega \end{aligned}$$

and the borrowing constraint (30).

In a special case when the tax his history independent, the dynamic program reduces to

$$V(a, z, \varepsilon, \kappa, \phi) = \max_{c,h,a'} \left\{ (1-\beta)\big(\ln c - \frac{\phi}{1+\eta} h^{1+\eta}\big) + \beta \mathbb{E}\left[V(a', z', \varepsilon', \kappa, \phi) | z\right] \right\}$$

subject to

$$\begin{aligned} c + qa' &\leq \lambda \left( e^{z\varepsilon\kappa} h \right)^{1-\tau} + a \\ z' &= \rho z + \omega \end{aligned}$$

and the borrowing constraint (30). The first-order condition in $h$ is

$$\frac{1-\tau}{\phi} \frac{\lambda \left( e^{z\varepsilon\kappa} \right)^{1-\tau}}{c} = h^{\tau+\eta}$$

Substitute the first-order condition back into the budget constraint, we have one nonlinear equation in consumption:

$$c + s = c^{-\frac{1-\tau}{\eta+\tau}} \lambda \left(\frac{1-\tau}{\phi}\right)^{\frac{1-\tau}{\tau+\eta}} e^{\frac{(1-\tau)(1+\eta)}{\eta+\tau} z\varepsilon\kappa},$$

where $s = qa' - a$ are savings. This is an equation of the form

$$x = \alpha x^{-\beta} - \gamma, \quad \alpha, \beta > 0,$$

where

$$\alpha = \lambda \left(\frac{1-\tau}{\phi}\right)^{\frac{1-\tau}{\tau+\eta}} e^{\frac{(1-\tau)(1+\eta)}{\eta+\tau} z\varepsilon\kappa}, \quad , \beta = \frac{1-\tau}{\eta+\tau}, \quad \gamma = s,$$

which needs to be solved numerically for consumption as a function of the shocks and of savings. Hours are then obtained from its first-order condition.

# 11 An Aiyagari Economy OLD

I will now relax the assumption that agents are not allowed to borrow and save by assuming that agents have access to a risk-free bond. That is, I will solve for an Aiyagari economy. To make the problem tractable, I will further restrict the tax function (4) by assuming that the history dependence parameters take the form $\theta_j = \theta_1 \mu^{j-1}$ for $j \geq 1$. One can again write the tax system recursively by

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

where $x_t = \prod_{k=0}^{t} y_{t-k}^{-\frac{\theta_1}{\theta_0}\mu^k}$ or, equivalently, $x_t = y_t^{\kappa} x_{t-1}^{\mu}$, where $\kappa = -\frac{\theta_1}{\theta_0}$. The normalization constraint (7) restricts the parameters of the tax function $\theta_0$, $\theta_1$ and $\mu$ by

$$\theta_0 + \theta_1 \frac{\beta\delta}{1-\beta\delta\mu} = 1.$$

Solving for $\theta_1 = -(\theta_0 - 1)\frac{1-\beta\delta\mu}{\beta\delta}$, we thus have a two parameter family of tax functions (in $\theta_0$ and $\mu$) that, together with the tax wedge $\tau$, determine the tax system. The parameter $\kappa$ is now $\kappa = \frac{\theta_0 - 1}{\theta_0}\frac{1-\beta\delta\mu}{\beta\delta}$. This tax system includes a history independent tax system (for

$\theta_0 = 1$) and the optimal tax system (for $\mu = \beta$).

The agents choose consumption, $c$ and next period savings $a'$, subject to the budget constraint

$$c + qa' \leq \lambda \left(\frac{zh}{x}\right)^{(1-\tau)\theta_0} + a, \tag{31}$$

where $q$ is the intertemporal price of consumption. When choosing hours worked, the agents take into account the fact that their choice will affect their tax liabilities in the future via an aggregate "past income" variable $x$:

$$x' = (hz)^{\kappa} x^{\mu}. \tag{32}$$

We also assume that the agents are subject to the borrowing constraint $\underline{b}(z) \leq 0$:

$$a' \geq \underline{b}(z). \tag{33}$$

We assume that the borrowing constraint has the following homogeneity property: $\underline{b}(z) = e^{-\psi \ln z}\underline{b}$ where $\psi = (1-\tau)\theta_0(1 - \frac{\kappa}{1-\mu})$. The agent's problem is to choose consumption, savings and hours worked to maximize

$$V(a, x, z, \phi) = \max_{c,h,a'} \left\{ (1 - \beta\delta) \left( \ln c - \frac{\phi}{1+\eta} h^{1+\eta} \right) + \beta\delta \mathbb{E}\left[ V(a', x', z', \phi) | z \right] \right\}$$

subject to (28) and (29). Given that the productivity shock follows a random walk, the Bellman equation satisfies the following normalization:

**Lemma 15.** *The value function satisfies*

$$V(a, x, z, \phi) = V\left( e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu} \ln z} x, 0, \phi \right) + \psi \ln z.$$

*where* $\psi = (1-\tau)\theta_0(1 - \frac{\kappa}{1-\mu})$.

*Proof.* Let $\mathcal{S}$ be a space of value functions satisfying the above property, and $\mathcal{T}$ be na operator defined by the right-hand side of the Bellman equation above. Let $V \in \mathcal{S}$.

$$
\begin{aligned}
&\mathcal{T}V(a,x,z,\phi) \\
&= \max_{c,h,a'} \left\{ (1-\beta\delta)\ln\left(\lambda\left(e^{\ln z}\frac{h}{x}\right)^{(1-\tau)\theta_0} + a - qa'\right) - (1-\beta\delta)\frac{\phi}{1+\eta}h^{1+\eta} \right. \\
&\left. + \beta\delta\mathbb{E}\left[V\left(e^{-\psi z'}a', e^{-\frac{\kappa}{1-\mu}\ln z'}x', 0, \phi\right)|z\right] \right\} + \beta\delta\psi\left(\ln z - \frac{\sigma_\omega^2}{2}\right) \\
&= (1-\beta\delta)\psi\ln z + \max_{c,h,a'} \left\{ (1-\beta\delta)\ln\left(\lambda\left(e^{\frac{\kappa}{1-\mu}\ln z}\frac{h}{x}\right)^{(1-\tau)\theta_0} + e^{-\psi\ln z}(a-qa')\right) - (1-\beta\delta)\frac{\phi}{1+\eta}h^{1+\eta} \right. \\
&\left. + \beta\delta\mathbb{E}\left[V\left(e^{-\psi\ln z'}a', e^{-\frac{\kappa}{1-\mu}\ln z'}x', 0, \phi\right)|z\right] \right\} + \beta\delta\psi\left(\ln z - \frac{\sigma_\omega^2}{2}\right).
\end{aligned}
$$

We can write the law of motion for past income as

$$
x' = h^\kappa e^{\kappa\ln z}x^\mu = h^\kappa e^{\left(\kappa+\frac{\kappa}{1-\mu}\mu\right)\ln z}\left(xe^{-\frac{\kappa}{1-\mu}\ln z}\right)^\mu = h^\kappa e^{\frac{\kappa}{1-\mu}\ln z}\left(xe^{-\frac{\kappa}{1-\mu}\ln z}\right)^\mu.
$$

Redefine savings by $\hat{a}' = e^{-\psi\ln z}a'$ and define $\hat{x}' = h^\kappa(xe^{-\frac{\kappa}{1-\mu}\ln z})^\mu$. Write the maximization problem as

$$
\begin{aligned}
\mathcal{T}V(a,x,z,\phi) = \max_{c,h,\hat{a}'} \left\{ \ln\left(\lambda\left(\frac{h}{e^{-\frac{\kappa}{1-\mu}\ln z}x}\right)^{(1-\tau)\theta_0} + e^{-\psi\ln z}a - q\hat{a}'\right) - \frac{\phi}{1+\eta}h^{1+\eta} \right. \\
\left. + \beta\delta\mathbb{E}\left[V\left(e^{-\psi\omega}\hat{a}', e^{-\frac{\kappa}{1-\mu}\omega}\hat{x}', 0, \phi\right)\right] \right\} + \psi\left(\ln z - \beta\delta\frac{\sigma_\omega^2}{2}\right).
\end{aligned}
$$

We also have that $a' \geq \underline{b}(z)$ if and only if $\hat{a}' \geq \underline{b}$. Hence $\mathcal{T}V(a,x,z,\phi) \in \mathcal{S}$ and $\mathcal{T}$ maps $\mathcal{S}$ onto itself. The fixed point of the Bellman operator, which exists and is unique by standard arguments, thus belong to $\mathcal{S}$.

□

The normalized value function $v(a,x,\phi) = V(a,x,0,\phi)$ then solves the following Bellman equation:

$$
v(a,x,\phi) = \max_{c,h,a'\geq\underline{b}} \left\{ \ln c - \frac{\phi}{1+\eta}h^{1+\eta} + \beta\delta\mathbb{E}\left[v\left(e^{-\psi\omega}a', e^{-\frac{\kappa}{1-\mu}\omega}x', \phi\right)\right] \right\} - \beta\delta\psi\frac{\sigma_\omega^2}{2}
$$

subject to

$$
\begin{aligned}
c + qa' &\leq \lambda \left(\frac{h}{x}\right)^{(1-\tau)\theta_0} + a \\
x' &= h^\kappa x^\mu.
\end{aligned}
$$

The constant at the end of the Bellman equation is irrelevant from individual's perspective, but matters for the government's optimization problem. Let $g_c(a, x, \phi)$, $g_h(a, x, \phi)$, $g_a(a, x, \phi)$ and $g_x(a, x, \phi)$ be the optimal policy functions in the normalized problem. The first-order conditions are

$$
\begin{aligned}
\frac{\partial v}{\partial a} &= \eta \\
x\frac{\partial v}{\partial x} &= \mu h^\kappa x^\mu \beta\delta\mathbb{E}\left[e^{-\frac{\kappa}{1-\mu}\omega}\frac{\partial v}{\partial x}\left(e^{-\psi\omega}g_a, e^{-\frac{\kappa}{1-\mu}\omega}g_x, \phi\right)\right] - \lambda(1-\tau)\theta_0\left(\frac{h}{x}\right)^{(1-\tau)\theta_0}\eta \\
\frac{1}{g_c} &= \eta \\
q\eta &= \beta\delta\mathbb{E}\left[e^{-\psi\omega}\frac{\partial v}{\partial a}\left(e^{-\psi\omega}g_a, e^{-\frac{\kappa}{1-\mu}\omega}g_x, \phi\right)\right] \\
\psi h^{1+\eta} &= \lambda(1-\tau)\theta_0\left(\frac{h}{x}\right)^{(1-\tau)\theta_0}\eta + \kappa h^\kappa x^\mu \beta\delta\mathbb{E}\left[e^{-\frac{\kappa}{1-\mu}\omega}\frac{\partial v}{\partial x}\left(e^{-\psi\omega}g_a, e^{-\frac{\kappa}{1-\mu}\omega}g_x, \phi\right)\right].
\end{aligned}
$$

Let $z_a(a, x) = \frac{\partial v}{\partial a}(a, x, \phi)$ and $z_x(a, x) = x\frac{\partial v}{\partial x}(a, x, \phi)$ (suppressing the dependence on $\phi$ for notational convenience). Then we can write the system of equations as

$$
\begin{aligned}
z_x &= \mu\beta\delta\mathbb{E}\left[z_x\left(e^{-\psi\omega}g_a, e^{-\frac{\kappa}{1-\mu}\omega}g_x, \phi\right)\right] - \lambda(1-\tau)\theta_0\left(\frac{h}{x}\right)^{(1-\tau)\theta_0} z_a \\
\frac{1}{g_c} &= z_a \\
qz_a &= \beta\delta\mathbb{E}\left[e^{-\psi\omega}z_a\left(e^{-\psi\omega}g_a, e^{-\frac{\kappa}{1-\mu}\omega}g_x, \phi\right)\right] \\
\psi h^{1+\eta} &= \lambda(1-\tau)\theta_0\left(\frac{h}{x}\right)^{(1-\tau)\theta_0} z_a + \kappa\beta\delta\mathbb{E}\left[z_x\left(e^{-\psi\omega}g_a, e^{-\frac{\kappa}{1-\mu}\omega}g_x, \phi\right)\right].
\end{aligned}
$$

The un-normalized policy functions are

$$
\begin{aligned}
G_c(a,x,z,\phi) &= g_c\left(e^{-\psi z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right)e^{\phi \ln z}\\
G_h(a,x,z,\phi) &= g_h\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right)\\
G_a(a,x,z,\phi) &= g_a\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right)e^{\psi \ln z}\\
G_x(a,x,z,\phi) &= g_x\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right)e^{\frac{\kappa}{1-\mu}\ln z}.
\end{aligned}
$$

The budget constraint can be written as

$$
\begin{aligned}
G_c(a,x,z,\phi) + qG_a(a,x,z,\phi) &\leq \lambda\left(e^{\ln z}\frac{G_h(a,x,z,\phi)}{x}\right)^{(1-\tau)\theta_0} + a\\
g_c\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right)e^{\phi \ln z} + qg_a\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right)e^{\psi \ln z} &\leq \lambda\left(e^{\left(1-\frac{\kappa}{1-\mu}\right)\ln z}\frac{g_h\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\right.}}{e^{-\frac{\kappa}{1-\mu}\ln z}x}\right.\\
g_c\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right) + qg_a\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right) &\leq \lambda\left(\frac{g_h\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right)}{e^{-\frac{\kappa}{1-\mu}\ln z}x}\right)^{(1}
\end{aligned}
$$

Evaluating at $\ln z = 0$ we get

$$
g_c\left(a,x,\phi\right) + qg_a\left(a,x,\phi\right) \leq \lambda\left(\frac{g_h\left(a,x,\phi\right)}{x}\right)^{(1-\tau)\theta_0} + a,
$$

as expected. As for the borrowing constraint, we write

$$
\begin{aligned}
G_a(a,x,z,\phi) &\geq \underline{b}(z)\\
g_a\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right)e^{\psi \ln z} &\geq e^{\psi \ln z}\underline{b}\\
g_a\left(e^{-\psi \ln z}a, e^{-\frac{\kappa}{1-\mu}\ln z}x, \phi\right) &\geq \underline{b}.
\end{aligned}
$$

**Aggregation.** Let $S(a,x,\phi)$ be the distribution of assets, income aggregates and preference parameters. The associated operator is

$$
\mathcal{T}S(\mathcal{A},\mathcal{X},\phi) = \int_{a,x:\tilde{a}'(a,x,\phi)\in\mathcal{A},\tilde{x}'(a,x,\phi)\in\mathcal{X}} S(a,x,\phi)\,da\,dx \quad \forall \phi, \mathcal{A}\subseteq A, \mathcal{X}\subseteq X.
$$

$S$ is stationary if $\mathcal{T}S = S$. In the aggregate, assets are zero:

$$\int_{a,x,\phi} g_a(a,x,\phi) S(a,x,\phi)\, da\, dx\, d\phi = 0, \tag{34}$$

and the budget constraint holds:

$$\int_{a,x,\phi} \left[g_c(a,x,\phi) - g_h(a,x,\phi)\right] S(a,x,\phi)\, da\, dx\, d\phi = 0. \tag{35}$$

For given tax parameters $\tau$, $\theta_0$ and $\mu$, the *recursive competitive equilibrium* consists of the tax parameter $\lambda$, price $q$, value function $v(a,x,\phi)$, policy functions $g_c(a,x,\phi)$, $g_h(a,x,\phi)$, $g_a(a,x,\phi)$ and $g_x(a,x,\phi)$, and distribution $S(a,x,\phi)$ such that i) $v$, $g_c$, $g_h$, $g_a$ and $g_x$ solve the dynamic program above, ii) $S$ is stationary, iii) aggregate assets are zero, i.e. (34) holds, and iv) the resource constraint (35) holds.

$$\begin{aligned} F(a,x,\phi,z) &= Pr(\tilde{a} \le a, \tilde{x} \le x, \phi, z) \\ &= Pr(e^{-\psi \ln z}\tilde{a} \le e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu}\ln z}\tilde{x} \le e^{-\frac{\kappa}{1-\mu}\ln z} x, \phi, \ln z) \end{aligned}$$

Transition function:

$$Q(\mathcal{A}, \mathcal{X}, \ln z' | a, x, \ln z, \phi) = \begin{cases} f_\omega(\ln z' - \ln z) & \text{if } G_a(a,x,\ln z,\phi) \in \mathcal{A} \text{ and } G_x(a,x,\ln z,\phi) \in \mathcal{X}. \\ 0 & \text{otherwise} \end{cases}$$

The law of motion for the distribution $S(a,x,z,\phi)$ is

$$\mathcal{T}S(\mathcal{A},\mathcal{X},z',\phi) = \int_{a,x,z} Q(\mathcal{A},\mathcal{X},z'|a,x,z,\phi) S(a,x,z,\phi)\, da\, dx\, dz \quad \forall \phi, \mathcal{A} \subseteq A, \mathcal{X} \subseteq X.$$

Now note that $G_a(a,x,z,\phi) \in \mathcal{A}$ if and only if $G_a\left(e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu}\ln z} x, 0, \phi\right) \in e^{-\psi \ln z}\mathcal{A}$ and that $G_x(a,x,z,\phi) \in \mathcal{X}$ if and only if $G_x\left(e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu}\ln z} x, 0, \phi\right) \in e^{-\frac{\kappa}{1-\mu}\ln z}\mathcal{X}$. Thus

$$Q(\mathcal{A},\mathcal{X},z'|a,x,z,\phi) = Q(e^{-\psi \ln z}\mathcal{A}, e^{-\frac{\kappa}{1-\mu}\ln z}\mathcal{X}, \ln z' - \ln z | e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu}\ln z} x, 0, \phi).$$

Guess that $S(a,x,z,\phi) = S(e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu}\ln z} x, 0, \phi) e^{r \ln z}$. Now, $\forall \phi, \mathcal{A} \subseteq A, \mathcal{X} \subseteq X$,

$$
\begin{aligned}
&\mathcal{T}S(\mathcal{A},\mathcal{X},z',\phi) \\
&= \int_{a,x,\ln z} Q(\mathcal{A},\mathcal{X},z'|a,x,z,\phi) S(a,x,z,\phi)\, da\, dx\, d\ln z \\
&= \int_{a,x,\ln z} Q(e^{-\psi \ln z}\mathcal{A}, e^{-\frac{\kappa}{1-\mu}\ln z}\mathcal{X}, \ln z' - \ln z | e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu}\ln z} x, 0, \phi) S(e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu}\ln z} x, 0, \phi) e^{r \ln z}\, da\, d \\
&= \int_{a,x,z} Q(e^{-\psi \ln z}\mathcal{A}, e^{-\frac{\kappa}{1-\mu}\ln z}\mathcal{X}, \ln z' - \ln z | \tilde{a}, \tilde{x}, 0, \phi) S(\tilde{a}, \tilde{x}, 0, \phi) e^{(r+\psi+\frac{\kappa}{1-\mu})\ln z}\, d\tilde{a}\, d\tilde{x}\, d\ln z \\
&= e^{(r+\psi+\frac{\kappa}{1-\mu})\ln z'} \int_{a,x} Q(e^{-\psi \ln z}\mathcal{A}, e^{-\frac{\kappa}{1-\mu}\ln z}\mathcal{X}, \omega | \tilde{a}, \tilde{x}, 0, \phi) S(\tilde{a}, \tilde{x}, 0, \phi) e^{-(r+\psi+\frac{\kappa}{1-\mu})\omega}\, d\tilde{a}\, d\tilde{x} \\
&= e^{(r+\psi+\frac{\kappa}{1-\mu})\ln z'} \mathcal{T}S(e^{-\psi \ln z}\mathcal{A}, e^{-\frac{\kappa}{1-\mu}\ln z}\mathcal{X}, \omega, \phi)
\end{aligned}
$$

where $\tilde{a} = e^{-\psi \ln z} a$ and $\tilde{x} = e^{-\frac{\kappa}{1-\mu}\ln z} x$.

## 11.1 No savings as a special case.

The solution to the model with no savings is a special case with $a = a' = 0$. Next Lemma characterizes the value function:

**Lemma 16.** *Suppose that $a = a' = 0$ (no savings). Then the value function $\hat{v}(x,\phi) = v(0,x,\phi)$ satisfies*

$$\hat{v}(x,\phi) = A \ln x + B \ln \phi + C,$$

*where*

$$
\begin{aligned}
A &= -(1-\tau)\theta_0 \frac{1-\beta\delta}{1-\beta\delta\mu} \\
B &= -\frac{1-\tau}{1+\eta} \\
C &= \ln \lambda + \frac{1-\tau}{1+\eta}\left[\ln(1-\tau) - 1\right] - \frac{\beta\delta}{1-\beta\delta}(1-\tau)\frac{\sigma_\omega^2}{2}.
\end{aligned}
$$

*Proof.* Using the guess, we have

$$\mathbb{E}\left[\hat{v}\left(e^{-\frac{\kappa}{1-\mu}\omega} x', \phi\right)\right] = A \ln x' + B \ln \phi + C + A \frac{\kappa}{1-\mu}\frac{\sigma_\omega^2}{2}.$$

Substituting away $c$ and $x'$, and taking the first-order condition in $h$, the right-hand side becomes

$$
\begin{aligned}
RHS = {} & (1-\beta\delta)\ln\lambda + (1-\beta\delta)(1-\tau)\theta_0(\ln h - \ln x) - (1-\beta\delta)\frac{\phi}{1+\eta}h^{1+\eta} \\
& + \beta\delta\left[A\mu\ln x + A\kappa\ln h + B\ln\phi + C + A\frac{\kappa}{1-\mu}\frac{\sigma_\omega^2}{2} - \psi\frac{\sigma_\omega^2}{2}\right].
\end{aligned}
$$

Taking a first-order condition w.r.t. $h$ yields

$$
\phi h^{1+\eta} = (1-\tau)\theta_0 + \frac{\beta\delta}{1-\beta\delta}A\kappa,
$$

which implies that $h$ is independent of $x$. We can then directly equate the terms involving $x$ to obtain the expression for $A$ given in the lemma. This in turn allows us to obtain an expression for hours worked:

$$
h = \left(\frac{1-\tau}{\phi}\right)^{\frac{1}{1+\eta}}.
$$

Rewriting the right-hand side yields

$$
\begin{aligned}
A\ln x + B\ln\phi + C = {} & (1-\beta\delta)\ln\lambda + (1-\beta\delta)\frac{1-\tau}{1+\eta}[\ln(1-\tau) - \ln\phi - 1] \\
& + A\ln x + \beta\delta B\ln\phi + \beta\delta C \\
& + \beta\delta\left(A\frac{\kappa}{1-\mu} - \psi\right)\frac{\sigma_\omega^2}{2}.
\end{aligned}
$$

This yields $B$ in the expression. Finally, equating the constants and rearranging (using $A\frac{\kappa}{1-\mu} - \psi = -1+\tau$) yields the expression for $C$. □

**Value and Policy Functions.** The optimal policy functions, after un-normalizing them, are simply

$$
\begin{aligned}
\ln h &= \frac{1}{1+\eta}\left[\ln(1-\tau) - \ln\phi\right] \\
\ln c &= \ln\lambda + (1-\tau)\theta_0\left(\frac{\ln(1-\tau) - \ln\phi}{1+\eta} - \ln x + \ln z\right) \\
\ln x' &= \kappa\ln z + \frac{\kappa}{1+\eta}\left[\ln(1-\tau) - \ln\phi\right] + \mu\ln x.
\end{aligned}
$$

The value function $\hat{V}(x,z,\phi) = \hat{v}\left(e^{-\frac{\kappa}{1-\mu}\ln z}x,\phi\right) + \psi \ln z$ can be written as

$$\hat{V}(x,z,\phi) = -(1-\tau)\theta_0 \frac{1-\beta\delta}{1-\beta\delta\mu}\ln x - \frac{1-\tau}{1+\eta}\ln\phi + (1-\tau)\ln z + C.$$

Assume that $\ln x \sim N(\mu_x, \sigma_x^2)$. Since

$$\mu_{x,t+1} = -\kappa\sigma_\omega^2 t + \frac{\kappa}{1+\eta}\ln(1-\tau) - \kappa\sigma_\phi^2 + \mu\mu_{xt} \tag{36}$$

**Optimal Sequences** We can recover a sequence of the $x$ values by repeatedly substituting to the law of motion:

$$\begin{aligned}
\ln h_j &= \frac{1}{1+\eta}\left[\ln(1-\tau) - \ln\phi\right] \\
\ln c_j &= \ln\lambda + (1-\tau)\theta_0\left(\ln h - \ln x_j + \ln z_j\right) \\
&= \ln\lambda + (1-\tau)\theta_0\left[\ln h - \kappa\ln h \sum\nolimits_{k=0}^{j-1}\mu^k - \kappa\sum\nolimits_{k=0}^{j-1}\mu^k \ln z_{j-k-1} + \ln z_j\right] \\
&= \ln\lambda + (1-\tau)\sum\nolimits_{k=0}^{j}\theta_k \ln h + (1-\tau)\sum\nolimits_{k=0}^{j}\theta_k \ln z_{j-k} \\
\ln x_j &= \kappa\ln h\sum\nolimits_{k=0}^{j-1}\mu^k + \kappa\sum\nolimits_{k=0}^{j-1}\mu^k \ln z_{j-k-1}.
\end{aligned}$$

Note that the expressions for $c_j$ and $h_j$ are identical to the ones derived in the sequence problem, verifying the alternative derivation.

**Expected values.** The expected values are:

$$\begin{aligned}
\mathbb{E}_0 h_j &= (1-\tau)^{\frac{1}{1+\eta}} \\
\mathbb{E}_0 c_j &= \lambda(1-\tau)^{\sum_{k=0}^{j}\frac{1-\tau}{1+\eta}\theta_k} B_\phi\left(-\sum\nolimits_{k=0}^{j}\frac{1-\tau}{1+\eta}\theta_k\right)\prod\nolimits_{k=0}^{j-1} B_\omega\left[\sum\nolimits_{l=0}^{k}(1-\tau)\theta_l\right] \\
\mathbb{E}_0 \ln x_j &= \kappa\left[\sum\nolimits_{k=0}^{j-1}\mu^k\left(\frac{\ln(1-\tau)}{1+\eta} - \frac{\sigma_\phi^2}{2}\right) - \sum\nolimits_{k=0}^{j-1}\mu^{j-k-1}k\frac{\sigma_\omega^2}{2}\right] \\
&= -\frac{1}{\theta_0}\left[\sum\nolimits_{k=1}^{j}\theta_k\left(\frac{\ln(1-\tau)}{1+\eta} - \frac{\sigma_\phi^2}{2}\right) - \sum\nolimits_{k=0}^{j-1}\theta_{j-k}k\frac{\sigma_\omega^2}{2}\right].
\end{aligned}$$

The expected value function is

$$\mathbb{E}_0 \hat{V}_j = -(1-\tau)\theta_0 \frac{1-\beta\delta}{1-\beta\delta\mu} \mathbb{E}_0 \ln x_j - (1-\tau)\frac{\sigma_\phi^2}{2} - (1-\tau) j \frac{\sigma_\omega^2}{2} + C.$$

**Aggregation** The aggregate resource constraint is

$$(1-\delta)\textstyle\sum_{j=0}^{\infty} \delta^j \mathbb{E}(c_j - \ln z_j h) = 0$$
$$\lambda A(\tau,\theta) = (1-\tau)^{\frac{1}{1+\eta}},$$

yielding $\ln\lambda = \frac{1}{1+\eta}\ln(1-\tau) - \ln A(\tau,\theta)$, the same expression as in the unrestricted sequence problem.

**Welfare** The aggregate welfare can be as follows:

$$\mathcal{W} = \frac{(1-\delta)(1-\beta)}{1-\beta\delta} \sum_{t=-\infty}^{\infty} \beta^t \mathbb{E}_0 \hat{V}_t = \frac{(1-\delta)(1-\beta)}{1-\beta\delta} \left[\textstyle\sum_{t=0}^{\infty} \beta^t \mathbb{E}_0 \hat{V}_0 + \sum_{t=1}^{\infty} \delta^t \mathbb{E}_0 \hat{V}_t\right]$$

We write

$$\mathbb{E}_0 \hat{V}_0 = -\ln A + \frac{1}{1+\eta}\left[\ln(1-\tau) - 1 + \tau\right] + (1-\tau)\left[\frac{\ln(1-\tau)}{1+\eta}\frac{\sigma_\phi^2}{2}\right] - (1-\tau)\frac{\beta(1-\delta)}{1-\beta\delta}\frac{\sigma_z^2}{2}$$

$$\begin{aligned}
\textstyle\sum_{t=1}^{\infty} \delta^t \mathbb{E}_0 \hat{V}_t &= \frac{\delta}{1-\delta}\mathbb{E}_0 \hat{V}_0 - \textstyle\sum_{t=1}^{\infty} \delta^t t \frac{\sigma_\omega^2}{2} \\
&\quad + (1-\tau)\frac{1-\beta\delta}{1-\beta\delta\mu}\left[\left(\frac{\ln(1-\tau)}{1+\eta} - \frac{\sigma_\phi^2}{2}\right)\textstyle\sum_{t=1}^{\infty} \delta^t \sum_{k=0}^{t-1} \theta_{k+1} - \frac{\sigma_\omega^2}{2}\sum_{t=1}^{\infty} \delta^t \sum_{k=0}^{t-1} \theta_{j-k} k\right] \\
&= \frac{\delta}{1-\delta}\mathbb{E}_0 \hat{V}_0 - \frac{1}{1-\delta}(1-\tau)\frac{\sigma_z^2}{2} \\
&\quad + \frac{1}{1-\delta}(1-\tau)\frac{1-\beta\delta}{1-\beta\delta\mu}\left(\frac{\ln(1-\tau)}{1+\eta} - \frac{\sigma_\phi^2 + \sigma_z^2}{2}\right)\textstyle\sum_{k=1}^{\infty} \delta^k \theta_k,
\end{aligned}$$

where the last equality uses $\sum_{t=1}^{\infty}\delta^t t = \frac{\delta}{(1-\delta)^2}$, $\sum_{t=1}^{\infty}\delta^t\sum_{k=0}^{t-1}\theta_{k+1} = \frac{1}{1-\delta}\sum_{k=1}^{\infty}\delta^k\theta_k$, and $\sum_{t=1}^{\infty}\delta^t\sum_{k=0}^{t-1}\theta_{j-k}k = \frac{1}{1-\delta}\frac{\delta}{1-\delta}\sum_{k=1}^{\infty}\delta^k\theta_k$. We can then write the expression for welfare as

$$\begin{aligned}\mathcal{W} &= \mathbb{E}_0\hat{V}_0 - \frac{1-\beta}{1-\beta\delta}(1-\tau)\frac{\sigma_z^2}{2} + (1-\tau)\frac{1-\beta}{1-\beta\delta\mu}\left(\frac{\ln(1-\tau)}{1+\eta} - \frac{\sigma_\phi^2+\sigma_z^2}{2}\right)\sum_{k=1}^{\infty}\delta^k\theta_k \\ &= -\ln A + \frac{1}{1+\eta}\left[\ln(1-\tau) - 1 + \tau\right] \\ &+ (1-\tau)\left(\frac{\ln(1-\tau)}{1+\eta} - \frac{\sigma_\phi^2+\sigma_z^2}{2}\right)\left[1 + \frac{1-\beta}{1-\beta\delta\mu}\sum_{k=1}^{\infty}\delta^k\theta_k\right],\end{aligned}$$

where the second equality uses the fact that $\frac{\beta(1-\delta)}{1-\beta\delta} + \frac{1-\beta}{1-\beta\delta} = 1$ and consolidates terms. Now write

$$\begin{aligned}1 + \frac{1-\beta}{1-\beta\delta\mu}\sum_{k=1}^{\infty}\delta^k\theta_k &= \sum_{k=1}^{\infty}\delta^k\theta_k + 1 + \left(\frac{1-\beta}{1-\beta\delta\mu} - 1\right)\sum_{k=1}^{\infty}\delta^k\theta_k \\ &= \sum_{k=1}^{\infty}\delta^k\theta_k + 1 - \frac{\beta(1-\delta\mu)}{1-\beta\delta\mu}\sum_{k=1}^{\infty}\delta^k\theta_k \\ &= \sum_{k=1}^{\infty}\delta^k\theta_k + 1 - \frac{\beta(1-\delta\mu)}{1-\beta\delta\mu}\frac{\delta\theta_1}{1-\delta\mu} \\ &= \sum_{k=1}^{\infty}\delta^k\theta_k + 1 - \frac{\beta\delta\theta_1}{1-\beta\delta\mu} \\ &= \sum_{k=1}^{\infty}\delta^k\theta_k + \theta_0,\end{aligned}$$

where the third equality sums the coefficients using $\theta_k = \theta_1\mu^{k-1}$, and the last equality uses the relationship between $\theta_0$ and $\theta_1$. Hence the welfare is

$$\mathcal{W} = -\ln A + \frac{1}{1+\eta}\left[\ln(1-\tau) - 1 + \tau\right] + (1-\tau)\left(\frac{\ln(1-\tau)}{1+\eta} - \frac{\sigma_\phi^2+\sigma_z^2}{2}\right)\sum_{k=0}^{\infty}\delta^k\theta_k,$$

which is identical to the corresponding expression in the sequence problem.

Transition function:

$$Q(\mathcal{X}, z'|x, z, \phi) = \begin{cases} f_\omega(\ln z' - \ln z) & \text{if } \left(\frac{1-\tau}{\phi}\right)^{\frac{\kappa}{1+\eta}} x^\mu e^{\kappa \ln z} \in \mathcal{X}. \\ 0 & \text{otherwise}\end{cases}$$

Hence $Q(\mathcal{X}, z'|x, z, \phi) = Q(e^{-\kappa\ln z}\mathcal{X}, \omega|x, 0, \phi)$.

The distribution function is

$$F(\mathcal{X}, z'|x, z, \phi) = Q(e^{-\psi \ln z}\mathcal{A}, e^{-\frac{\kappa}{1-\mu}\ln z}\mathcal{X}, \ln z' - \ln z | e^{-\psi \ln z} a, e^{-\frac{\kappa}{1-\mu}\ln z} x, 0, \phi).$$

The distribution satisfies

$$\begin{aligned}
&\mathcal{T}S(x', z', \phi) \\
&= \int_{x,z} Q(x', z'|x, z, \phi) S(x, z, \phi)\, dx\, dz \\
&= \int_{z} f(\ln z' - \ln z) S\left(\left(\frac{1-\tau}{\phi}\right)^{-\frac{\kappa}{\mu(1+\eta)}} e^{-\frac{\kappa}{\mu}\ln z} x', \ln z, \phi\right) dz \\
&= \int_{z} f(\ln z' - \ln z) S\left(\left(\frac{1-\tau}{\phi}\right)^{-\frac{\kappa}{\mu(1+\eta)}} e^{-\frac{\kappa}{\mu}\ln z} x', \ln z, \phi\right) dz.
\end{aligned}$$

# 12 Concluding Remarks

This paper studies the nature of history dependent income taxation, and the resulting welfare gains, in a parametric framework that is easy to analyze. The main finding is that the welfare gains are large, and that a substantial fraction of those welfare gains can be captured by allowing for only a limited history dependence, where only a small number of past incomes is included.

There are reasonable arguments why the welfare calculations in this paper might either understate or overstate the true welfare gains from history dependence. The welfare gains might be understated, because the functional forms considered in this paper are clearly restrictive. There is no reason to believe that geometric weighted average of past incomes is the best way to introduce history dependence into taxes. Heathcote and Tsujiyama (2019) show that, in a static framework, the constant-rate-of-progressivity generates about 80 percent of the welfare gains that can be obtained from the best possible Mirrleesian policy, but it is obviously hard to say whether similar conclusions hold here. Needless to say, computing the optimal Mirrleesean policy in the current framework, where the agents are heterogeneous along three dimensions (all of which are important), is a problem that is not likely to be fully solved in the near future.

The main reason why the welfare gains might be overstated is that agents are not allowed to save and smooth consumption on their own, since saving would potentially reduce the need for government provided insurance. While it is possible to support zero savings as an equilibrium outcome in the current framework[18], it is of course more interesting to consider scenarios where at least some agents can smooth consumption beyond what the government provides. I have computed the welfare gains from self-insurance in a simple Bewley-Ayiagari economy with liquid assets, where the steady-state ratio of mean liquid assets to mean earnings is calibrated to be 58.8 percent, as reported in Kaplan and Violante (2014). The welfare gains from self-insurance under the current U.S. tax code are 1.92 percent in consumption equivalents.[19] Thus, the welfare gains from history dependence would be reduced by about two thirds, still leaving a large welfare gain from history dependence of more than one percent. But the computed gains from self-insurance are likely to be substantially smaller than 1.92 percent. As reported by Kaplan and Violante (2014), about one third of U.S. households are hand-to-mouth agents with sizeable illiquid assets but zero liquid assets; those agents consume all their after-tax earnings just like the agents in this paper, but they are not replicated by a simple one-asset Bewley-Ayiagari model with lognormally distributed shocks. At the same time, there might be additional welfare gains under the history dependent tax system from Proposition 5, and even larger welfare gains from the history dependent tax system that would be optimal if the agents can save.

Where do the welfare gains from history dependence come from? A simplest case to see the welfare gains is one where only the persistent component is present. It is easy to verify that the stochastic process for the resulting consumption (21) is just another representation of the Inverse Euler equation: $c_{j-1} = (q/\beta)\mathbb{E}c_j$. The Inverse Euler equation is a hallmark of efficiency, and is obviously inconsistent with a standard Euler equation (Golosov et al. (2003), Farhi and Werning (2012a)). The case where only one component is present is a very special case in this respect. In general, if more than one shock is present, consumption produced by the optimal policy in this paper does not satisfy the Inverse Euler equation and, even though there are efficiency gains above self-insurance, not all potential welfare gains are exploited. The design of more sophisticated history dependent tax systems that exploit all the potential welfare gains is left for future research.

[18]Since the natural borrowing constraint is zero given that productivity is lognormally distributed, it is enough to assume that assets are in zero net supply.

[19]Given that the shocks are lognormally distributed, the natural borrowing constraint is zero. I have used Tauchen approximation to approximate the distribution of persistent, transitory and permanent shocks with, respectively, 51, 15 and 15 gridpoints. All agents are assumed to start with zero assets and save at an intertemporal price $q$. The desired quantity of liquid assets is achieved at $q = 0.988$.

# References


Albanesi, S. and C. Sleet (2006). Dynamic optimal taxation with private information. *The Review of Economic Studies 73*(1), 1–30. 5

Battaglini, M. and S. Coate (2008). Pareto efficient income taxation with stochastic abilities. *Journal of Public Economics 92*(3-4), 844–868. 5

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. 5, 7

Blomquist, S. and L. Micheletto (2003). Age related optimal income taxation. *Scandinavian Journal of Economics 110*, 45–71. 6

Conesa, J. C. and D. Krueger (2006). On the optimal progressivity of the tax code. *Journal of Monetary Economics 53*(7), 1425–1450. 5

Farhi, E. and I. Werning (2005). Inequality and social discounting. *Journal of Political Economy 115*(1), 365–402. 5

Farhi, E. and I. Werning (2012a). Capital taxation: Quantitative explorations of the inverse Euler equation. *Journal of Political Economy 120*(3), 398–445. 30, 36, 57

Farhi, E. and I. Werning (2012b). Insurance and taxation over the life-cycle. *Review of Economic Studies 80*(2), 596–635. 5

Golosov, M., N. R. Kocherlakota, and A. Tsyvinski (2003). Optimal indirect and capital taxation. *The Review of Economic Studies 70*(3), 569–587. 2, 5, 57

Golosov, M., M. Troshkin, and A. Tsyvinski (2016). Redistribution and social insurance. *American Economic Review 106*(2), 359–86. 5

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review 97*(3), 687–712. 4, 38, 40

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, 5, 24

Heathcote, J., K. Storesletten, and G. Violante (2016). Optimal tax progressivity: An analytical framework. Working paper, Federal Reserve Bank of Minneapolis. 5, 23, 29, 30

Heathcote, J., K. Storesletten, and G. Violante (2020). Optimal progressivity with age-dependent taxation. *Journal of Public Economics forthcoming*. 4, 6

Heathcote, J. and H. Tsujiyama (2019). Optimal income taxation: Mirrlees meets ramsey. Technical report, University of Minnesota. 56

Kapicka, M. (2019). History dependent income taxation in a dynastic economy. Working paper. 6

Kaplan, G. (2012). Inequality and the life cycle. *Quantitative Economics 3*(3), 471–525. 4, 23

Kaplan, G. and G. L. Violante (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica 82*(4), 1199–1239. 57

Kindermann, F. and D. Krueger (2017). High marginal tax rates on the top 1%? Technical report, University of Pennsylvania. 5

Kocherlakota, N. R. (2005). Zero expected wealth taxes: A Mirrless approach to dynamic optimal taxation. *Econometrica 73*(5), 1587–1621. 5

Kremer, M. (2002). Should taxes be independent of age,. Technical report. 6

Schmalbeck, R. (1984). Income averaging after twenty years: A failed experiment in horizontal equity. *Duke Law Journal*, 509 – 579. 2

Sleet, C. and H. Yazici (2017). Taxation, redistribution and frictional labor supply. Working paper. 7

Vickrey, W. (1939). Averaging of Income for Income-Tax Purposes. *Journal of Political Economy 47*, 379–379. 2

Vickrey, W. (1947). *Agenda for Progressive Taxation*. Columbia University theses. Ronald Press Company. 2, 19

Weinzierl, M. C. (2011). The surprising power of age-dependent taxes. *The Review of Economic Studies 78*(4), 1–29. 6

Werning, I. (2007). Optimal fiscal policy with redistribution. *Quarterly Journal of Economics 22*(3), 925967. 5

# Appendices

## A Proofs

The following lemma shows relationships that will be used in the proofs that follow. The proof of the lemma is straightforward and is omitted.

**Lemma 17.** *Suppose that a sequence of history dependence coefficients $\{\theta_k\}$ satisfies the incentive keeping constraint (7). Then, for any $\rho \in [0,1]$,*

$$\sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j} \rho^{j-k} \theta_k = \frac{1}{1-\beta\rho} \sum_{j=0}^{\infty} \beta^j \theta_j = \frac{1}{1-\beta\rho} \tag{37}$$

$$\sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j} \rho^{2(j-k)} \theta_k^2 = \frac{1}{1-\beta\rho^2} \sum_{j=0}^{\infty} \beta^j \theta_j^2 \tag{38}$$

$$\sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j} \sum_{l=0}^{k-1} \rho^{2j-k-l} \theta_k \theta_l = \frac{1}{1-\beta\rho^2} \sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j-1} \rho^{j-k} \theta_j \theta_k. \tag{39}$$

$$\sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j} (j+1-k) \theta_k = \frac{1}{(1-\beta)^2} \sum_{j=0}^{\infty} \beta^j \theta_j = \frac{1}{(1-\beta)^2} \tag{40}$$

$$\sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j} (j+1-k)^2 \theta_k^2 = \frac{1+\beta}{(1-\beta)^3} \sum_{j=0}^{\infty} \beta^j \theta_j^2 \tag{41}$$

$$\sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j} \sum_{l=0}^{k-1} (j+1-k)(j+1-l) \theta_k \theta_l = \frac{1+\beta}{(1-\beta)^3} \sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j-1} \theta_j \theta_k \left[1 + \frac{1-\beta}{1+\beta}(j-k)\right]. \tag{42}$$

*Proof of Proposition 1.* The government chooses the tax parameters $\{\lambda_j\}$, $\tau$ and $\{\theta_j\}$ to maximize $\mathcal{W}$ subject to $\mathcal{P} = 0$ and (7), taking (8) and (9) as given. Rewrite first the consumption function explicitly in terms of the underlying shocks:

$$\ln c_j = \ln \lambda_j + \frac{1-\tau}{1+\eta} \sum_{k=0}^{j} \theta_k \left[\ln(1-\tau) - \ln \phi\right] + (1-\tau) \sum_{k=0}^{j} \left[\theta_k \varepsilon_{j-k} + \theta_k \kappa + \left(\textstyle\sum_{l=0}^{k} \rho^{k-l} \theta_l\right) \omega_{j-k}\right].$$

Taking expectation of log consumption and of the period utility yields

$$\mathbb{E}_0(\ln c_j) = \ln \lambda_j + \frac{1-\tau}{1+\eta}\left[\ln(1-\tau) - \mathbb{E}\ln\phi\right]\sum_{k=0}^{j}\theta_k + (1-\tau)\sum_{k=0}^{j}\left(\theta_k\mathbb{E}\varepsilon + \theta_k\mathbb{E}\kappa + \frac{1-\rho^{j+1-k}}{1-\rho}\theta_k\mathbb{E}\omega\right)$$
$$\mathbb{E}_0(u_j) = \mathbb{E}_0(\ln c_j) - \frac{1-\tau}{1+\eta}.$$

Substitute $\mathbb{E}_0(u_j)$ into the objective function (1). After some algebra, the expression for the objective function can be written as

$$\mathcal{W} = (1-\beta)\sum\nolimits_{j=0}^{\infty}\beta^j\ln\lambda_j - \frac{1-\tau}{1+\eta} + (1-\tau)\left[\frac{\ln(1-\tau) - \mathbb{E}\ln\phi}{1+\eta} + \mathbb{E}\varepsilon + \frac{\mathbb{E}\omega}{1-\beta\rho} + \mathbb{E}\kappa\right], \tag{43}$$

where (37) was used to simplify the present values. The resource constraint is simplified similarly. The expected values of period consumption, and production are

$$\mathbb{E}_0(c_j) = \lambda_j e^{\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}\prod\nolimits_{k=0}^{j}\left(B_{\omega k}B_{\varepsilon k}\right)B_{\phi j}B_{\kappa j}$$
$$\mathbb{E}_0(y_j) = e^{\left(\frac{1-\rho^{2j+2}}{1-\rho^2} - \frac{1-\rho^{j+1}}{1-\rho}\right)\frac{\sigma_\omega^2}{2}}(1-\tau)^{\frac{1}{1+\eta}},$$

where $B_{\omega k}$, $B_{\varepsilon k}$, $B_{\phi j}$ and $B_{\kappa j}$ are as defined in in the text. Substituting $\mathbb{E}_0(c_j)$ and $\mathbb{E}_0(y_j)$ into the resource constraint (10) yields

$$\nu(1-\tau)^{1+\eta} = (1-q)\sum\nolimits_{j=0}^{\infty}q^j\lambda_j e^{\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}\prod\nolimits_{k=0}^{j}\left(B_{\omega k}B_{\varepsilon k}\right)B_{\phi j}B_{\kappa j} + G,$$

where $\nu = \exp(-\frac{(1-\rho)q\rho}{(1-q\rho)(1-q\rho^2)}\frac{\sigma_\omega^2}{2})$ is the present value of the persistent component. Let $\zeta$ be the Lagrange multiplier on the resource constraint. The first-order conditions in $\lambda_j$ yield

$$\frac{1}{\lambda_j} = \zeta\frac{1-q}{1-\beta}\left(\frac{q}{\beta}\right)^j e^{\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}\prod\nolimits_{k=0}^{j}\left(B_{\omega k}B_{\varepsilon k}\right)B_{\phi j}B_{\kappa j}.$$

Eliminating $\lambda_j$ from the resource constraint yields the value of the Lagrange multiplier $\zeta^{-1} = \nu\,(1-\tau)^{\frac{1}{1+\eta}} - G$. Rearranging implies (12).

The aggregate consumption in period $j$ must then satisfy

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

□

*Proof of Proposition 2.* With productivity shocks being lognormally distributed, the aggregate moments simplify to

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

Furthermore, $\mathbb{E}\omega = -\frac{\sigma_\omega^2}{2}$, $\mathbb{E}\varepsilon = -\frac{\sigma_\varepsilon^2}{2}$, $\mathbb{E}\kappa = -\frac{\sigma_\kappa^2}{2}$ and $\mathbb{E}\ln\phi = (1+\eta)\frac{\sigma_\phi^2}{2}$. Substituting those expressions into (13), simplifying the expressions by using (37)-(39) and cancelling terms yields (14).

□

*Proof of Proposition 5.* Since $s_\phi$ and $s_\kappa$ enter symmetrically, assume that $s_\kappa = 0$. Rearranging equation (18) shows that it is a fourth-order homogeneous linear difference equation

$$p_1\theta_{j+2} + p_2\theta_{j+1} + p_3\theta_j + p_4\theta_{j-1} + p_5\theta_{j-2} = 0,$$

where

$$\begin{aligned}
p_1 &= -\beta^2\rho s_\varepsilon \\
p_2 &= \beta(1-\beta\rho^2)s_\omega + \beta\left[(1+\beta)\rho + 1 + \beta\rho^2\right]s_\epsilon + \beta(1-\beta)\rho s_\phi \\
p_3 &= -(1-\beta\rho^2)(1+\beta)s_\omega - (1+\beta\rho)(1+\beta+\beta\rho)s_\varepsilon - (1-\beta)(1+\beta\rho^2)s_\phi \\
p_4 &= (1-\beta\rho^2)s_\omega + (1+\rho)(1+\beta\rho)s_\varepsilon + \rho(1-\beta)s_\phi \\
p_5 &= -\rho s_\varepsilon.
\end{aligned}$$

The characteristic equation is a quartic equation, and it can be written as (19). It has four roots, $\mu_1 \in (0, \rho)$, $\mu_2 \in (\rho, 1)$, $\mu_3 \in (\beta^{-1}, \beta^{-1}\rho^{-1})$ and $\mu_4 \in (\beta^{-1}\rho^{-1}, \infty)$. The last two roots are greater than $\beta^{-1}$ and so must have zero weight in the optimal solution, otherwise the incentive keeping constraint (7) cannot hold. So

$$\theta_j = c_1 \mu_1^j + c_2 \mu_2^j, \quad j > 0. \tag{44}$$

for some $c_1$ and $c_2$. To find $c_1$ and $c_2$, as well as $\theta_0$, substitute (44) into (18) for $j > 0$ and rearrange terms to write it as

$$\theta_0 s_\phi + \frac{c_1}{1-\mu_1} s_\phi + \frac{c_2}{1-\mu_1} s_\phi + \rho^j \left( \theta_0 + \frac{c_1}{\rho - \mu_1} + \frac{c_2}{\rho - \mu_2} \right) s_\omega = \zeta. \tag{45}$$

In order to hold for all $j$, the last term on the right-hand side must be zero, and so

$$\theta_0 + \frac{c_1}{\rho - \mu_1} + \frac{c_2}{\rho - \mu_2} = 0.$$

Eliminating $\zeta$ from (45) by using (18) for $j = 0$ and rearranging gives a second condition:

$$\theta_0 = \frac{c_1}{\mu_1} + \frac{c_2}{\mu_2}.$$

The final condition comes from the incentive keeping constraint(7). Using (44) yields

$$\theta_0 + \frac{\beta c_1}{1 - \beta\mu_1} + \frac{\beta c_2}{1 - \beta\mu_2} = 1.$$

Solving the last three equations for $c_1$ and $c_2$ and $\theta_0$ yields the result in the Proposition. □

*Proof of Proposition 10.* Let $\tilde{s}_\phi \geq s_\phi$. Let also $\theta^*$ and $\tilde{\theta}^*$ be the corresponding history dependence coefficients, and $P(\theta)$ and $\tilde{P}(\theta)$ be the corresponding risk loading functions. Finally, let $\varsigma = s_\omega / s_\varepsilon$ and $\tilde{\varsigma} = \tilde{s}_\omega / \tilde{s}_\varepsilon$. We have $\tilde{\varsigma} \geq \varsigma$. We have

$$\begin{aligned} \tilde{P}(\tilde{\theta}^*) &= (1 - \tilde{s}_\phi) \frac{P_0(\tilde{\theta}^*) + \tilde{\varsigma} P_\rho(\tilde{\theta}^*)}{1 + \tilde{\varsigma}} + \tilde{s}_\phi P_1(\tilde{\theta}^*) \\ &\geq (1 - s_\phi) \frac{P_0(\tilde{\theta}^*) + \varsigma P_\rho(\tilde{\theta}^*)}{1 + \varsigma} + s_\phi P_1(\tilde{\theta}^*) \geq P(\theta^*). \end{aligned}$$

The first equality is by definition of $\tilde{P}$. The second one follows from the fact that $P_0(\theta) < P_\rho(\theta) < P_1(\theta)$ for any strictly positive $\theta$. The last inequality follows from the fact that $\tilde{\theta}^*$ may not minimize $P(\theta)$, but $\theta^*$ does. The proof of the second part is analogous. $\square$

To prove Proposition 12, the following Lemma will be needed.

**Lemma 18.** *Let*

$$\Gamma = (1-q)\sum_{i=0}^{\infty} q^i \prod_{k=0}^{i} e^{-f_k}$$

*for some* $\{f_i\}$, $f_i \in \mathcal{R}$, $i \geq 0$. *Then* $\Gamma \approx e^{-\sum_{i=0}^{\infty} q^i f_i}$.

*Proof.* First truncate the infinite sum at some finite length $K$, and define $\Gamma_K = (1-q)\sum_{i=0}^{K} q^i \prod_{k=0}^{i} e^{-f_k}$. We can write it as $\Gamma_K = (1-q)e^{-f_0}A_1$, where $A_1$ is defined by a recursive relation

$$A_i = 1 + qe^{-f_i}A_{i+1}, \quad i = 1, 2 \ldots, K-1,$$

with terminal value $A_K = 1 + qe^{-f_K}$. The terminal value can be approximated by

$$A_K \approx 1 + q(1 - f_K) = (1+q)\left(1 - \frac{q}{1+q} f_K\right) \approx (1+q)e^{-\frac{q}{1+q} f_K},$$

where 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 assume that $A_{i+1}$ satisfies

$$A_{i+1} \approx \left(\sum_{k=0}^{K-i} q^k\right) e^{-\frac{q\sum_{k=0}^{K-i-1} q^k}{\sum_{k=0}^{K-i} q^k} f_{i+1} - \frac{q^2\sum_{k=0}^{K-i-2} q^k}{\sum_{k=0}^{K-i} q^k} f_{i+2} \cdots - \frac{q^{K-1}}{\sum_{k=0}^{K-i} q^k} f_K}$$

for some $i < K$. We have shown that $A_K$ takes this form. Then

$$\begin{aligned}
A_i &\approx 1 + q\left(\sum_{k=0}^{K-i} q^k\right) e^{-f_i - q\frac{\sum_{k=0}^{K-i-1} q^k}{\sum_{k=0}^{K-i} q^k} f_{i+1} - q^2\frac{\sum_{k=0}^{K-i-2} q^k}{\sum_{k=0}^{K-i} q^k} f_{i+2} \cdots - \frac{q^{K-1}}{\sum_{k=0}^{K-i} q^k} f_K} \\
&\approx 1 + q\left(\sum_{k=0}^{K-i} q^k\right)\left(1 - f_i - q\frac{\sum_{k=0}^{K-i-1} q^k}{\sum_{k=0}^{K-i} q^k} f_{i+1} - q^2\frac{\sum_{k=0}^{K-i-2} q^k}{\sum_{k=0}^{K-i} q^k} f_{i+2} \ldots - \frac{q^{K-1}}{\sum_{k=0}^{K-i} q^k} f_K\right) \\
&= \left(\sum_{k=0}^{K+1-i} q^k\right)\left(1 - q\frac{\sum_{k=0}^{K-i} q^k}{\sum_{k=0}^{K+1-i} q^k} f_i - q^2\frac{\sum_{k=0}^{K-i-1} q^k}{\sum_{k=0}^{K+1-i} q^k} f_{i+1} \ldots - \frac{q^K}{\sum_{k=0}^{K+1-i} q^k} f_K\right) \\
&\approx \left(\sum_{k=0}^{K+1-i} q^k\right) e^{-q\frac{\sum_{i=0}^{K-i} q^k}{\sum_{k=0}^{K+1-i} q^k} f_i - q^2\frac{\sum_{k=0}^{K-i-1} q^k}{\sum_{k=0}^{K+1-i} q^k} f_{i+1} \cdots - \frac{q^K}{\sum_{k=0}^{K+1-i} q^k} f_K},
\end{aligned}$$

where we have used the same approximations as in the first step. Thus, $A_i$ satisfies the functional form as well, and so

$$\Gamma_K \approx (1-q)\left(\sum_{k=0}^{K+1-i} q^k\right) e^{-f_0 - q\frac{\sum_{i=0}^{K-i} q^k}{\sum_{k=0}^{K-i} q^k} f_i - q^2 \frac{\sum_{k=0}^{K-i-1} q^k}{\sum_{k=0}^{K-i} q^k} f_{i+1} \cdots - \frac{q^K}{\sum_{k=0}^{K-i} q^k} f_K}.$$

Taking the limit as $K$ goes to infinity yields the expression in the Lemma. □

*Proof of Proposition 12.* The proof shows that the welfare function is approximately equal to (14). The welfare function can be written analogously to (43),

$$\mathcal{W} = \ln\lambda - \frac{1-\tau}{1+\eta} + (1-\tau)\frac{\ln(1-\tau)}{1+\eta} - \frac{1-\tau}{2}\left(\frac{\sigma_\omega^2}{1-\beta\rho} + \sigma_\varepsilon^2 + \sigma_\phi^2 + \sigma_\kappa^2\right), \quad (46)$$

where the (age independent) tax parameter $\lambda$ is obtained from the resource constraint (10) and is equal to

$$\lambda = \frac{\nu(1-\tau)^{\frac{1}{1+\eta}} - G}{\Gamma}, \quad (47)$$

where $\Gamma$ satisfies

$$\Gamma = (1-q)\sum_{j=0}^{\infty} q^j e^{\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k} \prod_{k=0}^{j}\left(B_{\omega k}B_{\varepsilon k}\right) B_{\phi j}.$$

Note that the aggregate moments can be written as

$$B_{\omega j} = e^{-\frac{1}{2}m_j(1-m_j)\sigma_\omega^2}, \qquad m_j = (1-\tau)\sum_{k=0}^{j}\rho^{j-k}\theta_k$$

$$B_{\varepsilon j} = e^{-\frac{1}{2}p_j(1-p_j)\sigma_\varepsilon^2}, \qquad p_j = (1-\tau)\theta_j$$

$$B_{\phi j} = e^{-\sum_{k=0}^{j} n_k \frac{\sigma_\phi^2}{2}}, \qquad n_j = (1-\tau)\theta_j - (1-\tau)^2\theta_j^2 - 2(1-\tau)^2\sum_{k=0}^{j-1}\theta_k\theta_j.$$

I will approximate $\Gamma$ by using Lemma 18. Write $\Gamma = (1-q)\sum_{j=0}^{\infty} q^j \prod_{k=0}^{j} e^{-f_k}$, where

$$f_j = -\frac{1-\tau}{1+\eta}\ln(1-\tau)\theta_j + n_j\frac{\sigma_\phi^2}{2} + m_j(1-m_j)\frac{\sigma_\omega^2}{2} + p_j(1-p_j)\frac{\sigma_\varepsilon^2}{2}.$$

Applying Lemma 18, it can be approximated by $\Gamma \approx e^{-\sum_{j=0}^{\infty} q^j f_j}$. That is,

$$\begin{aligned}\ln \Gamma &\approx \textstyle\sum_{j=0}^{\infty} q^j \left[ \frac{1-\tau}{1+\eta} \ln(1-\tau)\theta_j - n_j \frac{\sigma_\phi^2}{2} - m_j(1-m_j)\frac{\sigma_\omega^2}{2} - p_j(1-p_j)\frac{\sigma_\varepsilon^2}{2} \right] \\ &= \frac{1-\tau}{1+\eta} \ln(1-\tau) \textstyle\sum_{j=0}^{\infty} \beta^j \theta_j - \frac{1-\tau}{2} \left( \frac{\sigma_\omega^2}{1-q\rho} + \sigma_\varepsilon^2 + \sigma_\phi^2 \right) + \frac{(1-\tau)^2}{2} P(\theta)\sigma^2\end{aligned}$$

where the last equality uses using the fact that $P_\rho(\theta) = (1-\beta\rho^2) \sum_{j=0}^{\infty} \beta^j \left( \sum_{k=0}^{j} \rho^{j-k}\theta_k \right)^2$, and the assumption that $q = \beta$. Using (47), the approximation for $\ln \Gamma$ and cancelling terms in (46) yields the result. □

*Proof of Proposition 13.* The stochastic process for consumption is now given by

$$\begin{aligned}\ln c_j = \ln \lambda_j &+ \frac{1-\tau}{1+\eta} \sum_{k=0}^{j} \theta_k \left[ \ln(1-\tau) - \ln \phi \right] \\ &+ (1-\tau) \sum_{k=0}^{j} \left[ \theta_k \varepsilon_{j-k} + \theta_k \kappa + \left( \textstyle\sum_{l=0}^{k} \rho^{k-l} \theta_l \right) \omega_{j-k} + (j+1-k)\theta_k \gamma \right].\end{aligned}$$

Taking expectations and solving for $\{\lambda_j\}$ yields the following expression for the social welfare function:

$$\begin{aligned}\mathcal{W}(\tau,\theta) = \bar{u}(\tau) &+ (1-\tau) \left( \frac{\mathbb{E}\omega}{1-\beta\rho} + \mathbb{E}\varepsilon + \mathbb{E}\kappa - \frac{\mathbb{E}\ln\phi}{1+\eta} + \frac{\mathbb{E}\gamma}{1-\beta} \right) \\ &- \sum_{j=0}^{\infty} \beta^j \left[ \ln B_{\omega j} + \ln B_{\varepsilon j} + (1-\beta)(\ln B_{\kappa j} + \ln B_{\phi j} + \ln B_{\gamma_j}) \right]. \qquad (48)\end{aligned}$$

where the expression for $\mathbb{E}\gamma$ on the right-hand side uses the result that

$$(1-\beta) \textstyle\sum_{j=0}^{\infty} \beta^j \sum_{k=0}^{j} (j+1-k)\theta_k = \frac{1}{1-\beta} \sum_{j=0}^{\infty} \beta^j \theta_j.$$

The new term $B_{\gamma_j}$ satisfies $B_{\gamma_j} = \mathbb{E} e^{(1-\tau)\sum_{k=0}^{j} \theta_k (j+1-k)\gamma}$. Using the fact that $\gamma$ is normally distributed, we write

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

Substituting into (48), simplifying the expressions by using (37)-(42) and cancelling terms yields (14).

□

*Proof of Proposition 14.* Let $\zeta$ be the Lagrange multiplier on the incentive keeping constraint. The first-order condition in $\theta_j$ is

$$
\sum_{k=0}^{j} \theta_k \left( 1 + \frac{1-\beta}{1+\beta} (j-k) \right) + \sum_{k=j+1}^{\infty} \theta_k \beta^{k-j} \left( 1 + \frac{1-\beta}{1+\beta} (k-j) \right) = \zeta, \quad j \geq 0. \tag{49}
$$

Twice differencing the first-order condition yields

$$
0 = \frac{1+\beta}{1-\beta} \theta_j + \sum_{k=j+1}^{\infty} \theta_k \beta^{k-j} (k-j+1), \quad j \geq 2.
$$

This is a set of equations in $\theta_j$ for $j \geq 2$, and is solved by $\theta_j = 0$ for all $j \geq 2$. Evaluating the first-order condition (49) at $j = 0$ and $j = 1$, we obtain

$$
\theta_0 + \frac{2\beta}{1+\beta} \theta_1 = \frac{2}{1+\beta} \theta_0 + \theta_1 = \zeta.
$$

combining both equations yields $\theta_0 + \theta_1 = 0$. The incentive keeping constraint then implies that $\theta_0 = 1/(1-\beta)$ and $\theta_1 = -1/(1-\beta)$. The minimized value of $\hat{P}$ can be easily computed from the optimal history dependence parameters. □

# Appendix NP1

## Deriving the Aggregates when $\lambda$ is Age Dependent - more details

**Policy functions.** Taking the first-order conditions to the agent's problem yields the optimal hours worked and consumption

$$\begin{aligned}
\ln h &= \frac{1}{1+\eta}\left[\ln(1-\tau)-\ln\phi\right] \\
\ln c_j &= \ln\lambda_j + \frac{1-\tau}{1+\eta}\left[\ln(1-\tau)-\ln\phi\right]\sum_{k=0}^{j}\theta_k + (1-\tau)\sum_{k=0}^{j}\theta_k \ln w_{j-k} \\
&= \ln\lambda_j + \frac{1-\tau}{1+\eta}\left[\ln(1-\tau)-\ln\phi\right]\sum_{k=0}^{j}\theta_k + (1-\tau)\sum_{k=0}^{j}\theta_k\left(\kappa + z_{j-k} + \varepsilon_{j-k}\right) \\
&= \ln\lambda_j + \frac{1-\tau}{1+\eta}\left[\ln(1-\tau)-\ln\phi\right]\sum_{k=0}^{j}\theta_k + (1-\tau)\left[\sum_{k=0}^{j}(\kappa+\epsilon_{j-k}) + \sum_{k=0}^{j}\left(\sum_{l=0}^{k}\rho^{k-l}\theta_l\right)\omega_{j-k}\right].
\end{aligned}$$

The expected values are

$$\begin{aligned}
\mathbb{E}_0(\ln c_j) &= \ln\lambda_j + \frac{1-\tau}{1+\eta}\left[\ln(1-\tau)-\mathbb{E}\ln\phi\right]\sum_{k=0}^{j}\theta_k + (1-\tau)\sum_{k=0}^{j}\left(\theta_k\mathbb{E}\varepsilon + \theta_k\mathbb{E}\kappa + \frac{1-\rho^{j+1-k}}{1-\rho}\theta_k\mathbb{E}\omega\right) \\
\mathbb{E}_0(u_j) &= \mathbb{E}_0(\ln c_j) - \frac{1-\tau}{1+\eta} \\
\mathbb{E}_0(c_j) &= \lambda_j e^{\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}\prod_{k=0}^{j}\left(B_{\omega k}B_{\varepsilon k}\right)B_{\phi j}B_{\kappa j} \\
\mathbb{E}_0(y_j) &= e^{\left(\frac{1-\rho^{2j+2}}{1-\rho^2}-\frac{1-\rho^{j+1}}{1-\rho}\right)\frac{\sigma_\omega^2}{2}}(1-\tau)^{\frac{1}{1+\eta}},
\end{aligned}$$

**Aggregate welfare.** The government maximizes

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

or,

$$\mathcal{W} = (1-\beta)\sum_{j=0}^{J}\beta^j\ln\lambda_j - (1-\beta^{J+1})\frac{1-\tau}{1+\eta} + (1-\tau)\left[\frac{\ln(1-\tau)-\mathbb{E}\ln\phi}{1+\eta} + \mathbb{E}\varepsilon + \frac{\mathbb{E}\omega}{1-\beta\rho} + \mathbb{E}\kappa\right]$$

subject to the resource constraint

$$(1-q)\sum_{j=0}^{J} q^j \left[\mathbb{E}_0 c_j - \mathbb{E}_0 y_j\right] + G = 0. \tag{50}$$

Substituting $\mathbb{E}_0(c_j)$ and $\mathbb{E}_0(y_j)$ into the resource constraint (10) yields

$$\nu(1-\tau)^{1+\eta} = (1-q)\sum\nolimits_{j=0}^{\infty} q^j \lambda_j e^{\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}\Psi_j + G,$$

where $\nu = \exp(-\frac{(1-\rho)q\rho}{(1-q\rho)(1-q\rho^2)}\frac{\sigma_\omega^2}{2})$ is the present value of the persistent component and

$$\Psi_j = \prod\nolimits_{k=0}^{j}\left(B_{\omega k}B_{\varepsilon k}\right)B_{\phi j}B_{\kappa j}.$$

Let $\zeta$ be the Lagrange multiplier on the resource constraint. The first-order condition in $\lambda_j$ yields

$$\frac{1}{\lambda_j} = \zeta\frac{1-q}{1-\beta}\left(\frac{q}{\beta}\right)^j e^{\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}\Psi_j.$$

Eliminating $\lambda_j$ from the resource constraint yields the value of the Lagrange multiplier $\zeta^{-1} = \nu\left(1-\tau\right)^{\frac{1}{1+\eta}} - G$. Rearranging implies (12). The Lagrange multiplier is

$$\lambda_j = \frac{1-\beta}{1-q}\left(\frac{\beta}{q}\right)^j \frac{e^{-\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}}{\Psi_j}\left[\nu\left(1-\tau\right)^{\frac{1}{1+\eta}} - G\right]$$

The lifetime utility now becomes

$$\begin{aligned}\mathcal{W} = \bar{u}(\tau) + (1-\tau)\left(\frac{\mathbb{E}\omega}{1-\beta\rho} + \mathbb{E}\varepsilon + \mathbb{E}\kappa - \frac{\mathbb{E}\ln\phi}{1+\eta}\right)\\ - \sum_{j=0}^{\infty}\beta^j\left[\ln B_{\omega j} + \ln B_{\varepsilon j} + (1-\beta)(\ln B_{\kappa j} + \ln B_{\phi j})\right].\end{aligned} \tag{51}$$

with

$$\overline{u}(\tau) = \ln\left[\nu(1-\tau)^{\frac{1}{1+\eta}} - G\right] - \frac{1-\tau}{1+\eta} + \ln\left(\frac{1-\beta}{1-q}\right) + \frac{\beta}{1-\beta}\ln\left(\frac{\beta}{q}\right).$$

## 12.1 Lognormal distribution.

With productivity shocks being lognormally distributed, we have

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

Furthermore, $\mathbb{E}\omega = -\frac{\sigma_\omega^2}{2}$, $\mathbb{E}\varepsilon = -\frac{\sigma_\varepsilon^2}{2}$, $\mathbb{E}\kappa = -\frac{\sigma_\kappa^2}{2}$ and $\mathbb{E}\ln\phi = (1+\eta)\frac{\sigma_\phi^2}{2}$.

The welfare is

$$\mathcal{W} = \overline{u}(\tau) - \frac{1}{2}(1-\tau)^2\left[s_\omega P_\rho(\theta) + s_\varepsilon P_0(\theta) + (s_\kappa + s_\phi)P_1(\theta)\right]\sigma^2. \quad (52)$$

where

$$s_\omega = \frac{1}{1-\beta\rho^2}\frac{\sigma_\omega^2}{\sigma^2}, \quad s_\varepsilon = \frac{\sigma_\varepsilon^2}{\sigma^2}, \quad s_\kappa = \frac{\sigma_\kappa^2}{\sigma^2}, \quad s_\phi = \frac{\sigma_\phi^2}{\sigma^2}.$$

**History Independence.** When the tax is history independent, we have

$$\ln B_{\omega j} = -\left[(1-\tau)\rho^j\theta_0 - (1-\tau)^2\rho^{2j}\theta_0^2\right]\frac{\sigma_\omega^2}{2}$$
$$\ln B_{\varepsilon 0} = -\tau(1-\tau)\frac{\sigma_\varepsilon^2}{2}$$
$$\ln B_{\varepsilon j} = 0, \quad j > 0$$
$$\ln B_{\kappa j} = -\tau(1-\tau)\frac{\sigma_\kappa^2}{2}$$
$$\ln B_{\phi j} = -\tau(1-\tau)\frac{\sigma_\phi^2}{2}.$$

Thus, we write

$$\Psi_j = e^{-\sum_{k=0}^{j}\left[(1-\tau)\rho^k-(1-\tau)^2\rho^{2k}\right]\frac{\sigma_\omega^2}{2}-\tau(1-\tau)\left(\frac{\sigma_\varepsilon^2}{2}+\frac{\sigma_\kappa^2}{2}+\frac{\sigma_\phi^2}{2}\right)}$$

$$= e^{-\left[(1-\tau)\frac{1-\rho^{j+1}}{1-\rho}-(1-\tau)^2\frac{1-\rho^{2(j+1)}}{1-\rho^2}\right]\frac{\sigma_\omega^2}{2}-\tau(1-\tau)\left(\frac{\sigma_\varepsilon^2}{2}+\frac{\sigma_\kappa^2}{2}+\frac{\sigma_\phi^2}{2}\right)}$$

and the welfare as

$$\mathcal{W} = \overline{u}(\tau) - \frac{1}{2}(1-\tau)^2\left[\frac{1}{1-\beta\rho^2}\sigma_\omega^2 + \sigma_\varepsilon^2 + \sigma_\kappa^2 + \sigma_\phi^2\right]. \tag{53}$$

# Appendix NP2: Deriving the Aggregates when $\lambda$ is independent of age - more details.

We obtain $\lambda$ directly from the resource constraint:

$$\lambda = \frac{\nu(1-\tau)^{1+\eta} - G}{(1-q)\sum_{j=0}^{\infty} q^j e^{\frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{j}\theta_k}\Psi_j}$$

For history independence this reduces to

$$\lambda = \frac{\nu(1-\tau)^{1+\eta} - G}{(1-\tau)^{\frac{1-\tau}{1+\eta}}(1-q)\sum_{j=0}^{\infty} q^j \Psi_j}$$

Welfare is

$$\mathcal{W} = \ln\lambda - \frac{1-\tau}{1+\eta} + (1-\tau)\left[\frac{\ln(1-\tau) - \mathbb{E}\ln\phi}{1+\eta} + \mathbb{E}\varepsilon + \frac{\mathbb{E}\omega}{1-\beta\rho} + \mathbb{E}\kappa\right]$$

**Policy functions.** Taking the first-order conditions to the agent's problem yields the optimal hours worked and consumption

$$\begin{aligned}
\ln h_j &= \frac{1}{1+\eta}\left[\ln(1-\tau) - \ln\phi\right] \\
\ln c_j &= \ln\lambda + \frac{1-\tau}{1+\eta}\sum_{k=0}^{j}\theta_k\left[\ln(1-\tau)-\ln\phi\right] + (1-\tau)\sum_{k=0}^{j}\theta_k \ln z_{j-k} \\
&= \ln\lambda + \frac{1-\tau}{1+\eta}\sum_{k=0}^{j}\theta_k\left[\ln(1-\tau)-\ln\phi\right] + (1-\tau)\sum_{k=0}^{j-1}\left(\textstyle\sum_{l=0}^{k}\theta_l\right)\omega_{j-k}.
\end{aligned}$$

The expected values are

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

The corresponding residual variance is

$$\mathrm{Var}\left(\ln c_j^*\right) = (1-\tau)^2\textstyle\sum_{k=1}^{j}\theta_{j-k}^2\sigma_\phi^2 + (1-\tau)^2\textstyle\sum_{k=1}^{j}\left(\textstyle\sum_{l=0}^{j-k}\theta_l\right)^2\sigma_\omega^2.$$

**Aggregate welfare.** To obtain the aggregate welfare, aggregate over all ages into (**??**):

$$\begin{aligned}
\mathcal{W} &= (1-\delta\mu)\sum_{j=0}^{\infty}(\delta\mu)^j\,\mathbb{E}_0(u_j) \\
&= \ln\lambda - \frac{1-\tau}{1+\eta} + (1-\tau)(1-\delta\mu)\sum_{j=0}^{\infty}(\delta\mu)^j\left[\frac{\ln(1-\tau)-\mathbb{E}\ln\phi}{1+\eta}\sum_{k=0}^{j}\theta_k + \mathbb{E}\omega\sum_{k=0}^{j}(j-k)\theta_k\right] \\
&= \ln\lambda - \frac{1-\tau}{1+\eta} + \frac{1-\tau}{1+\eta}\left[\ln(1-\tau)-\mathbb{E}\ln\phi\right](1-\delta\mu)\sum_{j=0}^{\infty}(\delta\mu)^j\sum_{k=0}^{j}\theta_k \\
&+ (1-\tau)\mathbb{E}\omega\,(1-\delta\mu)\sum_{j=0}^{\infty}(\delta\mu)^j\sum_{k=0}^{j}(j-k)\theta_k \\
&= \ln\lambda - \frac{1-\tau}{1+\eta} + (1-\tau)\left[\frac{\ln(1-\tau)}{1+\eta} - \frac{1}{1+\eta}\mathbb{E}\ln\phi + \frac{\mu\delta}{1-\mu\delta}\mathbb{E}\omega\right]\sum_{k=0}^{\infty}(\mu\delta)^k\theta_k,
\end{aligned}$$

where the last equality uses the fact that $(1-\mu\delta)\sum_{j=0}^{\infty}(\mu\delta)^j\sum_{k=0}^{j}\theta_k = \sum_{j=0}^{\infty}(\mu\delta)^j\theta_j$ and that $(1-\mu\delta)\sum_{j=0}^{\infty}(\mu\delta)^j\sum_{k=0}^{j}(j-k)\theta_k = \mu\delta/(1-\mu\delta)\sum_{j=0}^{\infty}(\mu\delta)^j\theta_j$.

**Aggregate cost.** The expected value of period $j$ costs is

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

where the second term on the right-hand side uses $\mathbb{E}\left(e^{-\frac{1}{1+\eta}\ln\phi}\right) = 1$. Aggregating over all ages, one obtains that the aggregate resource constraint is

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

which is to be solved for $\lambda$:

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

which is repeated here from the main text only for convenience.

## 12.2 Lognormal distribution.

With productivity shocks being lognormally distributed , the function $\lambda^*$ is approximated by see appendix NP2)

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

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

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

## 12.3 Special Case: $\mu = \beta$

Suppose that the government discounts at a rate that equals to the agent's discount rate. Then the objective function becomes

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

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

## 12.4 Special Case: $\mu = q$

If the government discounts both utility and resources at the same rate,

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

### 12.5 Special Case: Benchmark General Equilibrium

In the general equilibrium version of the model we set $\mu = 1$ and $q = \delta$. The last two terms drop out and the objective function becomes

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

where $\sigma_z^2 = \frac{\delta}{1-\delta}\sigma_\omega^2$.

# Appendix NP3: Approximating the function $\lambda^*$

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

$$B_{\omega j} \equiv B_\omega\left[(1-\tau)\sum_{k=0}^{j} \theta_k\right] = e^{-\sum_{k=0}^{j}\left[(1-\tau)\theta_k - (1-\tau)^2\theta_k^2 - 2(1-\tau)^2\sum_{l=0}^{k-1}\theta_l\theta_k\right]\frac{\sigma_\omega^2}{2}} = e^{-\frac{1}{2}m_j(1-m_j)\sigma_\omega^2}$$

$$B_{\phi j} \equiv B_\phi\left[-\frac{1-\tau}{1+\eta}\sum_{k=0}^{j} \theta_k\right] = e^{-\sum_{k=0}^{j}\left[(1-\tau)\theta_k - (1-\tau)^2\theta_k^2 - 2(1-\tau)^2\sum_{l=0}^{k-1}\theta_l\theta_k\right]\frac{\sigma_\phi^2}{2}} = e^{-\sum_{k=0}^{j} n_k \frac{\sigma_\phi^2}{2}}$$

$$\bar{B}_j \equiv e^{\sum_{k=0}^{j}\frac{1-\tau}{1+\eta}\ln(1-\tau)\theta_k} = e^{\sum_{k=0}^{j} o_k},$$

and

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

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

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

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

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

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

$$A_k \approx e^{-\frac{\delta}{1-\delta}\left(m_k - m_k^2 + \sum_{l=1}^{K-k} \delta^l n_{k+l}\right) \frac{\sigma_\omega^2}{2} - \sum_{l=1}^{K-k} \delta^l n_{k+l} \frac{\sigma_\phi^2}{2} + \sum_{l=1}^{K-k} \delta^l o_{k+l}}. \tag{54}$$

Clearly, $A_K$ takes this form. Write

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

Thus, $A_{k-1}$ takes the form in (54) 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 and substituting in for $n_k$ and $o_k$ yields the desired result.

## 12.6 Accuracy of the Approximation

The approximate solution, as characterized by the welfare function (**??**) and by Proposition **??** 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 (**??**) 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 acheved 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 8 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 $\hat{\sigma}^2$.[20]

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

[20] To simplify exposition I only plot coefficients $\theta_k$ for $k = 1, \ldots K-1$ and do not show $\theta_0$ and $\theta_K$ that have different magnitudes.

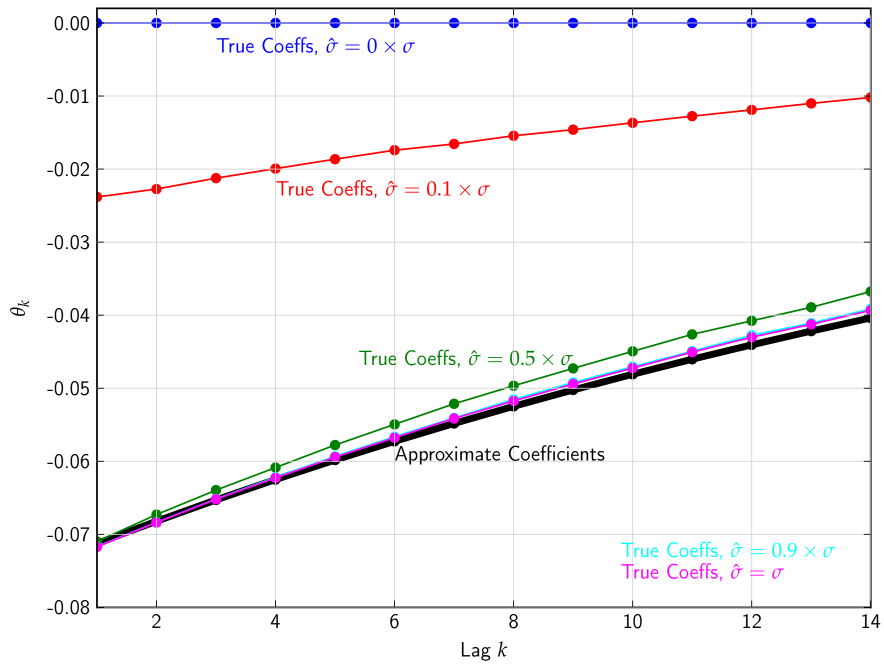


Figure 8: Accuracy of the Approximate Solution for $K = 15$, $\tau = 0.238$ and $\sigma^2 = \sigma_\phi^2 + \sigma_z^2 = 0.198$.

true solution. For realistic values of the

# Appendix NP4: Equilibrium Interest Rate with History Independence

With history independence, the consumption is given by

$$c_j^*(z_j) = \bar{c}(\tau) e^{\frac{\tau(1-\tau)}{(1+\eta)^2}\frac{\sigma_\phi^2}{2}} A_0(\tau)^{-1} e^{(1-\tau)\ln z}$$

The Euler equation is

$$\frac{1}{c_j(z_j)} = \frac{\beta\delta}{q}\mathbb{E}\left[\frac{1}{c_{j+1}(\ln z_{j+1})} | \ln z_j\right]$$

and substituting the equilibrium process,

$$1 = \frac{\beta\delta}{q}\mathbb{E}\left[e^{-(1-\tau)\omega}\right] = \frac{\beta\delta}{q}e^{(1-\tau)(2-\tau)\frac{\sigma_\omega^2}{2}}$$

Hence

$$q = \beta\delta e^{(1-\tau)(2-\tau)\frac{\sigma_\omega^2}{2}}$$

# Appendix NP5: No idiosyncratic shocks

If there are no idiosyncratic shocks, then approximation is not helpful. We have

$$\mathcal{W}(\tau,\theta) = \frac{1}{1+\eta}[\ln(1-\tau) - 1 + \tau] + \ln A(\tau,\theta) + \frac{1-\tau}{1+\eta}\ln(1-\tau)\sum_{k=0}^{\infty}\delta^k\theta_k$$
$$A(\tau,\theta) = -(1-\delta)\sum_{j=0}^{\infty}\delta^j e^{\sum_{k=0}^{j}\frac{1-\tau}{1+\eta}\ln(1-\tau)\theta_k}.$$

First-order condition in $\theta_k$:

$$1 + \frac{1}{A}\sum_{j=k}^{\infty}\delta^{j-k}e^{\sum_{k=0}^{j}\frac{1-\tau}{1+\eta}\ln(1-\tau)\theta_k} = \zeta\beta^k$$

Evaluating this for $k = 0$ yields $\zeta = 0$, which in turn implies $\theta_k = 0$ for $k > 0$, and so $\theta_0 - 1$. Thus, consumption smoothing applies.