

Insurance and Taxation over the Life Cycle

Emmanuel Farhi
Harvard University

Iván Werning
MIT

July 2010

Abstract

We consider a dynamic Mirrlees economy in a life cycle context and study the optimal insurance arrangement. Individual productivity evolves as a Markov process and is private information. We use a first order approach in discrete and continuous time and obtain novel theoretical and numerical results. Our main contribution is a formula describing the dynamics for the labor-income tax rate. When productivity is an AR(1) our formula resembles an AR(1) with a trend where: (i) the auto-regressive coefficient equals that of productivity; (ii) the trend term equals the covariance productivity with consumption growth divided by the Frisch elasticity of labor; and (iii) the innovations in the tax rate are the negative of consumption growth. The last property implies a form of short-run regressivity. Our simulations illustrate these results and deliver some novel insights. The average labor tax rises from 0% to 46% over 40 years, while the average tax on savings falls from 17% to 0% at retirement. We compare the second best solution to simple history independent tax systems, calibrated to mimic these average tax rates. We find that age dependent taxes capture a sizable fraction of the welfare gains. In this way, our theoretical results provide insights into simple tax systems.

1 Introduction

To a twenty five year old entering the labor market, the landscape must feel full of uncertainties. Will they land a good job relatively quickly or will they initially bounce from one job to another in search of a good match? What opportunities for on-the-job training and other forms of skill accumulation be they find? How well will they take advantage of these opportunities? Just how good are they? How high will they rise? Will they advance steadily within a firm or industry, or be laid off and have to reinvent themselves

elsewhere? For all these reasons, young workers must find it challenging to predict how much they will be making at, say, age fifty. More generally, they face significant uncertainty in their lifetime earnings.

This paper investigates the optimal design of a tax system that efficiently shares these risks. With a few notable exceptions, since [Mirrlees \(1971\)](#), optimal tax theory has mostly worked with a static model that treats heterogeneity and uncertainty symmetrically, since redistribution can be seen as insurance behind the “veil of ignorance”. More recently, there has been growing interest in the special role of uncertainty and insurance. To date, this more dynamic approach has focused on savings distortions, or considered special cases, such as two periods or i.i.d. shocks.¹ Little is known in more realistic settings about the pattern of labor income taxes when uncertainty is gradually revealed over time.

This paper aims to fill this gap and address the following questions. How are the lessons for labor income taxes from the static models (e.g. [Mirrlees \(1971\)](#), [Diamond \(1998\)](#), [Saez \(2001\)](#), [Werning \(2007b\)](#)) altered in a dynamic context? How is taxation with an insurance motive different from the redistributive motive? How does the optimal tax system compare to simpler systems? Are the welfare gains from a more elaborate system large? What lessons can we draw from the optimal tax structure for simpler tax systems?

We adapt the standard dynamic Mirrleesian framework to a life cycle context. Agents live for T years. They work and consume for T_E years and then retire, just consuming, for the remaining $T_R = T - T_E$ periods. During their working years, labor supply in efficiency units is the product of work effort and productivity. An agent’s productivity evolves as a persistent Markov process. Both effort and productivity are privately observed by the agent. The planner controls consumption and output, but cannot observe productivity nor work effort. Due to this private information, allocations must be incentive compatible. We study constrained efficient allocations and characterize the implicit marginal taxes or wedges implied by the allocation.

A direct attack on this problem is largely intractable, but we show that both theoretical and numerical progress can be made by using a first-order approach. A similar approach has proven useful in moral-hazard contexts with unobservable savings (see for example [Werning \(2002\)](#)). [Kapicka \(2008\)](#) spells out the first-order approach for a Mirrleesian setting, which we implement here. The basic idea is to relax the problem by imposing only local incentive constraints. Unlike the original problem, the relaxed problem has a recursive structure that makes it tractable. One can then check whether the solution to the relaxed problem is incentive compatible, and, hence, a solution to the original problem.

¹See for example [Diamond and Mirrlees \(1978\)](#); [Farhi and Werning \(2008b\)](#); [Golosov, Kocherlakota, and Tsyvinski \(2003\)](#); [Golosov, Tsyvinski, and Werning \(2006\)](#); [Albanesi and Sleet \(2006\)](#).

We find it useful to work in both discrete and continuous time (and relate the two by limiting arguments).

Our theoretical results are summarized by a novel formula for the dynamics of the labor wedge $\tau_{L,t}$. Although we derive the formula for a general stochastic process for productivity, it is most easily explained in the case where the logarithm of productivity follows an AR(1) with coefficient of mean-reversion ρ :

$$\log \theta_{t+1} = \rho \log \theta_t + (1 - \rho) \log \bar{\theta} + \varepsilon_{t+1}.$$

We require utility to be additively separable between consumption and labor and an isoelastic disutility function for labor. We then obtain

$$\mathbb{E}_t \left[\frac{\tau_{L,t+1}}{1 - \tau_{L,t+1}} \frac{1}{u'(c_{t+1})} \right] = \rho \frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u'(c_t)} + \left(\frac{1}{\varepsilon} + 1 \right) \text{Cov}_t \left(\log \theta_{t+1}, \frac{1}{u'(c_{t+1})} \right).$$

The first term captures mean-reversion and is simply the past labor wedge weighted by the coefficient of mean-reversion ρ in productivity. In this sense, the labor wedge inherits its degree of mean reversion from the stochastic process for productivity. The second term is zero if productivity or consumption are predictable. In this case, if $\rho = 1$, the formula specializes to a case of perfect tax-smoothing: the labor wedge remains constant between periods t and $t + 1$. If instead $\rho < 1$, then the labor wedge reverts to zero at rate ρ . When productivity and consumption are not predictable and are positively correlated, the second term on the right hand side is positive, contributing to higher average taxes. Intuitively, uncertainty in consumption creates a role for insurance, delivered by larger taxes. The covariance captures the marginal benefit of more insurance. The marginal cost depends on the elasticity of labor, which explains the role of the Frisch elasticity ε .

We also show that there is a sense in which our conditional expectation formula accurately describes the realized path. To do so, we consider a continuous time limit where productivity is a Brownian diffusion: $d \log \theta_t = -(1 - \rho)(\log \theta_t - \log \bar{\theta})dt + \sigma_t dW_t$, so that ρ controls the degree of mean reversion as above. We study this continuous time model directly and also approach it by taking the limit in the discrete time model. We show that

$$\lim_{\tau \downarrow 0} \frac{1}{\tau} \text{Var}_t \left[\frac{\tau_{L,t+\tau}}{1 - \tau_{L,t+\tau}} \frac{1}{u'(c_{t+\tau})} \right] = 0.$$

Thus, the conditional expectation above completely describes changes in $\frac{\tau_t}{1 - \tau_t} \frac{1}{u'(c_t)}$ over small intervals of time. In the limit, the process $\left\{ \frac{\tau_t}{1 - \tau_t} \frac{1}{u'(c_t)} \right\}$ is a diffusion that has no instantaneous volatility (it is of bounded variation):

$$d \left(\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u'(c_t)} \right) = \left[-(1 - \rho) \frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u'(c_t)} + \left(\frac{1}{\epsilon} + 1 \right) \text{Cov}_t \left(d \log \theta_t, d \left(\frac{1}{u'(c_t)} \right) \right) \right] dt.$$

The drift in the continuous time is essentially the counterpart of our conditional expectation formula. Thus, this equation states that there are no errors around the conditional expectation. Of course, this does not mean that the entire process for $\left\{ \frac{\tau_L}{1 - \tau_L} \frac{1}{u'(c)} \right\}$ is deterministic, since the drift itself may vary stochastically. It does imply that over short periods of time $\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u'(c_t)}$ varies a lot less than productivity θ_t .

Economically, this result is important because it describes a form of regressivity. When productivity rises, consumption rises, so the marginal utility of consumption falls. The labor wedge must then fall by the same amount, at least in the short run. This induces a negative short-run relation between productivity and the labor wedge. This force for regressive taxation is novel and due to the dynamic aspects of our model.

Finally, we extend the well-known zero taxation result at the top and bottom of the productivity distribution.² If the conditional distribution for productivity has a fixed support, then the labor wedge is zero at both extremes, just as in the static Mirrlees model. However, in our dynamic model, a moving support may be more natural, with the top and bottom, $\bar{\theta}_t(\theta_{t-1})$ and $\underline{\theta}_t(\theta_{t-1})$, being functions of the previous period's productivity, θ_{t-1} . With a moving support, we establish that the labor wedge is no longer zero at the top and bottom. An interesting example is when productivity is a geometric random walk, and innovations have a bounded support, the extremes $\bar{\theta}_t(\theta_{t-1})$ and $\underline{\theta}_t(\theta_{t-1})$ move proportionally with θ_{t-1} . In this case, the labor wedge at the top must be below the previous period's labor wedge. The reverse is true at the bottom: the labor wedge must be higher than in the previous period. This result is consistent with the short-run regressivity discussed in the previous paragraph. Note, however, that no limit argument is required.

For our numerical exploration, we adopt a random walk for productivity. This choice is motivated by two considerations. First, the evidence in [Storesletten, Telmer, and Yaron \(2004\)](#) points to a near random walk for labor earnings, which requires a near random walk for productivity. Second, by focusing on a random walk we are considering the opposite end of the spectrum of the well explored i.i.d. case ([Albanesi and Sleet, 2006](#)).

Our findings both serve to illustrate our theoretical results and provide novel insights. Indeed, although our numerical work is based the discrete time version of the model, with a period modeled as a year, the simulations show that the continuous time provides an

²The result at the bottom requires labor to be positive.

excellent explanations of our findings.

We find that the average labor wedge starts near zero and increases over time, asymptoting to around 40% precisely at retirement. The intertemporal wedge displays the opposite pattern, with its average starting around 0.9%, corresponding to a 17% tax on net interest, and falling to zero at retirement. Both results are easily explained by our theoretical formula. As retirement approaches the variance of consumption growth falls to zero, for standard consumption smoothing reasons. Our formulas then indicate that the labor wedge will rise over time and asymptote at retirement and that the intertemporal wedge will reach zero at retirement.

Our tax system comes out to be slightly regressive in the sense that marginal tax rates are higher for agents with currently low productivity shocks. Our short-run regressivity result seems to explain at least part of this regressivity. However, it should be noted that the tax system is progressive in the sense that average tax rates are increasing in productivity.

The second-best allocation we have characterized can be implemented with taxes, but, as is well known, it requires relatively elaborate, history-dependent tax instruments. We wish to investigate how our results translate to simpler systems, restricted to being history independent. Do our theoretical results provide guidance for such real-world tax systems? We find that they do: the second best turns out to be unexpectedly informative in the design of simpler policies.

Specifically, we compute the equilibrium with history-independent linear taxes on labor and capital income, and consider both age-dependent and age-independent taxes. When age-dependent taxes are allowed, we set the tax rate on labor or capital at a given age equal to the average for that given age of the corresponding wedge in the simulated second best. Surprisingly, the welfare loss of such a system, relative to the second best, is minuscule—around 0.1% of lifetime consumption. By implication, the pattern of tax rate we chose must be very close to optimal. In this way, our theoretical results do provide guidance for more restrictive tax systems. When we solve the model with age-independent tax rates, we optimize over the value of these tax rates. The results are as follows. First, we find a more significant loss in welfare, of around 0.7% of lifetime consumption. Age dependent tax rates are important in approaching the second best. Second, when taxes are age independent, the optimal tax rate on capital comes out to be essentially zero. This can be explained by the fact that a subsidy on capital helps imitate the missing age-dependent taxes on labor. This new effect cancels the desire for a positive tax on capital, obtained when more sophisticated taxes are allowed.

Related literature. Our paper contributes to the is the optimal taxation literature based on models with private information (see [Golosov, Tsyvinski, and Werning, 2006](#), and the references therein). The case where shocks are i.i.d. has been extensively studied [see for example [Albanesi and Sleet \(2006\)](#) and more recently [Ales and Maziero \(2009\)](#)]. Outside of the i.i.d. case few undertake a quantitative analysis. Persistent shocks significantly complicate the analysis. As emphasized by [Fernandes and Phelan \(2000\)](#), the efficient allocations have a recursive structure, but the dimensionality of the state is proportional to the number of possible shock values, severely limiting the possibilities for realistic numerical analyses.³⁴

This paper continues our efforts to quantify dynamic Mirrleesian models using more realistic assumptions about uncertainty. In [Farhi and Werning \(2008a\)](#) and [Farhi and Werning \(2009\)](#), our strategy was to focus on the welfare gains from savings distortions. We presented a simple method to do so, which allowed us to consider rich stochastic processes and was tractable enough to apply in general equilibrium settings, which proved to be important. However, these papers do not attempt anything regarding labor wedges, which are the main focus of the present paper.

Versions of the first-order approach on which we rely in this paper have been studied in other papers. [Werning \(2002\)](#) introduced this approach in a moral-hazard setting with unobservable savings to study optimal unemployment insurance with free-savings. [Pavan, Segal, and Toikka \(2009\)](#) characterize necessary and sufficient conditions for the first-order approach in very general dynamic environment. [Williams \(2008\)](#) studies a continuous-time economy with hidden income that follows a Brownian motion. [Garrett and Pavan \(2010\)](#) use a first-order approach to study managerial compensation. [Kapicka \(2008\)](#) spells out the first-order approach for a general Mirrleesian setting with persistent productivity shocks. He also simulates a simple example to illustrate the approach.⁵

[Fukushima \(2010\)](#) performs a numerical study of an overlapping generations economy, where each generation looks much like the ones in our model. He considers a special class of Markov chains with two discrete shocks that allow for a low dimensional representation of the state space. For a planning problem that seeks to maximize steady-state utility, he reports substantial welfare gains of the optimal tax system over a system combining a flat tax and an exemption. [Golosov, Troshkin, and Tsyvinski \(2010\)](#) use a first-

³Two exceptions are [Golosov and Tsyvinski \(2006\)](#) for disability insurance and [Shimer and Werning \(2008\)](#) for unemployment insurance. In both cases, the nature of the stochastic process for shocks allows for a low dimensional recursive formulation that is numerically tractable.

⁴See also [Battaglini and Coate \(2008\)](#). See as well [Tchisty \(2006\)](#) and [Battaglini \(2005\)](#) for applications in a non-taxation context.

⁵See also [Abraham and Pavoni \(2008\)](#) and [Jarque \(2008\)](#).

order approach to study a life-cycle economy with two periods and persistent shocks. The goal of their paper is to calibrate the distribution of shocks in both periods using the observed distribution of incomes, as [Saez \(2001\)](#) did for a single period in a static setting.⁶

An important implication of our results is that with persistent productivity shocks, labor taxes should on average increase with age. Our theoretical formula provides the underpinnings for this observation as well as insights into its origin; our numerical simulation explores its quantitative importance. This aspect of our contribution connects with a prior contributions focusing on the benefits of age-dependent taxes.⁷ Most closely related to our paper are [Kremer \(2002\)](#) and [Weinzierl \(2008\)](#). [Kremer \(2002\)](#) emphasized the potential benefits of age-dependent labor taxation, noting that the wage distribution is likely to become more dispersed with age and conjectured that labor taxes should generally rise depend on age. [Weinzierl \(2008\)](#) provides a more comprehensive treatment. He calibrates two- and three-period Mirrlees models. Like us, he finds important welfare gains from age dependent taxes.

2 The Insurance Problem

This section first describes the economic environment and its planning problem. We then explain our first order approach to solving this problem.

2.1 The Environment and Planning Problem

Preferences, Uncertainty and Information. The economy is populated by a continuum of agents who live for T periods. Their ex ante utility is

$$\mathbb{E}_0 \sum_{t=1}^T \beta^{t-1} u^t(c_t, y_t; \theta_t).$$

Here c_t represents consumption, y_t represents efficiency units of labor, and $\theta_t \in \Theta = [\underline{\theta}, \bar{\theta}]$ is a state variable with conditional density $f^t(\theta_t | \theta_{t-1})$. This state affects preferences over consumption and labor in efficiency units and can capture both taste and productivity fluctuations. In particular, an important case is when $u^t(c, y; \theta) = \tilde{u}^t(c, y/\theta)$, for some

⁶Both [Kapicka \(2008\)](#) and [Golosov, Troshkin, and Tsyvinski \(2010\)](#) rely on exponential utility and special shock specifications to make the problem tractable, by reducing the number of state variables.

⁷[Erosa and Gervais \(2002\)](#) analyze age-dependent linear labor taxation in Ramsey setting. In their model, optimal linear labor income taxes are indexed on age because the elasticity of labor supply varies, endogenously, with age.

utility function $\tilde{u}^t(c, n)$, defined over consumption and labor effort; then $y = \theta n$ and θ can be interpreted as productivity.

We allow the utility function and the density to depend on the period t to be able to incorporate life-cycle considerations. For example, an economy where agents work for T_E periods and then retire for T_R periods can be captured by setting $\tilde{u}(c, y/\theta)$ for $t \leq T_E$ and $\tilde{u}(c, 0)$ for $T_E < t \leq T$.

We make several smoothness assumptions. The utility function is twice continuously differentiable and the density is continuously differentiable in its second argument. To simplify the exposition, we start with the full support assumption that $f^t(\theta'|\theta) > 0$ for all $\theta, \theta' \in \Theta$.

The realization of the state θ_t for all $t = 1, 2, \dots, T$ is privately observed by the agent. To simplify and focus on the insurance motive, we assume that the initial state θ_{-1} is publicly known. It is easy to introduce initial heterogeneity and private information with respect to θ_{-1} , together with a redistributive motive.

More explicitly, an allocation is $\{c, y\} \equiv \{c(\theta^t), y(\theta^t)\}$ and utility is

$$U(\{c, y\}) \equiv \sum_{t=1}^T \beta^{t-1} \int u^t(c(\theta^t), y(\theta^t); \theta_t) f^t(\theta_t|\theta_{t-1}) f^{t-1}(\theta_{t-1}|\theta_{t-2}) \cdots f^1(\theta_1|\theta_0) d\theta_t d\theta_{t-1} \cdots d\theta_0$$

Incentive Compatibility. By the revelation principle, without loss of generality, we can focus on direct mechanisms, where agents make reports $r_t \in \Theta$ regarding θ_t . For any reporting strategy $\sigma = \{\sigma_t(\theta^t)\}$ we have a history of reports $\sigma^t(\theta^t) = (\sigma_1(\theta_0), \dots, \sigma_t(\theta^t))$ and a corresponding path $\{c^\sigma, y^\sigma\} = \{c(\sigma^t(\theta^t)), y(\sigma^t(\theta^t))\}$. Let Σ denote the set of all reporting strategies σ .

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

$$U(\{c, y\}) \geq U(\{c^\sigma, y^\sigma\}) \quad \forall \sigma \in \Sigma,$$

so that truth telling $\sigma^* = \{\sigma_t^*(\theta^t)\}$ with $\sigma_t^*(\theta_t) = \theta_t$, is optimal.

Planning Problem. To keep things simple, we work in partial equilibrium, that is, assuming a linear technology that converts labor into consumption goods one for one and a linear storage technology with gross rate of return q^{-1} (and a net rate of return equal to $q^{-1} - 1$). This allows us to study the contracting problem for a single cohort in

isolation. The relevant cost of an allocation is then its expected net present value:

$$\Psi(\{c, y\}) \equiv \sum_{t=1}^T q^{t-1} \int (c(\theta^t) - y(\theta^t)) f^t(\theta_t | \theta_{t-1}) \cdots f^1(\theta_1 | \theta_0) d\theta_t \cdots d\theta_1.$$

An allocation $\{c^*, y^*\}$ is efficient if there is no other incentive compatible allocation $\{c, y\}$ with $U(\{c, y\}) \geq U(\{c^*, y^*\})$ and $\Psi(\{c, y\}) \leq \Psi(\{c^*, y^*\})$, with at least one strict inequality. Efficient allocations solve the following program.

Program IC.

$$\begin{aligned} K_0(v) &\equiv \min_{\{c, y\}} \Psi(\{c, y\}) \\ \text{s.t. } &U(\{c, y\}) \geq v \\ &U(\{c, y\}) \geq U(\{c^\sigma, y^\sigma\}) \quad \forall \sigma \in \Sigma \end{aligned}$$

2.2 A Recursive First-Order Approach

In this section, we lay down our first-order approach, and explain how it leads to a relaxed version of Program IC. Previous papers using a similar approach include [Werning \(2002\)](#), [Kapicka \(2008\)](#), [Williams \(2008\)](#), and [Pavan, Segal, and Toikka \(2009\)](#).

Let $w(\theta^t)$ denote the equilibrium continuation utility after history θ^t . That is, w is the unique solution to

$$w(\theta^t) = u^t(c(\theta^t), y(\theta^t), \theta_t) + \beta \int w(\theta^t, \theta_{t+1}) f^{t+1}(\theta_{t+1} | \theta_t) d\theta_{t+1} \quad (1)$$

for all $t = 1, \dots, T$ with $w(\theta^{T+1}) \equiv 0$. Incentive compatibility can be represented with this notation as follows. For any strategy $\sigma \in \Sigma$, let continuation utility $w^\sigma(\theta^t)$ solve

$$w^\sigma(\theta^t) = u^t(c(\sigma^t(\theta^t)), y(\sigma^t(\theta^t)), \theta_t) + \beta \int w^\sigma(\theta^t, \theta_{t+1}) f^{t+1}(\theta_{t+1} | \theta_t) d\theta_{t+1}$$

with $w^\sigma(\theta^{T+1}) \equiv 0$. Then incentive compatibility is equivalent to

$$\int w(\theta_1) f^1(\theta_1 | \theta_0) d\theta_1 \geq \int w^\sigma(\theta_1) f^1(\theta_1 | \theta_0) d\theta_1$$

for all $\sigma \in \Sigma$. This condition requires

$$w(\theta^t) \geq w^\sigma(\theta^t)$$

for almost all θ^t .

A Necessary Condition. We now use this representation to sketch a derivation of a necessary condition for incentive compatibility. We omit the full details of the argument which have been derived by [Pavan, Segal, and Toikka \(2009\)](#).

For any incentive compatible allocation, consider the following one-shot deviation from truth telling:

$$\begin{aligned}\sigma_s(\theta^s) &= \theta_s & \theta^s &\neq \theta^t \\ \sigma_t(\theta^t) &= r\end{aligned}$$

Thus, the agent reports truthfully until θ^t , then reports r and reverts back to truth telling thereafter. Denote this strategy by σ^r . Continuation utility solves

$$w^r(\theta^t) = u^t(c(\theta^{t-1}, r), y(\theta^{t-1}, r), \theta_t) + \beta \int w^r(\theta^{t+1}) f^{t+1}(\theta_{t+1} | \theta_t) d\theta_{t+1}$$

Since truth telling is employed from $t + 1$ onwards we have that

$$w^r(\theta^{t+1}) = w(\theta^{t-1}, r, \theta_{t+1}).$$

Incentive compatibility requires

$$w(\theta^t) = \max_r w^r(\theta^t).$$

Putting the three conditions together yields

$$w(\theta^t) = \max_r \{u^t(c(\theta^{t-1}, r), y(\theta^{t-1}, r), \theta_t) + \beta \int w(\theta^{t-1}, r, \theta_{t+1}) f^{t+1}(\theta_{t+1} | \theta_t) d\theta_{t+1}\}$$

Differentiating and applying an envelope condition implies:

$$\frac{\partial}{\partial \theta_t} w(\theta^t) = u_{\theta_t}^t(c(\theta^t), y(\theta^t), \theta_t) + \beta \int w(\theta^{t+1}) f_{\theta_t}^{t+1}(\theta_{t+1} | \theta_t) d\theta_{t+1} \quad (2)$$

Equations (1)–(2) provide a set of necessary conditions for an allocation to be incentive compatible. [Pavan, Segal, and Toikka \(2009\)](#) derive these conditions rigorously in a more general setting.

It will be useful to collect these conditions and rewrite them as

$$w(\theta^t) = u^t(x(\theta^t), \theta_t) + \beta v(\theta^t)$$

$$\frac{\partial}{\partial \theta_t} w(\theta^t) = u_{\theta}^t(x(\theta^t), \theta_t) + \beta \Delta(\theta^t)$$

where

$$v(\theta^t) \equiv \int w(\theta^{t+1}) f^{t+1}(\theta_{t+1} | \theta_t) d\theta_{t+1}$$

$$\Delta(\theta^t) \equiv \int w(\theta^{t+1}) f_{\theta_t}^{t+1}(\theta_{t+1} | \theta_t) d\theta_{t+1}$$

The Relaxed Planning Problem. Define $C^t(y, u, \theta)$ denote the expenditure function—the inverse of the utility function for consumption $u^t(\cdot, y, \theta)$. Let K denote the minimized cost for the relaxed program that only imposes the necessary conditions derived above. This function satisfies the following Bellman equation:

Program FOA.

$$K(v, \Delta, \theta_-, t) = \min \int \{C^t(y(\theta), w(\theta) - \beta v(\theta), \theta) - y(\theta) \\ + q \int K(v(\theta), \Delta(\theta), \theta', t+1) f^{t+1}(\theta' | \theta) d\theta\} f^t(\theta | \theta_-) d\theta$$

$$v = \int w(\theta) f^t(\theta | \theta_-) d\theta$$

$$\Delta = \int w(\theta) f_{\theta_-}^t(\theta | \theta_-) d\theta$$

$$\dot{w}(\theta) = u_{\theta}^t(C^t(y(\theta), w(\theta) - \beta v(\theta), \theta), y(\theta), \theta) + \beta \Delta(\theta))$$

Verifying IC. Suppose that a solution to Program FOA has been computed. Then it is a solution to Program IC if and only if the proposed allocation is incentive compatible. One approach is to seek sufficient conditions on primitives that a priori guarantee that this will be the case. A practical alternative, is to verify the incentive compatibility of any candidate optimum. It turns out that doing so is relatively straightforward.

The solution to Program FOA yields policy functions that determine the allocation

recursively as a function of reports as follows

$$\begin{aligned}c_t &= g^c(v_t, \Delta_t, r_{t-1}, r_t, t) \\y_t &= g^y(v_t, \Delta_t, r_{t-1}, r_t, t) \\v_{t+1} &= g^v(v_t, \Delta_t, r_{t-1}, r_t, t) \\\Delta_{t+1} &= g^\Delta(v_t, \Delta_t, r_{t-1}, r_t, t)\end{aligned}$$

An agent takes these functions as a constraints in his optimal reporting problem, represented by the Bellman equation:

$$\begin{aligned}V(v, \Delta, r_-, \theta, t) &= \max_r \{u^t(g^c(v, \Delta, r_-, r, t), g^y(v, \Delta, r_-, r, t), \theta) \\&\quad + \beta \int V(g^v(v, \Delta, r_-, r, t), g^\Delta(v, \Delta, r_-, r, t), r, \theta', t + 1) f^{t+1}(\theta'|\theta) d\theta'\}\end{aligned}$$

Let

$$g^w(v_t, \Delta_t, r_{t-1}, r_t, t) = u^t(g^c(v_t, \Delta_t, r_{t-1}, r_t, t), g^y(v_t, \Delta_t, r_{t-1}, r_t, t), r_t) + \beta g^v(v_t, \Delta_t, r_{t-1}, r_t, t)$$

represent the optimal value of w in the planning problem given states $(v_t, \Delta_t, r_{t-1}, r_t, t)$. The proposed allocation is incentive compatible if and only if

$$V(v, \Delta, r_-, \theta, t) = g^w(v, \Delta, r_-, \theta, t)$$

Plugging this condition into into the Bellman equation we obtain

$$\begin{aligned}\theta \in \arg \max_r \{u^t(g^c(v, \Delta, r_-, r, t), g^y(v, \Delta, r_-, r, t), \theta) \\+ \beta \int g^w(\theta', g^v(v, \Delta, r_-, r, t), g^\Delta(v, \Delta, r_-, r, t), r, t + 1) f^{t+1}(\theta'|\theta) d\theta\} \quad (3)\end{aligned}$$

Interestingly, because we have the guess for the agent's value function, this verification does not require solving a dynamic optimization problem.

Initial Heterogeneity. We have interpreted the planning problem as confronting a single agent. However, it is simple to add heterogeneity to the analysis. Indeed, we can interpret the first shock θ_1 as initial heterogeneity—recall that we allow the conditional density to depend flexibly on the period t . If the social welfare function is Utilitarian,

then the analysis requires no change: insurance behind the veil of ignorance and utilitarian redistribution are equivalent. It is also possible to allow for a different social welfare function, or to characterize the entire set of constrained Pareto-efficient allocations, by treating the initial period in the planning problem differently.

3 Optimality Conditions

Given an allocation $\{c, y\}$, and a history θ^t , define the intertemporal wedge

$$\tau_K(\theta^t) = 1 - \frac{q}{\beta} \frac{u_c^t(c(\theta^t), y(\theta^t), \theta_t)}{\int u_c^{t+1}(c(\theta^{t+1}), y(\theta^{t+1}), \theta_{t+1}) f^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}}$$

and the labor wedge

$$\tau_L(\theta^t) \equiv 1 + \frac{u_y^t(c(\theta^t), y(\theta^t), \theta_t)}{u_c^t(c(\theta^t), y(\theta^t), \theta_t)}.$$

In this section, we characterize these wedges for allocations that solve Programs IC and FOA.

3.1 A Positive Intertemporal Wedge

Our first result restates the well-known inverse Euler condition. This result requires utility from consumption to be separable from the disutility of labor.

Assumption 1 *For every $t \geq 0$, the utility function $u^t(c, y, \theta)$ is separable so that there exists functions \hat{u}^t and \hat{h}^t such that $u^t(c, y, \theta) = \hat{u}^t(c) - \hat{h}^t(y, \theta)$.*

Proposition 1 *Suppose that Assumption 1 holds and that $\{c, y\}$ solves Program IC or Program FOA. Then for every $t \geq 1$ and history θ^{t-1} , the following Inverse Euler equation holds*

$$\frac{1}{\hat{u}^{t-1}(c(\theta^{t-1}))} = \frac{q}{\beta} \int \frac{1}{\hat{u}^t(c(\theta^t))} f^t(\theta_t|\theta_{t-1}) d\theta_t$$

and the intertemporal wedge satisfies

$$\tau_K(\theta^{t-1}) = 1 - \frac{\left[\int [\hat{u}^t(c(\theta^t))]^{-1} f^t(\theta_t|\theta_{t-1}) d\theta_t \right]^{-1}}{\int \hat{u}^t(c(\theta^t)) f^t(\theta_t|\theta_{t-1}) d\theta_t}. \quad (4)$$

Note that this result holds for any allocation that solves both Program IC or Program FOA and for any stochastic process for idiosyncratic shocks $\{\theta\}$. Applying Jensen's in-

equality to the second equation implies that the intertemporal wedge $\tau_K (\theta^{t-1})$ is positive. In other words, positive savings distortions are present at the constrained optimum.

3.2 Labor Wedge Dynamics: Tax Smoothing and Mean Reversion

We now seek an optimality condition for the labor wedge. The following isoelastic assumption is useful for this purpose. It has been used to prove perfect tax-smoothing results by [Werning \(2007a\)](#).

Assumption 2 *Assumption 1 holds and the disutility of work is isoelastic $\hat{h}^t (y, \theta) = (\kappa/\alpha) (y/\theta)^\alpha$ with $\kappa > 0$ and $\alpha > 1$.*

For any function Π , define

$$\phi_t^\Pi (\theta_{t-1}) \equiv \int \Pi (\theta_t) f^t (\theta_t | \theta_{t-1}) d\theta_t.$$

We then have the following proposition.

Proposition 2 *Suppose that Assumptions 1 and 2 hold, and that $\{c, y\}$ solves Program FOA. Consider a function $\pi (\theta)$ and let $\Pi (\theta)$ be a primitive of $\pi (\theta) / \theta$. Then the labor wedge satisfies the following equation for every $t \geq 1$ and history θ^{t-1}*

$$\begin{aligned} & \int \frac{\tau_L (\theta^t)}{1 - \tau_L (\theta^t)} \frac{q \hat{u}^{t-1'} (c (\theta^{t-1}))}{\beta \hat{u}^{t'} (c (\theta^t))} \pi (\theta_t) f^t (\theta_t | \theta_{t-1}) d\theta_t \\ &= \frac{\tau_L (\theta^{t-1})}{1 - \tau_L (\theta^{t-1})} \theta_{t-1} \frac{d\phi_t^\Pi (\theta_{t-1})}{d\theta_{t-1}} + \alpha \int \Pi (\theta_t) \left[\frac{q \hat{u}^{t-1'} (c (\theta^{t-1}))}{\beta \hat{u}^{t'} (c (\theta^t))} - 1 \right] f^t (\theta_t | \theta_{t-1}) d\theta_t. \quad (5) \end{aligned}$$

The fact that equation (5) holds for all possible $\pi(\theta_t)$ imposes restrictions on the stochastic $\left\{ \frac{\tau_L}{1 - \tau_L} \frac{1}{\hat{u}^{t'}} \right\}$. We shall come back to this point later, especially in its continuous time counterpart. For now, the case where $\pi (\theta) = 1$ and $\Pi (\theta) = \log (\theta)$ is of particular interest.

Corollary 1 *Suppose that Assumptions 1 and 2 hold, and that $\{c, y\}$ solves Program FOA. Then*

the labor wedge satisfies the following equation for every $t \geq 1$ and history θ^{t-1}

$$\begin{aligned} & \int \frac{\tau_L(\theta^t)}{1 - \tau_L(\theta^t)} \frac{q \hat{u}^{t-1'}(c(\theta^{t-1}))}{\beta \hat{u}^{t'}(c(\theta^t))} f^t(\theta_t | \theta_{t-1}) d\theta_t \\ &= \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}} + \alpha \int \log(\theta_t) \left[\frac{q \hat{u}^{t-1'}(c(\theta^{t-1}))}{\beta \hat{u}^{t'}(c(\theta^t))} - 1 \right] f^t(\theta_t | \theta_{t-1}) d\theta_t. \end{aligned} \quad (6)$$

To understand the role of $\theta_{t-1} d\phi_t^{\log}(\theta_{t-1}) / d\theta_{t-1}$ in equation (6), consider the generalized geometric AR(1) process:

$$\log(\theta_t) = \rho \log(\theta_{t-1}) + \bar{\theta}_t + \varepsilon_t,$$

where $\{\bar{\theta}_t\}$ is a deterministic sequence for the unconditional mean of θ_t , and ε_t are independent draws from a distribution $g^t(\varepsilon_t, \theta^{t-1})$, normalized so that $\int \varepsilon_t g^t(\varepsilon_t, \theta^{t-1}) d\varepsilon_t = 0$. Then $\phi_t^{\log}(\theta_{t-1}) = \rho \log(\theta_{t-1}) + \bar{\theta}_t$ so that

$$\theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}} = \rho.$$

For this AR(1) specification, equation (6) can be written more compactly as

$$\mathbb{E}_{t-1} \left[\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{q \hat{u}^{t-1'}(c_{t-1})}{\beta \hat{u}^{t'}(c_t)} \right] = \rho \frac{\tau_{L,t-1}}{1 - \tau_{L,t-1}} + \alpha \text{Cov}_{t-1} \left(\log(\theta_t), \frac{q \hat{u}^{t-1'}(c_{t-1})}{\beta \hat{u}^{t'}(c_t)} \right) \quad (7)$$

Proposition 1 implies that

$$\mathbb{E}_{t-1} \left[\frac{q \hat{u}^{t-1'}(c_{t-1})}{\beta \hat{u}^{t'}(c_t)} \right] = 1$$

so the term $(q/\beta) \hat{u}^{t-1'}(c_{t-1}) / \hat{u}^{t'}(c_t)$ on the left hand side of equation (7) represents a change of measure.

Thus, we have a formula for the conditional expectation of $\tau_{L,t} / (1 - \tau_{L,t})$ under a risk-adjusted probability measure. On the right hand side, $\tau_{L,t-1} / (1 - \tau_{L,t-1})$ is weighted by the coefficient of mean-reversion ρ . In this sense, $\{\tau_L / (1 - \tau_L)\}$ inherits its degree of mean reversion from the stochastic process for productivity. The second term provides a drift for $\{\tau_L / (1 - \tau_L)\}$.

It is useful to first consider the special cases where the drift is zero, which occurs when consumption at t is predictable at $t - 1$, so that $\text{Var}_{t-1}(c_t) = 0$. This would be the case if the productivity level θ_t were predictable at $t - 1$, so that $\text{Var}_{t-1}(\theta_t) = 0$. In this case, if

$\rho = 1$ equation (7) implies that the labor wedge remains constant between periods $t - 1$ and t , a form of perfect tax-smoothing. When $\rho < 1$ the labor wedge reverts to zero at rate ρ .⁸

The drift is positive whenever $\text{Var}_{t-1}(\theta_t) > 0$ provided that consumption is increasing in productivity. Compared to the case with $\text{Var}_{t-1}(\theta_t) = 0$, the additional shocks to productivity create an additional motive for insurance. This pushes the labor wedge up. Interestingly, the size of the drift is precisely the covariance of the log of productivity with the inverse growth rate in marginal utility, divided by $1/\alpha = \epsilon/(1 + \epsilon)$, where ϵ is the Frisch elasticity of labor supply. The covariance captures the benefit of added insurance, since it depends on the variability of consumption as well as on the degree of risk aversion. Insurance comes at the cost of lower incentives for work. This effect is stronger the more elastic is labor supply, explaining the role of the Frisch elasticity.

Returning to the more general statement in Proposition 2, equation (5) shares many ingredients with equation (6). Note however that, in general,

$$\int \frac{q \hat{u}^{t-1'}(c(\theta^{t-1}))}{\beta \hat{u}^{t'}(c(\theta^t))} \pi(\theta_t) f^t(\theta_t | \theta_{t-1}) d\theta_t$$

will not equal one, so that by contrast with equation (6), the right-hand side cannot be interpreted as a risk-adjusted conditional expectation of the labor wedge in period t . Another important case is $\pi(\theta_t) = (\beta/q) \hat{u}^{t'}(c(\theta^t)) / \hat{u}^{t-1'}(c(\theta^{t-1}))$, so that equation (5) provides a formula for the unadjusted conditional expectation for $\tau_{L,t}/(1 - \tau_{L,t})$. The corresponding expression is somewhat more involved than equations (6) and (7). Rather than develop the expression here, we present its neater continuous time counterpart in Section 4.

Equations (5) and (6) hold for any allocation that solves Program FOA. They do not necessarily hold for an allocation that solves Program IC when the two programs do not coincide. Nevertheless, we are able to show that Proposition 2 applies with a particular function π to any allocation that solves Program IC under the following assumption.

Assumption 3 *The process $\{\theta\}$ is a geometric random walk. That is, the growth rate θ_t/θ_{t-1} is independent of the history θ^{t-1} .*

⁸These special cases are consistent with the results in [Werning \(2007a\)](#), who studied a model where agent's private types are fixed (similar to $\text{Var}_{t-1}(\theta_t) = 0$ here). Productivity may still vary for each type, due to changes in inequality or aggregate shocks. At the optimum, the tax rate is constant with respect to aggregate shocks to productivity, but is an increasing function of the current degree of inequality. This relates to the analysis here, since when $\rho < 1$ and $\text{Var}_{t-1}(\theta_t) = 0$ we have a decreasing pattern for inequality and the tax rate.

Proposition 3 Suppose that Assumptions 1, 2, and 3 hold, and that $\{c, y\}$ solves Program IC. Then the labor wedge satisfies equation (5) for every $t \geq 1$ and history θ^{t-1} with $\pi(\theta) = \theta^{-\alpha}$.

3.3 Labor Wedge at the Top and Bottom

We now look at the labor wedge for the two extreme realizations of θ_t , top and bottom. As we shall see, when the support for current productivity is independent of previous productivity then standard zero-distortion results apply. However, it is important to consider the more general case of a moving support, where the upper and lower bounds, $\bar{\theta}_t(\theta_{t-1})$ and $\underline{\theta}_t(\theta_{t-1})$ vary with θ_{t-1} . We assume $\bar{\theta}_t(\theta_{t-1})$ and $\underline{\theta}_t(\theta_{t-1})$ are differentiable. For short, we often simply write $\bar{\theta}_t$ and $\underline{\theta}_t$ leaving the dependence on θ_{t-1} implicit.

The only modification to Program FOA is that Δ now incorporates two terms to capture the movements in the support:

$$\Delta = \int_{\underline{\theta}_t(\theta_-)}^{\bar{\theta}_t(\theta_-)} w(\theta) f_{\theta_-}^t(\theta, \theta_-) d\theta + \frac{d\bar{\theta}_t}{d\theta_-} w(\bar{\theta}_t) f^t(\bar{\theta}_t | \theta_-) - \frac{d\underline{\theta}_t}{d\theta_-} w(\underline{\theta}_t) f^t(\underline{\theta}_t | \theta_-).$$

Propositions 1, 2 and 3 extend without modification to the case of moving support.

Proposition 4 Consider an interior allocation that solves Program FOA:

i. if for a history θ^{t-1} , $d\bar{\theta}_t/d\theta_{t-1} = d\underline{\theta}_t/d\theta_{t-1} = 0$, then

$$\frac{\tau_L(\theta^{t-1}, \bar{\theta}_t)}{1 - \tau_L(\theta^{t-1}, \bar{\theta}_t)} = \frac{\tau_L(\theta^{t-1}, \underline{\theta}_t)}{1 - \tau_L(\theta^{t-1}, \underline{\theta}_t)} = 0;$$

ii. suppose that Assumptions 1 and 2 hold, then for every history θ^{t-1}

$$\begin{aligned} \frac{\tau_L(\theta^{t-1}, \bar{\theta}_t)}{1 - \tau_L(\theta^{t-1}, \bar{\theta}_t)} &= \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \frac{\beta \hat{u}^{t'}(c(\theta^{t-1}, \bar{\theta}_t))}{q \hat{u}^{t-1'}(c(\theta^{t-1}))} \frac{\theta_-}{\bar{\theta}_t} \frac{d\bar{\theta}_t}{d\theta_{t-1}}, \\ \frac{\tau_L(\theta^{t-1}, \underline{\theta}_t)}{1 - \tau_L(\theta^{t-1}, \underline{\theta}_t)} &= \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \frac{\beta \hat{u}^{t'}(c(\theta^{t-1}, \underline{\theta}_t))}{q \hat{u}^{t-1'}(c(\theta^{t-1}))} \frac{\theta_-}{\underline{\theta}_t} \frac{d\underline{\theta}_t}{d\theta_{t-1}}. \end{aligned}$$

With a fixed support, the no-distortion results from the static model extend to our dynamic setting. However, when the support is not fixed the labor wedge after an extreme realization is the product of the labor wedge in the previous period times the growth rate of marginal utility, and the elasticity of the corresponding bound of the support with respect to θ_{t-1} . For example, in the case where $\{\theta\}$ is a geometric random walk with bounded innovations, the elasticity of the bounds with respect to θ_{t-1} is equal to one.

In this case, provided that consumption $c(\theta^{t-1}, \theta_t)$ is increasing in θ_t , the Inverse Euler equation implies: $\tau_L(\theta^{t-1}, \bar{\theta}_t) \leq \tau_L(\theta^{t-1}) \leq \tau_L(\theta^{t-1}, \underline{\theta}_t)$. The fact that the tax rate may go both up or down illustrates a form of tax smoothing.

3.4 A Regressivity Result

In Section 4, we develop a continuous time formulation of Program FOA when the stochastic process for productivity follows a Brownian diffusion, and we use continuous time stochastic control to analyze this program. In this section, instead, we perform an analysis in discrete time, but taking the limit as the period length shrinks to zero. We show that, in the limit, the quantity $(1/\hat{u}''(c_t))(\tau_{L,t}/(1 - \tau_{L,t}))$ does not vary unpredictably, in response to productivity shocks. In this sense, the conditional expectation formula in equation (6) provides a very accurate prediction for this variable over short intervals of time.

Let τ be the length of a period. Instead of indexing periods by $t = 1, 2, 3 \dots$ we now take $t = \tau, 2\tau, 3\tau, \dots$. We assume that $\theta_{t+\tau}$ is log normally distributed so that

$$\log \theta_{t+\tau} \sim N(\log \theta_t + \mu_t^{\log}(\theta_t), \sigma_t^2) \quad (8)$$

We set the parameters of our model to scale as follows with τ :

$$\beta = e^{-\rho\tau}, \quad q = e^{-\rho\tau}, \quad \mu_t^{\log}(\theta_t) = \tau \left[\hat{\mu}_t(\theta_t) - \frac{1}{2}\theta_t\hat{\sigma}_t^2 \right], \quad \sigma_t = \hat{\sigma}_t\sqrt{\tau} \quad (9)$$

for some constants $\rho > 0$, some function of time and productivity $\hat{\mu}_t(\theta_t)$ and some function of time $\hat{\sigma}_t$. To simplify, we assume here that $q = \beta$.

The definition for μ_t^{\log} contains an adjustment term $-\tau\frac{1}{2}\theta_t\hat{\sigma}_t^2$ to ensure that $\mathbb{E}_t[\theta_{t+\tau}] = \theta_t e^{\tau\hat{\mu}_t(\theta_t)}$. Thus, $\hat{\mu}_t(\theta_t)$ can be interpreted as the (instantaneous) conditional expected growth rate in productivity, per unit of time. In the limit as $\tau \rightarrow 0$, it is well known that there exists a Brownian motion W_t such that the stochastic process $\{\theta\}$ converges to the continuous time diffusion $d\theta_t/\theta_t = \hat{\mu}_t(\theta_t) d\theta_t + \hat{\sigma}_t dW_t$. This is the limit that we study directly in Section 4.

Assumption 4 *The stochastic process $\{\theta\}$ satisfies (8)–(9).*

Proposition 5 *Suppose that Assumptions 1, 2, and 4 hold, and that $\{c, y\}$ solves Program FOA.*

Then for every t and history θ^t , then, conditional (on θ^t)

$$\text{plim}_{\tau \rightarrow 0} \frac{\partial}{\partial \theta_{t+\tau}} \left(\frac{\tau_L(\theta^t, \theta_{t+\tau})}{1 - \tau_L(\theta^t, \theta_{t+\tau})} \frac{1}{\hat{u}^{t'}(c(\theta^t, \theta_{t+\tau}))} \right) = 0,$$

where $\text{plim}_{\tau \rightarrow 0}$ denotes the probability limit (conditional on θ^t).

For short period lengths, this result implies that the unpredictable changes in $\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{\hat{u}^{t'}(c_t)}$ are small relative to those in θ_t . As a result, the conditional variance per unit of time $\frac{1}{\tau} \text{Var}_t \left[\frac{\tau_{L,t+\tau}}{1 - \tau_{L,t+\tau}} \frac{1}{\hat{u}^{t'}(c_{t+\tau})} \right]$ is small and vanishes as $\tau \rightarrow 0$. Another implication is that unpredictable changes in $\tau_{L,t}/(1 - \tau_{L,t})$ must mirror those in the marginal utility $\hat{u}^{t'}(c_t)$. This induces a negative conditional covariance between productivity and the labor wedge. Economically, this represents a form of regressivity: good productivity shocks lower both the marginal utility of consumption and the current labor wedge by the same amount, at least in the short run.

To make this result more palatable, note that

$$\frac{\tau_L(\theta^t)}{1 - \tau_L(\theta^t)} \frac{1}{\hat{u}^{t'}(c(\theta^t))} = \frac{1}{\hat{u}^{t'}(c(\theta^t))} - \frac{\theta_t}{\hat{h}'\left(\frac{y(\theta^t)}{\theta_t}\right)}.$$

The first-best allocation has both terms, $1/\hat{u}^{t'}(c_t)$ and $\theta_t/\hat{h}'(y_t/\theta_t)$, constant over time. This is not possible at the constrained-efficient allocation. However, although $1/\hat{u}^{t'}(c_t)$ is not constant, the Inverse Euler equation implies that its changes are unpredictable. Somewhat intuitively, our result implies that changes in $\theta_t/\hat{h}'(y_t/\theta_t)$ should track changes in $1/\hat{u}^{t'}(c_t)$ in an attempt to stabilize the difference.

Why should this be the case? Consider an allocation $\{c, y\}$ that satisfies the constraints of Program FOA, a history θ^t , and two numbers $\underline{\varepsilon}$ and $\bar{\varepsilon}$. Consider the following perturbed allocation $\{c^{\delta, \tau}, y^{\delta, \tau}\}$, which differs from the original allocation $\{c, y\}$ only at histories $(\theta^t, \theta_t + \underline{\varepsilon}\sqrt{\tau})$ and $(\theta^t, \theta_t + \bar{\varepsilon}\sqrt{\tau})$:

$$\begin{aligned} \hat{h}\left(y^{\delta, \tau}(\theta^t, \theta_t + \underline{\varepsilon}\sqrt{\tau})\right) &= \hat{h}\left(y(\theta^t, \theta_t + \underline{\varepsilon}\sqrt{\tau})\right) - (\theta_t + \underline{\varepsilon}\sqrt{\tau})^\alpha \delta, \\ \hat{u}^t\left(c^{\delta, \tau}(\theta^t, \theta_t + \underline{\varepsilon}\sqrt{\tau})\right) &= \hat{u}^t\left(c(\theta^t, \theta_t + \underline{\varepsilon}\sqrt{\tau})\right) - \delta, \end{aligned}$$

and

$$\begin{aligned}\hat{h}\left(y^{\delta,\tau}\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)\right) &= \hat{h}\left(y\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)\right)+\left(\theta+\bar{\varepsilon}\sqrt{\tau}\right)^\alpha\delta, \\ \hat{u}^t\left(c^{\delta,\tau}\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)\right) &= \hat{u}^t\left(c\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)\right)+\delta.\end{aligned}$$

The change in the cost of the allocation $\Psi\left(\left\{c^{\delta,\tau},y^{\delta,\tau}\right\}\right)-\Psi\left(\left\{c,y\right\}\right)$ from this perturbation is proportional to δ . Using the identity $\frac{1}{\hat{u}^{t'}(c)}-\frac{1}{\frac{1}{\theta}\hat{h}'\left(\frac{y}{\theta}\right)}=\frac{\tau_L}{1-\tau_L}\frac{1}{\hat{u}^{t'}(c)}$, the constant of proportionality can be expressed as

$$\frac{\tau_L\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)}{1-\tau_L\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)}\frac{1}{\hat{u}^{t'}\left(c\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)\right)}-\frac{\tau_L\left(\theta^t,\theta_t+\underline{\varepsilon}\sqrt{\tau}\right)}{1-\tau_L\left(\theta^t,\theta_t+\underline{\varepsilon}\sqrt{\tau}\right)}\frac{1}{\hat{u}^{t'}\left(c\left(\theta^t,\theta_t+\underline{\varepsilon}\sqrt{\tau}\right)\right)}.$$

If this perturbation were feasible, then we should set this expression to zero and conclude that a stronger form of Proposition 5 holds.

This argument is heuristic and incomplete, because the perturbed allocation $\left\{c^{\delta,\tau},y^{\delta,\tau}\right\}$ is not feasible: it delivers the same utility as $\left\{c,y\right\}$, but does not preserve the local incentive constraints of Program FOA. However, this basic perturbation can be modified to preserve these constraints. The key to our result is that the corrections are of order τ , while the distance between the two productivity levels $\theta_t+\underline{\varepsilon}\sqrt{\tau}$ and $\theta_t+\bar{\varepsilon}\sqrt{\tau}$ at which we perturb the allocation is of order $\sqrt{\tau}$, so we obtain

$$\begin{aligned}\frac{\tau_L\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)}{1-\tau_L\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)}\frac{1}{\hat{u}^{t'}\left(c\left(\theta^t,\theta_t+\bar{\varepsilon}\sqrt{\tau}\right)\right)} \\ =\frac{\tau_L\left(\theta^t,\theta_t+\underline{\varepsilon}\sqrt{\tau}\right)}{1-\tau_L\left(\theta^t,\theta_t+\underline{\varepsilon}\sqrt{\tau}\right)}\frac{1}{\hat{u}^{t'}\left(c\left(\theta^t,\theta_t+\underline{\varepsilon}\sqrt{\tau}\right)\right)}+O(\tau),\end{aligned}$$

which implies our result. In addition to a proof based on optimal control, the appendix contains a rigorous version of this more constructive perturbation argument, which provides a second proof to Proposition 5.

4 A Continuous Time Approach

In this section, we formulate Program FOA in continuous time and tackle this program using continuous time stochastic control. We assume that the stochastic process for pro-

ductivity is a Brownian diffusion with deterministic volatility:

$$\frac{d\theta_t}{\theta_t} = \hat{\mu}_t(\theta_t)d\theta_t + \hat{\sigma}_t dW_t \quad (10)$$

or equivalently that

$$d \log \theta_t = \hat{\mu}_t^{\log}(\theta_t)d\theta_t + \hat{\sigma}_t dW_t,$$

where $\{W\}$ is a Brownian motion, $\hat{\mu}_t$ is a function of current productivity θ_t which controls the drift of productivity, $\hat{\sigma}_t$ is deterministic function of time which determines the volatility of productivity, and $\hat{\mu}_t^{\log}(\theta_t) = \hat{\mu}_t(\theta_t) - \frac{1}{2}\theta_t^{-2}\hat{\sigma}_t^2$. We assume that the rate of time preferences and the instantaneous interest rates are both equal to $\rho > 0$ (this can be easily generalized to separate the two).

To formulate Program FOA in continuous time, we need to determine the laws of motions for Δ_t and v_t . Our approach here is to work with Assumption 4 and take the continuous time limit of the discrete time model.⁹

We start with Δ_t . Integrate $\Delta_t = \int w_{t+\tau} f_{\theta_t}^t(\theta_{t+\tau}|\theta_t) d\theta_{t+\tau}$ by parts, using Lemma 2 in the appendix, to obtain an expression for $f_{\theta_t}^t(\theta_{t+\tau}|\theta_t)$. Using $\mu_t(\theta_t) = \tau(\hat{\mu}_t(\theta_t) - \frac{1}{2}\theta_t\hat{\sigma}_t^2)$ we obtain

$$\theta_t \Delta_t = \int [\theta_{t+\tau} u_{\theta}^{t+\tau} + e^{-\rho\tau} \theta_{t+\tau} \Delta_{t+\tau}] \left(1 + \theta_t \tau \frac{d\hat{\mu}_t^{\log}}{d\theta} \right) f^t(\theta_{t+\tau}|\theta_t) d\theta_{t+\tau}.$$

This implies that in the continuous time limit, we can write

$$d(\theta_t \Delta_t) = \left[\rho - \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right] (\theta_t \Delta_t) - \theta_t u_{\theta}^t + \tilde{\sigma}_{\Delta,t} \hat{\sigma}_t \theta_t dW_t$$

for some function $\tilde{\sigma}_{\Delta,t}$ of the state variables $(v_t, \Delta_t, \theta_t, t)$. Applying Ito's lemma, we infer that $\{\Delta\}$ solves the following stochastic differential equation:

$$d\Delta_t = \left[\left(\rho - \hat{\mu}_t - \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right) \Delta_t - u_{\theta}^t - \sigma_{\Delta,t} \hat{\sigma}_t \right] dt + \sigma_t^{\Delta} \hat{\sigma}_t dW_t, \quad (11)$$

where $\sigma_{\Delta,t} = \tilde{\sigma}_{\Delta,t} - \Delta_t$.

Turning now to v_t , note that the definition of v_t as the net present value of utility implies that $\{v\}$ solves a differential equation of the form $dv_t = \rho v_t dt - u^t dt + \sigma_{v,t} \sigma_t dW_t$

⁹It is also possible to start with the model in continuous time and derive the relevant first order approach versions of the incentive constraints from scratch.

for some function $\sigma_{v,t}$ of the state variables $(v_t, \Delta_t, \theta_t, t)$. Finally, in the continuous time limit, the constraint $\dot{w}(\theta) = \theta u_\theta^t + \beta \theta \Delta(\theta)$ simply amounts to the requirement that the sensitivity of continuation utility to productivity changes be $\sigma_{v,t} = \theta_t \Delta_t$. Therefore, $\{v\}$ solves the following differential equation:

$$dv_t = \rho v_t dt - u^t dt + \theta_t \Delta_t \hat{\sigma}_t dW_t. \quad (12)$$

Having re-expressed the constraints in Program FOA as stochastic differential equations for the state variables, we can now write down the Hamilton-Jacobi-Bellman (HJB) equation for the cost function $K(v_t, \Delta_t, \theta_t, t)$. The states are $(v_t, \Delta_t, \theta_t, t)$ with laws of motion given by equations (12), (11), and (10). The controls are $(c_t, y_t, \sigma_{\Delta,t})$. The HJB equation is

$$\begin{aligned} \rho K = \max_{c_t, y_t, \sigma_{\Delta,t}} & \left\{ [c_t - y_t] + K_v [\rho v_t - u^t] + K_\Delta \left[\left(\rho - \hat{\mu}_t - \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right) \Delta_t - \sigma_{\Delta,t} \hat{\sigma}_t - u_\theta^t \right] \right. \\ & + K_\theta \theta_t \hat{\mu}_t + K_t + \frac{1}{2} K_{vv} \theta_t^2 \Delta_t^2 \hat{\sigma}_t^2 + \frac{1}{2} K_{\Delta\Delta} \sigma_{\Delta,t}^2 \hat{\sigma}_t^2 + \frac{1}{2} K_{v\theta} \theta_t^2 \Delta_t^2 \hat{\sigma}_t^2 + \frac{1}{2} K_{\Delta\Delta} \sigma_{\Delta,t}^2 \hat{\sigma}_t^2 \\ & \left. + \frac{1}{2} K_{\theta\theta} \theta^2 \hat{\sigma}_t^2 + K_{v\Delta} \theta_t \Delta_t \sigma_{\Delta,t} \hat{\sigma}_t^2 + K_{v\theta} \theta_t^2 \Delta_t \sigma_\theta^2 + K_{\Delta\theta} \theta_t \sigma_{\Delta,t} \hat{\sigma}_t^2 \right\}. \end{aligned}$$

It will prove convenient to introduce the dual variables of $(v_t, \Delta_t) : \lambda(v_t, \Delta_t, \theta_t, t) = K_v(v_t, \Delta_t, \theta_t, t)$ and $\gamma(v_t, \Delta_t, \theta_t, t) = K_\Delta(v_t, \Delta_t, \theta_t, t)$.

Proposition 6 *Suppose that Assumptions 1 and 2 hold, and that productivity evolves according to equation (10). Then:*

- i. *There exists a function $\sigma_\lambda(v_t, \Delta_t, \theta_t, t)$ such that the stochastic processes for $\{\lambda\}$ and $\{\gamma\}$ verify the following stochastic differential equations*

$$\frac{d\lambda_t}{\lambda_t} = \sigma_{\lambda,t} \hat{\sigma}_t dW_t \quad (13)$$

$$d\gamma_t = \left[-\theta_t \lambda_t \sigma_{\lambda,t} \hat{\sigma}_t^2 + \left(\hat{\mu}_t + \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right) \gamma_t \right] dt + \gamma_t \hat{\sigma}_t dW_t, \quad (14)$$

with $\gamma_0 = 0$.

- ii. *Consumption c_t and output y_t can be computed as follows:*

$$\frac{1}{\hat{u}^{t'}(c_t)} = \lambda_t \quad \text{and} \quad \frac{1}{\hat{u}^{t'}(c_t)} - \frac{\theta_t}{h^{t'}(y_t/\theta_t)} = -\alpha \frac{\gamma_t}{\theta_t}.$$

iii. The labor and intertemporal wedges, $\tau_{L,t}$ and $\tau_{K,t}$, can be computed as follows:

$$\frac{\tau_{L,t}}{1 - \tau_{L,t}} = -\alpha \frac{\gamma_t}{\lambda_t} \frac{1}{\theta_t} \quad \text{and} \quad \tau_{K,t} = \sigma_{\lambda,t}^2 \hat{\sigma}_t^2.$$

Part (i) may be used as follows. If the functions $\lambda(v_t, \Delta_t, \theta_t, t)$ and $\gamma(v_t, \Delta_t, \theta_t, t)$ can be inverted for (v_t, Δ_t) , then an alternative state space is $(\lambda_t, \gamma_t, \theta_t, t)$. In this case, we can write $\sigma_\lambda(\lambda_t, \gamma_t, \theta_t, t)$. Equations (13)–(14) then provide the evolution of these alternative state variables. Part (ii) and (iii) then offer a way to compute the allocation and wedges as a function of $(\lambda_t, \gamma_t, \theta_t, t)$.

Combining parts (i) and (iii) leads to the following corollary.

Corollary 2 *Suppose that Assumptions 1 and 2 hold, and that productivity evolves according to equation 10. Then the labor wedge verifies the following stochastic differential equation*

$$d \left(\lambda_t \frac{\tau_{L,t}}{1 - \tau_{L,t}} \right) = \left[\alpha \lambda_t \sigma_{\lambda,t} \hat{\sigma}_t^2 + \lambda_t \frac{\tau_{L,t}}{1 - \tau_{L,t}} \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta_t} \right] dt. \quad (15)$$

This result is the continuous time extension of Proposition 5. A bit more can be said here, because in continuous time the process $\left\{ \lambda \frac{\tau_L}{1 - \tau_L} \right\}$ is a diffusion. The fact that it has zero instantaneous volatility implies that the realized paths will vary much less than those for productivity $\{\theta\}$, in the sense that they are (a.s.) of bounded variation.

Combining (5) with (15) and applying Ito's lemma, yields the following formula:

$$d \left(\frac{\tau_{L,t}}{1 - \tau_{L,t}} \right) = \left[\alpha \sigma_{\lambda,t} \hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1 - \tau_{L,t}} \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} + \frac{\tau_{L,t}}{1 - \tau_{L,t}} \sigma_{\lambda,t}^2 \hat{\sigma}_t^2 \right] dt - \frac{\tau_{L,t}}{1 - \tau_{L,t}} \sigma_{\lambda,t} \hat{\sigma}_t dW_t. \quad (16)$$

This is the continuous time counterpart of equation (5) with the particular weighting function $\pi(\theta_t) = (\beta/q) \hat{u}^{t'}(c(\theta^t)) / \hat{u}^{t-1'}(c(\theta^{t-1}))$.¹⁰ It describes the dynamics for $\left\{ \frac{\tau_L}{1 - \tau_L} \right\}$ with the original, unadjusted probability. Unlike equation (5) which focuses on the conditional expectation, this equation describes not only the drift, but also the volatility of $\left\{ \frac{\tau_L}{1 - \tau_L} \right\}$.

The drift of $\left\{ \frac{\tau_L}{1 - \tau_L} \right\}$ contains the elements discussed after Proposition 2 (the covariance and auto-regressive term) and a new term given by $\frac{\tau_{L,t}}{1 - \tau_{L,t}} \sigma_{\lambda,t}^2 \hat{\sigma}_t^2$. One can interpret the term $-\theta_t \frac{d\hat{\mu}_t}{d\theta} - \sigma_{\lambda,t}^2 \hat{\sigma}_t^2$ as an instantaneous mean reversion coefficient. In this sense, this stochas-

¹⁰Corollary 2 also sheds light on the general discrete time formula in equation (5). Indeed, for any twice differentiable function $\pi(\theta_t)$, one can apply Ito's lemma to infer the drift of $\pi_t \lambda_t \frac{\tau_{L,t}}{1 - \tau_{L,t}}$ and obtain a continuous time counterpart of equation (5).

tic process features less mean-reversion than productivity. The higher is the variance of consumption growth, the larger is the discrepancy.

It is also possible to rearrange this equation as

$$d\left(\frac{\tau_{L,t}}{1-\tau_{L,t}}\right) = \left[\alpha\sigma_{\lambda,t}\hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1-\tau_{L,t}}\theta_t\frac{d\hat{\mu}_t^{\log}}{d\theta}\right]dt + \frac{\tau_{L,t}}{1-\tau_{L,t}}\lambda_t d\left(\frac{1}{\lambda_t}\right).$$

This version shows explicitly how the innovations in the labor wedge must be perfectly mirrored by those in the marginal utility of consumption $\hat{u}''(c_t) = \lambda_t^{-1}$. This formulation neatly highlights the regressivity result alluded to in subsection 3.4.

It may seem surprising that we obtain a regressivity result in our dynamic model, especially in light of the absence of such a result in a static setting. As is well understood, the skill distribution is key in shaping the tax schedule in the static model (Mirrlees (1971); Diamond (1998); Saez (2001)). In contrast, in our dynamic model, the regressivity result holds for a large class of productivity processes.

It is important to stress what this regressivity result does and does not say. Over short horizons, it induces a negative conditional correlation between consumption and the labor wedge. However, whether or not this translates into a negative correlation over longer horizons depends on the evolution of the drift term. In particular, the endogenous volatility term $\sigma_{\lambda,t}$ may play a central role. We investigate these dynamics more explicitly in Section 6.

Using part (iii) in Proposition 6, we can solve for the volatility $\sigma_{\lambda,t}$ in terms of the intertemporal wedge: $\sigma_{\lambda,t} = \frac{\sqrt{\tau_{K,t}}}{\hat{\sigma}_t}$ to rewrite these last three equations in terms of the labor and intertemporal wedges. In this way, optimality can be seen as imposing a joint restriction on the labor and intertemporal distortions.

We close by noting the key role played by the endogenous volatility process $\{\sigma_{\lambda}\}$ in all these equations. Firstly, it represents the loading of consumption changes on productivity innovations. Higher values for $\sigma_{\lambda,t}$ provide more incentives at the expense of insurance. Secondly, by part (i) in Proposition 6, it essentially summarizes the entire allocation by determining the evolution of the alternative endogenous states (λ_t, γ_t) . Finally, it is featured prominently in the equations describing the labor and savings wedges. It pins down the value $\sigma_{\lambda,t}^2\hat{\sigma}_t^2$ of the savings wedge $\tau_{K,t}$ and scales the drift $\alpha\lambda_t\sigma_{\lambda,t}\hat{\sigma}_t^2$ of $\{\lambda_{\frac{\tau_L}{1-\tau_L}}\}$. Section 6 exploits the fundamental role of $\{\sigma_{\lambda}\}$ to interpret our numerical findings.

5 General Preferences

In this section, we investigate what can be said for general utility functions $u^t(c, y, \theta)$. It is well known that when consumption and labor are not additively separable, the Inverse Euler condition does not necessarily hold—we lose Proposition 1. However, as we show next, we still can make progress generalizing Corollary 1.

Recall that the expenditure function $C^t(y, u, \theta)$ is the inverse of $u^t(\cdot, y, \theta)$. Define

$$\eta_t(y, w, \theta) \equiv \frac{-\theta C_{y\theta}^t(y, w, \theta)}{C_y^t(y, w, \theta)}.$$

Since $C_y^t = -u_y^t/u_c^t = |MRS_t| = 1 - \tau_{L,t}$ is the marginal rate of substitution, η_t represents the elasticity $-\frac{d \log |MRS_t|}{d \log \theta_t}$. It plays a key role below. Note that in the separable isoelastic utility case (Assumptions 1–2) that we studied above, this elasticity is constant with $\eta_t(y, w, \theta) = \alpha$.

Discrete Time. In order to generalize equation (6), we need to introduce the dual of the variable $v(\theta^t)$ defined by

$$\lambda(\theta^t) \equiv K_v(v(\theta^t), \Delta(\theta^t), \theta_t, t + 1).$$

At an optimum, we have the martingale relation

$$\lambda(\theta^{t-1}) = \frac{q}{\beta} \int \lambda(\theta^t) f^t(\theta_t | \theta_{t-1}) d\theta_t.$$

Below we adopt the shorthand notation of writing $x_t(\theta_t)$ for any function $x_t(c(\theta^t), y(\theta^t), \theta_t)$ (see η_t and u_c^t below).

Proposition 7 *Suppose that $\{c, y\}$ solves Program FOA. Then the labor wedge satisfies the following equation for every $t \geq 1$ and history θ^{t-1}*

$$\begin{aligned} \int \frac{1}{\eta_t(\theta^t)} \frac{\tau_L(\theta^t)}{1 - \tau_L(\theta^t)} \frac{q}{\beta} \frac{u_c^t(\theta^{t-1})}{u_c^t(\theta^t)} f^t(\theta_t | \theta_{t-1}) d\theta_t &= \frac{1}{\eta_{t-1}(\theta^{t-1})} \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}} \\ &+ u_c^t(\theta^{t-1}) \lambda(\theta^{t-1}) \int \log(\theta_t) \left[\frac{q}{\beta} \frac{\lambda(\theta^t)}{\lambda(\theta^{t-1})} - 1 \right] f^t(\theta_t | \theta_{t-1}) d\theta_t. \end{aligned} \quad (17)$$

This proposition generalizes equation (6). The martingale relation satisfied by $\lambda(\theta^t)$ implies that we can rewrite the second term on the right-hand side of equation (17) as a

covariance: $\text{Cov}_{t-1}(\log(\theta_t), \frac{q}{\beta} \frac{\lambda_t}{\lambda_{t-1}})$. Note that we could also generalize equation (5) along the exact same lines.

There are two important differences between equations (6) and (17). First, note that unless utility is separable—Assumption 1 holds—we no longer have $\frac{q}{\beta} \lambda(\theta^t) = \frac{1}{u_c^t(\theta^t)}$. As a result, $\frac{1}{u_c^t(\theta^t)}$ is no longer a martingale and, by contrast with equation (6), the term $\frac{q}{\beta} \frac{u_c^t(\theta^{t-1})}{u_c^t(\theta^t)}$ cannot be interpreted as a change of measure.

Second, note that $\frac{1}{\eta_t(\theta^t)} \frac{\tau_L(\theta^t)}{1-\tau_L(\theta^t)}$ replaces $\frac{\tau_L(\theta^t)}{1-\tau_L(\theta^t)}$. When Assumptions 1 and 5 hold, $\eta_t = \alpha$, we can multiply through by α , as in equation (6). Otherwise, the general equation indicates that changes in the elasticity should affect the labor wedge. To elaborate on this point, it will prove convenient to specialize the discussion to a class of generalized isoelastic preferences for which equation (17) takes a simpler form.

Assumption 5 *For every $t \geq 0$, there exists functions \tilde{u}^t , \hat{u}^t , and constants $\kappa_t > 0$, $\alpha_t > 1$, such that $u^t(c, y, \theta) = \tilde{u}^t\left(\hat{u}^t(c) - \frac{\kappa_t}{\alpha_t} \left(\frac{y}{\theta}\right)^{\alpha_t}\right)$.*

For this class of preferences, we have $\eta_t(y, w, \theta) = \alpha_t$.

Corollary 3 *Suppose that Assumption 5 holds and that $\{c, y\}$ solves Program FOA. Then the labor wedge satisfies the following equation for every $t \geq 1$ and history θ^{t-1}*

$$\int \frac{\tau_L(\theta^t)}{1-\tau_L(\theta^t)} \frac{q}{\beta} \frac{u_c^t(\theta^{t-1})}{u_c^t(\theta^t)} f^t(\theta_t|\theta_{t-1}) d\theta_t = \frac{\alpha_t}{\alpha_{t-1}} \frac{\tau_L(\theta^{t-1})}{1-\tau_L(\theta^{t-1})} \theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}} + \alpha_t u_c^t(\theta^{t-1}) \lambda(\theta^{t-1}) \int \log(\theta_t) \left[\frac{q}{\beta} \frac{\lambda(\theta^t)}{\lambda(\theta^{t-1})} - 1 \right] f^t(\theta_t|\theta_{t-1}) d\theta_t. \quad (18)$$

Other things being equal, we see that increases in α_t (decreases in the Frisch elasticity of labor supply) increase the autoregressive coefficient $\frac{\alpha_t}{\alpha_{t-1}} \theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}}$ of the labor wedge. In other words, a decrease in the Frisch elasticity of labor supply tends to increase the labor wedge. This is a manifestation of a standard inverse elasticity principle: other things equal, labor is taxed more in periods in which it is less elastic.

Continuous Time. Our continuous time analysis can also be extended to general preferences. In particular, we can generalize equation (19).

Proposition 8 *Suppose that productivity evolves according to equation (10). Then the labor wedge satisfies the following stochastic differential equation*

$$d \left(\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u_c^t} \frac{1}{\eta_t} \right) = \left[\lambda_t \sigma_{\lambda,t} \hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u_c^t} \frac{1}{\eta_t} \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta_t} \right] dt. \quad (19)$$

This expression is the continuous time analogue of equation (17). Note that our no-volatility result generalizes: the stochastic process $\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u_c^t} \frac{1}{\eta_t}$ has zero instantaneous volatility so that its realized paths will vary much less than those for productivity $\{\theta\}$, in the sense that they are (a.s.) of bounded variation. Equation (19) takes a simple and illuminating form when preferences are in the generalized isoelastic class defined by Assumption 5.

Corollary 4 *Suppose that Assumption 5 holds and that productivity evolves according to equation (10). Then the labor wedge satisfies the following stochastic differential equation*

$$d \left(\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u_c^t} \right) = \left[\alpha_t \lambda_t \sigma_{\lambda,t} \hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u_c^t} \left(\theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta_t} + \frac{1}{\alpha_t} \frac{d\alpha_t}{dt} \right) \right] dt. \quad (20)$$

Equation (20) clearly shows how a time-varying Frisch elasticity of labor supply impacts the autoregressive coefficient of $\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u_c^t}$. Increases in α_t (decreases in the Frisch elasticity of labor supply) tend to increase the labor wedge.

We can also derive a generalization of equation (16)

$$d \left(\frac{\tau_{L,t}}{1 - \tau_{L,t}} \right) = \left[\alpha_t \lambda_t \sigma_{\lambda,t} \hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u_c^t} \left(\theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta_t} + \frac{1}{\alpha_t} \frac{d\alpha_t}{dt} \right) \right] dt + \frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{1}{u_c^t} d(u_c^t).$$

This shows explicitly how innovations in the labor wedge must be perfectly mirrored by those in the marginal utility of consumption u_c^t .

6 The Model At Work: A Numerical Solution

In this section we parametrize the model and solve it numerically. This serves to illustrate some of our theoretical results, but also leads to some new insights.

A Life Cycle Economy. Agents live for $T = 60$ years, working for 40 years and then retiring for 20 years. Their period utility function is

$$\log(c_t) - \frac{\kappa}{\alpha} \left(\frac{y_t}{\theta} \right)^\alpha$$

with $\alpha > 1$ and $\kappa > 0$ during working years $t = 1, 2, \dots, 40$ and

$$\log(c_t)$$

during retirement $t = 41, 42, \dots, 60$. We set $\alpha = 3$ implying a Frisch elasticity for labor of 0.5, and $\kappa = 1$. We set the agent's and planner's discount factors equal to each other at $q = \beta = 0.95$.

For productivity, we follow [Storesletten, Telmer, and Yaron \(2004\)](#). They estimate earnings to fit an AR(1) plus white noise.¹¹ The coefficient of auto correlation is very close to one, leading them to adopt a geometric random walk:

$$\theta_t = \varepsilon_t \theta_{t-1},$$

with $\log \varepsilon \sim N(-\frac{\hat{\sigma}^2}{2}, \hat{\sigma}^2)$. We truncate the normal distribution: the density of $\log \varepsilon$ is proportional to the density of the normal over a finite interval $[\underline{\varepsilon}, \bar{\varepsilon}]$. We adjust $\hat{\sigma}^2$ so that $\text{Var}(\log \varepsilon) = 0.0161$ the value reported in [Storesletten, Telmer, and Yaron \(2004\)](#) for the variance of the log changes in earnings.

Under these conditions, the value function satisfies

$$K(v, \Delta, \theta_-, t) = \theta_- K(v - (1 + \beta + \dots + \beta^{T-t}) \log \theta_-, \theta_- \Delta, 1, t).$$

This holds because if $\{c_t, y_t\}$ is feasible given (v, Δ, θ_-) and has cost k , then, due to balanced growth preferences, it follows that $\{\phi c_t, \phi y_t\}$ is feasible given $(v + (1 + \beta + \dots + \beta^{T-t}) \log \phi, \phi^{-1} \Delta, \phi \theta_-)$ and costs ϕk . Setting $\phi = 1/\theta_-$ then yields the desired property for K . A similar homogeneity condition holds for the policy functions. These properties reduce the dimensionality of our problem.

After computing policy functions, we iterate on them to produce a Montecarlo simulation with 1 million agents evolving through periods $t = 1, 2, \dots, T$. For any given v_1 , we initialize Δ_1 at $t = 1$ to minimize cost

$$\Delta_1 \in \arg \min_{\Delta} K(v_1, \Delta, 1, 1).$$

We set the initial value for utility v_1 so that the resulting cost is zero, $K(v_1, \Delta_1, 1, 1) = 0$.

¹¹[Storesletten, Telmer, and Yaron \(2004\)](#) estimate earnings, not wages or productivity. calibrating the productivity process with the earnings process is an approximation. As we discuss below, in logs, earnings are linear in productivity at the first best solution as well as in autarky.

Two Benchmarks. Before discussing the results of our simulations, it is useful to consider two benchmark allocations, those corresponding to autarky and the first best.

Consider first an autarkic situation, where there are no taxes. Agents can consume their own production. They can neither borrow nor save. Thus, they solve the static maximization: $\max_y u(y, y; \theta)$. With logarithmic utility, or more generally with balanced growth preferences, this implies $c_t = y_t = \theta_t \bar{n}$ for some constant level of work effort, defined by the solution to $u_c(\bar{n}, \bar{n}; 1) = -u_y(\bar{n}, \bar{n}; 1)$. Consumption and output are geometric random walks: $c_t = \varepsilon c_{t-1}$ and $y_t = \varepsilon y_{t-1}$. The labor wedge is zero and the inter-temporal wedge is a positive constant equal to $1 - Rq = 1 - R\beta > 0$, where $R \equiv \beta^{-1}(\mathbb{E}_\varepsilon \frac{1}{\varepsilon})^{-1}$.¹²

Consider next the first-best planning problem given by:

$$\max_{\{c, y\}} \mathbb{E}_0 \sum_{t=1}^T u(c_t, y_t; \theta_t) \quad \text{s.t.} \quad \mathbb{E}_0 \sum_{t=1}^T q^t (c_t - y_t) \leq e,$$

for some constant $e \in \mathbb{R}$, representing outside resources available to the planner. The optimum features perfect insurance, with constant consumption $c_t = \bar{c}$ and constant marginal disutility of work, implying

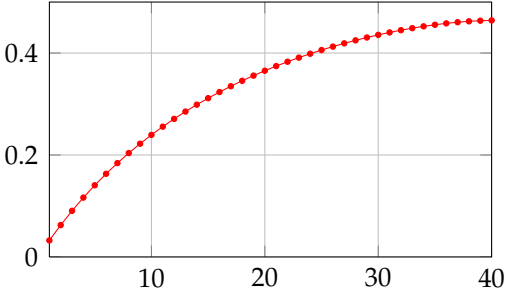
$$y_t = \left(\frac{\alpha}{\kappa \bar{c}} \theta_t \right)^{\frac{1}{\alpha-1}}.$$

Output is increasing in productivity, with Frisch elasticity $1/(\alpha - 1) > 0$. Both the labor and intertemporal wedges are zero.

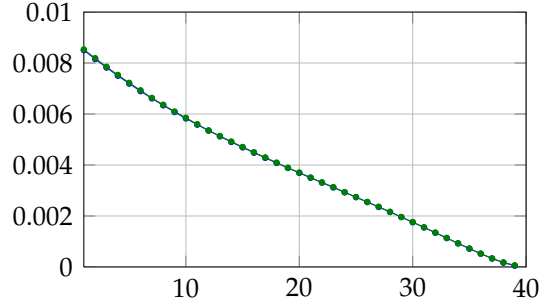
Findings from Simulation. Within each period t , we compute the average in the cross section for a number of variables of interest, such as consumption, output, and the labor and intertemporal wedges. During retirement each agent's consumption is constant, while output and wedges are zero. Thus, we focus on the working periods $t = 1, 2, \dots, 40$.

Although our simulations are for the discrete time model, with a period representing a year, our results from the continuous time version turn out to provide an excellent explanation for our findings. In particular, Proposition 6 shows that the optimum is summarized by the volatility process $\{\sigma_\lambda\}$, since this determines the laws of motion for wedges, consumption and output. With logarithmic utility, the instantaneous variance of consumption growth is given by $\sigma_{\lambda,t}^2 \hat{\sigma}^2$. Figure 1 panel (b) plots the average variance of

¹²Alternatively, in the case with no retirement, this allocation can also be sustained as an equilibrium where $q = R^{-1}$ (instead of $q = \beta$) and agents can freely save and borrow. The intertemporal wedge in this latter case is zero. This serves to make the point that the sign of the intertemporal wedge is somewhat uninteresting, because it depends on the value of various parameters, including q . Another way to proceed is to define autarky as allowing agents to borrow and save at rate q , in which case all wedges are zero by definition, but, unless there is no retirement and $q = R^{-1}$, we would be unable to solve the equilibrium in closed form.



(a) Average labor wedge



(b) Intertemporal wedge (blue) and variance of consumption growth (green)

Figure 1: Average wedges over time

consumption growth in our simulation $\text{Var}_t[c_{t+1}/c_t]$. This is the discrete time counterpart of $\sigma_{\lambda,t}^2 \hat{\sigma}^2$.

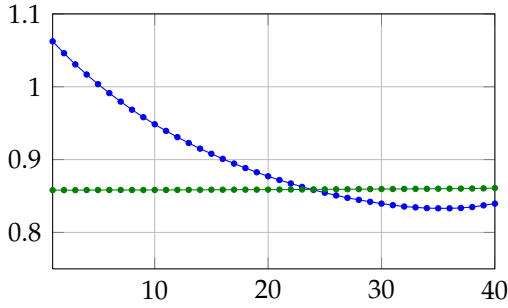
As the figure shows, the average variance of consumption growth falls over time and reaches zero at retirement. There are two key forces at play. First, as retirement nears, productivity shocks have a smaller effect on the present value of earnings, since they affect earnings for fewer periods. Since consumption is smoothed over the entire lifetime, including retirement, the impact of shocks on consumption falls and approaches zero at retirement. This is the usual permanent income mechanism. Indeed, this property would be present at an equilibrium with no taxes and free savings. Second, as we show below, the labor wedge is increasing over time. This provides increased insurance, in the sense of lowering the effect of productivity shocks on net earnings.

The decreasing pattern towards zero in the average variance of consumption growth will be key in understanding a number of results presented below.

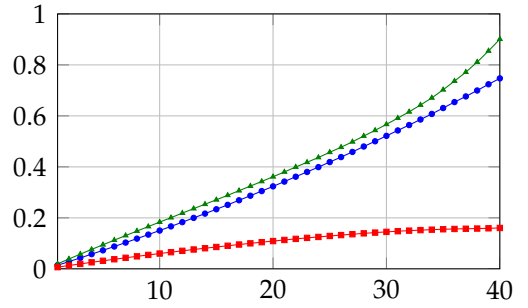
Turning to the wedges, panel (a) in Figure 1 shows that the labor wedge starts near zero and increases over time, asymptoting around 46% at retirement. Panel (b) displays the intertemporal wedge, which displays the reverse pattern. It is decreasing over time, starting around 1%—which represents an implicit tax on net interest of around 17%—and falling to zero at retirement.¹³ Both of these findings are easily explained by our theoretical results, together with the behavior of the average variance of consumption growth.

As shown in equations (7) and (16), when $\rho = 1$, the expected change in the labor

¹³To put these magnitudes in perspective, recall that the intertemporal wedge represents an implicit tax on the gross rate of return to savings. In this interpretation, agents perceive a gross interest of $(1 - \tau_{K,t})(1 + r)$ instead of $(1 + r)$, where $1 + r = q^{-1}$. An equivalent reduction in the gross interest rate can be obtained by an implicit tax $\hat{\tau}_{K,t}$ on net interest r given by $1 + (1 - \hat{\tau}_{K,t})r$. Setting, $1 + (1 - \hat{\tau}_{K,t})r = (1 - \tau_{K,t})(1 + r)$ gives $\hat{\tau}_{K,t} = \frac{1+r}{r} \tau_{K,t}$. In our case, $q = 0.95$, so that $\frac{1+r}{r} \approx \frac{1}{r} \approx 20$.



(a) Average for output (blue) and consumption (green)



(b) Variance of output (green), productivity (blue) and consumption (red)

Figure 2: Statistics for optimal allocation over time

tax is proportional to the covariance of consumption growth with the log of productivity, which is positive, in order to provide incentives. This explains the increasing pattern in the average wedge. The covariance equals $\sigma_{\lambda,t}\hat{\sigma}^2$ in the continuous time limit. Then, since $\sigma_{\lambda,t}^2\hat{\sigma}^2$ decreases over time to zero, so does $\sigma_{\lambda,t}\hat{\sigma}^2$, explaining the asymptote in the labor wedge at retirement.

As for the intertemporal wedge, equation (4) implies that it is increasing in the uncertainty of consumption growth, in the sense that a mean-preserving spread leads to an increase in the wedge. In the continuous time limit the intertemporal wedge equals the variance of consumption growth: $\tau_{K,t} = \sigma_{\lambda,t}^2\hat{\sigma}^2$. Indeed, although panel (b) plots both the variance of consumption growth and the intertemporal wedge, the two are indistinguishable to the naked eye. More generally, while we simulate the discrete time version of the model, with a period representing a year, the continuous time formulas turn out to provide excellent approximations for our findings.

Figure 2 shows the evolution over time for the cross-sectional means and variances of the allocation. Panel (a) shows that average consumption is perfectly flat. This is expected given the Inverse Euler condition, which with logarithmic utility is $(q/\beta)\mathbb{E}_{t-1}[c_t] = c_{t-1}$. Output, on the other hand, is mostly decreasing, consistent with the increasing pattern in the labor wedge.¹⁴

Panel (b) shows the cross-sectional variance for consumption, productivity and output. The variance of productivity grows, by assumption, linearly. The variance of output

¹⁴Note that average output can also be affected by the increasing dispersion in productivity. For example, in a first best solution, output would be proportional to $\theta_t^{\alpha-1}$. When $\alpha < 2$ this function is concave inducing a decreasing pattern. The reverse is true when $\alpha > 2$. In our case $\alpha = 2$ so the increasing dispersion in productivity would not have an effect on average output at the first best solution. An autarkic solution, without taxes and where agents consume their current output (i.e. with no savings or with q set at a level that induces no savings), would feature constant output regardless of the value of α .

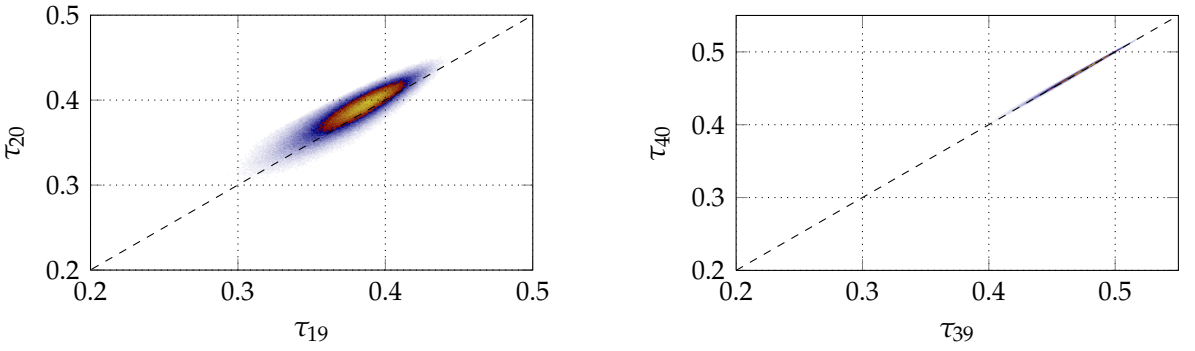


Figure 3: Scatter of $\tau_{L,t}$ vs. $\tau_{L,t-1}$

is higher and grows in a convex manner. The variance of consumption, on the other hand, is lower than the variance of productivity and grows in a concave manner. For reference, note that in autarky, with no taxes and no savings, since $c = y \sim \theta$, the variance for consumption, output and productivity are equal to each other. At the other end of the spectrum, the first best solution has zero variance in consumption and since $y_t \sim \theta_t^{\frac{1}{\alpha-1}}$, the variance for output is higher than that of productivity and grows in a convex manner. The planner's solution, in contrast, partially insures productivity shocks and lies between these two benchmarks.

The degree of insurance is nicely illustrated by the lower variance of consumption, relative to that of output and productivity. Over time, the variance for consumption rises, and does so in a concave fashion. Recall that consumption is a martingale, which implies that inequality must rise. As we discussed above, over time the variance in consumption growth falls and reaches zero at retirement, explaining the concave shape.

Figure 3 illustrates the intertemporal labor wedge formula by showing scatter plots of the current labor wedge against the previous period's labor wedge. In period $t = 20$, the average relationship is close to linear with a slope near one and lies above the 45 degree line. Both of these properties are consistent with our formula in equation (7). The average tax in the current period lies slightly above the previous period's, illustrating the positive drift in taxes.

In the last working period, $t = 40$, the scatter plot shows an almost perfect relationship between the previous tax and the current one, with a slope of one. Taxes on labor are almost perfectly smoothed near retirement. Recall that the variance of consumption growth drops to zero as retirement approaches. This explains why the average relationship is essentially the 45 degree line. The reason there is no dispersion around the average relationship is an implication of the results in Sections 4 and 3.4 that show that unpredictable changes in the labor wedge are related to unpredictable changes in marginal utility. Near

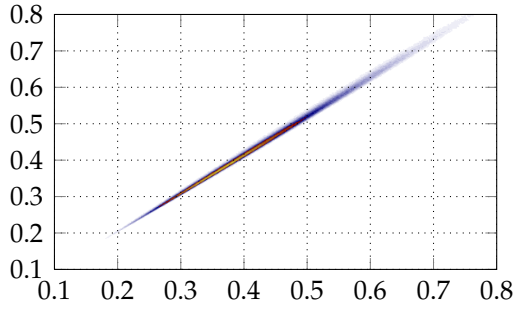


Figure 4: Scatter of $\frac{\tau_{L,t}}{1-\tau_{L,t}} u'(c_t)$ against $\frac{\tau_{L,t-1}}{1-\tau_{L,t-1}} u'(c_{t-1})$ for $t = 20$.

retirement, consumption becomes almost perfectly predictable, so the labor wedge does as well.

To illustrate this point further, Figure 4 plots $\frac{\tau_{L,t}}{1-\tau_{L,t}} u'(c_t)$ against $\frac{\tau_{L,t-1}}{1-\tau_{L,t-1}} u'(c_{t-1})$ for $t = 20$. The average relationship is slightly above the 45 degree line and the dispersion around this relationship is minimal. This illustrates the results in Sections 3.4 and 4, that there is no instantaneous volatility in $\left\{ \frac{\tau_L}{1-\tau_L} \frac{1}{u'(c)} \right\}$. In other words, unpredictable changes in the labor wedge $\left\{ \frac{\tau_L}{1-\tau_L} \right\}$ are entirely explained by unpredictable changes in the reciprocal of marginal utility $\left\{ \frac{1}{u'(c)} \right\}$.

Figure 5 panel (a) plots the current period's labor wedge $\tau_{L,t}$ against the productivity θ_t for period $t = 20$. On average, tax rates are higher for agents with low productivity. In this sense, the tax system is regressive. What accounts for this finding? In a static setting, it is well known that the pattern of taxes is dependent, among other things, on the distribution of productivity shocks (Diamond, 1998; Saez, 2001). We have assumed a log-normal distribution for the productivity shocks. In our dynamic context, however, it is less obvious whether this particular choices is responsible for the regressive pattern we find. Indeed, the results in Sections 3.4 and 4 point towards a negative correlation between the labor wedge and productivity, at least in the short run.

The figure also shows that, for any given level of current productivity, there is significant dispersion in the labor wedge. If the labor wedge were solely a function of current productivity, then there would be no dispersion. Thus, this dispersion illustrates the history dependence in the labor wedge. Recall that the allocation and wedges depend on the history of shocks as summarized by our two state variables v and Δ .

It is important to keep in mind, that a history independent tax system, with a fixed non-linear tax schedule that allows for savings, can also produces a history dependent labor wedge. The history of productivity shocks affects savings decisions. The accumulated wealth, in turn, affects the current labor choice, determining the position, and marginal tax rate, along the fixed non-linear tax schedule.

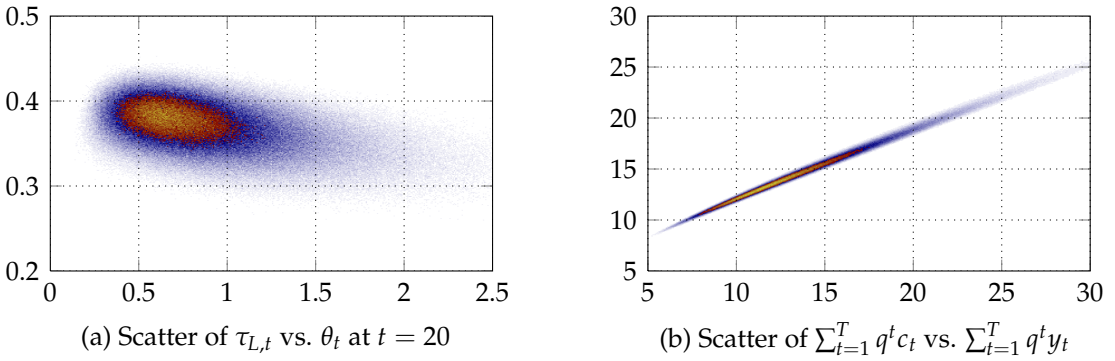


Figure 5: History dependence and Insurance

Figure 5 panel (b) gets at a measure of the overall degree of insurance by plotting the realized present value of consumption $\sum_{t=1}^T q^{t-1} c_t$ against the present value of output $\sum_{t=1}^T q^{t-1} y_t$ in the simulation. Without taxes there is no insurance and $\sum_{t=1}^T q^{t-1} c_t$ would vary one for one with $\sum_{t=1}^T q^{t-1} y_t$. Insurance makes the present value of consumption $\sum_{t=1}^T q^{t-1} c_t$ vary less than one for one with the present value of income $\sum_{t=1}^T q^{t-1} y_t$. The scatter shows that at the optimum there is a near linear relationship, with a slope around 0.67. For reference, a linear tax with a rate of 33% would produce an exact linear relationship with this slope.

We have performed some comparative statics and welfare analysis which we report briefly now.

A fundamental parameter in our exercise is the degree of uncertainty. Unfortunately, empirical estimates for the earnings process vary quite substantially, due to differences in methodologies, econometric specifications and data sets. In addition, an important consideration is the degree to which agents have better information than the econometrician, allowing them to forecast future earnings and productivity better. This suggests that agents face lower uncertainty than what is implied by the available empirical estimates.

Motivated by these considerations, we solved for an economy with lower uncertainty, with the same parameters as our baseline, but with the standard deviation $\hat{\sigma}^2 = 0.0061$. The results show the same qualitative patterns as the benchmark. Quantitatively, both the labor and intertemporal wedges are lower, with the labor wedge peaking at 30% and the intertemporal wedge starting at 0.45%—which represents an implicit tax on net interest of around 9%. With lower uncertainty the optimum features lower insurance and distortions. These results are consistent with our formulas, which stress the role that the degree of uncertainty, captured by $\hat{\sigma}$, has in determining both the labor and intertemporal wedges.

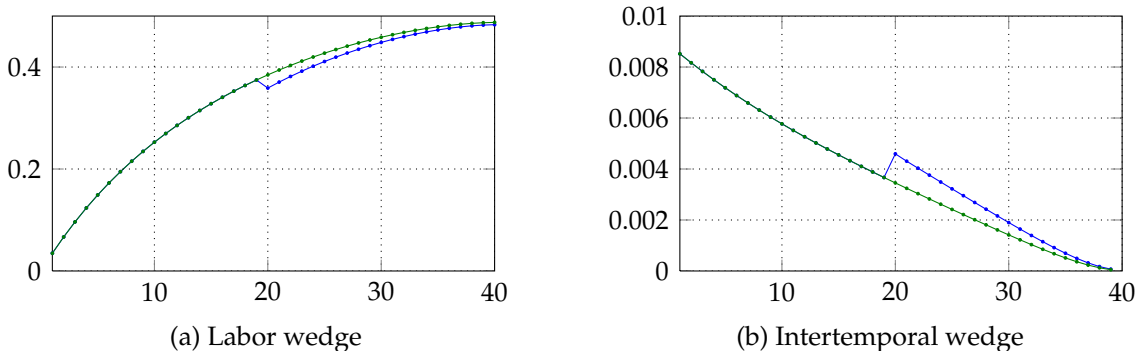


Figure 6: Impulse response to shock at $t = 20$.

Labor Wedge Dynamics: An Impulse Response. The scatter of the labor wedge at t against the labor wedge at $t - 1$ shown above illustrates the average short-run dynamics implied by our formula. Here, we wish to zoom in more and see how these dynamics play out over longer horizons. To this end, we follow an agent with a productivity realization given by $\varepsilon_t = F^{-1}(1/2)$ for $t \neq 20$ and $\varepsilon_{20} = F^{-1}(0.95)$. We compare this to an agent with $\varepsilon_t = F^{-1}(1/2)$ for all $t = 1, 2, \dots, 40$. We plot the evolution of the labor wedge, and other variables, for these two agents. The difference can be interpreted as the impulse response to a shock at $t = 20$.

Figure 6 shows the evolution of the wedges for these two realizations. Without a shock, the wedges behaves similarly to the averages shown in Figure 1. In contrast, with the shock, we see a downward jump on impact in the labor wedge (consumption, not shown, jumps upward). After the shock, the labor wedge displays a higher rate of growth. In the figure, the labor wedge remains below the path for the no-shock scenario. This feature is not general: we have found that for other values of λ_0 , the path with a shock may jump below but eventually cross and overtake the path without a shock. The higher growth rate in the labor wedge may be enough to overcome the initial jump downward. But why does the labor wedge grow faster after a shock? Panel (b) displays a partial answer: the intertemporal wedge jumps up on impact, due to an increase in the variance of consumption growth. Our formulas indicate that this increases the drift term in the labor wedge. Why does the variance of consumption growth rise? Intuitively, due to partial insurance, the shock raises consumption by less than productivity. As a result, the agent becomes poorer, relatively speaking, and, hence, more susceptible to the fluctuations in productivity.

Welfare. We now compute the welfare gains relative to a situation with no taxes. Our baseline is a market equilibrium without taxes, where agents can save and borrow freely

	$\hat{\sigma}^2 = 0.0161$	$\hat{\sigma}^2 = 0.0061$
second-best	3.43%	0.86%
first-best	13.04%	5.20%

Table 1: Welfare gains over free-savings, no-tax equilibrium.

in a risk-free asset with rate of return q^{-1} . This allocation is easily solved backwards starting at retirement by using the agent's first-order conditions, with zero wedges, and the budget constraints.

In [Table 1](#), we report the welfare gains for the second best, the solution to Program FOA, and the first-best, the optimum with publicly observable productivity. The numbers represent the constant percentage increase, at all dates and histories, in the baseline consumption required to achieve the same utility as the alternative allocation. The first column corresponds to our benchmark value for the conditional variance of productivity $\hat{\sigma}^2$, while the second reports a simulation with half this value. As expected, the welfare gains increase with $\hat{\sigma}^2$.

Comparison with Simple Policies. The second best requires sophisticated history-dependent taxes. If these are not available, how do our results inform us about simpler, history-independent ones? In welfare terms, how well can simpler policies do? These are the questions we explore next.

To this end, we consider history-independent taxes. To simplify the analysis and aid the interpretation, we further restrict taxes to be linear. Since the second best features an important age pattern for taxes, we consider both age-dependent and age-independent taxes.

Optimizing over age dependent taxes is not very tractable numerically, due to the large number of tax variables and the cost of computing the equilibrium for each tax arrangement. In this case, instead of optimizing, we take a hint from the second-best to formulate a sensible choice: we set the tax rates at each age to their cross-sectional averages in the second-best. In contrast, with age-independent taxes there are just two variables, so the problem is numerically tractable. In this case, we compute the optimal age-independent tax rates. There are also intermediate cases, such as age-dependent taxes on labor combined with an age-independent capital tax. In this case, we set the labor tax rates to the corresponding cross-sectional averages in the second-best, but optimize over the constant capital tax rate.

[Table 2](#) below reports the welfare gains over the zero-tax allocation of various simple policies. These are comparable to the numbers in [Table 1](#). Although we perform the

	$\hat{\sigma}^2 = 0.0161$	$\hat{\sigma}^2 = 0.0061$
age-dependent τ_L and τ_K	3.30%	0.71%
age-dependent τ_L , and $\tau_K = 0$	3.16%	0.66%
age-dependent τ_L , age-independent τ_K	3.29%	0.70%
age-independent τ_L and τ_K	2.71%	0.54%

Table 2: Welfare from simple tax policies.

exercises for two values of $\hat{\sigma}$, since the findings are qualitatively similar in both cases, we will focus our discussion on our benchmark reported in the first column.

The first row reports welfare for an age-dependent linear tax system, where tax rates at each age are set to the cross-sectional average obtained from the second-best simulation. It is surprising just how well this relatively simple policy performs. It delivers a welfare gain of 3.30% in lifetime consumption, compared to the 3.43% obtained by the second best. Remarkably, age-dependent linear taxes deliver 96% of the welfare gains of the second-best.

It is worth repeating that we have not optimized over the age-dependent tax rates. Instead, the tax rates are taken to be the cross-sectional average from the second-best simulation, as in [Figure 2](#). Of course, the fact that welfare comes out to be very close to that of the second best, suggests that this policy is very close to being optimal within the set of simple age-dependent tax policies.¹⁵ We think this illustrates that our characterization of the second best, theoretical and numerical, provides not only useful insights, but can also deliver detailed and surprisingly accurate guidance for simple tax systems.

Although our age-dependent policy is constructed to mimic the second best as much as possible, it lacks history dependence. In particular, it cannot implement the short-term regressivity property which we found to be optimal. At least for this simulation, it appears that history dependence is not crucial for welfare. At present, we do not know how robust this conclusion is.

As the second row indicates, preserving age-dependent linear labor taxes but setting capital taxes to zero delivers a welfare gain of 3.16%. The difference of 0.14% represents the gains from taxing capital. This magnitude is in line with [Farhi and Werning \(2008a,b\)](#), who find relatively modest gains, especially when incorporating general equilibrium effects which are absent here.

The third row maintains the same age-dependent labor tax, but allows for a non-zero, age-independent tax on capital. This improves welfare to 3.29%, very close to the welfare obtained by age dependent labor and capital taxes of 2.30% from the first row. The op-

¹⁵Other findings discussed below imply that the shape of the age-dependent tax does affect welfare.

timal age-independent intertemporal wedge is 0.33% (corresponding to tax rate of 6.60% on the net interest). Interestingly, this is close to the average wedge across ages from the second-best simulation, as displayed in [Figure 1](#).

The last row reports welfare for the simplest tax system we consider: age-independent linear labor and capital taxes. The optimal age-independent linear tax on labor is equal to 27.43%, quite close to the average across ages found in the second best simulation, or the calculation behind panel (b) in [Figure 5](#). This simplest of tax systems delivers welfare of 2.71%. Comparing this to the first row, we see that the cost of imposing an age-independent tax system is roughly 0.60% of lifetime consumption.

Not reported in the table is the fact that the optimal age-independent tax on capital comes out to be minuscule: an intertemporal wedge of 0.0138%, corresponding to a tax rate of around 0.28% on net interest. Given this, the cost of imposing a zero tax on capital constraint are minuscule, below 0.001% of lifetime consumption. Interestingly, taxing capital does not appear to be optimal unless the labor tax is somewhat sophisticated and features either age-dependence or the richer history-dependence of the second best.

With an age-dependent labor tax, an age-independent tax on capital provides modest but non-negligible benefits, equal to 0.13%. However, the addition of an age-dependent capital tax provides little extra benefit, equal to 0.01% of lifetime consumption. In contrast, age-dependent taxes on labor provide a sizable improvement of 0.60% over the completely age-independent tax system. Allowing for age-dependent labor taxes is more important in this simulation than allowing for age-dependent capital taxes.

Why is the optimal age-independent tax on capital significant when labor taxes are age-dependent, yet minuscule when labor taxes are age independent? There are two forces at play. The first pushes for a positive tax on capital to get closer to the Inverse Euler condition. This force is clearly at play in the second best, but also appears to be present in the simpler tax systems (rows 1–3 in the table). The second force occurs only when when labor taxes cannot be age-dependent (row 4). The reason is that a capital subsidy could help mimic an increasing age profile of labor taxes. Intuitively, labor income earned earlier in life, while taxed at the same rate as later in life, has the benefit that, when saved, it accrues a higher interest rate from the capital subsidy. This sort of mimicking effect is explained in [Erosa and Gervais \(2002\)](#) for a Ramsey framework.¹⁶ When we allow for

¹⁶They assume no uncertainty, so that the age-dependence of the desire path of labor taxes is entirely driven by the age-dependence of the Frisch elasticity of labor supply. In our simulation, instead, the Frisch elasticity of labor supply is constant, and it is the information structure that is responsible for the age-dependence of desired labor taxes, which is increasing. Restricting labor taxes to be age-independent calls for a mimicking capital subsidy. Instead, they focus on a specification where the elasticity of the disutility of labor varies, with a functional form that can lead to the reverse case, with decreasing labor taxes or a positive tax on capital to mimic them.

age-dependent labor taxes, the second force is absent leading to a positive tax on capital. When the labor tax cannot depend on age, both forces are present and roughly cancel each other out, resulting in a practically zero tax on capital.

Gains From the Inverse Euler Equation. We conclude our numerical investigations by linking our results to [Farhi and Werning \(2008a,b\)](#). In these papers, we propose a method to assess the importance of savings distortions from a welfare perspective. Starting with any incentive-compatible allocation, we consider the class of all allocations that share the same labor allocation, are incentive compatible, and deliver the same utility to agents. We show how these allocations can be obtained from the original allocation through a simple set of perturbations. In this class of allocations, the one with the least cost is the only one that satisfied the Inverse Euler equation. The corresponding cost-savings represent the gains from optimal savings distortions. In [Table 3](#), we report the welfare gains from allowing for optimal savings distortions for the different allocations discussed above. Overall, our findings are supportive of the results in [Farhi and Werning \(2008a,b\)](#).

We report the corresponding numbers both for the high and low volatility case. The gains from optimal savings distortions decrease with the volatility of productivity. For example, they decrease from 0.449% to 0.086% for the allocation with zero taxes. This is intuitive. Starting with an allocation that satisfies the Euler equation, the gains from optimal savings distortions are linked to the amount of precautionary savings. With log preferences, the strength of this motive is measured by the variance of consumption growth, which is an increasing function of the variance of productivity. In the rest of the discussion, we focus on the high volatility case.

Consistent with [Farhi and Werning \(2008a,b\)](#), we find that the gains from optimal savings distortions are never very large. The maximum number that we compute is 0.449%. Starting from the allocation with zero taxes, the gains from optimal savings distortions only contribute to a relatively small fraction (about 13%) of the gains from moving to an optimal tax system (3.43%).

That the the gains from optimal savings distortions are largest (0.449%) for the allocation with zero taxes is intuitive for two reasons. First, this allocation has no savings distortions to start with (it satisfies the Euler equation). Second, this allocation features the least insurance among the allocations that we consider. As a result, the variance of consumption growth is high, and so are the welfare gains from optimal savings distortions. The same reasons explain why the gains from optimal savings distortions for the allocation with age-independent labor and capital taxes are also larger (at 0.180%) than for the other allocations that we consider. Regarding this allocation, it is interesting to

	$\hat{\sigma}^2 = 0.0161$	$\hat{\sigma}^2 = 0.0061$
zero taxes	0.449%	0.086%
age-dependent τ_L and τ_K	0.005%	0.000%
age-dependent τ_L , and $\tau_K = 0$	0.095%	0.036%
age-dependent τ_L , age-independent τ_K	0.011%	0.026%
age-independent τ_L and τ_K	0.180%	0.056%

Table 3: Welfare from Optimal Savings Distortions.

note that the optimal age-independent capital tax is basically equal to zero (corresponding to zero savings distortions): the gains from savings distortions cannot be achieved by simply introducing a linear tax on capital.

The gains from optimal savings distortions for the allocation with age-dependent labor and capital taxes are absolutely minuscule (much less than the gains from moving to the second best). This allocation approximates the second best very well. Our exercise informs us that the small gains from the second best over this allocation (of the order of 0.1%) do not originate in better savings distortions.

The allocation with age-dependent labor taxes but zero capital taxes offers greater gains from optimal savings distortions (0.095%). Note also that the gains from optimal savings distortions are only about a third of the gains from moving to the second best allocation (0.270%): the bulk of these gains does not originate in savings distortions per se. The gains from optimal savings distortions for the allocation with age-dependent labor taxes but age-independent capital taxes are lower (0.011%): this allocation already features some savings distortions that partially replicate the savings distortions prescribed by the Inverse Euler equation.

References

- ABRAHAM, A., AND N. PAVONI (2008): “Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending: A Recursive Formulation,” *Review of Economic Dynamics*, 11(4), 781–803.
- ALBANESI, S., AND C. SLEET (2006): “Dynamic Optimal Taxation with Private Information,” *Review of Economic Studies*, 73(1), 1–30.
- ALES, L., AND P. MAZIERO (2009): “Accounting for Private Information,” mimeo.
- BATTAGLINI, M. (2005): “Long-Term Contracting with Markovian Consumers,” *American Economic Review*, 95(3), 637–658.

- BATTAGLINI, M., AND S. COATE (2008): "Pareto efficient income taxation with stochastic abilities," *Journal of Public Economics*, 92(3-4), 844–868.
- DIAMOND, P. A. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88(1), 83–95.
- DIAMOND, P. A., AND J. A. MIRRLEES (1978): "A model of social insurance with variable retirement," *Journal of Public Economics*, 10(3), 295–336.
- EROSA, A., AND M. GERVAIS (2002): "Optimal Taxation in Life-Cycle Economies," *Journal of Economic Theory*, 105(2), 338–369.
- FARHI, E., AND I. WERNING (2008a): "Optimal savings distortions with recursive preferences," *Journal of Monetary Economics*, 55(1), 21–42.
- (2008b): "The Political Economy of Non-Linear Capital Taxation," mimeo.
- (2009): "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," mimeo.
- FERNANDES, A., AND C. PHELAN (2000): "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory*, 91(2), 223–247.
- FUKUSHIMA, K. (2010): "Quantifying the Welfare Gains from Simple Tax Systems," mimeo.
- GARRETT, D., AND A. PAVAN (2010): "Dynamic Managerial Compensation: a Mechanism Design Approach," mimeo.
- GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.
- GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2010): "Optimal Dynamic Taxes," mimeo.
- GOLOSOV, M., AND A. TSYVINSKI (2006): "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 114(2), 257–279.
- GOLOSOV, M., A. TSYVINSKI, AND I. WERNING (2006): "New Dynamic Public Finance: A User's Guide," *forthcoming in NBER Macroeconomics Annual 2006*.
- JARQUE, A. (2008): "Repeated Moral Hazard with Effort Persistence," mimeo.

- KAPICKA, M. (2008): “Efficient Allocation in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach,” mimeo.
- KREMER, M. (2002): “Should Taxes be Independent of Age,” mimeo.
- MIRRELES, J. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 38(2), 175–208.
- PAVAN, A., I. SEGAL, AND J. TOIKKA (2009): “Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization, and Information Disclosure,” mimeo.
- SAEZ, E. (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68(1), 205–29.
- SHIMER, R., AND I. WERNING (2008): “Liquidity and Insurance for the Unemployed,” *American Economic Review*, 98(5), 1922–42.
- STORESLETTEN, K., C. I. TELMER, AND A. YARON (2004): “Consumption and risk sharing over the life cycle,” *Journal of Monetary Economics*, 51(3), 609–633.
- TCHISTYI, A. (2006): “Security Design with Correlated Hidden Cash-Flows: The Optimality of Performance Pricing,” mimeo.
- WEINZIERL, M. (2008): “The Surprising Power of Age-Dependent Taxes,” mimeo.
- WERNING, I. (2002): “Optimal Unemployment Insurance with Unobservable Savings,” Mimeo.
- (2007a): “Optimal Fiscal Policy with Redistribution,” *Quarterly Journal of Economics*, 122(3), 925–967.
- (2007b): “Pareto Efficient Income Taxation,” mimeo.
- WILLIAMS, N. (2008): “Persistent Private Information,” mimeo.

7 Appendix

Proof of Proposition 1. Consider an allocation $\{c, y\}$ that solves Program IC or Program FOA. Then consider a history θ^{t-1} and a neighborhood $|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon$ of this history

where $|\cdot|$ is the sup norm. Consider the following perturbed allocation $\{\tilde{c}^\delta, \tilde{y}^\delta\}$. Define for every $\tilde{\theta}^{t-1}$ such that $|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon$,

$$\hat{u}^t \left(\tilde{c}^\delta \left(\tilde{\theta}^t \right) \right) = \hat{u}^t \left(c \left(\tilde{\theta}^t \right) \right) - \delta,$$

$$\hat{u}^{t-1} \left(\tilde{c}^\delta \left(\tilde{\theta}^{t-1} \right) \right) = \hat{u}^{t-1} \left(c \left(\tilde{\theta}^{t-1} \right) \right) + \beta \delta,$$

and for every other θ^s

$$\hat{u}^s \left(c^\delta \left(\theta^s \right) \right) = \hat{u}^s \left(c^\delta \left(\theta^s \right) \right).$$

Finally for every θ^s , define

$$\tilde{y}^\delta \left(\theta^s \right) = y \left(\theta^s \right)$$

The perturbed allocation $\{\tilde{c}, \tilde{y}\}$ satisfies all the constraints (of either Program IC or Program FOA). A necessary condition for the initial allocation $\{c, y\}$ to be optimal is that it be the least cost allocation among the class of allocations $\{\tilde{c}^\delta, \tilde{y}^\delta\}$ indexed by δ . This implies that

$$\frac{d\Psi \left\{ \tilde{c}^\delta, \tilde{y}^\delta \right\}}{d\delta} = 0$$

which can be rewritten as

$$\begin{aligned} & \beta \int_{|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon} \frac{1}{\hat{u}^{t-1} \left(c \left(\tilde{\theta}^{t-1} \right) \right)} f^{t-1} \left(\tilde{\theta}_{t-1} | \tilde{\theta}_{t-2} \right) \dots f^0 \left(\tilde{\theta}_0 | \tilde{\theta}_{-1} \right) d\tilde{\theta}_{t-1} \dots d\tilde{\theta}_0 \\ & = q \int_{|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon} \int \frac{1}{\hat{u}^{t'} \left(c \left(\tilde{\theta}^t \right) \right)} f^t \left(\tilde{\theta}_t | \tilde{\theta}_{t-1} \right) d\theta_t f^{t-1} \left(\tilde{\theta}_{t-1} | \tilde{\theta}_{t-2} \right) \dots f^0 \left(\tilde{\theta}_0 | \tilde{\theta}_{-1} \right) d\tilde{\theta}_{t-1} \dots d\tilde{\theta}_0. \end{aligned}$$

Dividing by $\int_{|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon} f^{t-1} \left(\tilde{\theta}_{t-1} | \tilde{\theta}_{t-2} \right) \dots f^0 \left(\tilde{\theta}_0 | \tilde{\theta}_{-1} \right) d\tilde{\theta}_{t-1} \dots d\tilde{\theta}_0$ and taking the limit when $\varepsilon \rightarrow 0$ yields the result.

Proof of Proposition 2. We tackle Program FOA using optimal control. To clarify the origins of the results, we first only make Assumption 1. Then we introduce Assumption 2 in the proof only when it is needed. We attach multipliers λ and γ on the first and second constraints. We denote by $\mu \left(\theta \right)$ the co-state variable associated with $w \left(\theta \right)$. We then form the corresponding Hamiltonian. The Envelope conditions can be written as

$$K_v \left(v, \Delta, \theta_-, t \right) = \lambda \text{ and } K_\Delta \left(v, \Delta, \theta_-, t \right) = \gamma.$$

In line with these identities, we write

$$K_v \left(v \left(\theta \right), \Delta \left(\theta \right), \theta, t + 1 \right) = \lambda \left(\theta \right) \text{ and } K_\Delta \left(v \left(\theta \right), \Delta \left(\theta \right), \theta, t + 1 \right) = \gamma \left(\theta \right).$$

The boundary conditions are

$$\lim_{\theta \rightarrow \bar{\theta}} \mu(\theta) = 0 \text{ and } \lim_{\theta \rightarrow \underline{\theta}} \mu(\theta) = 0. \quad (21)$$

The law of motion for the co-state $\mu(\theta)$ is

$$\frac{d\mu(\theta)}{d\theta} = - \left[\frac{1}{\hat{u}^{t'}(c(\theta))} - \lambda - \gamma \frac{f_{\theta_-}^t(\theta|\theta_-)}{f^t(\theta|\theta_-)} \right] f^t(\theta|\theta_-) \quad (22)$$

The first order conditions for $\Delta(\theta)$, $v(\theta)$ and $y(\theta)$ can be rearranged as follows

$$\frac{\mu(\theta)}{\theta f^t(\theta|\theta_-)} = - \frac{q}{\beta} \frac{\gamma(\theta)}{\theta}, \quad (23)$$

$$\frac{1}{\hat{u}^{t'}(c(\theta))} = \frac{q}{\beta} \lambda(\theta), \quad (24)$$

and

$$\left(1 - \frac{\hat{h}_y(y(\theta), \theta)}{\hat{u}^{t'}(c(\theta))} \right) = \frac{\mu(\theta)}{f^t(\theta|\theta_-)} [h_{y\theta}(y(\theta), \theta)]. \quad (25)$$

Using equation (24) to replace λ by $(\beta/q)(1/\hat{u}^{t-1'}(c_-))$ in equation (22), and integrating and using equation (21) we get

$$0 = \int \left[\frac{1}{\hat{u}^{t'}(c(\theta))} - \frac{\beta}{q} \frac{1}{\hat{u}^{t-1'}(c_-)} \right] f^t(\theta|\theta_-)$$

which provides another proof of Proposition 1.

Let us now make Assumption 2. We can then simplify equation (25) as follows

$$\left(1 - \frac{\frac{1}{\theta} \kappa \left(\frac{y(\theta)}{\theta} \right)^{\alpha-1}}{\hat{u}^{t'}(c(\theta))} \right) = \alpha \frac{\mu(\theta)}{\theta f^t(\theta|\theta_-)} \hat{u}^{t'}(c(\theta)) \frac{\frac{1}{\theta} \kappa \left(\frac{y(\theta)}{\theta} \right)^{\alpha-1}}{\hat{u}^{t'}(c(\theta))}. \quad (26)$$

Replacing the expression for the labor wedge in this last condition, we get

$$\frac{\tau_L(\theta)}{1 - \tau_L(\theta)} \frac{1}{\hat{u}^{t'}(c(\theta))} \pi(\theta) f^t(\theta|\theta_-) = \alpha \mu(\theta) \frac{\pi(\theta)}{\theta}.$$

Integrating by parts this equality, we get

$$\int \frac{\tau_L(\theta)}{1 - \tau_L(\theta)} \frac{1}{\hat{u}^{t'}(c(\theta))} \pi(\theta) f^t(\theta|\theta_-) d\theta = \alpha \int \mu(\theta) \frac{\pi(\theta)}{\theta} d\theta$$

$$\begin{aligned}
&= \alpha [\mu(\theta) \Pi(\theta)]_{\theta}^{\bar{\theta}} \\
&\quad + \alpha \int \Pi(\theta) \left[\frac{1}{\hat{u}^{t'}(c(\theta))} - \lambda - \gamma \frac{f_{\theta_-}^t(\theta|\theta_-)}{f^t(\theta|\theta_-)} \right] f^t(\theta|\theta_-) d\theta \\
&= \alpha \int \Pi(\theta) \left[\frac{1}{\hat{u}^{t'}(c(\theta))} - \lambda \right] f^t(\theta|\theta_-) d\theta - \alpha \frac{\gamma}{\theta_-} \theta_- \frac{d\phi^{\Pi}(\theta_-)}{d\theta_-}
\end{aligned}$$

where we have used the fact that

$$\frac{d\phi^{\Pi}(\theta_-)}{d\theta_-} = \int \Pi(\theta) f_{\theta_-}^t(\theta|\theta_-) d\theta.$$

Now note that

$$\frac{\tau_L(\theta)}{1 - \tau_L(\theta)} \frac{1}{\hat{u}^{t'}(c(\theta))} = \alpha \frac{\mu(\theta)}{\theta f^t(\theta|\theta_-)} = -\alpha \frac{q}{\beta} \frac{\gamma(\theta)}{\theta}$$

so that we also have

$$\frac{\tau_{L-}}{1 - \tau_{L-}} \frac{1}{\hat{u}^{t-1'}(c_-)} = -\alpha \frac{q}{\beta} \frac{\gamma}{\theta_-}.$$

Similarly we have

$$\lambda = \frac{\beta}{q} \frac{1}{\hat{u}^{t-1'}(c_-)}.$$

This implies that

$$\begin{aligned}
&\int \frac{\tau_L(\theta)}{1 - \tau_L(\theta)} \frac{q}{\beta} \frac{\hat{u}^{t-1'}(c_-)}{\hat{u}^{t'}(c(\theta))} \pi(\theta) f^t(\theta|\theta_-) d\theta \\
&= \alpha \int \Pi(\theta) \left[\frac{q}{\beta} \frac{\hat{u}^{t-1'}(c_-)}{\hat{u}^{t'}(c(\theta))} - 1 \right] f^t(\theta|\theta_-) d\theta + \frac{\tau_{L-}}{1 - \tau_{L-}} \theta_- \frac{d\phi^{\Pi}(\theta_-)}{d\theta_-}.
\end{aligned}$$

This proves Proposition 2.

Proof of Proposition 3. The proof is very similar to that of Proposition 1. Define

$$\chi \equiv \int (\theta_{t-1}/\theta_t)^\alpha f^t(\theta_t|\theta_{t-1}) d\theta_t.$$

The idea is to consider a history θ^{t-1} , a neighborhood $|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon$ of this history, and the following perturbed allocation $\{\tilde{c}^\delta, \tilde{y}^\delta\}$. Define for every $\tilde{\theta}^{t-1}$ such that $|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon$,

$$\begin{aligned}
\left(\tilde{y}^\delta(\tilde{\theta}^t)\right)^\alpha &= \left(y(\tilde{\theta}^t)\right)^\alpha - \frac{\delta}{\chi}, \\
\left(\tilde{y}^\delta(\tilde{\theta}^{t-1})\right)^\alpha &= \left(y(\tilde{\theta}^{t-1})\right)^\alpha + \beta\delta,
\end{aligned}$$

and for every other θ^s

$$\tilde{y}^\delta(\theta^s) = y^\delta(\theta^s).$$

Finally for every θ^s , define

$$\tilde{c}^\delta(\theta^s) = c(\theta^s).$$

It is easy to see that the perturbed allocation is incentive compatible and delivers the same utility as the original allocation. As in the proof of Proposition 1, a necessary condition for the initial allocation $\{c, y\}$ to solve Program IC is that it be the least cost allocation among the class of allocations $\{\tilde{c}^\delta, \tilde{y}^\delta\}$ indexed by δ . The limit of the corresponding first-order condition when ε goes to zero delivers

$$\begin{aligned} & \int \frac{\tau_L(\theta^t)}{1 - \tau_L(\theta^t)} \frac{q \hat{u}^{t-1'}(c(\theta^{t-1}))}{\beta \hat{u}^{t'}(c(\theta^t))} \left(\frac{\theta_t}{\theta_{t+1}}\right)^\alpha f^t(\theta_t | \theta_{t-1}) d\theta_t \\ &= \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \int \left(\frac{\theta_t}{\theta_{t+1}}\right)^\alpha f^t(\theta_t | \theta_{t-1}) d\theta_t \\ & \quad + \int \left[1 - \frac{q \hat{u}^{t-1'}(c(\theta^{t-1}))}{\beta \hat{u}^{t'}(c(\theta^t))}\right] \left(\frac{\theta_t}{\theta_{t+1}}\right)^\alpha f^t(\theta_t | \theta_{t-1}) d\theta_t. \end{aligned}$$

This completes the proof.

Proof of Proposition 4. We use optimal control to analyze this program: we attach multipliers λ and γ on the first and second constraints; we denote by $\mu(\theta)$ the co-state variable associated with $w(\theta)$; and we then form the corresponding Hamiltonian. The Envelope conditions can be written as

$$K_v(v, \Delta, \theta_-, t) = \lambda \text{ and } K_\Delta(v, \Delta, \theta_-, t) = \gamma.$$

In line with these identities, we write

$$K_v(v(\theta), \Delta(\theta), \theta, t+1) = \lambda(\theta) \text{ and } K_\Delta(v(\theta), \Delta(\theta), \theta, t+1) = \gamma(\theta).$$

The boundary conditions for the co-state variable are

$$\begin{aligned} \lim_{\theta \rightarrow \bar{\theta}^t} \frac{\mu(\bar{\theta}_t)}{\bar{\theta}_t f^t(\bar{\theta}_t | \theta_-)} &= -\frac{\gamma}{\theta_-} \frac{\theta_-}{\bar{\theta}_t} \frac{d\bar{\theta}_t}{d\theta_-}, \\ \lim_{\theta \rightarrow \underline{\theta}^t} \frac{\mu(\underline{\theta}_t)}{\underline{\theta}_t f^t(\underline{\theta}_t | \theta_-)} &= -\frac{\gamma}{\theta_-} \frac{\theta_-}{\underline{\theta}_t} \frac{d\underline{\theta}_t}{d\theta_-}. \end{aligned}$$

The first-order condition for $y(\theta)$ can be rearranged as follows

$$\frac{\tau_L(\theta)}{1 - \tau_L(\theta)} = -\frac{\mu(\theta)}{\theta f^t(\theta|\theta_-)} \frac{1}{C_u} \frac{\theta C_{n\theta}}{C_n},$$

where for short, the argument $(y(\theta), w(\theta) - \beta v(\theta), \theta)$ of the function $\frac{1}{C_u} \frac{\theta C_{n\theta}}{C_n}$ is omitted. Combining the last three equations immediately yields part (i) of the proposition.

Turning to part (ii), we now make Assumptions 1 and 2. We can then simplify the first-order condition for $y(\theta)$ as

$$\frac{\tau_L(\theta)}{1 - \tau_L(\theta)} = \alpha \frac{\beta}{q} \hat{u}^{t'}(c(\theta)) \frac{\mu(\theta)}{\theta f^t(\theta|\theta_-)}.$$

Combining this with the first-order condition for $\Delta(\theta)$

$$\frac{\mu(\theta)}{\theta f^t(\theta|\theta_-)} = -\frac{q}{\beta} \frac{\gamma(\theta)}{\theta}$$

yields

$$\frac{\tau_L(\theta)}{1 - \tau_L(\theta)} = -\alpha \frac{\gamma(\theta)}{\theta} \hat{u}^{t'}(c(\theta)).$$

These conditions also hold in the previous period

$$\frac{\tau_{L-}}{1 - \tau_{L-}} = -\alpha \frac{\gamma}{\theta_-} \hat{u}^{t-1'}(c_-).$$

Together with the boundary conditions, this yields

$$\begin{aligned} \frac{\tau_L(\bar{\theta}_t)}{1 - \tau_L(\bar{\theta}_t)} &= \alpha \frac{\beta}{q} \hat{u}^{t'}(c(\bar{\theta}_t)) \frac{\mu(\bar{\theta}_t)}{\bar{\theta}_t f^t(\bar{\theta}_t|\theta_-)} \\ &= -\alpha \frac{\gamma}{\theta_-} \frac{\beta}{q} \hat{u}^{t'}(c(\bar{\theta}_t)) \frac{\theta_-}{\bar{\theta}_t} \frac{d\bar{\theta}_t}{d\theta_-} \\ &= \frac{\tau_{L-}}{1 - \tau_{L-}} \frac{\beta}{q} \hat{u}^{t'}(c(\bar{\theta}_t)) \frac{\theta_-}{\bar{\theta}_t} \frac{d\bar{\theta}_t}{d\theta_-}. \end{aligned}$$

A similar calculation yields

$$\frac{\tau_L(\underline{\theta}_t)}{1 - \tau_L(\underline{\theta}_t)} = \frac{\tau_{L-}}{1 - \tau_{L-}} \frac{\beta}{q} \hat{u}^{t'}(c(\underline{\theta}_t)) \frac{\theta_-}{\underline{\theta}_t} \frac{d\underline{\theta}_t}{d\theta_-}$$

Proof of Proposition 5. The proof follows immediately, by combining the following two lemmas.

Lemma 1 Suppose that Assumptions 1 and 2 hold, and that $\{c, y\}$ solves Program FOA. Then the labor wedge satisfies the following equation for every $t \geq 1$, history θ^{t-1} and shock θ_t :

$$\begin{aligned} \frac{\partial}{\partial \theta_t} \left[\frac{\tau_L(\theta^{t-1}, \theta_t)}{1 - \tau_L(\theta^{t-1}, \theta_t)} \frac{q}{\beta} \frac{1}{\hat{u}^{t'}(c(\theta^{t-1}, \theta_t))} \right] = & \\ & - \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \frac{1}{\hat{u}^{t-1'}(c(\theta^{t-1}))} \frac{\theta_{t-1} f_{\theta_{t-1}}^t(\theta_t | \theta_{t-1})}{\theta_t f^t(\theta_t | \theta_{t-1})} \\ & + \frac{\tau_L(\theta^{t-1}, \theta_t)}{1 - \tau_L(\theta^{t-1}, \theta_t)} \frac{q}{\beta} \frac{1}{\hat{u}^{t'}(c(\theta^{t-1}, \theta_t))} \frac{-f^t(\theta_t | \theta_{t-1}) - \theta_t f_{\theta_t}^t(\theta_t | \theta_{t-1})}{\theta_t f^t(\theta_t | \theta_{t-1})} \\ & - \frac{\alpha}{\theta_t} \left[\frac{q}{\beta} \frac{1}{\hat{u}^{t'}(c(\theta^{t-1}, \theta_t))} - \frac{1}{\hat{u}^{t-1'}(c(\theta^{t-1}))} \right]. \end{aligned}$$

We build on the proof of Proposition 4. We use the same notations and definitions. We start with the following equation

$$\frac{\tau(\theta)}{1 - \tau(\theta)} \frac{q}{\beta} \theta \lambda(\theta) f(\theta | \theta_-) = \alpha \mu(\theta).$$

Using the law of motion for μ and the boundary condition for μ at $\underline{\theta}_t$, we can transform this equation as follows

$$\begin{aligned} & \frac{\tau_L(\theta)}{1 - \tau_L(\theta)} \frac{q}{\beta} \theta \lambda(\theta) f^t(\theta | \theta_-) \\ & = -\alpha \gamma \frac{d\theta_t}{d\theta_-} f^t(\theta_t | \theta_-) - \alpha \int_{\underline{\theta}_t}^{\theta} \left[\frac{q}{\beta} \lambda(\hat{\theta}) - \lambda - \gamma \frac{f_{\hat{\theta}_-}^t(\hat{\theta} | \theta_-)}{f(\hat{\theta} | \theta_-)} \right] f^t(\hat{\theta} | \theta_-) d\hat{\theta} \\ & - \alpha \gamma \frac{d\theta_L(\theta_-)}{d\theta_-} f(\theta | \theta_-) + \alpha \gamma \int_{\underline{\theta}_t}^{\theta} f_{\theta_-}(\hat{\theta} | \theta_-) d\hat{\theta} - \alpha \int^{\theta} \left[\frac{q}{\beta} \lambda(\hat{\theta}) - \lambda \right] f^t(\hat{\theta} | \theta_-) d\hat{\theta} \\ & \frac{\tau_{L-}}{1 - \tau_{L-}} \theta_- \lambda \left[\frac{d\theta_t}{d\theta_-} f^t(\theta_t | \theta_-) - \int_{\underline{\theta}_t}^{\theta} f_{\theta_-}^t(\hat{\theta} | \theta_-) d\hat{\theta} \right] - \alpha \int^{\theta} \left[\frac{q}{\beta} \lambda(\hat{\theta}) - \lambda \right] f^t(\hat{\theta} | \theta_-) d\hat{\theta}. \end{aligned}$$

where

$$\frac{1}{\hat{u}^{t-1'}(c_-)} = \frac{q}{\beta} \lambda \text{ and } \frac{1}{\hat{u}^{t'}(c(\theta))} = \frac{q}{\beta} \lambda(\theta).$$

This implies that

$$\begin{aligned} \frac{d \left[\frac{\tau_L(\theta)}{1-\tau_L(\theta)} \frac{q}{\beta} \lambda(\theta) \right]}{d\theta} &= -\frac{\tau_{L-}}{1-\tau_{L-}} \lambda \frac{\theta_- f_{\theta_-}^t(\theta|\theta_-)}{\theta f^t(\theta|\theta_-)} \\ &\quad + \frac{\tau(\theta)}{1-\tau(\theta)} \frac{q}{\beta} \lambda(\theta) \frac{-f^t(\theta|\theta_-) - \theta f_{\theta}^t(\theta|\theta_-)}{\theta f^t(\theta|\theta_-)} - \frac{\alpha}{\theta} \left[\frac{q}{\beta} \lambda(\hat{\theta}) - \lambda \right] \end{aligned}$$

which proves the lemma.

Lemma 2 *Suppose that*

$$f^t(\theta|\theta_-) = \frac{1}{\theta \sigma_t \sqrt{2\pi}} e^{-\frac{\left(\log \left(\frac{\theta}{\theta_- e^{\mu_t \log(\theta_-)}} \right) \right)^2}{2\sigma_t^2}} \quad (27)$$

where $\mu_t(\theta_-)$ is an arbitrary function of θ_- and σ_t is a constant. Then

$$\theta_- f_{\theta_-}^t(\theta|\theta_-) = - \left(1 + \theta_- \frac{d\mu_t^{\log}}{d\theta_-} \right) (\theta f_{\theta}^t(\theta|\theta_-) + f^t(\theta|\theta_-)).$$

Alternative Proof of Proposition 5 based on a perturbation argument. Consider an allocation $\{c, y\}$ that satisfies the constraints of Program FOA, and a history θ^{t-1} . We construct a sequence $\{c^{n,\delta}, y^{n,\delta}\}$ of perturbed allocations as follows. Let

$$\pi^n(\theta) = -(\hat{\theta} - \varepsilon) \delta^n(\theta - (\hat{\theta} - \varepsilon)) + (\hat{\theta} + \varepsilon) \delta^n(\theta - (\hat{\theta} + \varepsilon))$$

where δ^n is a sequence of functions that converges to the Dirac distribution δ_{distr} at 0 where we write δ_{distr} to differentiate the notation for the Dirac distribution and the notation for the number δ . The sequence of functions π^n converges to the following distribution

$$\pi(\theta) = -(\hat{\theta} - \varepsilon) \delta_{distr}(\theta - (\hat{\theta} - \varepsilon)) + (\hat{\theta} + \varepsilon) \delta_{distr}(\theta - (\hat{\theta} + \varepsilon))$$

Let $\Pi^n(\theta)$ be a primitive of $\pi^n(\theta)/\theta$. The sequences of functions Π^n converges to the following function Π

$$\begin{aligned} \Pi(\theta) &= 0 \text{ if } \theta < \hat{\theta} - \varepsilon \text{ or } \theta > \hat{\theta} + \varepsilon \\ \Pi(\theta) &= -1 \text{ otherwise.} \end{aligned}$$

Denote by $\hat{h}(x)$ the function $\kappa/\alpha x^\alpha$. The allocation $\{c^{n,\delta}, y^{n,\delta}\}$ differs from $\{c, y\}$ only

at histories θ^{t-1} and (θ^{t-1}, θ_t) (for all possible θ_t). Define

$$\hat{h}\left(y^{n,\delta}\left(\theta^{t-1}\right)\right) = \hat{h}\left(y\left(\theta^{t-1}\right)\right) + \hat{\theta}_{t-1}^\alpha \gamma^{n,\delta}$$

and

$$\hat{h}\left(y^{n,\delta}\left(\theta^{t-1}, \theta_t\right)\right) = \hat{h}\left(y\left(\theta^{t-1}, \theta_t\right)\right) - \theta_t^\alpha \pi^n\left(\theta_t\right) \frac{\delta}{\beta}$$

and define

$$\hat{u}^t\left(c^{n,\delta}\left(\theta^{t-1}, \theta_t\right)\right) = \hat{u}^t\left(c^{n,\delta}\left(\theta^{t-1}, \theta_t\right)\right) + \zeta^{n,\delta}\left(\theta_t\right) - \pi^n\left(\theta_t\right) \frac{\delta}{\beta}$$

where

$$\zeta^{n,\delta}\left(\theta_t\right) = \alpha \frac{\delta}{\beta} \Pi^n\left(\theta_t\right) + \bar{\zeta}^{n,\delta}.$$

The perturbed allocations satisfy the constraints of Program FOA as long as

$$\gamma^{n,\delta} = \delta \theta_{t-1} \frac{d\phi^{\Pi^n}}{d\theta_{t-1}},$$

$$\bar{\zeta}^{n,\delta} = \frac{\delta}{\beta} \left(\theta_{t-1} \frac{d\phi^{\Pi^n}}{d\theta_{t-1}} - \alpha \phi^{\Pi^n} \right).$$

Note that the perturbed allocation incorporates the necessary corrections to the perturbation constructed in the text.

The first-order condition

$$\frac{d\Psi\left\{c^{n,\delta}, y^{n,\delta}\right\}}{d\delta} = 0$$

can be rewritten as equation (5) for $\pi = \pi^n$ and $\Pi = \Pi^n$. Taking the limit when n goes to infinity, and then taking the limit when ε goes to zero delivers Proposition 1.

Note that under Assumption 4, we can rewrite

$$\gamma^{n,\delta} = \delta \int \pi^{n,\delta}\left(\theta_t\right) f^t\left(\theta_t|\theta_{t-1}\right)$$

and

$$\bar{\zeta}^{n,\delta} = \frac{\delta}{\beta} \left(\int \pi^{n,\delta}\left(\theta_t\right) f^t\left(\theta_t|\theta_{t-1}\right) d\theta_t - \alpha \int \Pi^{n,\delta}\left(\theta_t\right) f^t\left(\theta_t|\theta_{t-1}\right) d\theta_t \right).$$

We can then take the continuous time limit and derive Proposition 5.

Proof of Proposition 6. The first-order conditions for c_t , y_t , and $\sigma_{\Delta,t}$ in the HJB equation can be written as

$$\begin{aligned}\lambda_t &= \frac{1}{\hat{u}'_t(c_t)}, \\ \frac{\tau_{L,t}}{1 - \tau_{L,t}} &= -\alpha \frac{\gamma_t}{\lambda_t} \frac{1}{\theta_t}, \\ \sigma_{\Delta,t} &= \frac{K_{\Delta} - K_{v\Delta}\theta_t\Delta_t - \theta_t K_{\Delta\theta}}{K_{\Delta\Delta}}.\end{aligned}$$

Applying Ito's lemma to $\lambda_t = K_v(v_t, \Delta_t, \theta_t, t)$, and differentiating the HJB equation with respect to v_t (using the Envelope theorem) immediately yields that the drift term of λ_t is equal to zero. Hence, λ_t is a martingale. We can therefore write

$$d\lambda_t = (K_{vv}\theta_t\Delta_t + K_{v\Delta}\sigma_{\Delta,t} + K_{v\theta}\theta_t) \hat{\sigma}_t dW_t.$$

Using the first-order condition for $\sigma_{\Delta,t}$, we obtain

$$\frac{d\lambda_t}{\lambda_t} = \sigma_{\lambda,t} \hat{\sigma}_t dW_t,$$

where

$$\sigma_{\lambda,t} = \frac{1}{K_v} \left(\frac{K_{vv}K_{\Delta\Delta} - K_{v\Delta}^2}{K_{\Delta\Delta}} \theta_t \Delta_t + K_{v\Delta} \frac{K_{\Delta} - \theta_t K_{\Delta\theta}}{K_{\Delta\Delta}} + K_{v\theta} \theta_t \right).$$

Applying Ito's lemma to $\gamma_t = K_{\Delta}(v_t, \Delta_t, \theta_t, t)$, and differentiating the HJB equation with respect to Δ_t (using the Envelope theorem) yields that the drift term of γ_t is equal to

$$- \left(K_{vv}\theta_t^2\Delta_t\hat{\sigma}_t^2 + K_{v\Delta}\theta_t\sigma_{\Delta,t}\hat{\sigma}_t^2 + K_{v\theta}\theta_t^2\hat{\sigma}_t^2 - \left(\hat{\mu}_t + \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right) K_{\Delta} \right) dt$$

which using the definition of γ_t , the first-order condition for $\sigma_{\Delta,t}$ and the expression for $\sigma_{\lambda,t}$, we get

$$-\theta\lambda_t\sigma_{\lambda,t}\hat{\sigma}_t^2 dt + \left(\hat{\mu}_t + \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right) \gamma_t dt.$$

Similarly, the volatility term of γ_t is given by

$$(K_{\Delta v}\theta_t\Delta_t + K_{\Delta\Delta}\sigma_{\Delta,t} + K_{\Delta\theta}\theta_t) \hat{\sigma}_t dW_t$$

which using the first order condition for $\sigma_{\Delta,t}$, we can rewrite this as

$$\gamma_t \hat{\sigma}_t dW_t.$$

Hence we have

$$d\gamma_t = \left[-\theta_t \lambda_t \sigma_{\lambda,t} \hat{\sigma}_t^2 + \left(\hat{\mu}_t + \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right) \gamma_t \right] dt + \gamma_t \hat{\sigma}_t dW_t.$$

Applying Ito's lemma to the first order condition for y_t then yields

$$d \left(\lambda_t \frac{\tau_{L,t}}{1 - \tau_{L,t}} \right) = \left[\alpha \lambda_t \sigma_{\lambda,t} \hat{\sigma}_t^2 + \lambda_t \frac{\tau_{L,t}}{1 - \tau_{L,t}} \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right] dt.$$