Insurance and Taxation over the Life Cycle

EMMANUEL FARHI
Harvard University

and

IVÁN WERNING
MIT

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 labour-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 labour; 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 labour tax rises from 0% to 37% over 40 years, whereas the average tax on savings falls from 12% 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.

Key words: Optimal taxation

JEL Codes: H21, H31

1. INTRODUCTION

To a twenty-five-year-old entering the labour 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 which is slowly resolved over time.

   This article 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 labour-income taxes when uncertainty is gradually revealed over time.

This article aims to fill this gap and address the following questions. How are the lessons for labour-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? Does the insurance arrangement imply a tax systems that is progressive or regressive? How does the fully 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. In our model, 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, labour 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 simple 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. We find it useful to work in both discrete and continuous time.

Our theoretical results are summarized by a novel formula for the dynamics of the labour 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 labour and an isoelastic disutility function for labour. We then obtain

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

The first term captures mean-reversion and is simply the past labour wedge weighted by the coefficient of mean-reversion ρ in productivity. In this sense, the labour 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 labour wedge remains constant between periods t and

---

1. See for example, Diamond and Mirrlees (1978); Farhi and Werning (2008b); Golosov *et al.* (2003); Golosov *et al.* (2006); Albanesi and Sleet (2006).

$t+1$. If instead $\rho < 1$, then the labour 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 labour, which explains the role of the Frisch elasticity ϵ.

In continuous time we confirm these results and also derive a tighter characterization.

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 show that in the limit, the process $\{\frac{\tau_L}{1-\tau_L}\}$ with

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

The drift in the continuous time (the terms multiplying dt) is the exact counterpart of the discrete-time expectation formula above. The new result here is that the innovations to the labour wedge are related one to one with innovations in the marginal utility of consumption. Economically, this result describes a form of regressivity. When productivity rises, consumption rises, so the marginal utility of consumption falls and the labour 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 labour wedge. This regressive taxation result is novel and due to the dynamic aspects of our model. In a static optimal taxation settings with a Utilitarian welfare function no general results on regressive or progressive taxation are available, since the optimal tax schedule depends delicately on the skill distribution (Mirrlees, 1971; Diamond, 1998; Saez, 2001).

Finally, we extend the well-known zero-taxation result at the top and bottom of the productivity distribution. If the conditional distribution for productivity has a fixed support and labour is not zero, then the labour 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 labour 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 labour wedge at the top must be below the previous period's labour wedge. The reverse is true at the bottom: the labour 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 *et al.* (2004) points to a near random walk for labour 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. In addition, although our numerical work is based the discrete-time version of the model, with a period modelled as a year, the simulations show that our continuous-time results provides excellent explanations for our findings.

We find that the average labour wedge starts near zero and increases over time, asymptoting to around 37% precisely at retirement. The intertemporal wedge displays the opposite pattern, with its average starting around 0.6%, corresponding to a 12% 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 labour 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. In terms of average tax rates the optimal tax system is progressive, the present value of taxes paid relative income is increasing in productivity. This captures the insurance nature of the solution.

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 investigate how our results translate to simpler systems that are restricted to being history independent. Do our theoretical results provide guidance for such real-world tax systems? We find that they do. In fact, 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 labour and capital income, and consider both age-dependent and age-independent taxes. When age-dependent linear taxes are allowed, the optimal tax rates come out to be indistinguishable from the average rate for each age group from the fully optimal (history dependent) marginal tax rates. Surprisingly, the welfare loss of such a system, relative to the fully optimal one, is minuscule—around 0.15% of lifetime consumption. In this way, our theoretical results do provide guidance for more restrictive tax systems.

We then solve for optimal age-independent linear tax rates. We find that welfare losses are more significant, around 0.3% of lifetime consumption. Thus, age-dependent tax rates are important. Second, when linear taxes are age independent, the optimal tax on capital is essentially zero, despite the fact that these are positive in the full optimum, or in the system that allows for age-dependent taxes. This can be explained by the fact that a linear subsidy on capital helps imitate the missing age-dependent linear taxes on labour: with a subsidy on savings, income earned, and saved in early periods count for more at retirement. This new effect cancels the desire for a positive linear tax on capital. This provides an interesting force that contrasts the conclusions in Erosa and Gervais (2002). Their model is a deterministic life-cycle model and found positive taxes on capital when restricting to age-independent taxes.

1.1. *Related literature*

Our article contributes to the optimal taxation literature based on models with private information (see Golosov *et al.*, 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.[2,3]

---

2. Two 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.

3. See also Battaglini and Coate (2008). See as well Tchistyi (2006) and Battaglini (2005) for applications in a non-taxation context.

This article 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 labour wedges, which are the main focus of the present article.

Versions of the first-order approach on which we rely in this article 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 *et al.* (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.[4]

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 *et al.* (2010) use a first-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.[5]

An important implication of our results is that with persistent productivity shocks, labour 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.[6] Most closely related to our article are Kremer (2002) and Weinzierl (2008). Kremer (2002) emphasized the potential benefits of age-dependent labour taxation, noting that the wage distribution is likely to become more dispersed with age and conjectured that labour 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.

---

4. See also Abraham and Pavoni (2008), Jarque (2008), and Kocherlakota (2004).

5. Both Kapicka (2008) and Golosov *et al.* (2010) rely on exponential utility and special shock specifications to make the problem tractable, by reducing the number of state variables.

6. Erosa and Gervais (2002) analyse age-dependent linear labour taxation in Ramsey setting. In their model, optimal linear labour-income taxes are indexed on age because the elasticity of labour supply varies, endogenously, with age.

2.1. *The environment and planning problem*

2.1.1. 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 labour, 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 labour 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 utility function $\tilde{u}^t(c, n)$, defined over consumption and labour 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 $T_E < t \leq T$.

The realization of the state θ_t for all $t = 1, 2, \ldots, T$ is privately observed by the agent. Without loss of generality, we initialize θ_0 to some arbitrary value. Note that this does not constrain in any way the initial density $f^1(\cdot | \theta_0)$.

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

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

2.1.2. Technical assumptions.   We make the following assumptions on the utility and density functions. The utility function is assumed to be bounded, twice continuously differentiable. Moreover we assume that the partial derivative $u_\theta(c, y; \theta)$ is bounded, so that $\left| u_\theta^t(c, y; \theta) \right| \leq b$ for some $b \in \mathbb{R}_+$.

To simplify, we start with the full support assumption that $f^t(\theta' | \theta) > 0$ for all $\theta, \theta' \in \Theta$. We relax this assumption in Section 3.3. We assume that the density function has a continuously differentiable derivative $g^t(\theta' | \theta) \equiv \partial f^t(\theta' | \theta) / \partial \theta$ with respect to its second argument. Moreover, this function is bounded, so that $\left| g^t(\theta' | \theta) \right| \leq A$ for some $A \in \mathbb{R}_+$.

2.1.3. 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 an implied history of reports $\sigma^t(\theta^t) = (\sigma_1(\theta_0), \ldots, \sigma_t(\theta^t))$. Let Σ denote the set of all reporting strategies σ.

Consider an allocation $\{c, y\}$. Let $w(\theta^t)$ denote the equilibrium continuation utility after history θ^t, defined as the unique solution to

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

for all $t = 1, \ldots, T$ with $w(\theta^{T+1}) \equiv 0$. 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 $w^\sigma(\theta^{T+1}) \equiv 0$.

We say that an allocation $\{c,y\}$ is incentive compatible if and only if

$$w\left(\theta^t\right) \geq w^\sigma\left(\theta^t\right),$$

for all θ^t. That is, an allocation is incentive compatible if truth telling, $\sigma^* = \{\sigma_t^*(\theta^t)\}$ with $\sigma_t^*(\theta_t) = \theta_t$, is optimal. Let *IC* denote the set of all incentive compatible allocations $\{c,y\}$.[7]

2.1.4. Planning problem.

To keep things simple, we work in partial equilibrium, that is, assuming a linear technology that converts labour 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:

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

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:

$$K^*(v) \equiv \min_{\{c,y\}} \Psi(\{c,y\})$$

$$\text{s.t.} \quad U(\{c,y\}) \geq v$$

$$\{c,y\} \in IC.$$

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 the planning problem. Such an approach is standard in static setting, but many papers also use a similar approach in dynamic contexts, *e.g.* Werning (2002), Kapicka (2008), Williams (2008), and Pavan *et al.* (2009).

2.2.1. A necessary condition.

A strategy $_t\sigma = (\sigma_t \sigma)$ where σ_t

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

We now derive a necessary condition for incentive compatibility. Fix a history θ^t. Consider a deviation strategy σ^r indexed by $r \in \Theta$ with the property

$$\sigma_t^r\left(\theta^{t-1},\theta_t\right) = r,$$

$$\sigma_t^r\left(\theta^{t-1},\tilde{\theta}\right) = \tilde{\theta}, \qquad \tilde{\theta} \neq \theta_t.$$

---

7. Our notion here of incentive compatibility is stronger than the ex-ante optimality of truth telling (ex-ante incentive compatibility). We are also requiring the ex-post optimality, after any history of shocks and reports, of subsequent truth telling (ex-post incentive compatibility). This is without loss of generality. To see this note that ex-ante incentive compatibility implies ex-post incentive compatibility almost everywhere. Then, note that one can always insist on ex-post incentive compatibility on the remaining set of measure zero histories, without any effect on welfare for the agent or costs for the planner.

Thus, the agent reports truthfully until θ^t, then report r in period t. We do not need to specify the reporting strategy thereafter, it may or may not involve truth telling. Continuation utility solves

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

Due to the Markov property,

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

Incentive compatibility requires

$$w\left(\theta^t\right)=\max_r w^{\sigma^r}\left(\theta^t\right).$$

Equivalently,

$$w\left(\theta^t\right)=\max_r\left\{u^t\left(c\left(\theta^{t-1},r\right),y\left(\theta^{t-1},r\right);\theta_t\right)+\beta\int w^{\sigma^r}\left(\theta^{t-1},r,\theta_{t+1}\right)f^{t+1}\left(\theta_{t+1}|\theta_t\right)d\theta_{t+1}\right\}.$$
(2)

Recall that we have defined $g^t(\theta'|\theta)=\partial f^t(\theta'|\theta)/\partial\theta$. An envelope condition then suggests that

$$\frac{\partial}{\partial\theta_t}w\left(\theta^t\right)=u_\theta^t\left(c\left(\theta^t\right),y\left(\theta^t\right);\theta_t\right)+\beta\int w\left(\theta^{t+1}\right)g^{t+1}\left(\theta_{t+1}|\theta_t\right)d\theta_{t+1},$$

or its integral version

$$w\left(\theta^t\right)=\int_{\underline{\theta}}^{\theta_t}\left(u_\theta^t\left(c\left(\theta^{t-1},\tilde{\theta}_t\right),y\left(\theta^{t-1},\tilde{\theta}_t\right);\tilde{\theta}_t\right)\right.$$
$$\left.+\beta\int w\left(\theta^{t-1},\tilde{\theta}_t,\theta_{t+1}\right)g^{t+1}\left(\theta_{t+1}|\tilde{\theta}_t\right)d\theta_{t+1}\right)d\tilde{\theta}_t.$$
(3)

Let $\tilde{IC}$ denote the set of allocations $\{c,y\}$ that satisfy equation (3) for all θ^t where $\{w\}$ is defined by equation (1). This suggests that these conditions are necessary for incentive compatibility. The next lemma, proved in the Appendix, states this formally.[8]

Lemma 1. *Suppose $\{c,y\}$ is incentive compatible, so that $\{c,y\}\in IC$ and define $\{w\}$ using equation (1). Then equation (3) holds for all θ^t, so that $\{c,y\}\in\tilde{IC}$. In other words, $IC\subseteq\tilde{IC}$.*

2.2.2. The relaxed planning problem.   We now define a relaxed planning problem, replacing the incentive compatibility conditions with the necessary conditions for incentive compatibility:

$$K(v)\equiv\min_{\{c,y\}}\Psi(\{c,y\})$$

$$\text{s.t.}\quad U(\{c,y\})\geq v$$

$$\{c,y\}\in\tilde{IC}.$$

8. Pavan *et al.* (2009) derive a related, but different, necessary condition in a more general setting.

2.2.3.   A Bellman equation.   We now consider a family of related problems that admit a recursive representation. For expositional purposes, it will prove useful to rewrite conditions (1)–(3) for a given period t as follows:

$$w\left(\theta^t\right)=u^t\left(c\left(\theta^t\right),y\left(\theta^t\right);\theta_t\right)+\beta v\left(\theta^t\right) \tag{4}$$

$$\frac{\partial}{\partial\theta_t}w\left(\theta^t\right)=u_\theta^t\left(c\left(\theta^t\right),y\left(\theta^t\right);\theta_t\right)+\beta\Delta\left(\theta^t\right) \tag{5}$$

$$v\left(\theta^{t-1}\right)=\int w\left(\theta^t\right)f^t\left(\theta_t|\theta_{t-1}\right)d\theta_t \tag{6}$$

$$\Delta\left(\theta^{t-1}\right)=\int w\left(\theta^t\right)g^t\left(\theta_t|\theta_{t-1}\right)d\theta_t. \tag{7}$$

As is standard in the literature on optimal control, the differential equation (5) should be interpreted as shorthand for its integral version.

As we shall see next, the new variables $v(\theta^t)$ and $\Delta(\theta^t)$ will serve as state variables. For any date t and past history θ^{t-1}, consider the continuation problem that minimizes the remaining discounted expected costs while taking as given some previous values for $v(\theta^{t-1})$ and $\Delta(\theta^{t-1})$ as given; denote these values by v and Δ, respectively. The optimization is subject to all the remaining necessary conditions for incentive compatibility. Formally, define

$$K\left(v,\Delta,\theta_{s-1},s\right)\equiv\min\sum_{t=s}^{T}q^{t-s}\int\left(c\left(\theta^t\right)-y\left(\theta^t\right)\right)f^t\left(\theta_t|\theta_{t-1}\right)\cdots f^s\left(\theta_s|\theta_{s-1}\right)d\theta_t\cdots d\theta_s,$$

where the minimization is over continuation plans $\{c,y,w,v,\Delta\}_{t\geq s}$ subject to $v(\theta^{s-1})=v$, $\Delta(\theta^{s-1})=\Delta$ and equations (4)–(7) for $t\geq s$. Note that once one conditions on the past shock θ_{s-1}, the entire history of shocks θ^{s-1} is superfluous because of our assumption that $\{\theta\}$ is a Markov process.

This problem is recursive with Bellman equation

$$K(v,\Delta,\theta_-,t)=\min\int\{c(\theta)-y(\theta)+qK(v(\theta),\Delta(\theta),\theta,t+1)\}f^t(\theta|\theta_-)d\theta \tag{8}$$

subject to

$$w(\theta)=u^t(c(\theta),y(\theta);\theta)+\beta v(\theta)$$

$$\dot{w}(\theta)=u_\theta^t(c(\theta),y(\theta);\theta)+\beta\Delta(\theta)$$

for all $\theta\in\Theta$ and

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

$$\Delta=\int w(\theta)g^t(\theta|\theta_-)d\theta.$$

These constraints are the recursive counterparts of equations (4)–(7), taking into account that we can drop the time subscript t and the dependence on history θ^{t-1}.

Finally, note that the relaxed problem defined earlier can be recovered by setting $t=1$ and treating Δ as a free variable:

$$K(v) = \min_{\Delta} K(v, \Delta, \theta_0, 1).$$

Thus, we can solve the relaxed problem by solving the Bellman equation and then performing a simple minimization to initialize Δ. Optimal plans can then be constructed iterating on the policy functions obtained from the Bellman equation.

Note that the Bellman equation embeds, in each period's iteration $t=1, 2, \ldots, T$, an optimal control problem across current productivity types θ with two integral constraints. To see this, use the first constraint to substitute out $v(\theta) = \frac{1}{\beta}(w(\theta) - u^t(c(\theta), y(\theta); \theta))$. We can then think of the state variable as $w(\theta)$ and the controls as $c(\theta)$, $y(\theta)$, and $\Delta(\theta)$. The two integral constraints can be included in the objective with respective Lagrange multipliers. The problem thus transformed then fits into a standard optimal control problem where the initial states $w(\underline{\theta})$ is free. Thus, we decouple the full optimization into a sequence of optimal control problems, each one comparable to those in static optimal taxation settings, as pioneered by Mirrlees (1971).

2.2.4. Verifying IC.

Suppose that a solution to the relaxed planning problem has been computed. Then this also represents a solution to the original planning problem if and only if the proposed allocation is incentive compatible.

One approach is to seek sufficient conditions that guarantee that the solution to the relaxed problem is incentive compatible. In static settings this has proved fruitful. In particular, in the Mirrlees model a single-crossing assumption on the utility function together with monotonicity of the allocation provide such conditions. Unfortunately, we know of no general sufficient conditions for the dynamic case that would be useful in our context.[9] A practical alternative, is to solve the relaxed problem and then verify the incentive compatibility directly. We discuss next how this can be done, exploiting the recursive nature of the solution.

The solution to the Bellman equation (8) yields policy functions $c(\theta) = g^c(v, \Delta, \theta_-, \theta, t)$, $y(\theta) = g^y(v, \Delta, \theta_-, \theta, t)$, $w(\theta) = g^w(v, \Delta, \theta_-, \theta, t)$, $v(\theta) = g^v(v, \Delta, \theta_-, \theta, t)$, and $\Delta(\theta) = g^\Delta(v, \Delta, \theta_-, \theta, t)$. An agent takes these functions as given and solves an optimal reporting problem that can be represented by another Bellman equation:

$$V(v, \Delta, r_-, \theta, t) = \max_r \left\{ u^t \left(g^c(v, \Delta, r_-, r, t), g^y(v, \Delta, r_-, r, t), \theta \right) \right.$$

$$\left. + \beta \int V \left(g^v(v, \Delta, r_-, r, t), g^\Delta(v, \Delta, r_-, r, t), r, \theta', t+1 \right) f^{t+1} \left(\theta' | \theta \right) d\theta' \right\}.$$

Here r_- and r represent the previous and current report, respectively, whereas θ is the current true shock. The agent must condition on the previous report r_- because the allocation depends on this report. That is under the direct mechanism the previous report is taken as truthful of the previous true shock.

---

9. Pavan *et al.* (2009) work with a general dynamic model and offer some conditions that ensure incentive compatibility of allocations that satisfy the necessary first-order conditions for incentive compatibility. However, their result requires making assumptions on exogenous primitives and also verifying conditions on the endogenous allocation. Unfortunately, in our context their result does not offer any advantage compared to checking incentive compatibility directly, in the way we describe below. In other words, verifying the conditions on the allocation required for their result is just as onerous as verifying incentive compatibility.

If the utility the agent can achieve coincides with the utility the planner had intended, so that

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

holds for all v,Δ,r_-,θ, and t, then truth telling is always optimal. Thus, the solution to the relaxed problem is incentive compatible, *i.e.* it solves the original planning problem.

Indeed, this verification does not require solving for the value function V. Instead, we can simply verify that

$$\theta \in \operatorname*{argmax}_{r}\{u^t(g^c(v,\Delta,r_-,r,t),g^y(v,\Delta,r_-,r,t),\theta)$$

$$+\beta\int g^w(\theta',g^v(v,\Delta,r_-,r,t),g^\Delta(v,\Delta,r_-,r,t),r,t+1)f^{t+1}(\theta'|\theta)d\theta', \quad (9)$$

holds for all v,Δ,r_-,θ, and t.

2.2.5. Initial heterogeneity and redistribution.

We have interpreted the planning problem as involving a single agent facing uncertainty. Under this interpretation, the planner problem is purely about social insurance and not about redistribution. However, it is simple to add initial heterogeneity and consider redistribution.

The simplest way to model heterogeneity is to reinterpret the first shock θ_1. Instead of thinking of the value of θ_1 as the realization of uncertainty, we now interpret θ_1 as indexing some initial hidden characteristic of an agent. The agent is not alive before the realization of θ_1 and faces uncertainty only regarding future shocks $\theta_2, \theta_3,\ldots$. Recall that we allow the density to depend flexibly on the period t, so that $f^1(\theta_1|\theta_0)$ could accomodate any initial desired dispersion in productivitity types.

If the social welfare function is Utilitarian, then the analysis requires no change: insurance behind the veil of ignorance and utilitarian redistribution are equivalent. Formally, the social welfare in this case coincides with the expected utility calculation when θ_1 is interpreted as uncertainty. Both integrate utility over θ_1 using the density f^1. Thus, the planning problem at $t=1$ remains unchanged.

However, when it comes to redistribution, a Utilitarian welfare function is a special case. Indeed, we can allow for any social welfare function, or, more generally, characterize the entire set of constrained Pareto-efficient allocations. This does require treating the planning problem in the initial period $t=1$ differently. It turns out that this only affects the optimal allocation at $t=1$, as well as the optimal values for the endogenous state variables $v_1(\theta_1)$ and $\Delta_1(\theta_1)$. These values for $v_1(\theta_1)$ and $\Delta_1(\theta_1)$ are inherited at $t=2$ by the planner, but given these values, the problem from $t=2$ onwards remains unchanged. Thus, the dynamics for the allocation and taxes for $t=2,3,\ldots$ remains unchanged.

Formally, at $t=1$ Pareto optima solve the cost minimization problem

$$\min\int\{c(\theta_1)-y(\theta_1)+qK(v(\theta_1),\Delta(\theta_1),\theta_1,2)\}f^1(\theta_1|\theta_0)d\theta_1$$

subject to

$$w(\theta_1)=u^1(c(\theta_1),y(\theta_1);\theta_1)+\beta v(\theta_1)\geq\bar{w}(\theta_1)$$

$$\dot{w}(\theta_1)=u^1_\theta(c(\theta_1),y(\theta_1);\theta_1)+\beta\Delta(\theta_1),$$

for all $\theta_1\in\Theta$. Here the function $\bar{w}(\cdot)$ parameterizes the position on the Pareto frontier. Note that from $t=2$ onward the planning problem is characterized by the same Bellman equations described

above. Thus, our results about the dynamics of c_t, y_t, v_t, and Δ_t, and hence the dynamics of the implied marginal taxes, which is are our focus, are preserved.

3. OPTIMALITY CONDITIONS

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

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

and the labour wedge

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

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 labour.

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 the original planning problem or the relaxed planning problem. Then for every $t \geq 1$ and history θ^{t-1}, the following Inverse Euler equation holds*

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

and the intertemporal wedge satisfies

$$\tau_K(\theta^{t-1}) = 1 - \frac{\left[\int \left[\hat{u}^{t\prime}\left(c\left(\theta^t\right)\right)\right]^{-1} f^t\left(\theta_t | \theta_{t-1}\right) d\theta_t\right]^{-1}}{\int \hat{u}^{t\prime}\left(c\left(\theta^t\right)\right) f^t\left(\theta_t | \theta_{t-1}\right) d\theta_t}. \tag{10}$$

Note that this result holds for any allocation that solves both the original or the relaxed planning problem and for any stochastic process for idiosyncratic shocks $\{\theta\}$. Applying Jensen's inequality to the second equation implies that the intertemporal wedge $\tau_K\left(\theta^{t-1}\right)$ is positive. In other words, positive savings distortions are present at the constrained optimum.

3.2. *Labour wedge dynamics: tax smoothing and mean reversion*

We now seek optimality conditions for the labour wedge for the relaxed planning problem. As explained above, if the solution of the relaxed planning problem satisfies the original constraints, then it also solves the original planning problem. There are both cases where the solutions of the two planning problems coincide and cases where they do not. We do not attempt to provide necessary or sufficient conditions on primitives for it to be the case. Instead we tightly characterize

the behaviour of the labour wedge for the relaxed planning problem. In Section 6, we solve the relaxed planning problem numerically for a number of empirically relevant parametrizations, and verify numerically that our solution indeed satisfies the original constraints, and is hence also the solution of the original planning problem.

We derive a set of formulas for the evolution of labour wedges over time. We derive one such formula for each weighting functions $\pi(\theta_t)$, linking the period $t-1$ labour wedge $\tau_L(\theta^{t-1})$ and a conditional expectation as of $t-1$ of the period t labour wedge $\tau_L(\theta^t)$ weighted by the weighting function $\pi(\theta_t)$. Taken together, these formulas conveniently encode not on only the evolution of labour wedges over time from $t-1$ to t, but also across the different possible realizations of θ_t at t. In Section 4, we proceed in continuous time with Brownian diffusions. There, the analogue of our formulas for different weighting functions is a stochastic differential equation for the labour 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 $\alpha > 1$.*

We then have the following proposition.

Proposition 2.   *Suppose that Assumptions 1 and 2 hold, and that $\{c, y\}$ solves the relaxed planning problem. Consider a function $\pi(\theta)$, let $\Pi(\theta)$ be a primitive of $\pi(\theta)/\theta$ and let $\phi_t^\Pi(\theta_{t-1}) \equiv \int \Pi(\theta_t) f^t(\theta_t|\theta_{t-1}) d\theta_t$. Then the labour wedge satisfies the following equation for every $t \geq 1$ and history θ^{t-1}*

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

$$= \frac{\tau_L(\theta^{t-1})}{1-\tau_L(\theta^{t-1})} \theta_{t-1} \frac{d\phi_t^\Pi(\theta_{t-1})}{d\theta_{t-1}} + \alpha \int \Pi(\theta_t) \left[\frac{q}{\beta} \frac{\hat{u}^{t-1\prime}(c(\theta^{t-1}))}{\hat{u}^{t\prime}(c(\theta^t))} - 1 \right] f^t(\theta_t|\theta_{t-1}) d\theta_t. \quad (11)$$

These formulas show that a weighted conditional expectation of the labour wedge $\tau_L(\theta^t)$ is a function of the previous period's labour wedge $\tau_L(\theta^{t-1})$. Different weighting functions $\pi(\theta)$ lead to different weighted expectations. The fact that equation (11) holds for every possible weighting function $\pi(\theta_t)$ imposes restrictions on the stochastic process $\{\frac{\tau_L}{1-\tau_L} \frac{1}{\hat{u}^{t\prime}}\}$.[10]

The weighting function $\pi(\theta) = 1$ with $\Pi(\theta) = \log(\theta)$ is of particular interest and gives us an easily interpretable formula for the evolution of expected labour wedges $\tau_L(\theta^t)$ conditional on a history θ^{t-1}.

---

10. In particular, for any $\theta^* \in \Theta$, one can apply this formula with a sequence of functions $\pi^{n,\theta^*}(\theta)$ that converges to a Dirac distribution π at θ^*. The corresponding sequence $\Pi^{n,\theta^*}(\theta)$ converges to a weighted Heaviside function Π at θ^* given by $\frac{1}{\theta^*} 1_{\{\theta \geq \theta^*\}}$. The corresponding formula for $\theta^* = \theta_t$ gives us

$$\frac{q}{\beta} \frac{\tau_L(\theta^t)}{1-\tau_L(\theta^t)} \frac{1}{\hat{u}^{t\prime}(c(\theta^t))} = \frac{\tau_L(\theta^{t-1})}{1-\tau_L(\theta^{t-1})} \frac{1}{\hat{u}^{t-1\prime}(c(\theta^{t-1}))} \frac{\theta_{t-1}}{\theta_t} \frac{\int_{\theta_t}^{\bar{\theta}} g^t(\theta_t|\theta_{t-1}) d\theta_t}{f^t(\theta_t|\theta_{t-1})}$$

$$+ \alpha \frac{1}{\theta_t f^t(\theta_t|\theta_{t-1})} \int_{\theta_t}^{\bar{\theta}} \left[\frac{q}{\beta} \frac{1}{\hat{u}^{t\prime}(c(\theta^t))} - \frac{1}{\hat{u}^{t-1\prime}(c(\theta^{t-1}))} \right] f^t(\theta_t|\theta_{t-1}) d\theta_t.$$

Corollary 1. *Suppose that Assumptions 1 and 2 hold, and that $\{c, y\}$ solves the relaxed planning problem. Then the labour wedge satisfies the following equation for every $t \geq 1$ and history θ^{t-1}*

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

$$= \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}} + \alpha \int \log(\theta_t) \left[\frac{q}{\beta} \frac{\hat{u}^{t-1\prime}(c(\theta^{t-1}))}{\hat{u}^{t\prime}(c(\theta^t))} - 1 \right] f^t(\theta_t|\theta_{t-1}) d\theta_t. \quad (12)$$

To understand the role of $\theta_{t-1} d\phi_t^{\log}(\theta_{t-1})/d\theta_{t-1}$ in equation (12), 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 $f^{\epsilon,t}(\varepsilon_t, \theta^{t-1})$, normalized so that $\int \varepsilon_t f^{\epsilon,t}(\varepsilon_t, \theta^{t-1}) d\varepsilon_t = 0$. 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 (12) can be written more compactly as

$$\mathbb{E}_{t-1} \left[\frac{\tau_{L,t}}{1 - \tau_{L,t}} \frac{q}{\beta} \frac{\hat{u}^{t-1\prime}(c_{t-1})}{\hat{u}^{t\prime}(c_t)} \right] = \rho \frac{\tau_{L,t-1}}{1 - \tau_{L,t-1}} + \alpha \text{Cov}_{t-1} \left(\log(\theta_t), \frac{q}{\beta} \frac{\hat{u}^{t-1\prime}(c_{t-1})}{\hat{u}^{t\prime}(c_t)} \right). \quad (13)$$

Proposition 1 implies that

$$\mathbb{E}_{t-1} \left[\frac{q}{\beta} \frac{\hat{u}^{t-1\prime}(c_{t-1})}{\hat{u}^{t\prime}(c_t)} \right] = 1,$$

so the term $(q/\beta)\hat{u}^{t-1\prime}(c_{t-1})/\hat{u}^{t\prime}(c_t)$ on the left-hand side of equation (13) 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 (13) implies that the labour wedge remains constant between periods $t-1$ and t, a form of perfect tax smoothing. When $\rho < 1$ the labour wedge reverts to zero at rate ρ.[11]

This result incorporates elements of tax smoothing and mean reversion. It can be compared to the tax smoothing results that have been derived in the context of Ramsey models where

---

11. These 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.

the government must finance a given stream of expenditures with taxes on labour and these taxes are exogenously restricted to be linear. These papers emphasize the importance of the completeness of markets. With complete markets, taxes inherit the serial correlation property of the shocks (Lucas and Stokey, 1983), whereas with incomplete markets, taxes inherit a random walk component (Barro, 1979; Aiyagari *et al.*, 2002). Our model is quite different. In particular, no exogenous restriction on tax instruments is imposed, and distortionary taxes arise endogenously out of a desire to provide social insurance. Our tax-smoothing formula has both differences and similarities with the corresponding results in the Ramsey literature. An important difference is that it applies to the marginal tax rate faced by a given individual in response to idiosyncratic shocks versus aggregate tax rates in response to aggregate shocks. An interesting similarity with Lucas and Stokey (1983) is that taxes inherit the serial correlation of the shocks.

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 labour 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 labour 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 labour supply, explaining the role of the Frisch elasticity.

Returning to the more general statement in Proposition 2, equation (11) shares many ingredients with equation (12). Note however that, in general,

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

will not equal one, so that in contrast with equation (12), the right-hand side cannot be interpreted as a risk-adjusted conditional expectation of the labour wedge in period t. Another important case is $\pi(\theta_t) = (\beta/q)\hat{u}^{t\prime}\big(c\big(\theta^t\big)\big)/\hat{u}^{t-1\prime}\big(c\big(\theta^{t-1}\big)\big)$, so that equation (11) 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 (12) and (13). Rather than develop the expression here, we present its neater continuous-time counterpart in Section 4.

Equations (11) and (12) hold for any allocation that solves the relaxed planning problem. They do not necessarily hold for an allocation that solves the original planning problem 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 the original planning problem 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}.*

Proposition 3. *Suppose that Assumptions 1, 2, and 3 hold, and that $\{c, y\}$ solves the original planning problem. Then the labour wedge satisfies equation (11) for every $t \geq 1$ and history θ^{t-1} with $\pi(\theta) = \theta^{-\alpha}$.*

The proof proceeds by constructing a class of perturbations of the solution of the original planning problem that satisfy all the constraints, and exploiting the condition that their associated resource cost is higher. These perturbation operate at a given history θ^{t-1}, increasing $y(\theta^{t-1})$ and decreasing $y(\theta^t)$ for all successor nodes in such a way that the perturbed allocation is incentive

compatible and delivers the same utility as the original allocation. The construction of these perturbations relies heavily on Assumptions 1, 2, and 3.

It is interesting to compare Propositions 2 and 3. Unlike Proposition 2, Proposition 3 characterizes the solution of the original planning problem even when it does not coincide with the solution of the relaxed planning problem. It places fewer restrictions on the process $\{\frac{\tau_L}{1-\tau_L}, \frac{1}{\tilde{u}''}\}$ since it only applies for a particular weighting function. One way to understand this is as follows. Optimality conditions can be obtained by perturbation arguments: one constructs perturbations that preserves the constraint set of the planning problem and exploit the fact that these perturbations can only increase the objective function to derive first-order conditions. The space of perturbations that preserve incentive compatibility and utility is smaller than the set of perturbations that preserve local incentive compatibility and utility. As a result, the solution of the original planning problem is in general characterized by fewer first-order conditions and more constraints than the solution of the relaxed planning problem, except when these two solutions coincide. Note also that Proposition 2 holds for all stochastic processes, and can actually be extended to general preferences (see Section 5). This is not the case of Proposition 3 which makes crucial use of Assumptions 1, 2, and 3.

3.3.   *Labour wedge at the top and bottom*

We now look at the labour wedge for the two extreme realizations of θ_t, top and bottom and generalize the results obtained by Mirrlees (1971) in a static setting to our dynamic environment. 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, $\overline{\theta}_t(\theta_{t-1})$ and $\underline{\theta}_t(\theta_{t-1})$ vary with θ_{t-1}, with $\Theta = [\underline{\theta}, \overline{\theta}]$ such that $[\underline{\theta}_t(\theta_{t-1}), \overline{\theta}_t(\theta_{t-1})] \subseteq \Theta$ for all t and θ_{t-1}. We assume these functions are differentiable and have bounded derivatives. For short, we often simply write $\overline{\theta}_t$ 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:

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

Intuitively, this is simply the envelope condition using Leibniz's rule. More formally, there are two equivalent ways of approaching the moving support case to justify this necessary condition.

First, one can define allocations only for the set of histories that are consistent with the moving support, restricting reports in the same way. That is, consumption and labour $c(\theta^t)$ and $y(\theta^t)$ are defined for histories θ^t with the property that $\theta_s \in [\underline{\theta}_s(\theta_{s-1}), \overline{\theta}_s(\theta_{s-1})]$ for all $s = 1, 2, \ldots, t$. Reports are also restricted to satisfy $r_s \in [\underline{\theta}_s(r_{s-1}), \overline{\theta}_s(r_{s-1})]$ for all $s = 1, 2, \ldots, t$. This restriction can make one-shot deviations impossible, invalidating our original derivation of the first-order necessary condition. However, in the Appendix we rederive this condition under this restriction using a different set of deviations.

The second way of proceeding is simpler. Without loss of generality one can work with an extended allocation, which specifies consumption and labour for all histories $\theta^t \in \Theta^t$. One then proceeds as in the full support case, imposing incentive compatibility after any history $\theta^t \in \Theta^t$ including those that lie outside the moving support. This is without loss of generality because we can always perform the extension by assigning bundles for consumption and labour that were already offered. Thus, it does not impose any additional constraints, nor does it affect the planning

problem. Using this extended-allocation approach, the derivation of our necessary condition is valid.

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

Proposition 4. *Consider an interior allocation that solves the relaxed planning problem:*

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

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

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

$$\frac{\tau_L\left(\theta^{t-1},\overline{\theta}_t\right)}{1-\tau_L\left(\theta^{t-1},\overline{\theta}_t\right)} = \frac{\tau_L\left(\theta^{t-1}\right)}{1-\tau_L\left(\theta^{t-1}\right)} \frac{\beta}{q} \frac{\hat{u}^{t\prime}\left(c\left(\theta^{t-1},\overline{\theta}_t\right)\right)}{\hat{u}^{t-1\prime}\left(c\left(\theta^{t-1}\right)\right)} \frac{\theta_-}{\overline{\theta}_t} \frac{d\overline{\theta}_t}{d\theta_{t-1}},$$

$$\frac{\tau_L\left(\theta^{t-1},\underline{\theta}_t\right)}{1-\tau_L\left(\theta^{t-1},\underline{\theta}_t\right)} = \frac{\tau_L\left(\theta^{t-1}\right)}{1-\tau_L\left(\theta^{t-1}\right)} \frac{\beta}{q} \frac{\hat{u}^{t\prime}\left(c\left(\theta^{t-1},\underline{\theta}_t\right)\right)}{\hat{u}^{t-1\prime}\left(c\left(\theta^{t-1}\right)\right)} \frac{\theta_-}{\underline{\theta}_t} \frac{d\underline{\theta}_t}{d\theta_{t-1}}.$$

With a fixed support, the no-distortion results from the static model derived by Mirrlees (1971) extend to our dynamic setting. However, when the support is not fixed the labour wedge after an extreme realization is the product of the labour 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: $\tau_L\left(\theta^{t-1},\overline{\theta}_t\right) \leq \tau_L\left(\theta^{t-1}\right) \leq \tau_L\left(\theta^{t-1},\underline{\theta}_t\right)$. The fact that the tax rate may go both up or down illustrates a form of tax smoothing.

4. A CONTINUOUS-TIME APPROACH

In this section, we formulate the relaxed planning problem in continuous time and tackle it using continuous-time stochastic control. There are several advantages to this continuous-time approach. First, Ito calculus provides an elegant counterpart to the set of formulas derived in Section 3. Second, we establish a new result. We show that in continuous-time innovations in the labour wedge are negatively correlated with innovations in consumption. In economic terms, this implies a form of short-term regressivity.

4.1. *A roadmap*

We first explain how to set things up in continuous time by taking a limit of our discrete-time model. We assume that productivity follows a geometric Brownian diffusion with drift. We then set up the planning problem as a stochastic control problem. To do this, we derive the laws of motions for the endogenous state variables v_t and Δ_t as a function of a set of control variables: consumption, c_t, output y_t, and a new variable $\sigma_{\Delta,t}$ representing the sensitivity of Δ_t to productivity shocks. The cost function $K(v_t, \Delta_t, \theta_t, t)$ solves a Hamilton–Jacobi–Bellman (HJB). Its first-order conditions allow us derive results for the optimum.

4.2. *The continuous-time model*

Our approach here is to take the continuous-time limit of the discrete-time model.[12] Let τ be the length of a period. Instead of indexing periods by $t = 1, 2, 3...$ we now take $t = \tau, 2\tau, 3\tau, ...$. We assume that $\theta_{t+\tau}$ is log normally distributed so that

$$\log\theta_{t+\tau} \sim N\left(\log\theta_t + \mu_t^{\log}(\theta_t), \sigma_t^2\right). \tag{14}$$

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

$$\beta = e^{-\rho\tau},\ q = e^{-\rho\tau},\ \mu_t^{\log}(\theta_t) = \tau\left[\hat{\mu}_t(\theta_t) - \frac{1}{2}\theta_t\hat{\sigma}_t^2\right],\ \sigma_t = \hat{\sigma}_t\sqrt{\tau}, \tag{15}$$

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 adjust the scale, we multiply utility and cost by the period length τ. To simplify, we assume here that $q = \beta$. This can be easily generalized to separate the two.

The definition for $\mu_t^{\log}$ 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 $\tau \to 0$, it is well known that there exists a Brownian motion W_t such that the stochastic process $\{\theta\}$ converges to the continuous-time Brownian diffusion with deterministic volatility:

$$\frac{d\theta_t}{\theta_t} = \hat{\mu}_t(\theta_t)d\theta_t + \hat{\sigma}_t dW_t, \tag{16}$$

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. Equivalently, expressed in logs,

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

where $\hat{\mu}_t^{\log}(\theta_t) \equiv \hat{\mu}_t(\theta_t) - \frac{1}{2}\theta_t\hat{\sigma}_t^2$.

To formulate the relaxed planning problem in continuous time, we need to determine the laws of motions for v_t and Δ_t which incorporate our first-order necessary condition for incentive compatibility. This is summarized in the following lemma.

Lemma 2. *There exists a process $\{\sigma_\Delta\}$ such that the state variables $\{v, \Delta\}$ satisfy the following stochastic differential equations:*

$$dv_t = \rho v_t dt - u^t dt + \theta_t\Delta_t\hat{\sigma}_t dW_t, \tag{17}$$

$$d\Delta_t = \left[\left(\rho - \hat{\mu}_t - \theta_t\frac{d\hat{\mu}_t^{\log}}{d\theta}\right)\Delta_t - u_\theta^t - \sigma_{\Delta,t}\hat{\sigma}_t\right]dt + \sigma_{\Delta,t}\hat{\sigma}_t dW_t. \tag{18}$$

Since v_t is the present value of utility it follows that $dv_t = \rho v_t dt - u^t dt + \sigma_{v,t}\sigma_t dW_t$ for some process $\{\sigma_v\}$. The lemma does two things. First, it provides the drift for $\{\Delta\}$. Second, it shows that the volatility $\sigma_{v,t}$ must be $\theta_t\Delta_t$. Intuitively, this follows from the continuous-time limit of our first-order necessary condition for incentive compatibility $\dot{w}(\theta) = \tau\theta u_\theta^t + \beta\theta\Delta(\theta)$, noting that $\tau \to 0$ and $\beta \to 1$.

---

12. It 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.

4.3. *A HJB equation*

Having re-expressed the constraints in the relaxed planning problem as stochastic differential equations for the state variables, we can write the 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 (17), (18), and (16). The controls are $(c_t, y_t, \sigma_{\Delta,t})$. The HJB equation is (suppressing the state $(v_t, \Delta_t, \theta_t, t)$ for notational convenience)

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

$$
+ K_\theta \theta_t \hat{\mu}_t + K_t + \frac{1}{2} K_{vv} \theta_t^2 \Delta_t^2 \hat{\sigma}_t^2 + \frac{1}{2} K_{\Delta\Delta} \sigma_{\Delta,t}^2 \hat{\sigma}_t^2 + \frac{1}{2} K_{vv} \theta_t^2 \Delta_t^2 \hat{\sigma}_t^2 + \frac{1}{2} K_{\Delta\Delta} \sigma_{\Delta,t}^2 \hat{\sigma}_t^2
$$

$$
\left. + \frac{1}{2} K_{\theta\theta} \theta^2 \hat{\sigma}_t^2 + K_{v\Delta} \theta_t \Delta_t \sigma_{\Delta,t} \hat{\sigma}_t^2 + K_{v\theta} \theta_t^2 \Delta_t \sigma_\theta^2 + K_{\Delta\theta} \theta_t \sigma_{\Delta,t} \hat{\sigma}_t^2 \right\}.
$$

4.4. *Optimality conditions*

It will prove convenient to introduce the dual variables of (v_t, Δ_t): $\lambda(v_t, \Delta_t, \theta_t, t) = K_v(v_t, \Delta_t, \theta_t, t)$ and $\gamma(v_t, \Delta_t, \theta_t, t) = K_\Delta(v_t, \Delta_t, \theta_t, t)$. Economically, these variables represent the marginal increase of the cost function when promised utility v_t or Δ_t are marginally increased. As we show below, there exists a simple invariant mapping between these dual variables and easily interpretable features of the allocation: the marginal utility of consumption $u^{t\prime}(c_t)$ and the labour wedge $\tau_{L,t}$.

Proposition 5.   *Suppose that Assumptions 1 and 2 hold, and that productivity evolves according to equation (16). 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*

$$
\frac{d\lambda_t}{\lambda_t} = \sigma_{\lambda,t} \hat{\sigma}_t dW_t \tag{19}
$$

$$
d\gamma_t = \left[-\theta_t \lambda_t \sigma_{\lambda,t} \hat{\sigma}_t^2 + \left(\hat{\mu}_t + \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} \right) \gamma_t \right] dt + \gamma_t \hat{\sigma}_t dW_t, \tag{20}
$$

  with $\gamma_0 = 0$.

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

$$
\frac{1}{\hat{u}^{t\prime}(c_t)} = \lambda_t \qquad and \qquad \frac{1}{\hat{u}^{t\prime}(c_t)} - \frac{\theta_t}{h^{t\prime}(y_t/\theta_t)} = -\alpha \frac{\gamma_t}{\theta_t}.
$$

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

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

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 (19) and (20) 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 $(\lambda_t, \gamma_t, \theta_t, t)$.

An interesting feature of this alternative parametrization of the state space is the existence of a sufficient statistic, the volatility process $\{\sigma_\lambda\}$. This volatility controls how much innovations to productivity are passed through to consumption. It can therefore be thought of as a local proxy for the amount of insurance that is provided at the optimal allocation. Higher values for $\sigma_{\lambda,t}$ provide more incentives at the expense of insurance. Section 6 exploits the fundamental role of $\{\sigma_\lambda\}$ to interpret our numerical findings.

Combining parts (i) and (iii) and using Ito's lemma leads to the following corollary.

Corollary 2.   *Suppose that Assumptions 1 and 2 hold, and that productivity evolves according to equation (16). Then the labour wedge verifies the following stochastic differential equation*

$$d\left(\lambda_t \frac{\tau_{L,t}}{1-\tau_{L,t}}\right) = \left[\lambda_t \frac{\tau_{L,t}}{1-\tau_{L,t}} \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta_t} + \alpha\lambda_t\sigma_{\lambda,t}\hat{\sigma}_t^2\right]dt. \tag{21}$$

This lemma shows that the process $\left\{\lambda \frac{\tau_L}{1-\tau_L}\right\}$ is a diffusion with a particular drift and no volatility. The drift matches its discrete-time counterpart, formula (12). The first-term captures the tax smoothing and mean reverting forces. The second term is an expression for the instantaneous covariance between $\log\theta_t$ and λ_t scaled by α—just as in the discrete-time case.

Interestingly, in continuous time, we get the additional result that this diffusion has zero instantaneous volatility (*i.e.* there is no dW_t term in equation (21)). This implies that the realized paths are of bounded variation (a.s.). This means that the paths vary much less than those for productivity $\{\theta\}$. To draw out more economic implications of this result, apply Ito's lemma using (19) and (21) to obtain:

$$d\left(\frac{\tau_{L,t}}{1-\tau_{L,t}}\right) = \left[\frac{\tau_{L,t}}{1-\tau_{L,t}} \theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta} + \alpha\sigma_{\lambda,t}\hat{\sigma}_t^2\right]dt + \frac{\tau_{L,t}}{1-\tau_{L,t}}\lambda_t d\left(\frac{1}{\lambda_t}\right). \tag{22}$$

This shows explicitly how the innovations in the labour wedge must be perfectly mirrored by those in the marginal utility of consumption $\hat{u}^{t\prime}(c_t) = \lambda_t^{-1}$. This induces a negative instananeous covariance between consumption and the labour wedge. Economically, this represents a form of regressivity, in that good productivity shocks raise consumption and lower the labour wedge, at least in the short run.

Our regressivity result contrasts with the absence of such results in static settings. 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 with virtually no restrictions for a large class of productivity processes.

It is important to emphasize what our regressivity result does and does not say. Over short enough horizons, it guarantees a negative conditional correlation between consumption and the labour wedge. However, this may not translate into a negative correlation over longer horizons. This depends on the evolution of the drift terms in our forumla. In particular, the endogenous volatility term $\sigma_{\lambda,t}$ may vary endogenously and play a central role. We investigate these dynamics more explicitly in Section 6.

Finally, note that using part (iii) in Proposition 5, we can solve for the volatility $\sigma_{\lambda,t}$ in terms of the intertemporal wedge, $\sigma_{\lambda,t} = \frac{\sqrt{\tau_{K,t}}}{\hat{\sigma}_t}$. We can then use this to rewrite these last two equations in terms of the labour and intertemporal wedges. In this way, optimality can be seen as imposing

a joint restriction on the labour and intertemporal distortions:

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

5. GENERAL PREFERENCES

In this section, we investigate what can be said for general utility functions $u^t(c,y,\theta)$. In particular, we want to allow for non-separabilities between consumption and leisure, and also allow for an elasticity of labour supply that varies over the life cycle. Both of these features have been argued by some authors (see, *e.g.*, Saez, 2002; Conesa *et al.*, 2009) to be important to think labour and capital taxation.

It is well known that when consumption and labour are not additively separable, the Inverse Euler equation does not hold. Actually, even when there is no additional uncertainty between $t-1$ and t, so that the Euler and the Inverse Euler equation coincide, the Euler equation might not hold. As is well known, with non-separable preferences, the no capital tax result of Atkinson and Stiglitz (1976) does not hold. The reason for this is that income and productivity now directly affect the intertemporal rate of substitution for consumption. Taxing or subsidizing capital therefore helps separating types. Saez (2002) argues that these non-separabilities are relevant in practice. In particular, he suggests that poor agents have a lower propensity to save, and shows in that context that optimal capital taxes are positive. These forces also upset the Inverse Euler equation when there is additional uncertainty between $t-1$ and t.

Conesa *et al.* (2009) argue that the elasticity of labour supply falls over the life cycle. To capture that, they use preferences that are isoelastic over leisure (instead of labour). This implies that labour supply is more elastic when labour is low. Since labour decreases over the life cycle, labour supply is more elastic for older agents. This produces a force for decreasing labour taxes with age. Conesa *et al.* (2009) make the additional point that if age-dependent taxes are not available, then capital taxes emerge as a partial substitute.

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 $\eta_t(y,w,\theta) = \alpha$.

5.1. *Discrete time*

In order to generalize equation (12), we need to introduce the dual of the variable $v(\theta^t)$ defined by

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

At an optimum, we have the martingale relation

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

When utility is separable—Assumption 1 holds—we have $\frac{q}{\beta}\lambda(\theta^t) = \frac{1}{u_c^t(\theta^t)}$. The martingale relation then directly implies the Inverse Euler equation. This simple link between $\lambda(\theta^t)$ and $u_c^t(\theta^t)$ no longer holds when preferences are not separable.

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 6. *Suppose that $\{c, y\}$ solves the relaxed planning problem. Then the labour wedge satisfies the following equation for every $t \geq 1$ and history θ^{t-1}*

$$\int \frac{1}{\eta_t(\theta^t)} \frac{\tau_L(\theta^t)}{1 - \tau_L(\theta^t)} \frac{q}{\beta} \frac{u_c^t(\theta^{t-1})}{u_c^t(\theta^t)} f^t(\theta_t | \theta_{t-1}) d\theta_t = \frac{1}{\eta_{t-1}(\theta^{t-1})} \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}}$$

$$+ u_c^t(\theta^{t-1}) \lambda(\theta^{t-1}) \int \log(\theta_t) \left[\frac{q}{\beta} \frac{\lambda(\theta^t)}{\lambda(\theta^{t-1})} - 1 \right] f^t(\theta_t | \theta_{t-1}) d\theta_t. \quad (23)$$

This proposition generalizes equation (12). The martingale relation satisfied by $\lambda(\theta^t)$ implies that we can rewrite the second term on the right-hand side of equation (23) as a covariance: $\text{Cov}_{t-1}(\log(\theta_t), \frac{q}{\beta} \frac{\lambda_t}{\lambda_{t-1}})$. We could also generalize equation (11) along the exact same lines.

There are two important differences between equations (12) and (23). First, we have already noted that unless utility is separable, 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, in contrast with equation (12), 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, $\frac{1}{\eta_t(\theta^t)} \frac{\tau_L(\theta^t)}{1 - \tau_L(\theta^t)}$ replaces $\frac{\tau_L(\theta^t)}{1 - \tau_L(\theta^t)}$. When Assumptions 1 and 2 hold, $\eta_t = \alpha$, we can multiply through by α, as in equation (12). Otherwise, the general equation indicates that changes in the elasticity should affect the labour wedge. To elaborate on this point, it will prove convenient to specialize the discussion to a class of generalized isoelastic preferences for which equation (23) takes a simpler form.

Assumption 4. *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 4 holds and that $\{c, y\}$ solves the relaxed planning problem. Then the labour wedge satisfies the following equation for every $t \geq 1$ and history θ^{t-1}*

$$\int \frac{\tau_L(\theta^t)}{1 - \tau_L(\theta^t)} \frac{q}{\beta} \frac{u_c^t(\theta^{t-1})}{u_c^t(\theta^t)} f^t(\theta_t | \theta_{t-1}) d\theta_t = \frac{\alpha_t}{\alpha_{t-1}} \frac{\tau_L(\theta^{t-1})}{1 - \tau_L(\theta^{t-1})} \theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}}$$

$$+ \alpha_t u_c^t(\theta^{t-1}) \lambda(\theta^{t-1}) \int \log(\theta_t) \left[\frac{q}{\beta} \frac{\lambda(\theta^t)}{\lambda(\theta^{t-1})} - 1 \right] f^t(\theta_t | \theta_{t-1}) d\theta_t. \quad (24)$$

Recall that $1/(\alpha_t - 1)$ is the the Frisch elasticity of labour supply. This formula makes clear that both the level α_t and the growth rate $\frac{\alpha_t}{\alpha_{t-1}}$ matter. A higher growth rate of α_t increase the auto-regressive coefficient $\frac{\alpha_t}{\alpha_{t-1}} \theta_{t-1} \frac{d\phi_t^{\log}(\theta_{t-1})}{d\theta_{t-1}}$ of the labour wedge. This is a manifestation of a standard inverse elasticity principle: labour is taxed more in periods in which it is less elastic. A high level of α_t also increases the drift term. This is because when labour is less elastic, increases

in uncertainty lead to larger increases in taxes—the marginal cost of increasing taxes is lower, although the marginal benefit of increasing taxes is unchanged.

5.2. *Continuous time*

Our continuous-time analysis can also be extended to general preferences. In particular, we can generalize equation (25).

Proposition 7. *Suppose that productivity evolves according to equation (16). Then the labour wedge satisfies the following stochastic differential equation*

$$d\left(\frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{u_c^t}\frac{1}{\eta_t}\right) = \left[\lambda_t \sigma_{\lambda,t}\hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{u_c^t}\frac{1}{\eta_t}\theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta_t}\right]dt. \qquad (25)$$

This expression is the continuous-time analogue of equation (23). 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 (25) takes a simple and illuminating form when preferences are in the generalized isoelastic class defined by Assumption 4.

Corollary 4. *Suppose that Assumption 4 holds and that productivity evolves according to equation (16). Then the labour wedge satisfies the following stochastic differential equation:*

$$d\left(\frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{u_c^t}\right) = \left[\alpha_t\lambda_t \sigma_{\lambda,t}\hat{\sigma}_t^2 + \frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{u_c^t}\left(\theta_t \frac{d\hat{\mu}_t^{\log}}{d\theta_t} + \frac{1}{\alpha_t}\frac{d\alpha_t}{dt}\right)\right]dt. \qquad (26)$$

Equation (26) clearly illustrates the impact of a time-varying α_t (and hence a time-varying Frisch elasticity of labour supply). The growth rate $\frac{1}{\alpha_t}\frac{d\alpha_t}{dt}$ increases the auto-regressive coefficient and the level α_t increases the drift of $\frac{\tau_{L,t}}{1-\tau_{L,t}}\frac{1}{u_c^t}$. The intuition is similar to the one given above for the discrete-time case.

We can also derive a generalization of equation (22):

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

This shows explicitly how innovations in the labour 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. We verify numerically that the solution of the relaxed planning problem satisfies the constraints of the original planning problem. Therefore, the solution that we characterize is the solution of the original planning problem.

6.1.   *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 $t = 1, 2, \ldots, 40$ and

$$\log(c_t)$$

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

A fundamental primitive in our exercise is the stochastic process for productivity. Most empirical studies estimate an AR(1) plus white noise, where the white noise is sometimes interpreted as measurement error. Typically, the coefficien of auto-correlation is estimated to be very close to one. We therefore adopt a geometric random walk:

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

with $\log \varepsilon \sim N(-\frac{\hat{\sigma}^2}{2}, \hat{\sigma}^2)$.[13] A key parameter is the degree of uncertainty $\hat{\sigma}^2$. Empirical estimates vary quite substantially, due to differences in methodologies, econometric specifications, and data sets. Typically, this number is estimated by matching the increase in the cross-sectional variance of wages or earnings in a given cohort as this cohort ages. The estimate for $\hat{\sigma}^2$ depends on whether time fixed effects (smaller estimates) or cohort fixed effects (larger estimates) are imposed, and on the time period (larger estimates in the 1980's). Using time fixed effects over the period 1967–2005, Heathcote, Perri, and Violante (2010) find $\hat{\sigma}^2 = 0.00625$ for the wages of male individuals. Using cohort fixed effects over the period 1967–1996, Heathcote *et al.* (2005) find $\hat{\sigma}^2 = 0.0095$ for the wages of male individuals. Using cohort fixed effects over the period 1980–1996, Storesletten *et al.* (2004) find $\hat{\sigma}^2 = 0.0161$ for household earnings. Instead of trying to settle on one particular estimate to base our numerical exploration, we performed three calibrations based on these three value for $\hat{\sigma}^2$: a low risk calibration with $\hat{\sigma}^2 = 0.00625$, a medium risk calibration with $\hat{\sigma}^2 = 0.0095$ and a high risk calibration with $\hat{\sigma}^2 = 0.0161$. In the figures and discussions below, to avoid repetition, we focus on the medium risk calibration, but the qualitative results are similar and our two tables offer the quantitative conclusions for all three calibrations.

The value function satisfies

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

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 $(v + (1 + \beta + \cdots + \beta^{T-t})\log\phi, \phi^{-1}\Delta, \phi\theta_-)$ 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 $t = 1, 2, \ldots, T$. For any given v_1, we initialize Δ_1 at $t = 1$ to minimize cost

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

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

---

13. For our numerical simulation, we truncate the normal distribution: the density of $\log \varepsilon$ is proportional to the density of the normal over a finite interval $[\underline{\varepsilon}, \bar{\varepsilon}]$.

6.2. *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 labour wedge is zero and the intertemporal wedge is a positive constant equal to $1 - Rq = 1 - R\beta > 0$, where $R \equiv \beta^{-1}(\mathbb{E}\frac{1}{\varepsilon})^{-1}$.[14]

Consider next the first-best planning problem given by

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

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

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

so that output increases with productivity. Both the labour and intertemporal wedges are zero.

6.3. *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 labour and intertemporal wedges. During retirement each agent's consumption is constant, while output and wedges are zero. Thus, we focus on the working periods $t = 1, 2, \ldots, 40$.

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 5 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(b) plots the average variance of 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

---

14. Alternatively, 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.
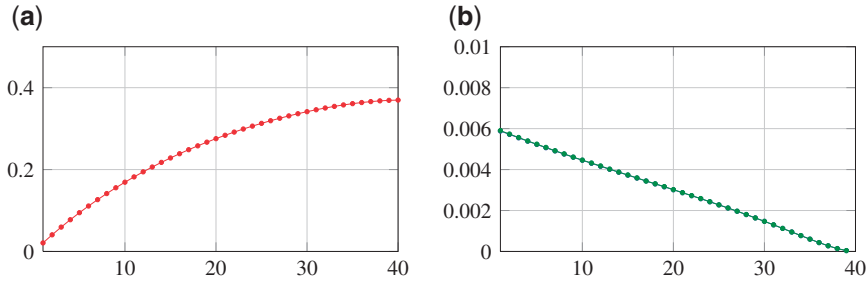
FIGURE 1

Average wedges over time. (a) Average labour wedge over time. (b) Intertemporal wedge and variance of consumption growth over time—both series are indistinguishable to the naked eye

income mechanism. Indeed, this property would be present at an equilibrium with no taxes and free-savings. Second, as we show below, the labour 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 labour wedge starts near zero and increases over time, asymptoting around 37% at retirement. Panel (b) displays the intertemporal wedge, which displays the reverse pattern. It is decreasing over time, starting around 0.6%—which represents an implicit tax on net interest of around 12%—and falling to zero at retirement.[15] Both of these findings are easily explained by our theoretical results, together with the behaviour of the average variance of consumption growth.

As shown in equations (13) and (21), when $\rho = 1$, the expected change in the labour 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 labour wedge at retirement.

As for the intertemporal wedge, equation (10) 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 labour wedge.[16]

---

15. To 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 $1 + (1-\hat{\tau}_{K,t})r = (1-\tau_{K,t})(1+r)$. 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 $\frac{1+r}{r} \approx \frac{1}{r} \approx 20$.

16. Note 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
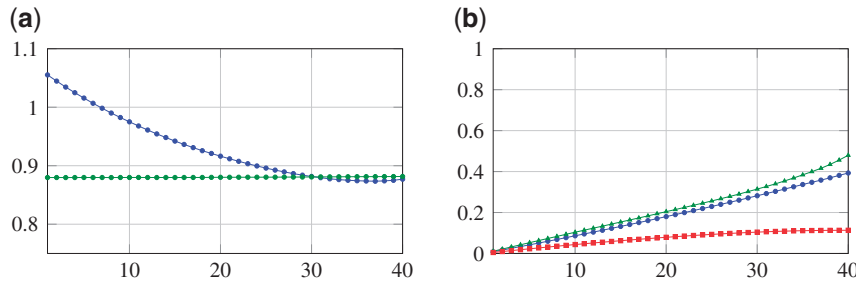
(a)

(b)

FIGURE 2

Statistics for optimal allocation over time. (a) Average for output (declining line) and consumption (constant line) over time. (b) Variance of output (top line), productivity (middle line) and consumption (bottom line) over time
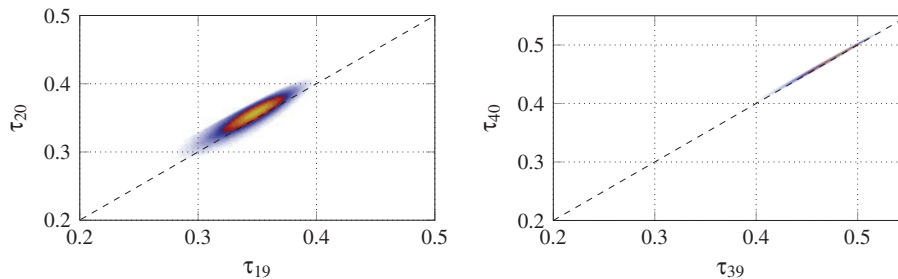

FIGURE 3

Scatter plot of $\tau_{L,t}$ versus $\tau_{L,t-1}$ at $t=20$ and $t=40$

Panel (b) shows the cross-sectional variance for consumption, productivity, and output. The variance of productivity grows, by assumption, linearly. The variance of output is higher and grows in a convex manner. The variance of consumption, on the other hand, it 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 labour wedge formula by showing scatter plots of the current labour wedge against the previous period's labour wedge. In period $t=20$, the average

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.
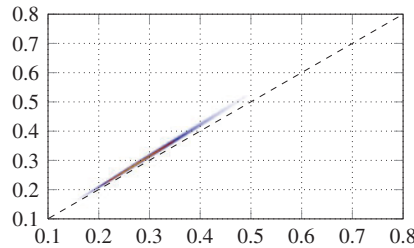
FIGURE 4

Scatter plot 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 $t=20$
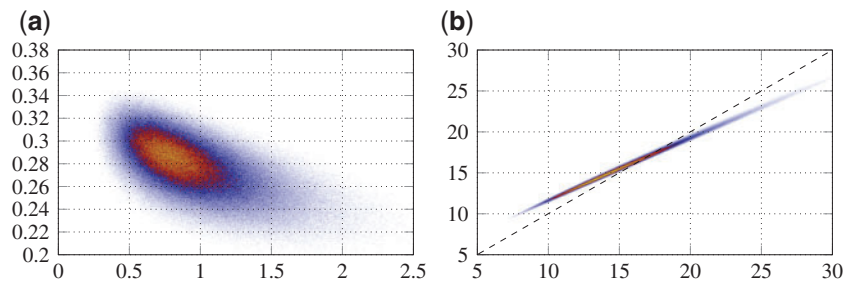

FIGURE 5

History dependence and insurance. (a) Scatter plot of $\tau_{L,t}$ versus θ_t at $t=20$. (b) Scatter plot of $\sum_{t=1}^{T}q^t c_t$ versus $\sum_{t=1}^{T}q^t y_t$

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 (13). 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 labour 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° line. The reason there is no dispersion around the average relationship is an implication of the results in Section 4 that show that unpredictable changes in the labour wedge are related to unpredictable changes in marginal utility. Near retirement, consumption becomes almost perfectly predictable, so the labour 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° line and the dispersion around this relationship is minimal. This illustrates the results in Section 4, that there is no instantaneous volatility in $\{\frac{\tau_L}{1-\tau_L}\frac{1}{u'(c)}\}$. In other words, unpredictable changes in the labour wedge $\{\frac{\tau_L}{1-\tau_L}\}$ are entirely explained by unpredictable changes in the reciprocal of marginal utility $\{\frac{1}{u'(c)}\}$.

Figure 5(a) plots the current period's labour 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 Section 4 point towards a negative correlation between the labour 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 labour wedge. If the labour wedge were solely a function of current productivity, then there would be no dispersion. Thus, this dispersion illustrates the history dependence in the labour 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 labour wedge. The history of productivity shocks affects savings decisions. The accumulated wealth, in turn, affects the current labour choice, determining the position, and marginal tax rate, along the fixed non-linear tax schedule.

Figure 5(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.

With $\hat{\sigma}^2 = 0.0061$ or $\hat{\sigma}^2 = 0.0161$, the results show the same qualitative patterns as the benchmark. Quantitatively, both the labour and intertemporal wedges are lower with $\hat{\sigma}^2 = 0.0061$, with the labour 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 labour and intertemporal wedges. Conversely, both the labour and intertemporal wedges are higher with $\hat{\sigma}^2 = 0.0161$, with the labour wedge peaking at 40% and the intertemporal wedge starting at 1%—which represents an implicit tax on net interest of around 17%.

6.4. *Labour wedge dynamics: an impulse response*

The scatter of the labour wedge at t against the labour 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 $t = 1, 2, \ldots, 40$. We plot the evolution of the labour 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 labour wedge (consumption, not shown, jumps upward). After the shock, the labour wedge displays a higher rate of growth. In the figure, the labour 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 labour wedge may be enough to over come the initial jump downward. But why does the labour wedge grow faster after a shock? Panel (b)
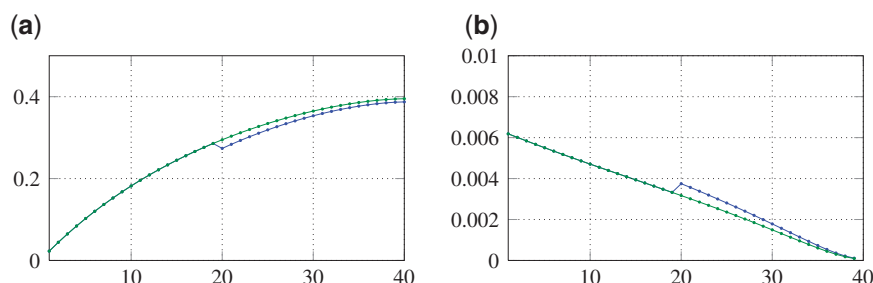
(a)

(b)

FIGURE 6

Impulse response to shock at $t = 20$. (a) Labour wedge over time. (b) Intertemporal wedge over time

TABLE 1

Welfare gains over free-savings, no-tax equilibrium

	$\hat{\sigma}^2 = 0.0061$	$\hat{\sigma}^2 = 0.0095$	$\hat{\sigma}^2 = 0.0161$
Second-best (%)	0.86	1.56	3.43

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 labour 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.

6.5. *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 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 the relaxed planning problem. 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 $\hat{\sigma}^2$, whereas the second and third report simulations with a lower value and a higher value, respectively. As expected, the welfare gains increase with $\hat{\sigma}^2$.

6.6. *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,

TABLE 2

Welfare from simple tax policies: history-independent (linear) but possibly age-dependent taxes

	$\hat{\sigma}^2 = 0.0061$	$\hat{\sigma}^2 = 0.095$	$\hat{\sigma}^2 = 0.0161$
Age-dependent τ_L and τ_K (%)	0.71	1.47	3.30
Age-dependent τ_L, and $\tau_K = 0$ (%)	0.66	1.38	3.16
Age-dependent τ_L, age-independent τ_K (%)	0.70	1.46	3.29
age-independent τ_L and τ_K (%)	0.54	1.14	2.71

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 labour combined with an age-independent capital tax. In this case, we set the labour tax rates to the corresponding cross-sectional averages in the second-best, but optimize over the constant capital tax rate.

Table 2 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 exercises for three values of $\hat{\sigma}$, since the findings are qualitatively similar in both cases, we will focus our discussion on our benchmark reported in the middle 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 1.47% in lifetime consumption, compared to the 1.56% obtained by the second best. Remarkably, age-dependent linear taxes deliver 95% 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.[17] 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 simpler 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 labour taxes but setting capital taxes to zero delivers a welfare gain of 1.38%. The difference of 0.09% 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 labour tax, but allows for a non-zero, age-independent tax on capital. This improves welfare to 1.46%, very close to the welfare obtained by age-dependent labour and capital taxes of 1.47% from the first row. The optimal age-independent intertemporal wedge is 0.27% (corresponding to tax rate of 5.40% on the net interest). Interestingly, this is close to the average wedge across ages from the second-best simulation, as displayed in Figure 1.

---

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

The last row reports welfare for the simplest tax system we consider: age-independent linear labour and capital taxes. The optimal age-independent linear tax on labour is equal to 21.74%, 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 1.14%. Comparing this to the first row, we see that the cost of imposing an age-independent tax system is roughly 0.33% 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 0.068%, corresponding to a tax rate of around 1.36% 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 labour tax is somewhat sophisticated and features either age-dependence or the richer history-dependence of the second best.

With an age-dependent labour tax, an age-independent tax on capital provides modest but non-negligible benefits, equal to 0.08%. 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 labour provide a sizable improvement of 0.33% over the completely age-independent tax system. Allowing for age-dependent labour 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 labour taxes are age-dependent, yet minuscule when labour 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 labour taxes cannot be age dependent (row 4). The reason is that a capital subsidy could help mimic an increasing age profile of labour taxes. Intuitively, labour-income earned earlier in life, whereas 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.[18] When we allow for age-dependent labour taxes, the second force is absent leading to a positive tax on capital. When the labour tax cannot depend on age, both forces are present and roughly cancel each other out, resulting in a practically zero tax on capital.

7. CONCLUSION

In this article, we consider a dynamic Mirrlees economy in a life-cycle context and study the optimal insurance arrangement. Individual productivity evolves as a general Markov process and is private information. We allow for a very general class of preferences. 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 labour-income tax rate. When productivity is an AR(1) our formula resembles an AR(1) with a trend. The auto-regressive coefficient equals that of productivity. The trend term equals the covariance productivity with consumption growth divided by the Frisch elasticity of labour. The innovations in the tax rate are the negative of consumption growth. The last property implies a form of short-run regressivity.

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

Our simulations illustrate these results and deliver some novel insights. The average labour tax rises from 0% to 37% over 40 years, whereas the average tax on savings falls from 12% 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. Hence, it seems that numerically, the history dependence of taxes that are required to implement the full optimum is not an important feature in terms of welfare. Moreover, our simulations emphasize that from a welfare perspective, labour taxes play a more important role than capital taxes (setting capital taxes to zero does not lead to a large deterioration of welfare).

In future work, we plan to enrich the model to incorporate important life-cycle considerations that are absent in our present model: human capital accumulation, endogenous retirement, a more realistic life-cycle profile of earnings, etc. We also plan to continue our numerical explorations by thoroughly investigating the quantiative comparative statics of our model with respect to the stochastic process of earnings, preference parameters, and tastes for initial redistribution.

APPENDIX

Proof of Lemma 1. Define

$$M\left(\tilde{\theta}_t\right) \equiv \int w\left(\theta^{t-1}, r, \theta_{t+1}\right) f^{t+1}\left(\theta_{t+1} | \tilde{\theta}_t\right) d\theta_{t+1}.$$

We argue that the derivative of M exists and can be computed by differentiating under the integral. Since u is bounded, w is bounded. This implies that the derivative of the integrand, $w(\theta^{t-1}, r, \theta_{t+1}) g^{t+1}(\theta_{t+1} | \theta_t)$, is bounded. It then follows that M is differentiable and

$$M'\left(\tilde{\theta}_t\right) = \int w\left(\theta^{t-1}, r, \theta_{t+1}\right) g^{t+1}\left(\theta_{t+1} | \tilde{\theta}_t\right) d\theta_{t+1}.$$

Note that M' is bounded.

All the conditions for Theorem 2 in Milgrom and Segal (2002) are satisfied for the maximization problem in equation (2) and the result follows.   ‖

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 neighbourhood $|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon$ 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

$$\beta \int_{|\tilde{\theta}^{t-1} - \theta^{t-1}| \leq \varepsilon} \frac{1}{\hat{u}^{t-1'}\left(c\left(\tilde{\theta}^{t-1}\right)\right)} f^{t-1}\left(\tilde{\theta}_{t-1} | \tilde{\theta}_{t-2}\right) ... f^0\left(\tilde{\theta}_0 | \tilde{\theta}_{-1}\right) d\tilde{\theta}_{t-1} ... d\tilde{\theta}_0$$

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

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

Proof of Proposition 2. We tackle the Bellman equation satisfied by the relaxed planning problem using optimal control. Define $C^t(y,u,\theta)$ denote the expenditure function—the inverse of the utility function for consumption $u^t(\cdot,y,\theta)$. We first rewrite this Bellman equation as follows:

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

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

$$\Delta=\int w(\theta)g^t(\theta|\theta_-)d\theta,$$

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

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. 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

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

We denote by $\mu(\theta)$ the co-state variable associated with $w(\theta)$. We then form the corresponding Hamiltonian:

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

$$+\lambda\left[v-w(\theta)f^t(\theta|\theta_-)\right]+\gamma\left[\Delta-w(\theta)g^t(\theta|\theta_-)\right]+\mu(\theta)\left[u_\theta^t(C^t(y(\theta),w(\theta)-\beta v(\theta),\theta),y(\theta),\theta)+\beta\Delta(\theta)\right].$$

The boundary conditions are

$$\lim_{\theta\to\bar\theta}\mu(\theta)=0\text{ and }\lim_{\theta\to\underline\theta}\mu(\theta)=0.\qquad\text{(A.2)}$$

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

$$\frac{d\mu(\theta)}{d\theta}=-\left[\frac{1}{\hat u^{t\prime}(c(\theta))}-\lambda-\gamma\frac{g^t(\theta|\theta_-)}{f^t(\theta|\theta_-)}\right]f^t(\theta|\theta_-).\qquad\text{(A.3)}$$

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

$$\frac{\mu(\theta)}{\theta f^t(\theta|\theta_-)}=-\frac{q}{\beta}\frac{\gamma(\theta)}{\theta},\qquad\text{(A.4)}$$

$$\frac{1}{\hat u^{t\prime}(c(\theta))}=\frac{q}{\beta}\lambda(\theta),\qquad\text{(A.5)}$$

and

$$\left(1-\frac{h_y(y(\theta),\theta)}{\hat u^{t\prime}(c(\theta))}\right)=\frac{\mu(\theta)}{f^t(\theta|\theta_-)}\left[-h_{y\theta}(y(\theta),\theta)\right].\qquad\text{(A.6)}$$

Using equation (A.5) to replace λ by $(\beta/q)\left(1/\hat u^{t-1\prime}(c_-)\right)$ in equation (A.3), and integrating and using equation (A.2) we get

$$0=\int\left[\frac{1}{\hat u^{t\prime}(c(\theta))}-\frac{\beta}{q}\frac{1}{\hat u^{t-1\prime}(c_-)}\right]f^t(\theta|\theta_-)d\theta,$$

which provides another proof of Proposition 1 .

These first-order conditions generalize the first-order conditions of the static Mirrlees model. Indeed, using the expression for the labour wedge, equation (A.6) can be rewritten as

$$\tau_L(\theta)=\frac{\mu(\theta)}{f^t(\theta|\theta_-)}\left[-h_{y\theta}(y(\theta),\theta)\right].$$

This equation relates the labour wedge distortion to the co-state $\mu(\theta)$ and the cross-partial $h_{y\theta}(y(\theta),\theta)$. It is familiar from the static Mirrlees model. The labour wedge is positive when the incentive constraints bind downwards and when higher types have a lower marginal disutility of labour income than lower types.

Equation (A.3) is the evolution equation for the co-state $\mu(\theta)$ and is also familiar from the static Mirrlees model. Indeed the exact same equation holds in the static Mirrlees model, with $\gamma=0$. Combining it with equation (A.5) and the Inverse Euler equation, it implies that in the static Mirrlees model, the derivative of the co-state $\mu(\theta)$ is positive for low θ and negative for high θ, which together with the boundary conditions (A.2) ensures that $\tau_L(\theta)$ is always positive.

A key difference between our dynamic model and the static Mirrlees model is that we can have $\gamma \neq 0$. Indeed equation (A.4) shows that the value of γ is (negatively) related to the past value of the co-state.

We now make Assumption 2 and manipulate these first-order conditions to obtain formulas describing the evolution of labour wedges. Actually, we derive a whole set of such formulas, one per weighting functions $\pi(\theta)$. The formulas corresponding to different weighting functions conveniently encode not on only the evolution of labour wedges over time, but also across states.

We can simplify equation (A.6) as follows:

$$1 - \frac{\frac{1}{\theta}\kappa\left(\frac{y(\theta)}{\theta}\right)^{\alpha-1}}{\hat{u}^{t\prime}(c(\theta))} = \alpha \frac{\mu(\theta)}{\theta f^{t}(\theta|\theta_{-})}\hat{u}^{t\prime}(c(\theta))\frac{\frac{1}{\theta}\kappa\left(\frac{y(\theta)}{\theta}\right)^{\alpha-1}}{\hat{u}^{t\prime}(c(\theta))}. \tag{A.7}$$

Replacing the expression for the labour wedge in this last condition, and multiplying both sides by $\pi(\theta)$, we get

$$\frac{\tau_{L}(\theta)}{1-\tau_{L}(\theta)}\frac{1}{\hat{u}^{t\prime}(c(\theta))}\pi(\theta)f^{t}(\theta|\theta_{-}) = \alpha\mu(\theta)\frac{\pi(\theta)}{\theta}.$$

Integrating by parts this equality, we get

$$\int \frac{\tau_{L}(\theta)}{1-\tau_{L}(\theta)}\frac{1}{\hat{u}^{t\prime}(c(\theta))}\pi(\theta)f^{t}(\theta|\theta_{-})d\theta = \alpha\int \mu(\theta)\frac{\pi(\theta)}{\theta}d\theta = \alpha[\mu(\theta)\Pi(\theta)]_{\underline{\theta}}^{\overline{\theta}}$$

$$+\alpha\int \Pi(\theta)\left[\frac{1}{\hat{u}^{t\prime}(c(\theta))}-\lambda-\gamma\frac{g^{t}(\theta|\theta_{-})}{f^{t}(\theta|\theta_{-})}\right]f^{t}(\theta|\theta_{-})d\theta$$

$$=\alpha\int \Pi(\theta)\left[\frac{1}{\hat{u}^{t\prime}(c(\theta))}-\lambda\right]f^{t}(\theta|\theta_{-})d\theta - \alpha\frac{\gamma}{\theta_{-}}\theta_{-}\frac{d\phi^{\Pi}(\theta_{-})}{d\theta_{-}},$$

where we have used the fact that

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

Now note that

$$\frac{\tau_{L}(\theta)}{1-\tau_{L}(\theta)}\frac{1}{\hat{u}^{t\prime}(c(\theta))} = \alpha\frac{\mu(\theta)}{\theta f^{t}(\theta|\theta_{-})} = -\alpha\frac{q}{\beta}\frac{\gamma(\theta)}{\theta},$$

so that we also have

$$\frac{\tau_{L-}}{1-\tau_{L-}}\frac{1}{\hat{u}^{t-1\prime}(c_{-})} = -\alpha\frac{q}{\beta}\frac{\gamma}{\theta_{-}}.$$

Similarly we have

$$\lambda = \frac{\beta}{q}\frac{1}{\hat{u}^{t-1\prime}(c_{-})}.$$

This implies that

$$\int \frac{\tau_{L}(\theta)}{1-\tau_{L}(\theta)}\frac{q}{\beta}\frac{\hat{u}^{t-1\prime}(c_{-})}{\hat{u}^{t\prime}(c(\theta))}\pi(\theta)f^{t}(\theta|\theta_{-})d\theta = \alpha\int \Pi(\theta)\left[\frac{q}{\beta}\frac{\hat{u}^{t-1\prime}(c_{-})}{\hat{u}^{t\prime}(c(\theta))}-1\right]f^{t}(\theta|\theta_{-})d\theta + \frac{\tau_{L-}}{1-\tau_{L-}}\theta_{-}\frac{d\phi^{\Pi}(\theta_{-})}{d\theta_{-}}.$$

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 neighbourhood $|\tilde{\theta}^{t-1}-\theta^{t-1}|\leq\varepsilon$ of this history, and the following perturbed allocation $\{\tilde{c}^{\delta},\tilde{y}^{\delta}\}$. Define for every $\tilde{\theta}^{t-1}$ such that $|\tilde{\theta}^{t-1}-\theta^{t-1}|\leq\varepsilon$,

$$\left(\tilde{y}^{\delta}\left(\tilde{\theta}^{t}\right)\right)^{\alpha} = \left(y\left(\tilde{\theta}^{t}\right)\right)^{\alpha}-\frac{\delta}{\chi},$$

$$\left(\tilde{y}^{\delta}\left(\tilde{\theta}^{t-1}\right)\right)^{\alpha} = \left(y\left(\tilde{\theta}^{t-1}\right)\right)^{\alpha}+\beta\delta,$$

and for every other θ^{s}

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

Finally for every θ^{s}, define

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

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

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

$$+ \int \left[1 - \frac{q}{\beta} \frac{\hat{u}^{t-1\prime}(c(\theta^{t-1}))}{\hat{u}^{t\prime}(c(\theta^t))}\right] \left(\frac{\theta_t}{\theta_{t+1}}\right)^\alpha f^t(\theta_t|\theta_{t-1})d\theta_t.$$

This completes the proof.   ‖

A. DERIVATION OF NECESSARY CONDITION FOR INCENTIVE COMPATIBILITY WITH MOVING SUPPORT

In this Appendix, we reconsider the case with a moving support for productivity and provide an alternative derivation of the same necessary condition for incentive compatibility. In the text we justified the same necessary condition by arguing that one can, without loss of generality, consider mechanisms that allow any report in Θ, regardless of past reports. Instead, here, we assume the agent is confronted with a direct mechanism that restricts reports to lie in the support implied by the previous period's report, so that $r_t \in [\underline{\theta}_t(r_{t-1}), \overline{\theta}_t(r_{t-1})]$. This restriction implies that the agent may not be able to tell the truth after a lie, *i.e.* we may have $\theta_t \notin [\underline{\theta}_t(r_{t-1}), \overline{\theta}_t(r_{t-1})]$ with positive probability if $r_{t-1} \neq \theta_{t-1}$.

To proceed it is useful to have a more forward-looking and recursive notation for reporting strategies. After any history of reports and true shocks (r^{t-1}, θ^t) the agent must make current and future reports. Thus, a strategy requires specifying the current report and the strategy for the next period, as a function of the new shock realization θ_{t+1}. Note that the current report must lie in the support implied by the previous report r_{t-1}. We denote the set of all possible reporting strategies by $\tilde{\Sigma}_t(r_{t-1})$. Note that this set only depends on the previous period's report r_{t-1}, and not on past reports r^{t-2} or the history of true productivity θ^t.

A strategy can be written recursively as

$$\tilde{\sigma}_t = (r_t, S_t) \in \tilde{\Sigma}_t(r_{t-1}),$$

where $r_t \in [\underline{\theta}_t(r_{t-1}), \overline{\theta}_t(r_{t-1})]$ and $S_t : \Theta \to \tilde{\Sigma}_{t+1}(r_t)$ is a measurable function which determines the continuation strategy $\tilde{\sigma}_{t+1}$ as a function of θ_{t+1}. Starting in the last period and working backwards one can use this relation to define $\tilde{\Sigma}_t(r_{t-1})$ for all nodes.

Consider an allocation $\{c, y\}$. At any node (r^{t-1}, θ^t), given strategy $\tilde{\sigma}_t = (r_t, S_t) \in \tilde{\Sigma}_t(r_{t-1})$, we can consider the agent's continuation utility. Note that this utility is independent of θ^{t-1}, but does depend on the history of reports r^{t-1} and current productivity θ_t and satisfies the following recursive relation:

$$w(r^{t-1}, \theta_t; \tilde{\sigma}_t) = u^t(c(r^{t-1}, r_t), y(r^{t-1}, r_t); \theta_t) + \beta \int_{\underline{\theta}_{t+1}(\theta_t)}^{\overline{\theta}_{t+1}(\theta_t)} w((r^{t-1}, r_t), \theta_{t+1}; S_t(\theta_{t+1})) f^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}$$

with $w(r^T, \theta^{T+1}; \tilde{\sigma}_{T+1}) \equiv 0$.

Define the truth-telling strategy $\tilde{\sigma}_t^*$ is defined as $r_t = \theta_t$ and $S(\theta_{t+1}) = \tilde{\sigma}_{t+1}^*$ for all θ_{t+1}. This strategy is always available if $r_{t-1} = \theta_{t-1}$. Denote continuation utility along the equilibrium with truth telling as

$$w(\theta^t) = w(\theta^{t-1}, \theta_t; \tilde{\sigma}_t^*).$$

With this notation for strategies our notion of incentive compatibility is

$$w(\theta^t) \geq w(\theta^{t-1}, \theta_t; \tilde{\sigma}_t) \qquad \forall \tilde{\sigma}_t \in \tilde{\Sigma}_t(\theta_{t-1}),$$

for all histories θ^t. Equivalently

$$w(\theta^{t-1}, \theta_t; \tilde{\sigma}_t^*) = \max_{(r_t, S_t) \in \tilde{\Sigma}_t(\theta_{t-1})} \{u^t(c(\theta^{t-1}, r_t), y(\theta^{t-1}, r_t); \theta_t) + \beta \int_{\underline{\theta}_{t+1}(\theta_t)}^{\overline{\theta}_{t+1}(\theta_t)} w((\theta^{t-1}, r_t), \theta_{t+1}; S_t(\theta_{t+1})) f^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1}\}.$$

Recall that we have defined $g^t(\theta'|\theta) = \partial f^t(\theta'|\theta)/\partial\theta$. An envelope condition then suggests that

$$\frac{\partial}{\partial\theta_t} w(\theta^t) = u_\theta^t(c(\theta^t), y(\theta^t); \theta_t) + \beta \int_{\underline{\theta}_{t+1}(\theta_t)}^{\overline{\theta}_{t+1}(\theta_t)} w(\theta^{t+1}) g^{t+1}(\theta_{t+1}|\theta_t) d\theta_{t+1} + \frac{d\overline{\theta}_{t+1}}{d\theta_t} w(\theta^t, \overline{\theta}_{t+1}(\theta_t)) f^{t+1}(\overline{\theta}_{t+1}(\theta_t)|\theta_t)$$

$$- \frac{d\underline{\theta}_{t+1}}{d\theta_t} w(\theta^t, \underline{\theta}_{t+1}(\theta_t)) f^{t+1}(\underline{\theta}_{t+1}(\theta_t)|\theta_t)$$

The corresponding integral version can be derived formally using the results of Milgrom and Segal (2002) exactly as in the proof of 1.

Proof of Proposition 4. We use optimal control to analyse the modified version of Bellman equation (A.1):

$$K(v, \Delta, \theta_-, t) = \min \int_{\underline{\theta}_t(\theta_-)}^{\overline{\theta}_t(\theta_-)} \{ C^t(y(\theta), w(\theta) - \beta v(\theta), \theta) - y(\theta)$$

$$+ q \int_{\underline{\theta}_{t+1}(\theta)}^{\overline{\theta}_{t+1}(\theta)} K(v(\theta), \Delta(\theta), \theta', t+1) f^{t+1}(\theta'|\theta) d\theta' \} f^t(\theta|\theta_-) d\theta, \tag{A.8}$$

$$v = \int_{\underline{\theta}_t(\theta_-)}^{\overline{\theta}_t(\theta_-)} w(\theta) f^t(\theta|\theta_-) d\theta,$$

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

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

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

$$\lim_{\theta \to \overline{\theta}^t} \frac{\mu(\overline{\theta}_t)}{\overline{\theta}_t f^t(\overline{\theta}_t|\theta_-)} = -\frac{\gamma}{\theta_-} \frac{\theta_-}{\overline{\theta}_t} \frac{d\overline{\theta}_t}{d\theta_-},$$

$$\lim_{\theta \to \overline{\theta}^t} \frac{\mu(\underline{\theta}_t)}{\underline{\theta}_t f^t(\underline{\theta}_t|\theta_-)} = -\frac{\gamma}{\theta_-} \frac{\theta_-}{\underline{\theta}_t} \frac{d\underline{\theta}_t}{d\theta_-}.$$

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

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

$$= -\alpha \frac{\gamma}{\theta_-} \frac{\beta}{q} \hat{u}^{t'}(c(\overline{\theta}_t)) \frac{\theta_-}{\overline{\theta}_t} \frac{d\overline{\theta}_t}{d\theta_-}$$

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

A similar calculation yields

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

B. A USEFUL LEMMA

Lemma 3. *Suppose that*

$$f^t(\theta|\theta_-) = \frac{1}{\theta\sigma_t\sqrt{2\pi}} e^{-\frac{\left(\log\left(\frac{\theta}{\theta_- e^{\mu_t^{\log}(\theta_-)}}\right)\right)^2}{2\sigma_t^2}}, \tag{A.9}$$

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

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

Proof of Lemma 2. We start with Δ_t. Integrate $\Delta_t = \int w_{t+\tau} g^t(\theta_{t+\tau}|\theta_t) d\theta_{t+\tau}$ by parts, using Lemma 3 to obtain an expression for $g^t(\theta_{t+\tau}|\theta_t)$. Using $\mu_t(\theta_t) = \tau(\hat{\mu}_t(\theta_t) - \frac{1}{2}\theta_t\hat{\sigma}_t^2)$ we obtain

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

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:

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

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 $dv_t = \rho v_t dt - u^t dt + \sigma_{v,t}\hat{\sigma}_t dW_t$ for some $\sigma_{v,t}$. Finally, in the continuous-time limit, the constraint $\dot{w}(\theta) = \tau\theta u_\theta^t + \beta\theta\Delta(\theta)$ 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:

$$dv_t = \rho v_t dt - u^t dt + \theta_t\Delta_t\hat{\sigma}_t dW_t. \qquad \|$$

Proof of Proposition 5. The first-order conditions for c_t, y_t, and $\sigma_{\Delta,t}$ in the HJB equation can be written as

$$\lambda_t = \frac{1}{\hat{u}_t'(c_t)},$$

$$\frac{\tau_{L,t}}{1-\tau_{L,t}} = -\alpha\frac{\gamma_t}{\lambda_t}\frac{1}{\theta_t},$$

$$\sigma_{\Delta,t} = \frac{K_\Delta - K_{v\Delta}\theta_t\Delta_t - \theta_t K_{\Delta\theta}}{K_{\Delta\Delta}}.$$

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

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

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

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

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

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

REFERENCES

ABRAHAM, A. and PAVONI, N. (2008), "Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending: A Recursive Formulation", *Review of Economic Dynamics*, **11**, 781–803.

AIYAGARI, S. R., MARCET, A., SARGENT, T. J. and SEPPALA, J. (2002), "Optimal Taxation without State-Contingent Debt", *Journal of Political Economy*, **110**, 1220–1254.

ALBANESI, S. and SLEET, C. (2006), "Dynamic Optimal Taxation with Private Information", *Review of Economic Studies*, **73**, 1–30.

ALES, L. and MAZIERO, P. (2009), "Accounting for Private Information" (Mimeo).

ATKINSON, A. and STIGLITZ, J. (1976), "The Design of Tax Structure: Direct vs. Indirect Taxation", *Journal of Public Economics*, **6**, 55–75.

BARRO, R. J. (1979), "On the Determination of the Public Debt", *Journal of Political Economy*, **87**, 940–971.

BATTAGLINI, M. (2005), "Long-Term Contracting with Markovian Consumers", *American Economic Review*, **95**, 637–658.

BATTAGLINI, M. and COATE, S. (2008), "Pareto Efficient Income Taxation with Stochastic Abilities", *Journal of Public Economics*, **92**, 844–868.

CONESA, J. C., KITAO, S. and KRUEGER, D. (2009), "Taxing Capital? Not a Bad Idea after All!", *American Economic Review*, **99**, 25–48.

DIAMOND, P. A. (1998), "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, **88**, 83–95.

DIAMOND, P. A. and MIRRLEES, J. A. (1978), "A Model of Social Insurance with Variable Retirement", *Journal of Public Economics*, **10**, 295–336.

EROSA, A. and GERVAIS, M. (2002), "Optimal Taxation in Life-Cycle Economies", *Journal of Economic Theory*, **105**, 338–369.

FARHI, E. and WERNING, I. (2008a), "Optimal Savings Distortions with Recursive Preferences", *Journal of Monetary Economics*, **55**, 21–42.

FARHI, E. and WERNING, I. (2008b), "The Political Economy of Non-Linear Capital Taxation" (Mimeo).

FARHI, E. and WERNING, I. (2009), "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation" (Mimeo).

FERNANDES, A. and PHELAN, C. (2000), "A Recursive Formulation for Repeated Agency with History Dependence", *Journal of Economic Theory*, **91**, 223–247.

FUKUSHIMA, K. (2010), "Quantifying the Welfare Gains from SImple Tax Systems" (Mimeo).

GARRETT, D. and PAVAN, A. (2010), "Dynamic Managerial Compensation: A Mechanism Design Approach" (Mimeo).

GOLOSOV, M., KOCHERLAKOTA, N. and TSYVINSKI, A. (2003), "Optimal Indirect and Capital Taxation", *Review of Economic Studies*, **70**, 569–587.

GOLOSOV, M., TROSHKIN, M. and TSYVINSKI, A. (2010), "Optimal Dynamic Taxes" (Mimeo).

GOLOSOV, M. and TSYVINSKI, A. (2006), "Designing Optimal Disability Insurance: A Case for Asset Testing", *Journal of Political Economy*, **114**, 257–279.

GOLOSOV, M., TSYVINSKI, A. and WERNING, I. (2006), "New Dynamic Public Finance: A User's Guide", *NBER Macroeconomics Annual 2006*, forthcoming.

HEATHCOTE, J., PERRI, F. and VIOLANTE, G. L. (2010), "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006", *Review of Economic Dynamics*, **13**, 15–51.

HEATHCOTE, J., STORESLETTEN, K. and VIOLANTE, G. L. (2005), "Two Views of Inequality Over the Life Cycle", *Journal of the European Economic Association*, **3**, 765–775.

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).

KOCHERLAKOTA, N. (2004), "Figuring out the Impact of Hidden Savings on Optimal Unemployment Insurance", *Review of Economic Dynamics*, **7**, 541–554.

KREMER, M. (2002), "Should Taxes be Independent of Age" (Mimeo).

LUCAS, R. J. and STOKEY, N. L. (1983), "Optimal Fiscal and Monetary Policy in an Economy without Capital", *Journal of Monetary Economics*, **12**, 55–93.

MILGROM, P. and SEGAL, I. (2002), "Envelope Theorems for Arbitrary Choice Sets", *Econometrica*, **70**, 583–601.

MIRRLEES, J. (1971), "An Exploration in the Theory of Optimum Income Taxation", *Review of Economic Studies*, **38**, 175–208.

PAVAN, A., SEGAL, I. and TOIKKA, J. (2009), "Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization, and Infomation Disclosure" (Mimeo).

SAEZ, E. (2001), "Using Elasticities to Derive Optimal Income Tax Rates", *Review of Economic Studies*, **68**, 205–229.

SAEZ, E. (2002), "The Desirability of Commodity Taxation under Non-Linear Income Taxation and Heterogeneous Tastes", *Journal of Public Economics*, **83**, 217–230.

SHIMER, R. and WERNING, I. (2008), "Liquidity and Insurance for the Unemployed", *American Economic Review*, **98**, 1922–1942.

STORESLETTEN, K., TELMER, C. I. and YARON, A. (2004), "Consumption and Risk Sharing over the Life Cycle", *Journal of Monetary Economics*, **51**, 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).

WERNING, I. (2007a), "Optimal Fiscal Policy with Redistribution", *Quarterly Journal of Economics*, **122**, 925–967.

WERNING, I. (2007b), "Pareto Efficient Income Taxation" (Mimeo).

WILLIAMS, N. (2008), "Persistent Private Information" (Mimeo).