UNIVERSITÄT BERN

Faculty of Business, Economics and Social Sciences

Department of Economics

Evaluating pay-as-you-go social security systems

Andreas Bachmann
Kaspar Wüthrich

15-07

July 2015

DISCUSSION PAPERS

Schanzeneckstrasse 1
Postfach 8573
CH-3001 Bern, Switzerland
http://www.vwi.unibe.ch

Evaluating pay-as-you-go social security systems *

Andreas Bachmann † — University of Bern
Kaspar Wüthrich‡ — University of Bern

July 7, 2015

Revised version of Discussion Paper 13-10

Abstract

This paper proposes a method for the welfare analysis of pay-as-you-go social security systems. We derive a formula for the welfare consequences of a permanent marginal change in the payroll tax rate that is valid under weak assumptions about the deep structure of the economy. Our approach requires neither a full specification of preferences and technology, nor knowledge of the individual savings behavior. Instead of parameterizing and calibrating the deep model structure, we implement our formula based on reduced form estimates of a VAR model. We apply our method to evaluate the social security system in the United States.

JEL Classification: E62, H55

Keywords: social security system; overlapping generations; optimal payroll taxes; welfare analysis; reduced form VAR

*We are grateful to Fabrice Collard, Gita Gopinath, Nils Herger, Stefan Leist, Klaus Neusser, Dirk Niepelt, Philipp Wegmüller, anonymous referees, and seminar participants at the Annual Congress of the European Economic Association in Toulouse, the Ski & Labor Workshop in Laax, the Young Swiss Economists Meeting in Bern, the University of Bern, and the Study Center in Gerzensee, for helpful suggestions and comments. The manuscript was written while both authors were working as employed researchers at a public university. The research was not funded by any third party. There are neither financial nor non-financial conflicts.

†Corresponding author. University of Bern, Department of Economics, Schanzeneckstrasse 1, CH-3001 Bern, Switzerland, phone: +41 31 631 4777, email: andreas.bachmann@vwi.unibe.ch.

‡University of Bern, Department of Economics, Schanzeneckstrasse 1, CH-3001 Bern, Switzerland, phone: +41 31 631 3189, email: kaspar.wuethrich@vwi.unibe.ch.

1 Introduction

Unfunded pay-as-you-go (PAYGO) social security systems play an important role in many developed countries' social insurance programs. Since demographic changes and the associated growing fraction of retirees in the population cause increasing financial stress for these systems, the question of how to design social security systems optimally becomes more and more relevant.

Social security systems are typically studied in the context of structural overlapping generations (OLG) models (examples include: Auerbach and Kotlikoff, 1987; Imrohoroglu et al., 1995; Kotlikoff et al., 1999; Krueger and Kubler, 2006; Kotlikoff et al., 2007; Nickel et al., 2008; Fehr et al., 2012; McGrattan and Prescott, 2013; Gahramanov and Tang, 2013, among many others).[1] Welfare analysis in these models proceeds in two steps: first, the deep structure of the model (e.g., preferences and technology) is parameterized and the structural parameters are calibrated or estimated. Second, the effect of different policies and alternative social security systems on social welfare is computed using simulation methods. This approach provides a flexible framework for the welfare analysis of competing social security systems. However, it features two main drawbacks: first, even flexible functional form assumptions might be arbitrary and hard to justify and second, it is typically difficult to identify and estimate all deep parameters in an empirically compelling manner.

This paper contributes to the literature by proposing a complementary method for welfare analysis of social security systems. Based on the Ramsey problem in an OLG model (cf. Diamond, 1965) featuring endogenous labor supply and idiosyncratic longevity, we derive a simple formula for the welfare consequences of permanent changes in payroll taxes used to finance transfers in PAYGO systems. Our formula reveals that changes in the payroll tax affect welfare through three distinct channels: (i) the direct effect of receiving more transfers and paying more taxes, (ii) the general equilibrium effect through changes in factor prices, and (iii) the change in transfers due to the labor adjustment of the subsequent generations. This decomposition is related to the literature on generational accounting, in particular to Fehr et al. (1999) who parameterize and calibrate an OLG model to assess the welfare effect through these different channels.

Our main contribution is to propose an approach to identification and estimation of marginal welfare changes that does not rely on parameterizing and calibrating the deep structure of the model. We exploit the fact that our formula depends on few high level quantities (such as

[1] An important focus of these studies has been on shifts from a primarily unfunded system towards mixed systems that combine PAYGO with investment based personal retirement accounts.

future growth rates and impulse responses of wages and labor with respect to tax changes) and marginal utilities only. This allows us to identify and estimate marginal welfare changes under fewer parametric assumptions. In particular, we do not need to specify the functional form of the aggregate production function nor to fully parameterize household preferences. Moreover, relying on the envelope theorem, our formula for the welfare consequences does not depend on individuals' savings behavior.

We propose two different approaches to implementation of our formula, both of which rely on the reduced form nature of the formula, but differ in their respective treatment of marginal utilities. First, we consider approximate consumption equivalent impacts on each generation. This approach does not require any additional assumptions regarding marginal utilities. However, it is uninformative about the overall welfare change associated with a marginal increase in the payroll tax. We therefore develop a second approach for evaluating the overall effect. This is achieved by obtaining a money metric of the welfare effect through an appropriate standardization. Because the overall effect inherently requires a comparison of weighted marginal utilities of different generations, we need to impose arguably weak assumptions on preferences and on the generation's welfare weights. We show that for both approaches to implementation of our formula, welfare changes can be stated as functions of impulse response functions and predictions of future growth rates only. This allows for an empirical implementation based on the reduced form estimates of a vector autoregressive (VAR) model. Our approach can be extended along various dimensions. In particular, we show that it may incorporate additional taxes that are used to finance the PAYGO system and changes in the retirement age.

In addition to the literature cited earlier, the analysis in this paper is related to studies focusing on globally optimal PAYGO systems (e.g., Feldstein, 1985; Imrohoroglu et al., 1995). However, this paper has a somewhat different focus on local welfare improvements due to small changes in the payroll tax.

Because our formula is a function of high level quantities rather than the deep structure of the model, it can be interpreted as a sufficient statistic (in the sense of Chetty, 2009). The sufficient statistic approach to welfare analysis has recently become important in the public economics literature (see e.g. Chetty, 2009, for a review). It provides a middle course between structural models and reduced-form methods. From the structural approach, it borrows the ability to make predictions about welfare, but avoids the problem of having to estimate or calibrate the deep parameters of the model. From the reduced-form approach, it borrows the advantage of transparent and credible identification. Our analysis differs from the sufficient statistics literature with respect to the structure of the model and the implementation strategy. We implement our central formula based on the reduced form estimates of a VAR because of the dynamic general equilibrium nature of our evaluation problem. This contrasts existing studies which mostly consider static partial equilibrium models and therefore rely on cross sectional estimates as sufficient statistic; for example Card et al. (2007) and Chetty (2008, 2009).

We illustrate our approach by assessing the PAYGO system of the United States. Our results suggest that, in terms of approximate consumption equivalents, a marginal increase in the payroll tax raises welfare of today's retirees and decreases welfare of today's workers and future generations.

The sign of the overall effect depends on the structure of the welfare weights of the Ramsey planner. We first consider "politician's" weights that only reflect the size of the current old and young generations, with zero weights for future generations. We find that the estimated overall welfare effect of a marginal increase in the payroll tax is negative for a broad range of values for the coefficient of relative risk aversion. In contrast, if the welfare weights reflect the size of all generations, aging, and discounting, then the sign of the overall effect is positive except for high values of the Ramsey planner's discount factor, i.e., the weight the planner attaches to future generations. A decomposition by theoretical channels reveals that the direct effect and the factor price effects (i.e., induced changes in wage and interest rates) are important determinants of welfare changes while the effect through the adjustments in labor is negligible. A scenario analysis confirms the robustness of our empirical findings.

The remainder of the paper is structured as follows. Section 2 presents the model and derives a formula for the welfare analysis of a change in the payroll tax. In section 3, we use this formula to assess the welfare consequences of a change in the payroll tax for the United States. Section 4 concludes.

2 Theory

We consider an OLG model with endogenous labor supply and idiosyncratic longevity risk. Our framework is closely related to the setups considered by Breyer and Straub (1993), Nourry (2001), Fanti and Spataro (2006), Gonzalez-Eiras and Niepelt (2008), Lopez-Garcia (2008), and Gonzalez-Eiras and Niepelt (2012). First, we discuss the problems of the household and the representative firm. Second, we analyze the Ramsey problem of the benevolent government. We derive a formula for the welfare consequences of a change in the payroll tax as a function of reduced form quantities only. Third, we discuss extensions of our formula to richer economic environments. Finally, we propose two approaches to empirically implement our formula.

2.1 Demographics, preferences and technology

We consider a perfectly competitive economy inhabited by an infinite sequence of overlapping generations. Each generation lives for two periods. In the first period, households supply labor elastically, $0 \leq n_t \leq \bar{n}$.[2] In the second period, they retire. Population grows at an exogenous rate. Let L_t denote the size of the labor force (i.e., the size of the young generation) in period t and define $\chi_{t,z} \equiv L_z/L_t - 1$ as the working age population growth rate between two periods t and z. We break the tight link between population growth and the ratio of workers to retirees by assuming that households face idiosyncratic longevity (e.g., Gonzalez-Eiras and Niepelt, 2008, 2012): with exogenous probability $p_{t+1} \in (0, 1]$, households born in t survive to become old households in period $t + 1$.[3]

Households have preferences over consumption in both periods and leisure $l_t = \bar{n} - n_t$. Consumption in the first period, c_t^y, equals post tax labor income, $n_t w_t(1-\tau_t)$, where τ_t denotes the payroll tax, minus savings s_{t+1}. In the second period, households consume c_{t+1}^o, which is equal to the gross returns on savings, $s_{t+1}R_{t+1}/p_{t+1}$,[4] plus lump sum social security benefits, T_{t+1}.[5] Preferences are summarized by the utility function $u(c_t^y, \bar{n} - n_t, c_{t+1}^o)$ with $u_{c^y}(\cdot) > 0$, $u_l(\cdot) > 0$, $u_{c^o}(\cdot) > 0$, $u_{c^y c^y}(\cdot) < 0$ and $u_{c^o c^o}(\cdot) < 0$. Note that the utility function $u(\cdot)$ generally depends on the exogenous survival probability p_{t+1}. To ease the notation, we suppress this dependence until section 2.5, where we impose additional structure on the preferences. The intertemporal decision of the household solves:

$$
\begin{aligned}
\max_{n_t, c_t^y, c_{t+1}^o, s_{t+1}} \quad & u(c_t^y, \bar{n} - n_t, c_{t+1}^o) \\
\text{s.t.} \quad & c_t^y + s_{t+1} = n_t w_t (1 - \tau_t), \\
& c_{t+1}^o = \frac{R_{t+1}}{p_{t+1}} s_{t+1} + T_{t+1}, \\
& c_t^y, c_{t+1}^o > 0.
\end{aligned}
$$

[2] $\bar{n}$ denotes the number of available hours in a time period that can be split between leisure and work.

[3] The ratio of workers to retirees in period t is given by $1 + \chi_{t-1,t}/p_t$. Thus, p_t can be inferred from observed working age population growth and the ratio of workers to retirees.

[4] Savings of young households who die before reaching old age are distributed among their surviving peers, leaving them with a gross interest rate of R_{t+1}/p_{t+1}.

[5] In principle, the budget of the retired households also includes profits of the firms. However, they will turn out to be zero in equilibrium due to constant returns to scale. Therefore, we drop them in the households' problem for notational simplicity.

Let λ_t^y and λ_{t+1}^o be the Lagrange multipliers associated with the budget constraint of the household when young and old, respectively. The first order conditions of the household maximization problem read:

$$\lambda_t^y = u_{c^y}(c_t^y, \bar{n} - n_t, c_{t+1}^o), \tag{1}$$

$$\lambda_t^y w_t(1-\tau_t) = u_l(c_t^y, \bar{n} - n_t, c_{t+1}^o), \tag{2}$$

$$\lambda_{t+1}^o = u_{c^o}(c_t^y, \bar{n} - n_t, c_{t+1}^o), \tag{3}$$

$$\lambda_t^y = \frac{R_{t+1}}{p_{t+1}} \lambda_{t+1}^o. \tag{4}$$

The savings decision of a household in cohort t is given by the usual consumption Euler equation:

$$u_{c^y}(c_t^y, \bar{n} - n_t, c_{t+1}^o) = \frac{R_{t+1}}{p_{t+1}} u_{c^o}(c_t^y, \bar{n} - n_t, c_{t+1}^o). \tag{5}$$

The labor supply is described by:

$$u_{c^y}(c_t^y, \bar{n} - n_t, c_{t+1}^o) w_t(1-\tau_t) = u_l(c_t^y, \bar{n} - n_t, c_{t+1}^o). \tag{6}$$

When making their decisions, households consider both current after-tax wages and future interest rates R_{t+1}/p_{t+1}, idiosyncratic longevity p_{t+1}, and benefits T_{t+1}. (5) and (6) combined with the household's budget constraints map $w_t(1-\tau_t)$, p_{t+1}, R_{t+1}, and T_{t+1} into savings and labor supply:

$$s_{t+1} = S(w_t(1-\tau_t), p_{t+1}, R_{t+1}, T_{t+1}), \tag{7}$$

$$n_t = N(w_t(1-\tau_t), p_{t+1}, R_{t+1}, T_{t+1}). \tag{8}$$

The firm sector is characterized by a set of competitive firms that can be represented by an aggregate production function $F(K_t, H_t E_t)$, which maps inputs of capital K_t, exogenous labor efficiency E_t, and hours worked $H_t = L_t n_t$ into output. The problem of the firm is static. In each period, the representative firm solves

$$\max_{K_t, H_t} \quad F(K_t, H_t E_t) - w_t H_t - r_t K_t.$$

The first order conditions of the firm problem imply:

$$w_t = F_{HE}(K_t, H_tE_t)E_t, \tag{9}$$

$$r_t = F_K(K_t, H_tE_t). \tag{10}$$

We impose the following standard assumption on the aggregate production function.

Assumption 1. $F(K_t, H_tE_t)$ *exhibits constant returns to scale.*

Note that Assumption 1 and the Euler theorem imply zero profits in equilibrium.

In addition to conditions from household and firm optimization, the following market clearing and feasibility conditions hold in general equilibrium: $K_t = s_tL_{t-1}$, $H_t = n_tL_t$ and $R_t = 1-\delta+r_t$.

2.2 Ramsey program

We consider the program of a benevolent government that seeks to maximize social welfare, W, subject to technological and competitive equilibrium constraints under commitment – the Ramsey program. In contrast to the standard procedure, we do not solve for the optimal sequence of payroll taxes. Instead, our goal is to derive an empirically implementable expression for the welfare impact of a permanent change in the payroll tax, $dW/d\tau$, that is a function of empirically estimable high level elasticities.[6]

Under Assumption 1, the Ramsey program at $t = 0$ for a given sequence of welfare weights $\{\xi_t\}$ reads:

[6]We consider a permanent change in the payroll tax τ as a more realistic and important case than a one time change in the tax rate, which would be a straightforward alternative to consider. Usually, we do not observe that a government decides on a future path for τ_t, but rather that τ is set to some specific value that is supposed to hold for the future, until important developments make another change in τ necessary. Most observed changes to social security tax rates are permanent in the sense that, at the time of change, the government has no intention to undo the change.

$$\max_{0\le\{\tau_t\}_{t=0}^{\infty}\le 1} W=\sum_{t=0}^{\infty}\xi_t u(c_t^y,\bar{n}-n_t,c_{t+1}^o)+\xi_{-1}u\left(c_{-1}^y,\bar{n}-n_{-1},c_0^o\right)$$

$$\text{subject to}\quad\begin{cases} s_0,\chi_{-1,0},p_0,E_0,c_{-1}^y,n_{-1}\quad\text{given},\\ s_{t+1}=S(w_t(1-\tau_t),p_{t+1},R_{t+1},T_{t+1})\quad t\ge 0,\\ n_t=N(w_t(1-\tau_t),p_{t+1},R_{t+1},T_{t+1})\quad t\ge 0,\\ \text{household budget constraints},\\ F(s_t,n_t(1+\chi_{t-1,t})E_t)=(R_t-1+\delta)s_t+w_tn_t(1+\chi_{t-1,t})\quad t\ge 0,\\ w_t=F_{HE}(s_t,n_t(1+\chi_{t-1,t})E_t)E_t\quad t\ge 0,\\ R_t=1-\delta+F_K(s_t,n_t(1+\chi_{t-1,t})E_t)\quad t\ge 0,\\ T_t=n_tw_t\tau_t\frac{(1+\chi_{t-1,t})}{p_t}.\end{cases}\tag{11}$$

We assume throughout that the sequence of welfare weights $\{\xi_t\}$ is declining sufficiently fast for the problem to be well defined. The last condition in (11) describes the PAYGO character of the social security system in which retiree pensions are equal to the taxes collected divided by the fraction p_t of households that reach retirement age. Our framework does not require the pension system to be balanced in every period. We show in section 2.3 and in appendix B that our analysis can be extended to include additional taxes to refinance the PAYGO system.[7] However, allowing for government debt is challenging because the timing of taxes and transfers that redistribute resources across generations matters in OLG models in which Ricardian equivalence does not hold in general. While it is in principal possible to extend our framework to setups where a known rule governs the evolution of government debt across periods, we consider such an ad hoc approach as unsatisfactory. A general treatment of government debt in our framework is beyond the scope of this paper and left for future research.

Using the envelope conditions of the household maximization problem, the effect of a marginal permanent increase in the payroll tax rate τ on the benevolent government's objective function is given by

$$\frac{dW}{d\tau}=\sum_{t=0}^{\infty}w_tn_t(1+\chi_{t-1,t})u_{c^o}(c_{t-1}^y,\bar{n}-n_{t-1},c_t^o)\xi_{t-1}\frac{1}{p_t}\\ -w_tn_tu_{c^y}(c_t^y,\bar{n}-n_t,c_{t+1}^o)\xi_t+\Psi,\tag{12}$$

[7]The issue of refinancing PAYGO systems using general (non-payroll) taxes is subject to intensive debates among researchers and policy makers, see for example the recent discussion and analysis in Gahramanov and Tang (2013).

where Ψ summarizes the general equilibrium effects of a change in τ. The first term in (12) measures the direct welfare gain of old generations due to an increase in social security transfers. The second term reflects the direct welfare loss of young generations caused by an increase in tax payments. The net social benefit of transferring resources in period t from young to old is proportional to

$$(1+\chi_{t-1,t})u_{c^o}(c^y_{t-1}, \bar{n}-n_{t-1}, c^o_t)\xi_{t-1}\frac{1}{p_t} - u_{c^y}(c^y_t, \bar{n}-n_t, c^o_{t+1})\xi_t.$$

Besides the direct redistributive effects, a change in τ has general equilibrium effects that are captured in Ψ. The policy change causes changes in labor and savings and, thus, in the capital stock, which in turn affects future wages, interest rates and, consequently, social welfare (for the ease of notation, we replace the marginal utilities of the households by the respective multipliers):

$$\begin{aligned}\Psi = \sum_{t=0}^{\infty} \xi_t \Bigg(& \lambda^y_t \frac{dw_t}{d\tau} n_t(1-\tau) \\ & + \frac{\lambda^o_{t+1}}{p_{t+1}} \left(\frac{dR_{t+1}}{d\tau} s_{t+1} + (1+\chi_{t,t+1})\tau \left(\frac{dn_{t+1}}{d\tau} w_{t+1} + \frac{dw_{t+1}}{d\tau} n_{t+1} \right) \right) \Bigg) \\ & + \xi_{-1} \frac{\lambda^o_0}{p_0} \left(\frac{dR_0}{d\tau} s_0 + (1+\chi_{-1,0})\tau \left(\frac{dn_0}{d\tau} w_0 + \frac{dw_0}{d\tau} n_0 \right) \right).\end{aligned}$$

Due to constant returns to scale, there is a direct relation between $dR_t/d\tau$ and $dw_t/d\tau$ that the Ramsey planner takes into account. Totally differentiating the general equilibrium constraint $F(s_t, n_t(1+\chi_{t-1,t})E_t) = (R_t - 1 + \delta)s_t + w_t n_t(1+\chi_{t-1,t})$ with respect to τ implies[8]

$$\frac{dR_t}{d\tau} = -\frac{dw_t}{d\tau}\frac{H_t}{K_t} = -\frac{dw_t}{d\tau}\frac{n_t(1+\chi_{t-1,t})}{s_t}. \tag{13}$$

Using the relationship between $dR_t/d\tau$ and $dw_t/d\tau$ as well as

$$\lambda^y_t = \frac{R_{t+1}}{p_{t+1}}\lambda^o_{t+1},$$

[8] Labor efficiency E_t is exogenous and therefore $dE_t/d\tau = 0$.

we get the following overall welfare effect:

$$
\begin{aligned}
\frac{dW}{d\tau} =& \sum_{t=0}^{\infty} w_t n_t \left(-\lambda^o_{t+1} \frac{R_{t+1}}{p_{t+1}} \xi_t + (1+\chi_{t-1,t}) \lambda^o_t \xi_{t-1} \frac{1}{p_t} \right) \\
&+ \sum_{t=0}^{\infty} \frac{dw_t}{d\tau} n_t (1-\tau) \left(\lambda^o_{t+1} \frac{R_{t+1}}{p_{t+1}} \xi_t - (1+\chi_{t-1,t}) \lambda^o_t \xi_{t-1} \frac{1}{p_t} \right) \\
&+ \sum_{t=0}^{\infty} \frac{dn_t}{d\tau} w_t \tau (1+\chi_{t-1,t}) \lambda^o_t \xi_{t-1} \frac{1}{p_t}.
\end{aligned}
\tag{14}
$$

It is worthwhile discussing the different channels of the overall welfare effect in more detail. There are three basic components, each of which corresponds to a line in equation (14). First, there is the direct effect of receiving more transfers and paying higher taxes. The initial old generation benefits by receiving more transfers without having paid more taxes. All other generations are affected by both higher tax rates when young and higher transfers when old. Second, there are indirect welfare effects owing to the impact of the policy change on factor prices. The effect of factor prices consists of three components: changes in the wage for the young household, changes in the interest rate and changes in the wage that affect the old households through transfers. Third, for each generation, there is an indirect welfare effect due to the labor adjustment of the subsequent generation. Our channels are closely related to the decomposition in Fehr et al. (1999), who apply generational accounting techniques in a setup that is similar to ours. However, in sharp contrast to Fehr et al. (1999), we do not proceed by parameterizing and calibrating the deep structure of the model to carry out welfare analysis.

Grouping the overall welfare gain by generations and expressing λ^o_{t+1} in terms of λ^y_t using the Euler equation, the overall effect (14) rewrites as

$$
\begin{aligned}
\frac{dW}{d\tau} =& \xi_{-1} \lambda^y_{-1} \frac{1+\chi_{-1,0}}{R_0} \left(w_0 n_0 - \frac{dw_0}{d\tau} n_0 (1-\tau) + \frac{dn_0}{d\tau} w_0 \tau \right) \\
&+ \sum_{t=0}^{\infty} \xi_t \lambda^y_t \Bigg(-w_t n_t + \frac{dw_t}{d\tau} n_t (1-\tau) \\
&+ \frac{1+\chi_{t,t+1}}{R_{t+1}} \left[w_{t+1} n_{t+1} - \frac{dw_{t+1}}{d\tau} n_{t+1} (1-\tau) + \frac{dn_{t+1}}{d\tau} w_{t+1} \tau \right] \Bigg).
\end{aligned}
\tag{15}
$$

$1+\chi_{t,t+1}/R_{t+1}$ is often used to evaluate the effectiveness of social security systems, because it relates the return of the social security system, $1+\chi_{t,t+1}$, to the return on the capital market R_{t+1}. A more detailed discussion of the different components of the overall welfare effect and their relation to the concept of dynamic inefficiency is provided in appendix A.

Given the welfare effect in (14) or (15), the standard approach in the literature is to proceed by parameterizing the deep structure of the model, i.e., impose functional form assumptions on preferences and the aggregate production function and express $dW/d\tau$ as a function of a set of primitives that need to be calibrated or estimated from the data. In this paper, we follow an alternative strategy. We make use of the fact that equations (14) and (15) are a function of (i) $\{dw_t/d\tau, dn_t/d\tau\}_{t=0,\ldots,\infty}$, (ii) marginal utilities and (iii) economic quantities such as hours worked and wages. Put differently, in order to identify and estimate $dW/d\tau$, it is sufficient to know quantities (i) to (iii). The key point is that knowledge of the deep parameters that generate these quantities is not required and, thus, a full specification of the deep structure of the model can be avoided.

We argue in section 3 that (i) can be identified and estimated using a reduced form VAR and that empirical predictions can be used to compute (iii). The remaining challenge is to identify the marginal utilities empirically. We propose two alternative approaches. In section 2.4, we show that it is possible to compute first order approximations for the consumption equivalent impact of a marginal change in the payroll tax for each generation without further assumptions on preferences. Identification and estimation of the overall effect $dW/d\tau$ requires more assumptions because comparisons of marginal utilities between different generations are involved. In particular, we need to partly parameterize household preferences as discussed in section 2.5.

2.3 Extensions

In the empirical part of this paper, we focus on implementing the formula for the overall welfare effect given in equation (14). This formula is based on a simple model framework. However, our framework can be extended along various dimensions. We consider two such extensions in the appendix. In appendix B, we show how our analysis can be augmented to include additional taxes (e.g., consumption and capital taxes) that are used to refinance the social security system. Specifically, we consider a government that levies a consumption tax, τ^c, in addition to the payroll tax, τ^w. Social security transfers are given by

$$T_t = n_t w_t \tau_t^w (1 + \chi_{t-1,t}) + c_t^y \tau_t^c (1 + \chi_{t-1,t}) + c_t^o \tau_t^c,$$

where we set $p_t = 1$ to alleviate the exposition. We show that one can use similar arguments as in the previous section to derive formulas for the overall welfare effects dW/d_{τ^c} and dW/d_{τ^w}.

In Appendix C, we show how to incorporate changes in the retirement age into the analysis. Such changes can be introduced in the model by assuming that old households work for a fraction ϱ of a model period before they retire (e.g., Gonzalez-Eiras and Niepelt, 2012). We show that $dW/d\tau$ is given by a similar expression as (14). The difference is that labor supply decisions of the old household need to be taken into account, while the direct and general equilibrium effects on the working old cancel with the corresponding increase in transfers.

Moreover, instead of a permanent change, we could also look at a one time change in the tax rate τ. In this case, the analysis could proceed by extending the theoretical results by Gonzalez-Eiras and Niepelt (2007, appendix A). As argued before, we consider a permanent change to be the more realistic and interesting scenario and we do not pursue the one time change further.

2.4 Consumption equivalent impact on each generation

In this section, we are interested in estimating the consumption equivalent effect of a marginal increase in τ. Suppose that there is a hypothetical increase in c^o by ϕ percent. Utility of the generation born in t would then be given by $u(c_t^y, \bar{n} - n_t, (1+\phi_t)c_{t+1}^o)$. A first order Taylor approximation around $\phi_t = 0$ yields

$$u(c_t^y, \bar{n} - n_t, (1+\phi_t)c_{t+1}^o) \approx u(c_t^y, \bar{n} - n_t, c_{t+1}^o) + u_{c^o}(c_t^y, \bar{n} - n_t, c_{t+1}^o)c_{t+1}^o\phi_t.$$

The change in utility can therefore be approximated by

$$u(c_t^y, \bar{n} - n_t, (1+\phi_t)c_{t+1}^o) - u(c_t^y, \bar{n} - n_t, c_{t+1}^o) \approx \lambda_{t+1}^o c_{t+1}^o \phi_t.$$

It follows from equations (14) and (15) that the change in a generation's utility due to the change in τ is linear in this generation's marginal utility. In particular, the change in utility can be written as $\lambda_{t+1}^o \Omega_t$, where Ω_t is defined in Proposition 1. For each generation, we can approximatively calculate the (hypothetical) percentage change in consumption when retired which would make this generation equally well off as the policy change:

$$\lambda_{t+1}^o c_{t+1}^o \phi_t \approx \lambda_{t+1}^o \Omega_t. \tag{16}$$

Since $\lambda_{t+1}^o > 0$, it follows that

$$\phi_t \approx \frac{\Omega_t}{c_{t+1}^o}. \tag{17}$$

This yields the approximative welfare effect of the policy change in terms of consumption for each generation.

Proposition 1. *Consider an OLG economy as described in section 2.1 with a PAYGO social security system. Suppose that Assumption 1 is satisfied. Then, the impact of a permanent marginal change in* τ *on the welfare of the generation born in* t *is equivalent (up to a first-order approximation) to an increase in this generation's consumption when retired by* ϕ_t*, where* $\phi_t = \frac{\Omega_t}{c^o_{t+1}}$ *and*

$$\Omega_{-1} = (1+\chi_{-1,0})\frac{1}{p_0}\left(w_0 n_0 - \frac{dw_0}{d\tau}n_0(1-\tau) + \frac{dn_0}{d\tau}w_0\tau\right), \tag{18}$$

$$\begin{aligned}\Omega_t = \frac{1}{p_{t+1}}\Bigg(&- w_t n_t R_{t+1} + w_{t+1}n_{t+1}(1+\chi_{t,t+1}) + R_{t+1}\frac{dw_t}{d\tau}n_t(1-\tau) \\ &- \frac{dw_{t+1}}{d\tau}n_{t+1}(1-\tau)(1+\chi_{t,t+1}) + \frac{dn_{t+1}}{d\tau}w_{t+1}\tau(1+\chi_{t,t+1})\Bigg),\end{aligned} \tag{19}$$

for $t = 0, 1, 2, \ldots$.

Observe that equations (18) and (19) contain impulse response functions with respect to a permanent change in the payroll tax rate, $\{dw_t/d\tau, dn_t/d\tau\}_{j=0,\ldots,\infty}$, predictions of economic quantities, projected working age population growth, and the projected ratio of workers to retirees. This allows for an implementation based on a reduced form VAR model. We provide a more detailed discussion of the implementation in section 3.

2.5 Overall welfare effect

Proposition 1 allows for an approximate welfare analysis by generation in terms of consumption equivalents. However, for a thorough policy evaluation, this information is not sufficient because the Ramsey planner cares about a weighted sum of all future generations' utilities. Thus, knowledge of the overall effect of a change in the payroll tax is essential. Because utility is not assumed to be quasi-linear, we need to convert $dW/d\tau$ into a money metric (Chetty, 2009). We obtain an intuitive metric by normalizing the welfare change given an increase in the payroll tax rate by the welfare gain from a hypothetical additional unit of income, a_0, of the initial old household ($dW/da_0 = \xi_{-1}\lambda_0^o$):

$$\frac{\frac{dW}{d\tau}}{\frac{dW}{da_0}} = \sum_{t=0}^{\infty} w_t n_t \left(-\frac{\lambda_{t+1}^o \xi_t}{\lambda_0^o \xi_{-1}} \frac{R_{t+1}}{p_{t+1}} + (1+\chi_{t-1,t}) \frac{\lambda_t^o \xi_{t-1}}{\lambda_0^o \xi_{-1}} \frac{1}{p_t} \right)$$
$$+ \sum_{t=0}^{\infty} \frac{dw_t}{d\tau} n_t (1-\tau) \left(\frac{\lambda_{t+1}^o \xi_t}{\lambda_0^o \xi_{-1}} \frac{R_{t+1}}{p_{t+1}} - (1+\chi_{t-1,t}) \frac{\lambda_t^o \xi_{t-1}}{\lambda_0^o \xi_{-1}} \frac{1}{p_t} \right) \quad (20)$$
$$+ \sum_{t=0}^{\infty} \frac{dn_t}{d\tau} w_t \tau (1+\chi_{t-1,t}) \frac{\lambda_t^o \xi_{t-1}}{\lambda_0^o \xi_{-1}} \frac{1}{p_t}.$$

An assessment of the overall effect $dW/d\tau$ requires aggregating generation-specific welfare effects. This aggregation inherently includes a comparison of weighted marginal utilities between different generations and, therefore, requires some additional structure on marginal utilities, λ_{t+1}^o, and welfare weights, ξ_t. We impose the following assumption on the household utility function.

Assumption 2. *Household preferences are additively separable over time and flow utility of the old households is of CRRA type, i.e.,*

$$u(c_t^y, \bar{n} - n_t, c_{t+1}^o) = u(c_t^y, \bar{n} - n_t) + \beta p_{t+1} \frac{\left(c_{t+1}^o\right)^{1-\gamma} - 1}{1-\gamma}, \quad (21)$$

where β is the individual discount factor and γ is the coefficient of relative risk aversion.

Assumption 2 implies that the ratio of marginal utilities is a function of quantities that can be predicted empirically, namely gross consumption growth and p, which depends on working age population growth and the ratio of workers to retirees:

$$\frac{u_c(c_{t+1}^o)}{u_c(c_0^o)} = \frac{p_{t+1}}{p_0} \left(\frac{c_{t+1}^o}{c_0^o} \right)^{-\gamma}.$$

As a final step towards implementing equation (20), we need to impose some structure on the welfare weights, ξ_t. In principle, any sequence of welfare weights can be chosen, provided that the sequence is declining sufficiently fast for the problem to be well defined. Following the literature (e.g., Gonzalez-Eiras and Niepelt, 2008) we make the following assumption.

Assumption 3. *The social planner's welfare weights for different generations reflect the size of the generations, aging, and discounting (with discount factor $\kappa < 1$):*

$$\xi_j = \kappa(1+\chi_{j-1,j}) \frac{1+\kappa p_{j+1}}{1+\kappa p_j} \xi_{j-1},$$

with $\xi_{-1} = 1$.

The generation born in t is of size L_t when young and of size $L_t p_{t+1}$ when old. With discounting, this generation's weight is proportional to $L_t(1 + \kappa p_{t+1})$. These considerations imply relative weights across generations as described in Assumption 3. In appendix D, we consider a different set of weights that allows us to estimate the welfare cost solely caused by increasing the distortive tax and excluding redistribution effects.

We are now in the position to summarize the empirically implementable formula for the welfare consequences of a change in the payroll tax in the following proposition.

Proposition 2. *Consider an OLG economy as described in section 2.1 with a PAYGO social security system. Suppose that Assumptions 1, 2 and 3 hold. Then, the overall welfare gain from a permanent increase in the payroll tax* τ *relative to a \$1.00 increase in the income of the initial old household is given by*

$$\begin{aligned}\frac{\frac{dW}{d\tau}}{\frac{dW}{da_0}} =& \sum_{t=0}^{\infty} w_t n_t \kappa^t (1+\chi_{-1,t}) \frac{1}{p_0} \left(\left(\frac{c_t^o}{c_0^o} \right)^{-\gamma} \frac{1+\kappa p_t}{1+\kappa p_0} - \kappa \left(\frac{c_{t+1}^o}{c_0^o} \right)^{-\gamma} \frac{1+\kappa p_{t+1}}{1+\kappa p_0} R_{t+1} \right) \\ &+ \sum_{t=0}^{\infty} \frac{dw_t}{d\tau} n_t (1-\tau) \kappa^t (1+\chi_{-1,t}) \frac{1}{p_0} \left(\kappa \left(\frac{c_{t+1}^o}{c_0^o} \right)^{-\gamma} \frac{1+\kappa p_{t+1}}{1+\kappa p_0} R_{t+1} \right. \\ &\qquad\qquad \left. - \left(\frac{c_t^o}{c_0^o} \right)^{-\gamma} \frac{1+\kappa p_t}{1+\kappa p_0} \right) \\ &+ \sum_{t=0}^{\infty} \frac{dn_t}{d\tau} w_t \tau \kappa^t (1+\chi_{-1,t}) \left(\frac{c_t^o}{c_0^o} \right)^{-\gamma} \frac{1+\kappa p_t}{1+\kappa p_0} \frac{1}{p_0}. \end{aligned} \tag{22}$$

Apart from γ and κ, the quantities in equation (22) are either observable or can be estimated. As for the approximate consumption equivalent formulas in Proposition 1, equation (22) contains impulse response functions, predictions of economic quantities, projected population growth, and the projected ratio of workers to retirees. In addition, consumption growth must be projected to infer the ratios of marginal utilities. Given these similarities, implementation can be based on the same reduced form VAR estimates. We refer to section 3 for more details.

3 Empirical implementation

We illustrate our method for computing the welfare consequences of a change in the payroll tax by analyzing the social security system in the United States.[9] First, the empirical implementation of Propositions 1 and 2 is discussed. Second, we describe the data and the aggregation to the frequency of the OLG model. Finally, we present the results and some robustness checks.

[9]See, e.g., Feldstein (2005) for a description of the social security system in the United States.

3.1 Methodology

Proposition 1 provides formulas for the percentage change in consumption when retired which would make a generation equally well off as the change in the payroll tax rate. Proposition 2 provides a formula for the overall welfare effect of a change in the payroll tax rate. The quantities needed to empirically estimate the equations in these propositions can be divided into four groups.

Payroll tax. In the United States, payroll taxes for social security can be split into the designated purposes of Old-Age and Survivors Insurance (OASI), Disability Insurance (DI) and Hospital Insurance (HI). Given the setup of our model, only payroll taxes used to fund the PAYGO system are considered. Thus, we take τ to be the payroll tax for OASI. We evaluate equations (18), (19) and (22) at the current level of $\tau = 0.106$.

Impulse responses of hours worked and factor prices. The welfare effect depends on the dynamic responses of labor and wages with respect to a change in the payroll tax rate. A natural way to estimate these dynamic responses consists in estimating impulse response functions based on a reduced form VAR model:

$$Y_t = c + \Phi_1 Y_{t-1} + \ldots + \Phi_p Y_{t-p} + \theta_0 \Delta\tau_t + \ldots + \theta_q \Delta\tau_{t-q} + \sum_i \alpha_i D_{i,t} + \varepsilon_t. \quad (23)$$

Y_t is a vector of endogenous variables. Φ_j, c, θ_j and α_i are matrices and vectors of coefficients. ε_t is a vector of error terms which is multivariate white noise. We estimate a four-dimensional VAR model with $p = q = 3$ using quarterly data on growth rates of hours worked per capita, real wages, real consumption per capita and real GDP per capita. These growth rates are stationary. We include a set of dummy variables $D_{i,t}$ to control for the potential impact of important changes in the US economy or the social security system.[10]

The VAR order of 3 is chosen based on residual autocorrelation tests and the Akaike information criterion. We include the change in the payroll tax rate as an exogenous variable. Our identifying assumption is that $\mathbb{E}[\varepsilon_t \Delta\tau_t | Y_{t-1}, \ldots, Y_{t-p}, \Delta\tau_{t-1}, \ldots, \Delta\tau_{t-q}, D_{i,t}] = 0$, i.e., $\Delta\tau_t$ is exogenous conditional on lagged economic quantities. Clearly, the payroll tax rate depends on factors such as life expectancy, retirement age and the shape of the age pyramid. However,

[10]The dummies are equal to zero before a certain event and equal to one thenceforward. The following economic, political and social security related events are considered: Cost-of-living allowance (1972q2), expansion of the social security program (1972q4), oil crisis (1973q4), social security amendments (1977q4), oil crisis (1979q1), disability amendments (1980q2), social security and medicare amendments (1983q2), Great Moderation (1984q1), Social Security Disability Benefits Reform Act (1984q4), Dissolution of the Soviet Union (1991q4), Contract With America Advancement Act (1996q1), Ticket to Work and Work Incentives Improvement Act (1999q4), burst of dot-com bubble (2000q1), Senior Citizens' Freedom to Work Act (2000q2), financial crisis (2007q3).

the factors influencing $\Delta\tau_t$ are unlikely to have a direct impact on Y_t conditional on the lagged values of Y. Even if the payroll tax rate is adjusted to macroeconomic conditions, our identifying assumption is unlikely to be violated. Because we have quarterly data and $\Delta\tau_t$ is determined by legislation, which takes some time to adjust, $\Delta\tau_t$ is likely to respond with a lag. Thus, we are confident that $\Delta\tau_t$ is mean independent of ε_t conditional on lagged values of Y. To provide some evidence for our identifying assumption, we perform a Granger causality test by including $\Delta\tau_t$ as an endogenous variable in the VAR model. We find that $\Delta\tau_t$ is exogenous in the sense that it cannot be predicted using lagged values of Y.

A permanent change in τ_t corresponds to a one time change in $\Delta\tau_t$. The impulse response to such a payroll tax rate change is given by

$$\frac{dY_{t+j}}{d\Delta\tau_t} = \begin{cases} \Phi_1 \frac{dY_{t-1+j}}{d\Delta\tau_t} + \ldots + \Phi_p \frac{dY_{t-p+j}}{d\Delta\tau_t} + \theta_j & \text{for } j = 0, \ldots, q \\ \Phi_1 \frac{dY_{t-1+j}}{d\Delta\tau_t} + \ldots + \Phi_p \frac{dY_{t-p+j}}{d\Delta\tau_t} & \text{for } j = q+1, q+2, \ldots \end{cases} \tag{24}$$

where $\frac{dY_k}{d\Delta\tau_t} = 0$ for $k < t$.

Forecasts. The empirical implementation of equations (18), (19) and (22) requires forecasts for present and future generations on real wages, real interest rates, labor, working age population growth, the ratio of workers to retirees, and consumption of retirees. Because long-term forecasting is a very delicate issue, we conduct a scenario analysis to account for the uncertainty of future developments (see section 3.3.4).[11] The baseline scenario is constructed as follows. Real wages, hours worked and real consumption of retirees are projected using the (geometric) mean growth rate of the observed data. The projected real interest rate is set to the mean of the observed data. For the working age population, we base the projections on the (geometric) mean growth rate according to the national population projections released by the U.S. Census Bureau.[12] Regarding the ratio of workers to retirees, we also rely on the national population projections, which are available up to 2060, and we assume the ratio to remain constant afterwards.

Parameters. We avoid identification and estimation of the parameters κ and γ. Instead, we

[11] At first sight, the dependency on forecasts seems to be a limitation pertaining specifically to our proposed method. However, a similar difficulty also exists for specific parameterized and calibrated models. A permanent change in τ_t leads to a new steady state. Once this steady state is reached, the economy grows along the balanced growth path. Thus, the welfare evaluation of a payroll tax change in a parameterized and calibrated model hinges on technology and population growth for current and future generations, which is similarly hard to assess as to forecast the quantities needed to implement equations (18), (19) and (22).

[12] http://www.census.gov/population/projections/data/national/2012/downloadablefiles.html, last accessed on May 6, 2014.

estimate equation (22) for various plausible parameter values.

3.2 Data and aggregation

This section provides a brief description of the data used to implement our formula. More details are relegated to appendix E. To estimate our VAR, we use growth rates on hours worked per capita, real wages, real consumption per capita, and real GDP per capita. The series on hours worked per capita is constructed from an index of hours worked in the business sector divided by the working population. In accordance to hours worked, we compute real wages from real hourly compensation in the business sector. We measure the consumption growth rate by the growth rate of real private final consumption expenditure divided by the total population size. The data on real GDP come from the U.S. Bureau of Economic Analysis. To implement our formula, we also need data on consumption of the old generation and population quantities (working age population growth and the ratio of workers to retirees). The former series is constructed from yearly data on total average expenditures of people over age 65. Data and projections on the latter quantities are provided by the Bureau of Labor Statistics and the U.S. Census Bureau.

We cannot directly plug these data into equations (18), (19) and (22) in order to evaluate the PAYGO system. Particular attention has to be paid to the aggregation of the data in order to fit the framework of the OLG model. In the OLG model, two time periods correspond to an entire lifespan. Following the literature (e.g. Gonzalez-Eiras and Niepelt, 2008; Song, 2011), we take one period in the model to be 30 years. Period 0 in equations (18), (19) and (22) corresponds to the years 2013 to 2042, period 1 to the years 2043 to 2072, etc. We carefully explain our strategy on how to aggregate the data to fit the theoretical model in appendix E.

3.3 Results

3.3.1 VAR

Figure 1 plots the responses of hours worked and real wages to a permanent increase in the payroll tax rate by one percentage point. After some quarters of adjustment, both hours worked and real wages stabilize at a lower level.

The responses are in line with theoretical predictions from a standard OLG model. A permanent increase in taxes reduces the income of young households while the associated rise in transfers increases the income of old households. Thus, savings and, hence, the capital stock decrease, which has a negative impact on real wages. Given the decrease in real pre-tax wages and the even more pronounced drop in after tax wages, the substitution effect calls for a reduction in labor. If the substitution effect outweighs the income effect, hours worked drop.

In order to implement equations (18), (19) and (22), we need the impulse responses $dw_t/d\tau$ and $dn_t/d\tau$ at the model frequency of 30 years. Using our estimated quarterly impulse responses, we first compute quarterly level responses of wages and hours by subtracting the projected path for these variables absent any shock from the projected path given the increase in the payroll tax rate. Consistently with data aggregation (cf. table 6 in appendix E), the quarterly level responses are then aggregated to match the OLG model frequency. Figure 2 shows the resulting level responses at model frequency. Hours worked decline by a small amount. The response amounts to minus 180 hours for the initial generation, which means a reduction of labor by 0.5 hours per month. The real hourly wage decreases by 19 cents for the initial generation. After some adjustment period, the level response declines in absolute terms (for hours worked) or grows (for wages) in accordance with the long-run growth rates of these variables.

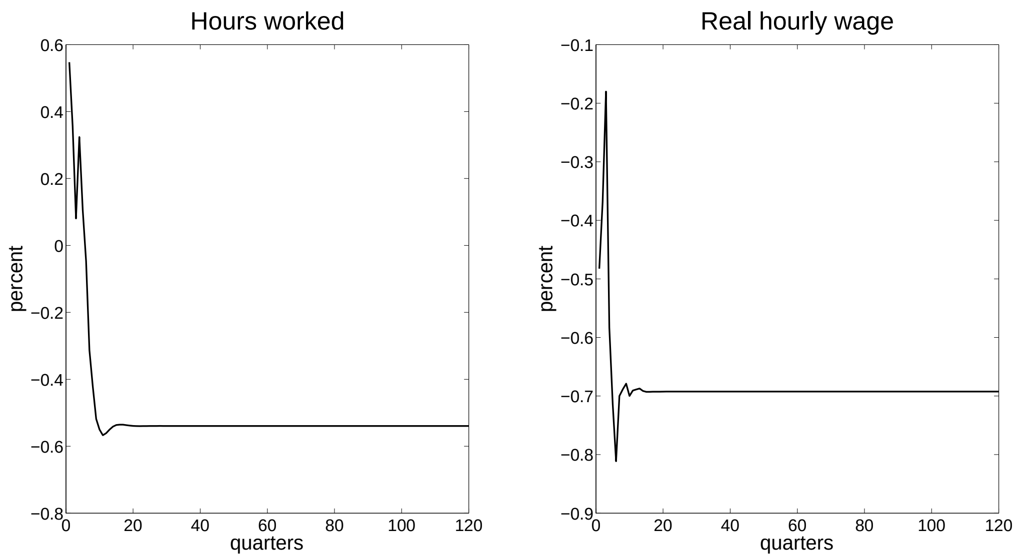

Figure 1: Impulse response of hours worked and real wages (in percent) to a permanent increase in the payroll tax rate by one percentage point.

3.3.2 Consumption equivalent impact

The impulse response functions estimated above and predictions of economic quantities allow for calculating the impact of the payroll tax change on each generation. Proposition 1 shows how the change in utility of each generation can be approximately expressed in terms of a percentage change in consumption during retirement. The results are summarized in Table 1.

Several results deserve closer attention. First, the largest part of the impact (in absolute

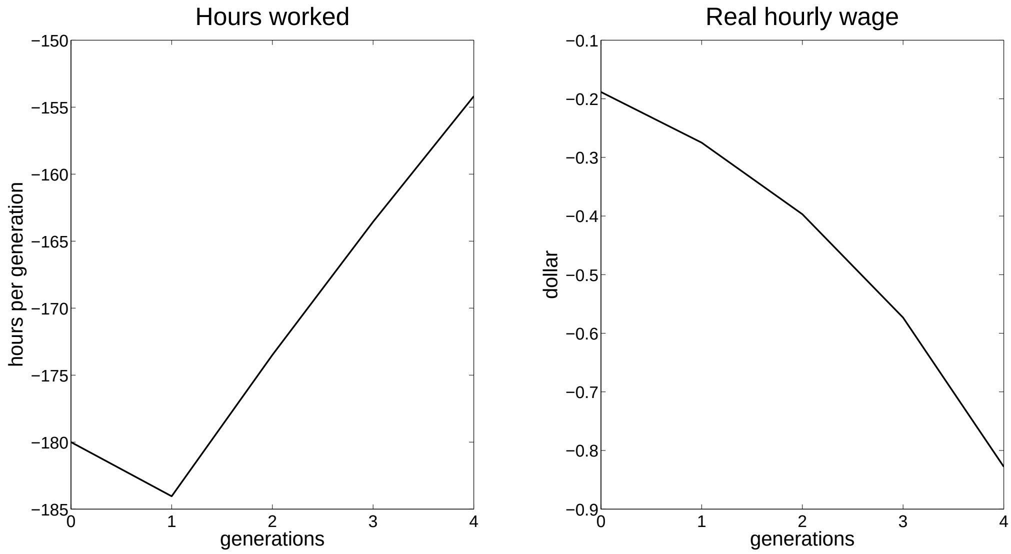

Figure 2: Impulse response of hours worked and real wages to a permanent increase in the payroll tax rate by one percentage point.

values) is due to the direct effect. The change in the payroll tax rate has a direct consumption equivalent effect of 0.95% for the initial old generation. The corresponding direct effect for the initial young and the future generations is negative. Thus, receiving more transfers when retired does not fully compensate for having to pay more taxes when working. This result matches the data which indicate that the US economy is dynamically efficient. Second, the factor price effect is substantial. For all generations, this effect amounts to about 60% of the direct effect's size. Note that the factor price effect is of the same sign and, hence, amplifies the direct effect. The factor price effect can be further decomposed into the effect due to changes of wages when young, wages when old and interest rates. This decomposition reveals that changes in the interest rate and changes in the wage level during the working age are important. In contrast, changes in the wage level when retired, which affect the level of transfers paid to retirees, are of minor importance. Third, the payroll tax change has a negative impact on labor and therefore on transfers, but this labor effect seems to be of minor quantitative relevance.

The findings in Table 1 have important implications for the welfare evaluation of PAYGO systems. The results indicate a distributional conflict across generations. We find that the initial old generation benefits, while all other generations are worse off. As the factor price effects go in the same direction as the direct effects, the distributional conflict due to differences in direct benefits and costs across generations is amplified. The retirees get more transfers and they also benefit from higher interest rates on their savings. The other generations are worse off because receiving more transfers does not fully compensate for paying more taxes, and they

Table 1: Percentage retiree's consumption change with equivalent welfare effect as the policy change.

	generation 0	generation 1	generation 2	generation 3
direct effect	0.95	-0.34	-0.36	-0.39
factor price effect	0.58	-0.20	-0.22	-0.24
labor effect	-0.05	-0.05	-0.06	-0.06
total effect	1.47	-0.59	-0.65	-0.70

The numbers represent the *average* percentage change in consumption during retirement *over a generation*, of which a fraction $(1-p)$ does not survive to become old households while the remaining fraction p gets $1/p$ times the numbers in the table.

also suffer from overall unfavorable factor price changes. In addition to the direct and the factor price effects, there is a negative labor effect for all generations. This can be interpreted as a consumption equivalent cost of increasing a distorting tax. Overall, we conclude that it is not sufficient to consider only the direct redistribution effects of a policy change in the PAYGO system because there are indirect effects of the tax change which seem to have substantial welfare impacts.

3.3.3 Overall welfare effect

The previous section presents an analysis of the approximate welfare impact for each generation. Since one generation is better and others are worse off, it is important to have a measure which aggregates the utility changes of all generations in order to evaluate the policy change. Proposition 2 provides such a measure. In this section, we use equation (22) to evaluate the PAYGO social security system of the United States.

In principle, we can calculate the overall welfare impact for any given sequence of welfare weights $\{\xi_t\}$. The first column in Table 2 shows the results for a "politician's welfare weights", i.e., welfare weights reflecting the size of the current old and young generation, with zero weight for all generations not yet born. All other columns in Table 2 show the results if we put the structure provided in Assumption 3 on $\{\xi_t\}$. The results depend on the unobservable parameters γ and κ. As there is some disagreement on γ in the literature, we estimate equation (22) for a variety of values covering the range of parameter values commonly used in the literature. The decision on the social planner's discount factor across generations, κ, is inherently and necessarily the researchers' choice. To avoid an arbitrary choice, we assume that the planner's discount factor is similar to the one of individual households. In particular, we analyze the welfare change for values $\gamma \in [0,2]$ and $\kappa \in [0.9^{30}, 0.985^{30}]$, where the latter interval stems from assuming $\kappa = \beta$ with an individual yearly discount factor β of at least 0.9 and at most 0.985.[13]

Table 2 shows the result for a selection of parameter values. We emphasize that each entry in the table corresponds to a different problem from the planner's point of view. The planner knows the utility function of the households including a specific value of γ. The parameter γ matters for the overall welfare evaluation because it affects marginal utility. As we project consumption to grow, future generations will have lower marginal utility. How much marginal utility shrinks across generations depends on γ. If γ is large, the consumption loss of future generations is of low relative importance for overall welfare compared to the consumption gain of the initial old generation. However, as the numbers in Table 2 indicate, exact knowledge of the households' coefficient of relative risk aversion γ seems to be not that important for the overall welfare effect, except for large values of κ. The larger κ, the more relative weight is given to future generations. As Table 1 shows, the current generation of retirees benefits from the policy change whereas the other generations are negatively affected. Thus, the overall welfare effect crucially depends on κ. The overall effect is positive for low values of κ, in which case γ is of minor relevance. For large values of κ, the overall effect becomes negative provided that γ is not too large.

Table 2: Overall welfare effect (in 10000) according to (22) depending on the households' parameter γ and the planner's weights $\{\xi_t\}_{t=0}^{\infty}$.

	"Politician's $\{\xi_t\}$"	$\kappa = 0.9^{30}$	$\kappa = 0.95^{30}$	$\kappa = 0.985^{30}$
$\gamma = 0.0$	-3.5	5.0	4.2	-80.6
$\gamma = 0.5$	-2.6	5.1	4.3	-9.3
$\gamma = 1.0$	-1.7	5.1	4.4	-2.3
$\gamma = 1.5$	-1.0	5.1	4.5	0.3
$\gamma = 2.0$	-0.3	5.1	4.6	1.7

The overall welfare change includes both redistribution and efficiency effects. In appendix D, we estimate the magnitude of the efficiency effect based on a hypothetical government authority in the spirit of the *lump-sum redistribution authority* (LSRA) in Auerbach and Kotlikoff (1987).

To understand the driving factors of the numbers in Table 2, it is instructive to decompose the values by channels and generations. In section 2, we have shown that the welfare consequence of a change in the payroll tax is the sum of three different components: the direct change in taxes and transfers, the welfare impact via changes in the wage and interest rate, and the

[13]For the planner's problem to be well defined, we need the sequence of welfare weights $\{\xi_t\}$ to decline sufficiently fast. For this reason, κ cannot be too close to one. We verify that the expression in (22) indeed converges for our upper boundary of κ in our empirical application.

welfare impact via changes in labor. As a benchmark case, Table 3 shows the results of this decomposition for $\kappa = 0.95^{30}$ and $\gamma = 1$. Of course, the sign of the numbers in Table 3 is identical with the sign of the numbers in Table 1. However, in contrast to Table 1, future generations are now downweighted because $\kappa < 1$ and because of the projection $\frac{c_t^o}{c_0^o} > 1$ for $t > 0$, which implies lower marginal utility for future generations.

The sum of the labor effects across generations is negative for each considered combination of parameter values for γ and κ. The sum of the factor price effects is positive except for $\kappa = 0.985^{30}$ and $\gamma < 1.5$. The same holds for the sum of the direct welfare effects. Thus, the sign of the overall welfare effect depends crucially on the direct and on the factor price effect, which may be positive or negative depending on the parameter values.

Table 3: Decomposition of the overall welfare effect (in 1000) by generations and channels, based on parameter values $\gamma = 1$ and $\kappa = 0.95^{30}$.

	gen. 0	gen. 1	gen. 2	gen. 3	total
direct effect	33.3	-3.1	-0.8	-0.2	29.1
factor price effect	20.3	-1.9	-0.5	-0.1	17.8
labor effect	-1.8	-0.5	-0.1	-0.0	-2.4
total	51.9	-5.5	-1.5	-0.4	44.5

It is important to recall that our results are only locally valid. Equation (22) captures the marginal effect of a change in the payroll tax as a function of macroeconomic variables and, in particular, of τ. Thus, the results presented in this section provide a welfare evaluation of the current PAYGO system in the United States (with $\tau = 0.106$). This is especially important when comparing our result to the findings in the literature based on structural models, which typically do not analyze marginal changes. For example, Auerbach and Kotlikoff (1987) analyze the welfare effects of the introduction of an unfunded social security system with 60 percent benefit-to-earnings replacement rate under different tax regimes. Similar to our results, they find gains for the older generations and losses for the younger and future generations. Using the same model, Fehr et al. (1999) study a 25% increase in social security benefits starting with a PAYGO system with a 40% benefit replacement rate and decompose the overall effect into different channels. Despite the differences in the underlying policy experiment and the modeling framework, their decomposition shows a qualitatively similar pattern as reported in tables 1 and 3. Using an applied general equilibrium model, Imrohoroglu et al. (1995) find the optimal replacement rate of an unfunded social security system to be 30% (as opposed the empirically more realistic rate of 60%). Moreover, their results indicate that even with an empirically realistic replacement rate of 60% a social security system can be welfare enhancing. Kotlikoff et al. (2007) consider different alternative policies to mitigate the problems of the demographic transition in the United States. One such policy consists of a 50% benefit reduction which helps to limit the (endogenous) growth in the payroll tax. Their simulations show welfare losses for the older and the present generations and welfare gains for the future generations. Because in our model a benefit reduction is directly linked to a tax cut through the government budget constraint, these results are qualitatively comparable to our welfare projections. At this point it is noteworthy that instead of focusing on the payroll tax rate as the policy instrument, our analysis could alternatively be based on the benefit rates.

3.3.4 Robustness checks

Knowledge of the future development of real wages, hours worked, real interest rates, consumption of retirees, working age population growth, and the ratio of workers to retirees is crucial for the evaluation of PAYGO systems based on equations (18), (19) and (22). As a baseline scenario, we use means or mean growth rates of the available data sample for forecasting.[14] In the light of structural breaks, this might be an inappropriate forecast to use given that we need projections for several future generations. Therefore, this section conducts a sensitivity analysis on the dependence of our findings on the projected paths for the relevant variables. Using scenarios to account for the uncertainty of future developments is quite common in the literature on social security systems (e.g., Pecchenino and Utendorf, 1999; Kotlikoff et al., 2007; Imrohoroglu and Kitao, 2009; McGrattan and Prescott, 2013).

In addition to the baseline scenario, we consider a higher and a lower future development for each variable of interest. Figure 3 depicts the alternative paths. We cover a wide range of possible developments. For time series with a positive (negative) trend in levels,[15] the high (low) scenario is constructed assuming a 50% higher growth rate compared to the baseline. The low (high) scenario consists in eliminating the trend and assuming the series to be constant. For the working age population, the U.S. Census Bureau provides a low, middle and high projection based on alternative net international migration series. We use the (geometric) mean growth rate according to the low and high projection in order to construct our low and high scenario for

[14] There are two exceptions: the working age population growth rate and the ratio of workers to retirees, for which the forecast is based on the national population projections from the U.S. Census Bureau.

[15] Real wages and real consumption per capita of retirees show a positive trend, hours worked per capita and the ratio of workers to retirees show a negative trend.

the working age population growth rate.[16] Finally, since the real interest rate has no trend, the high (low) scenario is constructed by adding (subtracting) 0.1 percentage points to the (yearly) baseline real interest rate per generation.[17]

For each scenario, we compute the consumption equivalent impact (equations (18) and (19)) and the overall welfare effect (equation (22)) for various parameter values κ and γ (i.e., we recompute the Tables 1 and 2).[18]

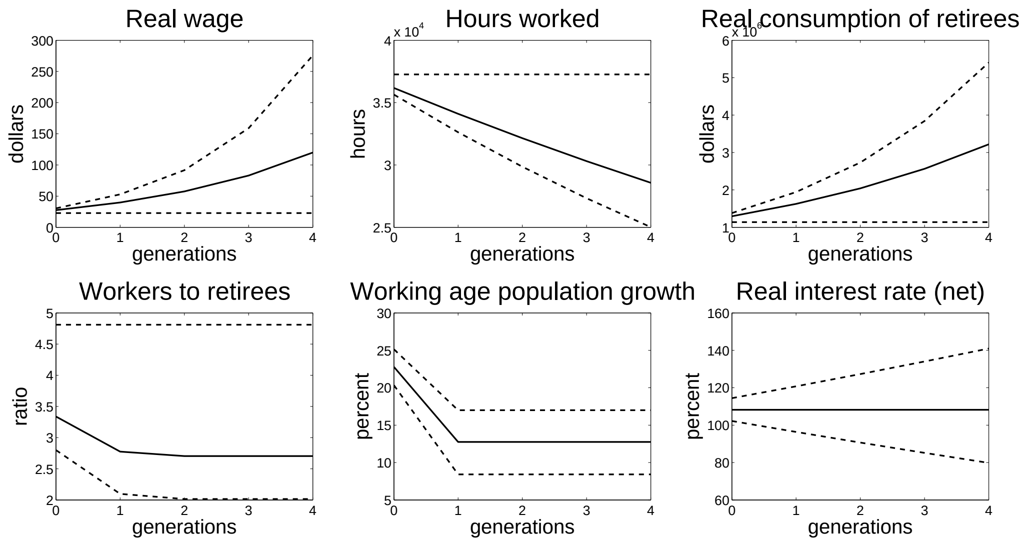

Figure 3: Different scenarios for the future development of variables affecting the welfare change. The solid line represents the baseline, the dashed lines indicate the high and low scenarios.

Table 4: Overall welfare effect (in 10000) according to (22) for different scenarios, based on parameter values $\kappa = 0.95^{30}$ and $\gamma = 1$.

	real wages	hours worked	retirees' consumption	workers to retirees	population growth	real interest rate
high	5.3	4.7	4.5	6.4	4.5	4.3
baseline	4.4	4.4	4.4	4.4	4.4	4.4
low	3.3	4.3	4.2	3.7	4.4	4.5

Overall, our results are robust regarding changes in the future development of macroeconomic variables. For all scenarios, the sign of the consumption equivalent impact is identical to Table 1 for each channel and generation (up to 11 generations). Moreover, although there are some quantitative differences, the overall welfare effect is stable over many of the scenarios for many parameter combinations. Table 4 summarizes the overall welfare effect for the different scenarios

[16]Note that Figure 3 depicts working age population growth over generations of 30 years. For example, the long-run baseline forecast of 12.8% would translate into an average yearly growth rate of 0.4%.

[17]Note that Figure 3 depicts the net real interest rate over generations of 30 years. For example, the long-run baseline forecast of net 108.2% interest would translate into an average yearly net interest rate of 2.5%.

[18]For some of the high scenarios, we cannot conduct the analysis for the combination of parameter values $\kappa = 0.985^{30}$ and $\gamma \leq 0.5$ because, in these cases, the sequence of welfare weights does not converge fast enough to zero for the problem to be well defined.

for the case of $\kappa = 0.95^{30}$ and $\gamma = 1$. For this combination of parameter values, all scenarios yield a very similar result. The future evolution of wages and the ratio of workers to retirees seem to be more relevant for the overall welfare effect than the development of the other variables.

For large values of κ, the overall welfare effect is less stable across the scenarios. In particular, there are some cases in which the sign changes. In the low scenario for the real interest rate, the overall welfare effect becomes positive for all parameter values of γ considered. In the low scenario for consumption of retirees, the overall welfare effect becomes negative.

If we use the "politician's welfare weights" instead of Assumption 3, the overall welfare effect is also more sensitive to the future development of the relevant variables. In particular, the overall effect gets positive in case of high real wage growth and, for some values of γ, also in a few other scenarios.

4 Conclusion

Old-age provision constitutes an essential element of many developed countries' social insurance programs. As demographic changes cause increasing financial stress for PAYGO systems, reforms of existing systems become more and more relevant.

This paper proposes a complementary method for the welfare analysis of PAYGO social security systems. Based on an OLG model featuring endogenous labor supply and idiosyncratic longevity, we derive a simple formula for the local welfare consequences of a permanent change in the payroll tax. In addition, we show how to extend this formula to incorporate different taxes to finance the PAYGO system and changes in the pension age.

We propose two different approaches to implement our formula, both of which are based on predictions for different key quantities of the model and the reduced form estimates of a VAR. Using data for the United States, we estimate a positive effect of a marginal increase in the payroll tax for most of the parameters values we consider. The sign of the overall effect stems from positive effects for today's retirees that outweigh the welfare losses for today's workers and future generations. To this end, our findings indicate that changing the payroll tax induces a distributional conflict across generations. A detailed decomposition by channels and generations sheds light on the driving forces behind this result. The direct effect through changes in the tax rate and social security benefits as well as the general equilibrium effects through changes in factor prices are the predominant determinants of the overall welfare effect while the effect of adjustment in individual labor supply is of minor importance. Robustness checks based on a scenario analysis confirm the robustness of our results.

Compared to the traditional approach to welfare analysis based on calibrated and estimated structural models, our method does neither require knowledge of the deep structure of the model nor does it rely on the estimation of this structure. In particular, we do not require a full parameterization of household preferences, a specification of the aggregate production function, or knowledge on individuals' savings behavior, nor do we attempt to estimate the structural parameters. Instead, the welfare consequences of a change in the payroll tax can be deduced from reduced form estimates under weaker, more transparent, and arguably more credible assumptions.

Regarding the sufficient statistics literature in microeconomics, our analysis extends the range of applications to macroeconomic dynamic general equilibrium models and highlights the challenges associated with deriving and implementing sufficient statistics in these models. The basic idea of deriving sufficient statistics in dynamic general equilibrium models and estimating them using time series models can be applied in different settings and is an interesting direction for further research.

Our approach shares two important limitations: First, our results are only locally valid. In particular, the analysis of real world payroll tax changes would require additional assumptions due to the discrete (and not infinitesimal) nature of these policy changes. Second, a new formula needs to be derived for every research question. In particular, it is not possible to use the same formula for the analysis of mixed social security systems or to compare different pension systems (e.g., funded and unfunded systems). In the light of these limitations, we consider our method to be complementary to the structural approach, because it allows for a weakening of some of the required assumptions on the one hand, but it only applies to the specific question of welfare consequences of payroll tax changes in PAYGO systems on the other hand.

References

Auerbach, Alan, J., Kotlikoff, Laurence, J., 1987. Dynamic fiscal policy. Cambridge University Press.

Breyer, F., Straub, M., 1993. Welfare effects of unfunded pension systems when labor supply is endogenous. Journal of Public Economics 50, 77–91.

Card, D., Chetty, R., Weber, A., 2007. Cash-on-hand and competing models of intertemporal behavior: New evidence from the labor market. Quarterly Journal of Economics 122(4), 1511–1560.

Chetty, R., 2008. Moral hazard vs. liquidity and optimal unemployment insurance. Journal of Political Economy 116, 173–234.

Chetty, R., 2009. Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. Annual Review of Economics 1, 451–488.

Diamond, P. A., 1965. National debt in a neoclassical growth model. The American Economic Review 55 (5), pp. 1126–1150.

Fanti, L., Spataro, L., 2006. Endogenous labor supply in Diamond's (1965) OLG model: A reconsideration of the debt role. Journal of Macroeconomics 28, 428–438.

Fehr, H., Kallweit, M., Kindermann, F., 2012. Pension reform with variable retirement age - a simulation analysis for Germany. Journal of Pension Economics and Finance 11(3), 389–417.

Fehr, H., Kotlikoff, L. J., Leibfritz, W., January 1999. Generational Accounting in General Equilibrium. University of Chicago Press, pp. 43–72.

Feldstein, M., 1985. The optimal level of social security benefits. The Quarterly Journal of Economics 100(2), 303–320.

Feldstein, M., 2005. Structural reform of social security. The Journal of Economic Perspectives 19 (2), 33–55.

Gahramanov, E., Tang, X., 2013. Should we refinance unfunded social security? Economica 80 (319), pp. 532–565.

Gonzalez-Eiras, M., Niepelt, D., 2007. The future of social security. Study Center Gerzensee Working Paper.

Gonzalez-Eiras, M., Niepelt, D., 2008. The future of social security. Journal of Monetary Economics 55, 197–218.

Gonzalez-Eiras, M., Niepelt, D., 2012. Ageing, government budgets, retirement, and growth. European Economic Review 56 (1), 97 – 115.

Imrohoroglu, A., Imrohoroglu, S., Joines, D. H., 1995. A life cycle analysis of social security. Economic Theory 6 (1), pp. 83–114.

Imrohoroglu, S., Kitao, S., 2009. Labor supply elasticity and social security reform. Journal of Public Economics 93 (7-8), 867–878.

Kotlikoff, Laurence, J., Smetters, K., Walliser, J., 1999. Privatizing social security in the United States - comparing the options. Review of Economic Dynamics 2, 532–574.

Kotlikoff, L. J., Smetters, K., Walliser, J., 2007. Mitigating America's demographic dilemma by pre-funding social security. Journal of Monetary Economics 54, 247–266.

Krueger, D., Kubler, F., 2006. Pareto-improving social security reform when financial markets are incomplete!? The American Economic Review 96 (3), pp. 737–755.

Lopez-Garcia, M.-A., 2008. On the role of public debt in an OLG model with endogenous labor supply. Journal of Macroeconomics 30, 1323–1328.

McGrattan, E., Prescott, Edward, C., 2013. On financing retirement with an aging population. NBER working paper series, No. 18760.

Nickel, C., Rother, P., Theophilopoulou, A., 2008. Population ageing and public pension reform in a small open economy. European Central Bank Working Paper Series, No. 863.

Nourry, C., 2001. Stability of equilibria in the overlapping generations model with endogenous labor supply. Journal of Economic Dynamics & Control 25, 1647–1663.

Pecchenino, R. A., Utendorf, K. R., 1999. Social security, social welfare and the aging population. Journal of Population Economics 12 (4), pp. 607–623.

Song, Z., 2011. The dynamics of inequality and social security in general equilibrium. Review of Economic Dynamics 14, 613–635.

A Components of the overall welfare effect and relation to dynamic inefficiency

As shown in section 2.2, a change in the payroll tax rate affects welfare through three distinct channels: a direct redistribution effect, an effect through changes in factor prices, and an effect due to the labor adjustment of the subsequent generations.

The sign of the direct redistribution component of the overall welfare effect is related to the dynamic inefficiency of the economy. If the economy is dynamically inefficient, then the redistribution from young to old households, which is induced by the increase in the payroll tax rate, benefits not only the initial old, but also the other generations. Neglecting the response of w_t and n_t in equation (15) for the moment and setting $p_t = 1$, the welfare effect amounts to

$$\xi_{-1}\lambda_{-1}^y \frac{(1+\chi_{-1,0})}{R_0} w_0 n_0 + \sum_{t=0}^{\infty} \xi_t \lambda_t^y w_t n_t \Bigg(-1 + \frac{1+\chi_{t,t+1}}{R_{t+1}} \frac{w_{t+1} n_{t+1}}{w_t n_t} \Bigg). \tag{25}$$

If the expression in parentheses is positive, then generations other than the initial old are also better off due to the direct redistribution. In a steady state with $\frac{n_{t+1}}{n_t} = 1$ and $\frac{w_{t+1}}{w_t} = 1 + g$,[19] this is the case if

$$\begin{aligned} -1 + \frac{(1+g)(1+\chi)}{R} &> 0 \\ \Leftrightarrow -(1-\delta+f_K) + (1+g)(1+\chi) &> 0 \\ \Leftrightarrow \delta + \chi + g + \chi g &> f_K. \end{aligned} \tag{26}$$

This is exactly the condition for dynamic inefficiency, which is given by the marginal product of capital being smaller than the depreciation rate plus the growth rate of the economy.

The sign of the factor price component is also related to the dynamic inefficiency of the economy. Neglecting the direct redistribution effects and the response of n_t in equation (15) for the moment, the welfare effect amounts to

$$\begin{aligned} &- \xi_{-1}\lambda_{-1}^y \frac{1+\chi_{-1,0}}{R_0} \frac{dw_0}{d\tau} n_0 (1-\tau) \\ &+ \sum_{t=0}^{\infty} \xi_t \lambda_t^y (1-\tau) n_t \Bigg(\frac{dw_t}{d\tau} - \frac{1+\chi_{t,t+1}}{R_{t+1}} \frac{dw_{t+1}}{d\tau} \frac{n_{t+1}}{n_t} \Bigg). \end{aligned} \tag{27}$$

In a steady state with $\frac{n_{t+1}}{n_t} = 1$ and $\frac{dw_{t+1}}{d\tau} = (1+g)\frac{dw_t}{d\tau}$, the expression in parentheses is given

[19] In steady state, wages grow with the same rate as productivity E, whose growth rate is denoted by g.

by

$$\left(1 - \frac{(1+g)(1+\chi)}{R}\right) \frac{dw}{d\tau}. \tag{28}$$

If the economy is dynamically inefficient, the factor price effect is of opposite sign than $dw/d\tau$.

Finally, the labor adjustment component in equation (15) is given by

$$\tau \left(\xi_{-1} \lambda^y_{-1} \frac{1+\chi_{-1,0}}{R_0} \frac{dn_0}{d\tau} w_0 + \sum_{t=0}^{\infty} \xi_t \lambda^y_t \frac{1+\chi_{t,t+1}}{R_{t+1}} \frac{dn_{t+1}}{d\tau} w_{t+1} \right).$$

Thus, the sign of the labor adjustment component equals the sign of $dn_t/d\tau$. Unsurprisingly, there is a negative impact on welfare if an increase in the distorting tax on labor leads to a reduction in hours worked.

B Extension I: consumption tax

Our analysis can directly be extended to include additional taxes (e.g., consumption and capital taxes) to refinance the PAYGO system. Consider, for example, adding a consumption tax to the model. To clarify the notation, denote the payroll tax as τ^w and the consumption tax as τ^c. Social security transfers are given by

$$T_t = n_t w_t \tau^w_t (1 + \chi_{t-1,t}) + c^y_t \tau^c_t (1 + \chi_{t-1,t}) + c^o_t \tau^c_t,$$

where we set $p_t = 1$ to simplify the exposition.

Similar arguments as before yield the following formula for the welfare effect of a permanent change in the payroll tax,

$$\begin{aligned}
\frac{dW}{d\tau^w} = & \sum_{t=0}^{\infty} w_t n_t (-\lambda^o_{t+1} R_{t+1} \xi_t + (1+\chi_{t-1,t}) \lambda^o_t \xi_{t-1}) \\
& + \sum_{t=0}^{\infty} \frac{dw_t}{d\tau^w} n_t (1 - \tau^w)(\lambda^o_{t+1} R_{t+1} \xi_t - (1+\chi_{t-1,t}) \lambda^o_t \xi_{t-1}) \\
& + \sum_{t=0}^{\infty} \frac{dn_t}{d\tau^w} w_t \tau^w (1+\chi_{t-1,t}) \lambda^o_t \xi_{t-1} \\
& + \sum_{t=0}^{\infty} \left(\frac{dc^y_t}{d\tau^w} (1+\chi_{t-1,t}) + \frac{dc^o_t}{d\tau^w} \right) \tau^c_t \lambda^o_t \xi_{t-1},
\end{aligned}$$

and the corresponding formula for the welfare consequences of a permanent change in the consumption tax:

$$\begin{aligned}\frac{dW}{d\tau^c} = &\sum_{t=0}^{\infty} c_t^y(-\lambda_{t+1}^o R_{t+1}\xi_t + (1+\chi_{t-1,t})\lambda_t^o \xi_{t-1}) \\ &+ \sum_{t=0}^{\infty} \frac{dw_t}{d\tau^c} n_t(1-\tau_t^w)(\lambda_{t+1}^o R_{t+1}\xi_t - (1+\chi_{t-1,t})\lambda_t^o \xi_{t-1}) \\ &+ \sum_{t=0}^{\infty} \frac{dn_t}{d\tau^c} w_t \tau_t^w (1+\chi_{t-1,t})\lambda_t^o \xi_{t-1} \\ &+ \sum_{t=0}^{\infty} \left(\frac{dc_t^y}{d\tau^c}(1+\chi_{t-1,t}) + \frac{dc_t^o}{d\tau^c}\right)\tau^c \lambda_t^o \xi_{t-1}.\end{aligned}$$

These formulas differ from equation (14) through the last sum that also includes the dynamic reactions of consumption with respect to the tax rates, which was absent in the analysis before. Moreover, the direct effect is scaled by c_t^y in the case of a change in the consumption tax whereas it is scaled by $w_t n_t$ in the formula for a change in the payroll tax rate.

In our empirical application for the US, we do not implement the above formulas because there is no consumption tax on the federal level. We stress, however, that implementing extensions of our formula in equation (14) is an interesting topic for future research.

C Extension II: retirement age

Our analysis can be extended to allow for changes in retirement age. The household problem is modified in the following way. Old households work for a fraction ϱ_t of the model period before they reach retirement age. An increase in ϱ_t therefore corresponds to an increase in the retirement age. This setup includes the main model of this paper as a special case: If $\varrho_t = 0$, then old households do not work at all. To simplify the exposition, we set $p_t = 1$.

The household solves the following maximization problem:

$$\begin{aligned}\max_{n_t^y, n_{t+1}^o, c_t^y, c_{t+1}^o, s_{t+1}} \quad & u(c_t^y, \bar{n} - n_t^y, c_{t+1}^o, \bar{n} - n_{t+1}^o) \\ \text{s.t.} \quad & c_t^y + s_{t+1} = n_t^y w_t(1-\tau_t), \\ & c_{t+1}^o = \varrho_{t+1} n_{t+1}^o w_{t+1}(1-\tau_{t+1}) + R_{t+1}s_{t+1} + T_{t+1}, \\ & c_t^y, c_{t+1}^o > 0.\end{aligned}$$

The first order conditions read:

$$\lambda_t^y = u_{c^y}(c_t^y, \bar{n} - n_t^y, c_{t+1}^o, \bar{n} - n_{t+1}^o), \tag{29}$$

$$\lambda_t^y w_t(1-\tau_t) = u_{l^y}(c_t^y, \bar{n} - n_t^y, c_{t+1}^o, \bar{n} - n_{t+1}^o), \tag{30}$$

$$\lambda_{t+1}^o = u_{c^o}(c_t^y, \bar{n} - n_t^y, c_{t+1}^o, \bar{n} - n_{t+1}^o), \tag{31}$$

$$\lambda_{t+1}^o \varrho_{t+1} w_{t+1}(1-\tau_{t+1}) = u_{l^o}(c_t^y, \bar{n} - n_t^y, c_{t+1}^o, \bar{n} - n_{t+1}^o), \tag{32}$$

$$\lambda_t^y = R_{t+1}\lambda_{t+1}^o. \tag{33}$$

Transfer payments to old households are equal to the tax revenues:

$$T_t = w_t \tau_t (n_t^y(1+\chi_{t-1,t}) + n_t^o \varrho_t). \tag{34}$$

Total hours worked amount to:

$$H_t = n_t^y L_t + n_t^o L_{t-1}\varrho_t. \tag{35}$$

Similar arguments as in the main text yield the following formula for the welfare effect of a permanent change in the payroll tax:

$$\begin{aligned}
\frac{dW}{d\tau} =& \sum_{t=0}^{\infty} w_t n_t^y(-\lambda_t^y \xi_t + (1+\chi_{t-1,t})\lambda_t^o \xi_{t-1}) \\
&+ \sum_{t=0}^{\infty} \xi_t \Bigg(\lambda_t^y \frac{dw_t}{d\tau} n_t^y (1-\tau) + \lambda_{t+1}^o \Bigg(\frac{dR_{t+1}}{d\tau} s_{t+1} \\
&\qquad + \frac{dn_{t+1}^y}{d\tau} w_{t+1}\tau(1+\chi_{t,t+1}) + \frac{dn_{t+1}^o}{d\tau} w_{t+1}\tau\varrho_{t+1} \\
&\qquad + \frac{dw_{t+1}}{d\tau}\tau(n_{t+1}^y(1+\chi_{t,t+1}) + n_{t+1}^o\varrho_{t+1}) + \frac{dw_{t+1}}{d\tau}\varrho_{t+1} n_{t+1}^o (1-\tau) \Bigg) \Bigg) \\
&+ \xi_{-1}\lambda_0^o \Bigg(\frac{dR_0}{d\tau} s_0 + \frac{dn_0^y}{d\tau} w_0\tau(1+\chi_{-1,0}) + \frac{dn_0^o}{d\tau} w_0 \tau \varrho_0 \\
&\qquad + \frac{dw_0}{d\tau}\tau(n_0^y(1+\chi_{-1,0}) + n_0^o\varrho_0) + \frac{dw_0}{d\tau}\varrho_0 n_0^o(1-\tau) \Bigg).
\end{aligned} \tag{36}$$

Using

$$\frac{dR_{t+1}}{d\tau} = -\frac{dw_{t+1}}{d\tau}\frac{n_{t+1}^y(1+\chi_{t,t+1}) + n_{t+1}^o \varrho_{t+1}}{s_{t+1}}$$

and

$$\lambda_t^y = R_{t+1}\lambda_{t+1}^o,$$

we get the following overall welfare effect:

$$\begin{aligned}\frac{dW}{d\tau} =& \sum_{t=0}^{\infty} w_t n_t^y(-\lambda_{t+1}^o R_{t+1}\xi_t + (1+\chi_{t-1,t})\lambda_t^o \xi_{t-1}) \\ &+ \sum_{t=0}^{\infty} \frac{dw_t}{d\tau} n_t^y (1-\tau)(\lambda_{t+1}^o R_{t+1}\xi_t - (1+\chi_{t-1,t})\lambda_t^o\xi_{t-1}) \\ &+ \sum_{t=0}^{\infty}\left(\frac{dn_t^y}{d\tau}(1+\chi_{t-1,t}) + \frac{dn_t^o}{d\tau}\varrho_t\right) w_t \tau \lambda_t^o \xi_{t-1}.\end{aligned} \tag{37}$$

This formula is almost identical to equation (14). There is only one additional term,

$$\sum_{t=0}^{\infty} \frac{dn_t^o}{d\tau}\varrho_t w_t \tau \lambda_t^o \xi_{t-1},$$

which pertains to the labor response of old households. The explanation for the small difference between the equations (14) and (37) is very intuitive. In the extended model, old households have to pay payroll taxes until they reach the retirement age, but these tax payments are refunded to the same old households after retirement. Thus, there is no redistributive effect of a change in τ beyond the one in the main model. However, old households may adjust their hours of work in response to a change in the distorting payroll tax rate, which has an impact on welfare.[20]

In our empirical application for the US, we do not implement formula (37) due to data limitations. In particular, we would need reliable data on hours worked by age groups in order to estimate $\{dn_t^y/d\tau\}_{t=0,\ldots,\infty}$ and $\{dn_t^o/d\tau\}_{t=0,\ldots,\ \infty}$. Nevertheless, (37) allows for analyzing how the welfare effect of a change in τ depends on the retirement age ϱ_t. If $\{dn_t^o/d\tau\}_{t=0,\ldots,\infty}$ is negative, then an increase in the retirement age negatively affects both overall $dW/d\tau$ and the welfare effect on each generation (and vice versa if $\{dn_t^o/d\tau\}_{t=0,\ldots,\infty}$ is positive). Intuitively, the more years people work before they retire, the more important the welfare cost of a reduction in hours worked due to the increase in the distortionary payroll tax rate becomes.

[20]This adjustment is a consequence of the fact that households are small, i.e., they do not take into account that their choice of n_{t+1}^o affects their own pension transfers T_{t+1}.

D Lump-sum redistribution authority

The overall welfare change includes both redistribution and efficiency effects. To assess the importance of the efficiency effects, we consider an additional, hypothetical government authority in the spirit of the *lump-sum redistribution authority* (LSRA) in Auerbach and Kotlikoff (1987). Suppose that, in each period t, the LSRA makes a lump-sum transfer of $w_t n_t - \frac{dw_t}{d\tau} n_t (1-\tau)$ to each young household and finances this transfer by imposing a lump-sum tax of $\frac{1+\chi_{t-1,t}}{p_t}\left(w_t n_t - \frac{dw_t}{d\tau} n_t (1-\tau)\right)$ on each old household. In this hypothetical case, only the terms related to $\frac{dn}{d\tau}$ remain in equation (15). To make each generation as well off as before the policy change, the LSRA would additionally need to pay a lump-sum transfer of $\frac{dn_t}{d\tau} w_t \tau \frac{1+\chi_{t-1,t}}{p_t}$ to each old household in period t. The present discounted value of these additional transfer payments can be interpreted as efficiency loss caused by the policy change. Normalizing $L_{-1} = 1$, it amounts to

$$\frac{dn_0}{d\tau} w_0 \tau (1+\chi_{-1,0}) + \sum_{k=1}^{\infty} \frac{\frac{dn_k}{d\tau} w_k \tau (1+\chi_{-1,k})}{\Pi_{j=1}^{k} R_j}. \tag{38}$$

Such a hypothetical LSRA can be mimicked by choosing planner's welfare weights equal to:

$$\xi_{-1} = \frac{R_0}{\lambda_{-1}^{y}}, \tag{39}$$

$$\xi_t = \frac{R_0}{\lambda_t^{y}} \frac{1+\chi_{-1,t}}{\Pi_{k=0}^{t} R_k} \qquad \text{for } t = 0, 1, 2, \ldots. \tag{40}$$

With these welfare weights, $dW/d\tau$ is exactly equal to the quantity in (38), which provides a measure of the efficiency costs of the change in τ.

In our empirical application for the US, we can estimate the efficiency cost of the change in τ by using the welfare weights (39) and (40). In present value terms, this efficiency cost amounts to USD 4449 per household (both young and old) in $t = 0$. This cost is cleaned up from redistribution effects and solely caused by increasing the distortive payroll tax.

E Data and aggregation

We use quarterly data from 1964 to 2010 to estimate the VAR described in section 3.1. The years 2011 to 2013 are excluded because there were adjustments in the payroll tax rate for which it is unclear how, if at all, they fit into our analysis. For 2011 and 2012, the payroll tax rate was temporarily reduced, but the reduction in tax revenue was made up by transfers from the general fund of the Treasury. Thus, there was a reduction in payroll taxes, but no corresponding change in the receipts of the social security system. As our model does not include a government besides the social security system, we cannot properly cover this temporary reduction in τ_t financed by the general fund of the Treasury. Therefore, we exclude this period from our estimation sample and use data until 2010 only.

Table 5 contains a description of the data that are used for estimating the VAR model or needed to implement equations (18), (19) and (22). Table 6 explains the data aggregation to the OLG model frequency for the variables appearing in equations (18), (19) and (22).

Table 5: Description of the Data

Variable	Description
n_t	Number of hours worked per capita per quarter. The series is constructed using an index of hours worked in the business sector (source: Bureau of Labor Statistics), scaled up to match total hours worked in the base year and divided by the number of people between 15 and 65 years (source: Bureau of Labor Statistics).
w_t	Real hourly wage (in 2010 dollars). In correspondence with n_t, we use an index of real hourly compensation in the business sector (source: Bureau of Labor Statistics), scaled to match average hourly earnings of all employees in the total private sector in 2010.
R_t	Real interest rate. The series is constructed using the *10-Year Treasury Constant Maturity Rate* (source: Board of Governors of the Federal Reserve System) as nominal long-term interest rate. Our inflation measure is based on the *Consumer Price Index for All Urban Consumers* (source: Bureau of Labor Statistics).
c_t^o	Real consumption of retirees (in 2010 dollars). The series is constructed by deflating yearly data on total average expenditures of people over age 65 (source: Bureau of Labor Statistics, Consumer Expenditure Survey).
$\chi_{t-1,t}$	Growth rate of the working age population (16 to 64 years) (source: Bureau of Labor Statistics, and Census Bureau).
Ratio of workers to retirees	Size of the working age population (16 to 64 years) divided by the size of the retiree population (65 and over) (source: Bureau of Labor Statistics, and Census Bureau).
real consumption per capita growth	Growth rate of real *Private Final Consumption Expenditure in United States* (source: OECD National Accounts Statistics) divided by total population (source: U.S. Department of Commerce, Census Bureau).
real GDP per capita growth	Growth rate of real GDP (source: U.S. Bureau of Economic Analysis).

Table 6: Aggregation of higher frequency data to OLG frequency of 30 years

Variable	Aggregation
n_t	Quarterly data are added up over 30 years in order to get the number of hours worked per capita over 30 years.
w_t	The average real hourly wage over 30 years is computed.
R_t	The real long-term gross interest rate per annum is projected to 30 years (i.e., $R_{30y} = R_{p.a.}^{30}$).
c_t^o	Annual data are added up over 30 years in order to get total real consumption per capita over 30 years.
$\chi_{t-1,t}$	For each 30 year window, the average size of the working age population is computed. This series is then used to compute the growth rate of the working age population.
p_t	p_t is computed by dividing $(1 + \chi_{t-1,t})$ by the average ratio of workers to retirees for each 30 year window.