The adverse and beneficial effects of front-loaded pension contributions

Ed Westerhout*

May 4, 2019

Abstract

Collective pension schemes typically apply age-independent contribution and accrual rates. This implies that pension contributions are front-loaded. If pension contributions and accruals relate to earned labour income, this affects labour market efficiency. For it implies that the labour supply of younger workers is implicitly taxed and that of older workers implicitly subsidized. This paper confirms conventional wisdom that front-loading may be welfare-reducing, both through the implicit tax and the implicit subsidy. It also shows that the welfare loss is weakened however if one accounts for government spending that is financed with a labour income tax. Front-loading may even be welfare-increasing. In particular, the more elastic is the labour supply of older workers relative to that of younger workers, the more likely it is that front-loading of pension contributions produces a welfare gain.

*Ed Westerhout is affiliated with the department Tax Economics at Tilburg University and with CPB Netherlands Bureau for Economic Policy Analysis. He thanks Jurre de Haan, Marcel Lever, Siert Vos and Sweder van Wijnbergen for discussion and comments on an earlier version of this paper.
1 Introduction

Linda, 25 years old, has an annual income of EUR 20,000. She considers increasing her number of working hours per week from 20 to 24. If she does, she will have to contribute EUR 1,866.40 extra to her collective pension scheme. Her pension rights will increase less, namely EUR 1,226.40. Sara, 60 years old, happens to be in the same position as Linda, except for her age. Hence, Sara participates in the same pension scheme and faces the same contribution rate as Linda. But if she adds 4 hours to her working week, she will pay only EUR 933.20 more contributions to the pension scheme, half as much as Linda, whereas her pension rights will increase as much as Linda’s, EUR 1,226.40.\(^1\)

Worldwide, pension reform is at the top of the policy agenda. Against the background of ageing populations and changing labour market patterns, many countries have reformed their pension schemes or are considering pension reform. In particular, some countries have made the move to individual defined contribution schemes (Chile) or have switched more gradually away from defined benefit towards defined contribution schemes (Australia, US, UK). Other countries have transformed their traditional pay-as-you-go schemes into notional defined contribution schemes (Italy, Latvia, Poland, Sweden). Thirdly, some countries introduced collective defined contribution schemes to replace traditional collective defined benefit schemes (Canada, Denmark, the Netherlands). Furthermore, many countries have adopted measures to stepwise increase the pension retirement age (see Bonenkamp et al. (2017) for a more detailed overview).

The current coalition government in the Netherlands is considering further pension reform (Regering, 2017; Ministerie van SZW, 2019). Among that is a quite uncommon element, namely the replacement of the current scheme of linear, age-independent accrual rates with a degressive scheme in which accrual rates decline with age. The aim is to make the second pillar of the Dutch pension scheme more actuarially fair. This would help to end the redistribution between workers of different age that is implied by a linear accrual scheme and that many consider as unfair. As underlined in Ministerie van SZW (2019), it would also improve the connection with the

\(^{1}\)This calculation assumes that both persons work from the age of 25 to 65, that they receive a pension from the age of 65 to 85, that labour income is constant, pension benefits are 75 percent of labour income, pensions are fully indexated against price inflation and the annual interest rate is 2 percent. Furthermore, all amounts in the example are expressed in terms of euros at the age of retirement.
current dynamic labour market, ease the introduction of more flexibility in the accumulation and pay-out phase and make it more attractive for the self-employed to opt in into the pension scheme that is currently obligatory only for salary workers.

This paper argues that this type of pension reform may have important labour market effects. The argument rests on two observations. The first is that a linear accrual scheme implies redistribution from young to old workers. As pension contributions increase in value over time because of the interest earned, pension contributions are front-loaded relative to accruals. Hence, young workers pay more than they earn in terms of pension accruals, whereas for old workers the reverse holds true. The second observation is that both pension contributions and accruals typically relate to earned labour income. That means that the income transfers between young and old workers may be perceived as implicit taxes (in case of young workers) and implicit subsidies (in case of old workers). If the pension reform studied here implies that these taxes and subsidies disappear, this may affect the labour market behaviour of both young and old workers.

One may speculate that the reform will be welfare-increasing as it removes two distortions on labour markets. As we will show, this idea is correct, but only if two conditions do not apply. One is that the government explicitly taxes labour income in order to finance a given amount of government services. The second is that the labour supply behaviour of old workers is more elastic than that of young workers. If both these conditions apply - which we will argue seems more to be the rule than the exception, the reform in case may deteriorate social welfare rather than improving it. That the front-loading of pension contributions may be welfare-increasing may be considered an important result, not only for the Netherlands, but also for a number of other countries. According to Whitehouse (2006), second-pillar pension schemes in many OECD countries feature linear accrual rates, whereas contribution rates are typically linear as well (Van Vuuren, 2014).

The combination of linear accrual rates and linear contribution rates can be found in defined benefit (DB) schemes and in collective defined contribution schemes like those in the Netherlands. It is not characteristic of individual defined contribution (DC) schemes, in which contributions and accruals match each other by definition. Overtime, DB schemes are losing market share worldwide, whereas DC schemes become gradually more important. This is illustrated in Willis Towers Watson (2018) who show that for the seven largest pension markets that together hold 91 percent of pension assets in the world, the share of DC has increased about 50 percent in only 20 years time. However, the collective DB scheme is still a major type of pension scheme, figuring in 17 out of 30 OECD countries (Whitehouse, 2006).

Our paper does not focus on PAYG schemes in which typically life-time contributions
This paper builds upon earlier literature in the field of pensions. In particular, Ippolito (1985), Bodie et al. (1988) and Kotlikoff and Wise (1988) discuss the valuation of pension liabilities, whereas Bonenkamp (2009) and Bovenberg and Gradus (2015) discuss the same front-loading of pension contributions as we focus upon. It also relates to a large literature on the distortionary effects of taxation in general and age-dependent labour income taxation in particular (Atkinson and Sandmo, 1980; Erosa and Gervais, 2002; Fenge et al., 2006; Bastani et al., 2013). Further, the paper is connected to the literature on the role of financial incentives in labour supply and retirement (Chan and Stevens, 2004; Asch et al., 2005; Hanel, 2010).

This paper is structured as follows. Section 2 is a starter that demonstrates how the combination of linear contribution and accrual rates makes pension contributions front-loaded relative to accruals. Sections 3 and 4 use the insights obtained from section 2 to explore the labour market and welfare effects of front-loading. Both sections adopt a stylized two-period model of the labour market, without (section 3) or with (section 4) government spending. Up till then, the paper describes an individual pension scheme. Section 5 follows up by exploring how our results would change if we assumed a collective pension scheme instead. Section 6 discusses some other qualifications. Finally, section 7 offers concluding remarks.

2 The implicit taxes and subsidies implied by front-loaded pension contributions

In general, collective pension schemes apply linear contribution and accrual rates, i.e. pension contribution rates and accrual rates do not differentiate with respect to participant characteristics like age. This type of uniformity is intrinsically linked with the collective nature of the pension scheme. In many cases, uniformity is also legally required.

This uniformity creates an imbalance between contributions and accruals. A euro of pension contributions made by a young worker has a greater economic value than the same euro contributed by an old worker as the former euro has a longer time to accumulate and earn interest on it. However, the economic value of a euro pension accrual for a young worker is equal to that for an old worker. Hence, uniformity implies that pension contributions and accruals are different. Indeed, Fenge et al. (2006) show that a PAYG scheme may imply implicit taxation over the whole working phase of the life cycle. The two types of schemes do have something in common, however. The PAYG scheme studied in Fenge et al. (2006) features declining implicit tax rates over the life cycle.
are front-loaded relative to pension accruals.\textsuperscript{4} The story is slightly different if pension benefits are indexed to price or wage inflation for then the economic value of pension accruals for young workers is also larger than that for old workers. However, as long as the rate of indexation is smaller than the rate of return on pension wealth, the imbalance between contributions and accruals remains.

This imbalance in turn implies redistribution between young and old workers. Young workers pay more than would be actuarially fair, whereas old workers pay less than what is actuarially fair. But there is more. If both contributions and accruals are related to individual labour income, the difference between the two will be similar to a tax or subsidy and may thus have labour market implications. How large this tax or subsidy is, is determined by the factors that drive the imbalance between contributions and accruals.

Let us establish these claims more formally. We focus on the life cycle of a representative worker that starts when the worker enters the labour market and ends when the worker retires. Let us use $T_R$ to denote the length of this life cycle. Upon retirement, the worker aims to have accumulated a given amount of pension wealth. We will not go into the question how this wealth is allocated over the retirement phase, as this is not relevant for the purposes of this paper.

As to the pension scheme, we make an assumption that may seem odd at first sight. We take the pension scheme to be an individual scheme. This implies that the representative individual plays a zero sum game with his or her pension scheme: accruals, pension benefits and contributions are equal in present-value terms. This route is a useful device to demonstrate the effects of uniform contribution and accrual rates. This individual scheme is not identical to collective schemes that feature the uniform contribution and accrual rates that we want to explore, however. Collective schemes and individual schemes differ in an important way and we will explore in section 5 in detail how this difference impacts upon our analysis.

We use $r$ to denote the rate of return on the capital market, $g$ to denote the rate of productivity growth and $\mu$ to denote the rate of indexation of pension accruals. All three are constant through time. Furthermore, we

\textsuperscript{4}Bovenberg and Gradus (2015) refer to the same phenomenon when they talk about back-loaded pension schemes. Indeed, front-loaded contributions and back-loaded accruals amount to the same thing. Note that this front-loading or back-loading is different from the type of back-loading discussed in the seminal paper by Ippolito (1985). That type of back-loading refers to final wage defined benefit schemes under the so-called legal theory of pension liabilities.
assume that the rate of return exceeds the rate of indexation: \( r > \mu \). This assumption does not seem unrealistic. If we impose that our economy is dynamically efficient, a feature that applies often if not always (Abel et al., 1989; Piketty, 2014), \( r \) will exceed \( g \). If we take \( \mu \) equal to \( g \), then dynamic efficiency is sufficient to validate our assumption \( r > \mu \). In the real world, pension benefits are not always fully indexated against wage inflation, however. Indexation is often against the lower rate of price inflation (Whitehouse, 2006) and indexation is often less than full. This means that even in a dynamically inefficient economy, \( r > \mu \) is likely to apply.

Given that the pension contribution rate is linear, we can express the total of pension contributions that an individual has made at the time of retirement as follows,

\[
\Pi_{TR} = \sum_{i=1}^{T_R} \pi y_i (1 + r)^{T_R - i} = \sum_{i=1}^{T_R} \pi y_{TR} \left( \frac{1 + r}{1 + g} \right)^{T_R - i}
\] (1)

where \( \pi \) denotes the (age-independent) contribution rate. The most RHS expression of equation (1) reflects that productivity grows at rate \( g \).

Let us denote the amount of pension wealth that the worker wants to have accumulated at the date of retirement as \( B_{TR} \). Absent any intergenerational transfers, \( \Pi_{TR} \) must then be equal to \( B_{TR} \). In other words, the individual’s pension benefits in the retirement phase should equal his or her contributions made in the working phase. This principle implies the following expression for the contribution rate:

\[
\pi = \frac{B_{TR}}{\sum_{i=1}^{T_R} \left( \frac{1 + r}{1 + g} \right)^{T_R - i} y_{TR}}
\] (2)

The accrual period covers \( T_R \) years, just like the contribution period. However, different from contributions, accruals accumulate at rate \( \mu \). Hence, we have the following expression for the total of pension accruals at the time of retirement,

\[
A_{TR} = \sum_{i=1}^{T_R} \alpha y_i (1 + \mu)^{T_R - i} = \sum_{i=1}^{T_R} \alpha y_{TR} \left( \frac{1 + \mu}{1 + g} \right)^{T_R - i}
\] (3)

where \( \alpha \) denotes the (age-independent) accrual rate. As in the case of \( \Pi_{TR} \), the most RHS expression of equation (3) expresses \( A_{TR} \) as proportional with \( y_{TR} \).
In order to calculate the accrual rate that ensures that the worker will receive the benefits that he has accumulated, we equate $A_{TR}$ to $B_{TR}$. This gives the following expression for the accrual rate:

$$\alpha = \frac{B_{TR}}{\sum_{i=1}^{TR} \left(\frac{1 + \mu}{1 + g}\right)^{TR-i} y_{TR}}$$

Combining equation (4) with equation (2) yields an expression for the contribution rate in terms of the accrual rate:

$$\pi = \alpha \left[ \frac{\sum_{i=1}^{TR} \left(\frac{1 + \mu}{1 + g}\right)^{TR-i} y_{TR}}{\sum_{i=1}^{TR} \left(\frac{1 + r}{1 + g}\right)^{TR-i} y_{TR}} \right]$$

This expression shows that the pension contribution rate is lower than the accrual rate (recall that $r > \mu$). This reflects the front-loaded nature of pension contributions.

Another way to show the front-loading is by elaborating how pension contributions and accruals evolve with age. We focus on pension contributions and accruals in terms of contemporaneous labour income in euros dated at time $T_R$. Pension contributions in terms of income at age $j$ ($j = 1, ..., T_R$) are defined as $\partial \Pi_{TR}/\partial y_j$ and read as $\pi(1 + r)^{TR-j}$. During the working phase, pension contributions thus decline with the factor $1 + r$: the holding period gets one year shorter when the person gets one year older. Pension accruals in terms of income at age $j$ ($j = 1, ..., T_R$) are defined as $\partial A_{TR}/\partial y_j$ and read as $\alpha(1 + \mu)^{TR-j}$. Pension accruals decline during the working phase with the factor $1 + \mu$. Recalling that $r > \mu$, pension contributions thus decline faster over the working phase than pension accruals.

Now, we can establish formally that pension contributions exceed (are smaller than) accruals at younger (older) ages. We define net benefits from the pension scheme, $Z_{TR}$, as accumulated pension rights minus pension contributions, $A_{TR} - \Pi_{TR}$, and differentiate this with respect to labour income at age $j$ ($j = 1, ..., T_R$). This gives $\partial Z_{TR}/\partial y_j$, which can be written as $\alpha(1 + \mu)^{TR-j} - \pi(1 + r)^{TR-j}$. Using the derived expressions for $\pi$ and $\alpha$ in equations (2) and (4), this can be elaborated into the following,

$$\frac{\partial Z_{TR}}{\partial y_j} = \kappa \left[ \sum_{i=1}^{TR} \left(\frac{1 + r}{1 + g}\right)^{TR-i} \left(\frac{1 + \mu}{1 + g}\right)^{TR-j} - \sum_{i=1}^{TR} \left(\frac{1 + \mu}{1 + g}\right)^{TR-i} \left(\frac{1 + r}{1 + g}\right)^{TR-j} \right]$$

(6)
where \( \kappa = \alpha (1 + g) T^r_j / \sum_{t=1}^{T_R} ((1 + r)/(1 + g))^{T_R-t} \).

One can easily derive that the expression between accolades is zero if the worker is \( j^\ast \) years old, where \( j^\ast \) is implicitly defined by the following condition:

\[
\sum_{i=1}^{T_R} \left( \frac{1 + \mu}{1 + g} \right)^{j-i} = \sum_{i=1}^{T_R} \left( \frac{1 + r}{1 + g} \right)^{j-i}
\]

(7)

One can read from equation (6) that for values of \( j \) larger than \( j^\ast \), \( \partial Z_{T_R} / \partial y_j \) is positive and that \( \partial Z_{T_R} / \partial y_j \) is negative for values of \( j \) smaller than \( j^\ast \). This formally shows that pension contributions exceed (are smaller than) accruals at younger (older) ages.

Let us visualize this. Figures 1 and 2 show the results of an illustrative calculation, based on \( r = 0.02, g = 0.01, \mu = 0.01, T_R = 45 \) (age 20 to 64) and \( \beta = 1 \). Figure 1 shows the calculated development of pension contributions and accruals, where, as before, contributions are defined as \( \pi (1 + r)^{T_R-j} \) and accruals as \( \alpha (1 + \mu)^{T_R-j} \). The figure shows that both variables decline over the life cycle, that contributions are front-loaded relative to accruals and that, for young workers, contributions exceed accruals, and for old workers, accruals exceed contributions.

As discussed above, the difference between contributions and accruals can be considered an implicit pension tax. Figure 2 displays the annual and cumulative values of implicit pension tax revenues, defined as \( (-\partial Z_{T_R} / \partial y_j) y_j \) and \( \sum_{i=1}^{T_R} (-\partial Z_{T_R} / \partial y_i) y_i \) respectively. As figure 2 shows, cumulative implicit pension tax revenues are zero at the beginning and at the end of the working phase, positive at all other ages, and peaking somewhere in the middle of the working phase.

It may be worth stressing that figures 1 and 2 can only illustrate the implications of uniform contribution and accrual rates. They should not be used for calculating the taxes paid (or subsidies received) by a real-world worker of a certain age as typical assumptions made here will generally not hold true. In the real world, for example, the rate of return on the capital market and the rate of productivity growth are not constant, the individual productivity profile differs from the macroeconomic profile and the participation profile is not flat.

Having illustrated the implications of linear accrual and contribution rates, we must now take a step back, however, as the model is too general in terms of the model we will elaborate in the next two sections. Indeed, a two-period model (with \( T_R = 2 \)) is more appropriate. This is not worrisome as a two-period model captures all the basic characteristics of the more general model studied thus far. Now, it is straightforward to derive expressions for
all variables in this two-period model, but we will not do so to save space. We do however present the expression for the implicit pension subsidies in the second period of the working phase. These subsidies are defined as \((\partial Z_2 / \partial y_2) y_2\), but I prefer to introduce symbol \(S_2\):

\[
S_2 \equiv (\partial Z_2 / \partial y_2) y_2 = \beta y_2 \left[ \left( \frac{1}{1 + r} \right) + \left( \frac{1 + r}{1 + g} \right) \right] - \left( \frac{1 + \mu}{1 + r} \right) \left( \frac{1 + r}{1 + g} + 1 \right)
\] (8)

Equation (8) nicely demonstrates that the implicit pension subsidy (one may easily verify that \(S_2 > 0\)) received by the older worker is proportional with the replacement rate \(\beta\) and relates in a complex way to the interest rate, the rate of indexation and the rate of productivity growth. The same holds true for the subsidy rate, \(S_2/y_2\). Similarly, I introduce symbol \(T_1\) to denote the implicit taxes paid in the first period of the life cycle, defined as \(-(\partial Z_2 / \partial y_1) y_1\). As will be clear by now, these period-1 taxes equal the period-2 subsidies: \(T_1 = S_2\).

3 A three-period life-cycle model

The previous section has demonstrated that the combination of linear contribution and accrual rates implies implicit taxes and subsidies. To explore the labour market and welfare effects of these what we will call front-loading policies, we set up a life-cycle model. We take the angle of an optimal tax problem in a world with identical workers who can change their hours of work but whose productivity is taken as given. We thus abstract from positive or negative effects of front-loading on productivity.\(^{5}\)

The most simple life-cycle model that enables an analysis of front-loading policies is a non-stochastic three-period model. This model describes a consumer in three periods: when he is young and working, say between age 25 and 45, when he is old and working, say between age 45 and 65, and when he is retired, say, between age 65 and 85. The consumer chooses his labour supply in the first two periods and consumption in the third period such as to maximize his lifetime utility. The front-loading of pension contributions

\(^{5}\)Kotlikoff and Wise (1988) argue that front-loading may induce workers to remain with their firms. To the extent that this increases the return on job-specific human capital, front-loading may increase labour productivity. On the other hand, front-loading makes it unattractive for salary workers to become self-employed later in their career, as we will demonstrate in this paper. To the extent that such a change of job market status raises productivity, front-loading may lower productivity.
enters this decision problem by changing the price of leisure in the first and second period. As discussed in section 2, front-loading does not affect lifetime wealth. Furthermore, in this section, there is no government spending and no labour income taxation. Adding these two features will be relegated to the next section.

We describe the preferences of the worker by the following intertemporal utility function,

\[ U = u_1(v_1) + \frac{u_2(v_2)}{1+\delta} + \frac{c_3}{(1+\delta)^2} \]  

where \( v_i \) denotes leisure in period \( i=1,2 \), \( c_3 \) denotes consumption when retired and \( \delta \) denotes the individual discount rate. The utility function features positive and decreasing marginal utility with respect to leisure in both periods. Marginal utility with respect to period-3 consumption is a constant. Generalizing the utility function in order to allow for decreasing marginal utility of consumption would not change the analysis substantially.

The worker needs to finance his consumption during retirement out of the labour income earned in the two working periods,

\[ c_3 = (1+r)^2w_1l_1 + (1+r)w_2l_2 \]  

where \( l_i = 1 - v_i \) denotes labour supply in period \( i = 1,2 \) and \( w_i \) denotes the wage rate in period \( i = 1,2 \).

Maximization by households of their intertemporal utility with respect to the associated budget constraint gives rise to demand functions for leisure in periods 1 and 2,

\[ v_i = u_i^{-1}(\hat{w}_i) \equiv x_i(\hat{w}_i) \]  

where we introduce the functions \( x_i \) for convenience, \( \hat{w}_1 \) is a shortcut for \( w_1((1+r)/(1+\delta))^2 \) and \( \hat{w}_2 \) is a shortcut for \( w_2((1+r)/(1+\delta)) \).

Throughout the analysis, we will assume that leisure in periods 1 and 2 is strictly between zero and one.

Use \( t_1 \) to denote the implicit tax rate in period 1 and use \( s_2 \) to denote the implicit subsidy rate in period 2. Hence, we can express the wage rate in period 1 as a function of \( t_1 \) and the wage rate in period 2 as a function of \( s_2 \),

\[ w_1 = q_1(1 - t_1) \]  
\[ w_2 = q_2(1 + s_2) \]  

where we use \( q_i \) \( i = 1,2 \) to denote productivity at age \( i \).
Combining things, we can express $U$ as a function of $t_1$ and $s_2$,

\[
U = u_1(x_1(\hat{q}_1(1 - t_1))) + u_2(x_2(\hat{q}_2(1 + s_2))) \\
+ \hat{q}_1(1 - t_1)(1 - x_1(\hat{q}_1(1 - t_1))) + \hat{q}_2(1 + s_2)(1 - x_2(\hat{q}_2(1 + s_2)))
\]  

(14)

where $\hat{q}_1$ is a shortcut for $q_1((1 + r)/(1 + \delta))^2$ and $\hat{q}_2$ is a shortcut for $q_2((1 + r)/(1 + \delta))$.

As demonstrated in section 2, the revenues of the period-1 implicit tax and the spending on the period-2 implicit subsidy are equal if measured in terms of euros at the time of retirement ($T_1 = S_2$). Noting that tax revenues, $T_1$, are defined as $(1 + r)^2q_1t_1l_1$ and subsidies, $S_2$, as $(1 + r)q_2s_2l_2$, we can formulate the following pension fund constraint:

\[
(1 + r)q_1t_1(1 - x_1(\hat{q}_1(1 - t_1))) = q_2s_2(1 - x_2(\hat{q}_2(1 + s_2)))
\]

(15)

Total differentiation of this pension fund constraint gives an expression for $dt_1/ds_2$,

\[
\frac{dt_1}{ds_2} = \frac{q_2l_2(1 + \tilde{s}_2\eta_2)}{(1 + r)q_1l_1(1 - t_1\eta_1)}
\]

(16)

where we add subscript $F$ to refer to the case of a first-best world. Furthermore, $\tilde{s}_2$ and $\tilde{t}_1$ are shorthand notations for $s_2/(1 + s_2)$ and $t_1/(1 - t_1)$ respectively. As will be clear by now, $l_1$ and $l_2$ relate to $t_1$ and $s_2$, so we suppress this in equation (16) and henceforth.

In equation (16), we use $\eta_i i = 1, 2$ to denote the elasticity of labour supply in period $i$ with respect to that period’s wage rate, $w_i$: $\eta_i = -x_i'(\hat{w}_i)/((1 - x_i(\hat{w}_i))/\hat{w}_i)$ $i = 1, 2$. These two labour supply elasticities are positive on account of our assumed utility function.

Note that the numerator of the fraction at the RHS of equation (16) corresponds to the derivative of subsidy outlays to the subsidy rate. Similarly, the denominator of this fraction corresponds to the derivative of tax revenues to the tax rate. The former is always positive. The latter is only positive if the economy is on the left side of its Laffer curve. We will assume that this is the case, which implies that $(dt_1/ds_2)_F > 0$.

Now we are ready to assess the welfare properties of front-loading policies. Note that, as the implicit tax and subsidy rate are related, $dU/ds_2$ must be elaborated as $\partial U/\partial s_2 + \partial U/\partial t_1(dt_1/ds_2)$. Upon using equations (14) and (16), we then derive the following expression for $dU/ds_2$:

\[
\left(\frac{dU}{ds_2}\right)_F = \frac{(1 + r)}{(1 + \delta)^2} q_2 l_2 \left(1 - \frac{1 + \tilde{s}_2\eta_2}{1 - t_1\eta_1}\right)
\]

(17)
Equation (17) allows us to derive two propositions that distinguish between a marginal reform and a discrete (more than marginal) reform. A marginal reform pertains to a marginal increase in $s_2$ (and $t_1$) from the initial point $s_2 = t_1 = 0$. Given that real-world pension policies imply values for $s_2$ and $t_1$ that are very different from zero, this marginal reform has no real-world counterpart. But it is a useful device to understand the discrete reform which does have relevance for real-world pension policies.

**Proposition 1:**
In a first-best world, the introduction of a marginal degree of front-loading in pension contribution policies exerts no effect upon welfare.

**Proof**
The welfare effect of a marginal reform is zero, as can be seen by inspection of equation (17).

**Proposition 2:**
In a first-best world, the introduction of a discrete degree of front-loading in pension contributions is strictly welfare-decreasing.

**Proof**
We view a discrete reform as a succession of marginal changes. The welfare effect of the first marginal change, that is the increase in $s_2$ (and $t_1$) from the initial point $s_2 = t_1 = 0$, is zero (see proposition 1). Let us denote the subsidy rate and tax rate after this first marginal change as $s_2^1$ and $t_1^1$. The second marginal change concerns again a marginal increase in $s_2$ (and $t_1$), but now from the point $s_2 = s_2^1 > 0$ and $t_1 = t_1^1 > 0$. Inspection of equation (17) shows that $\frac{dU}{ds_2} < 0$ if $s_2, t_1 > 0$. The welfare effect of the second change is thus negative and the same holds true for subsequent marginal changes. This proves proposition 2.

Equation (17) shows that the welfare loss due to front-loading is higher, the higher the implicit subsidy rate and implicit tax rate that are associated with front-loading. Furthermore, higher labour supply elasticities for young and old workers contribute to a higher welfare loss.

A question that cannot be answered by equation (17) is how the implicit tax and the implicit subsidy contribute to the welfare loss from front-loading. To answer this question, we split front-loading policies into two hypothetical policies: i) a distortionary tax in period 1 of which the revenues are rebated to households in a lump-sum fashion - to be denoted as H1; and ii)
a distortionary subsidy in period 2 that is financed by imposing lump-sum transfers upon households - to be denoted as H2. We express the welfare losses of these two policies in equations (18) and (19):

\[
\left(\frac{dU}{dt_1}\right)_{H1} = -\eta_1 \tilde{t}_1 l_1 \hat{q}_1
\]

\[
\left(\frac{dU}{ds_2}\right)_{H2} = -\frac{1}{1+\delta} \eta_2 \tilde{s}_2 l_2 \hat{q}_2
\]

These two equations reflect the famous Harberger formula (Harberger, 1964). They indicate that the marginal welfare cost of taxation (subsidization) is proportional to the initial tax (subsidy) rate \(\tilde{t}_1\) (\(\tilde{s}_2\)) and to the associated elasticity \(\eta_1\) (\(\eta_2\)).

We now combine these two types of policies, using the factor \((dt_1/ds_2)_F\) as calculated in equation (16) to ensure that the lump-sum taxes associated with H1 cancel exactly against the lump-sum transfers associated with H2. This produces a welfare loss equal to \((dU/ds_2)_{H1H2}:

\[
\left(\frac{dU}{ds_2}\right)_{H1H2} = \left(\frac{dU}{dt_1}\right)_{H1} \left(\frac{dt_1}{ds_2}\right)_F + \left(\frac{dU}{ds_2}\right)_{H2}
\]

Appendix A shows that this effect is exactly the welfare effect of front-loading as we derived above (equation (17)).

As indicated above, the next section will introduce government spending. But before moving to that section, we take a look at the effect of front-loading upon aggregate labour income. For this effect, measured as \(d(q_1 l_1 + q_2 l_2)/ds_2\), we derive the following form:

\[
\left(\frac{d(q_1 l_1 + q_2 l_2)}{ds_2}\right)_F = -q_2 l_2 \left(\frac{\eta_1}{1-\tilde{t}_1} \left(\frac{1+\tilde{s}_2 \eta_2}{1+\delta}\right) - \frac{\eta_2}{1+\delta}\right)
\]

\[\text{In order to demonstrate the equivalence of equation (18) with the Harberger formula, denote the Harberger formula in our notation: } \frac{dU}{dt_1} = -1/2 \eta_1 \tilde{t}_1 l_1 \hat{q}_1 \text{. This formula gives an expression for the total welfare cost of taxation. The corresponding expression for the marginal welfare cost can be found by differentiation: } \frac{dU}{dt_1} = -\eta_1 l_1 \hat{q}_1 \text{. This expression differs from equation (18) only in the tax term, which is } \tilde{t}_1 \text{ here and } l_1 \text{ in equation (18). The explanation for the difference relates to the subject of analysis. Harberger was interested to know the welfare cost of taxation in general, so assumed an initial tax rate of zero. If we also impose the assumption that the tax rate is initially zero, the expression for } dU/dt_1 \text{ in equation (18) becomes identical to that for } \frac{dU}{dt_1} \text{ above.} \]
This tells us three things. First, the effect upon aggregate labour income can generally not be signed as it combines the opposite effects upon the labour income of young workers (negative) and of old workers (positive). Second, in case the reform is marginal, the effect in equation (21) is proportional with $(\eta_1/(1 + r) - \eta_2)$. Without restricting the values of $\eta_1$, $\eta_2$ and $r$, this effect can also not be signed. Thirdly, the higher the initial values of $t_1$ and $s_2$, the larger the weight of the negative term at the RHS of equation (21), hence the less positive or the more negative will be the effect of front-loading upon aggregate labour income.

4 General taxation

The previous section analysed the effects of front-loading in a first-best world: apart from front-loading, no other distortions prevailed. This section adds a distortion to the model. In particular, we assume that the government levies labour income taxes in order to finance a certain amount of government spending. Aggregate labour income serve as the tax base. As we will see, addition of labour income taxation which is common in many countries may turn the conclusions of the previous section on its head.

Let us use $p$ to denote the rate of general taxation and $P$ to denote the financing requirement of the government. The expressions for the wage rate in period 1 and 2 are natural generalizations of the corresponding equations in the previous section:

$$w_1 = q_1(1 - p - t_1)$$  \hspace{1cm} (22)  \\
$$w_2 = q_2(1 - p + s_2)$$  \hspace{1cm} (23)

Different from the previous section, the pension fund constraint in equation (15) is now a function of three policy variables, namely $t_1$, $s_2$ and $p$. Total differentiation yields the following,

$$(1 + r)q_1l_1(1 - \eta_1\tilde{t}_1)dt_1 - q_2l_2(1 + \eta_2\tilde{s}_2)ds_2$$  \hspace{1cm} (24)  \\
$$- (\eta_2\tilde{s}_2l_2q_2 + \eta_1\tilde{t}_1l_1(1 + r)q_1)dp = 0$$

where $\tilde{t}_1$ is a shortcut for $t_1/(1 - p - t_1)$ and $\tilde{s}_2$ is a shortcut for $s_2/(1 - p + s_2)$.

Next to this pension fund constraint, we now have a government budget constraint, which relates the labour income tax rate to the financing requirement $P$:

$$P = p(q_1l_1 + q_2l_2)$$  \hspace{1cm} (25)
Total differentiation of this government budget constraint gives a second equation in the three policy variables,

\[-q_1 l_1 \ddot{\tilde{p}}_1 \eta_1 dt_1 + q_2 l_2 \ddot{\tilde{p}}_2 \eta_2 ds_2 \]
\[+ \left( (1 - \ddot{\tilde{p}}_1 \eta_1) q_1 l_1 + (1 - \ddot{\tilde{p}}_2 \eta_2) q_2 l_2 \right) dp = 0 \tag{26} \]

where \( \ddot{\tilde{p}}_1 \) is a shortcut for \( p/(1 - \tilde{p}_1 \eta_1) \) and \( \ddot{\tilde{p}}_2 \) is a shortcut for \( p/(1 - \tilde{p}_2 \eta_2) \).

Combining equations (24) and (26) gives expressions for \( dt_1/ds_2 \) and \( dp/ds_2 \),

\[
\left( \frac{dt_1}{ds_2} \right)_S = \left( \frac{q_2 l_2}{(1 + r)q_1 l_1} \right) \Omega_B \tag{27} \\
\left( \frac{dp}{ds_2} \right)_S = \frac{q_2 l_2 \left( \frac{\ddot{\tilde{p}}_1 \eta_1}{(1 + r) \Omega_B - \ddot{\tilde{p}}_2 \eta_2} \right)}{(1 - \ddot{\tilde{p}}_1 \eta_1) q_1 l_1 + (1 - \ddot{\tilde{p}}_2 \eta_2) q_2 l_2} \tag{28} \\
\]

where

\[
\Omega_B \equiv \left( \frac{(1 + \ddot{s}_2 \eta_2) - \Omega_A \ddot{\tilde{p}}_2 \eta_2}{(1 - \ddot{t}_1 \eta_1) - \Omega_A \ddot{\tilde{p}}_1 \eta_1} \right) \tag{29} \\
\]

and

\[
\Omega_A \equiv \left( \frac{(1 + r)q_1 l_1 \ddot{\tilde{t}}_1 \eta_1 + q_2 l_2 \ddot{\tilde{s}}_2 \eta_2}{(1 - \ddot{\tilde{p}}_1 \eta_1) q_1 l_1 + (1 - \ddot{\tilde{p}}_2 \eta_2) q_2 l_2} \right) \tag{30} \\
\]

and where the subscript \( S \) refers to the casus of a second-best world.

To evaluate the welfare effect of front-loading in this second-best model, we derive an expression for \( dU/ds_2 \), which must now be elaborated as \( \partial U/\partial s_2 + \partial U/\partial t_1 (dt_1/ds_2) + \partial U/\partial p(dp/ds_2) \),

\[
\left( \frac{dU}{ds_2} \right)_S = \left( \frac{1 + r}{1 + \delta} \right)^2 q_2 l_2 (1 - \Omega_B) - \left( \frac{(q_1 l_1 (1 + r)^2 + q_2 l_2 (1 + r)) q_2 l_2}{(1 + \delta)^2 ((1 - \ddot{\tilde{p}}_1 \eta_1) q_1 l_1 + (1 - \ddot{\tilde{p}}_2 \eta_2) q_2 l_2)} \right) \left( \frac{\ddot{\tilde{p}}_1 \eta_1}{(1 + r) \Omega_B - \ddot{\tilde{p}}_2 \eta_2} \right) \tag{31} \\
\]

where \( \Omega_B \) is defined in equation (29).

**A benchmark case**

We first evaluate the welfare effect in equation (31) for a particular case, defined by age-independent labour supply elasticities and a zero interest
rate: $\eta_1 = \eta_2$, $r = 0$. The case is not very realistic as we will explain below, but is useful as a benchmark case. As before, we look at a marginal reform and a discrete reform and state our results in the form of two propositions.

**Proposition 3:**
In the benchmark case of positive public spending ($P > 0$, $\eta_1 = \eta_2$, $r = 0$), the introduction of a marginal degree of front-loading in pension contribution policies exerts no effect upon welfare.

**Proof**
Note that $s_2 = t_1 = 0$ implies that $\tilde{\delta}_2 = \tilde{t}_1 = 0$. Hence, $\Omega_A = 0$ and $\Omega_B = 1$. Given that $\eta_1 = \eta_2$ and $r = 0$, we find that $dU/ds_2$ equals zero.

This result is identical to that in proposition 1: the addition of public spending exerts no effect under the assumptions of the benchmark case.

**Proposition 4:**
In the benchmark case of positive public spending ($P > 0$, $\eta_1 = \eta_2$, $r = 0$), a discrete degree of front-loading in pension contribution policies is strictly welfare-reducing.

**Proof**
As in the proof of proposition 2, we view a discrete policy reform as a succession of marginal changes. The welfare effect of the first marginal change, that is the increase in $s_2$ (and $t_1$) from the initial point $s_2 = t_1 = 0$, is zero (see proposition 3). The second and subsequent marginal changes have $s_2, t_1 > 0$ as a starting point. Combined with the definitions of $\tilde{t}_1$, $\tilde{s}_2$, $\tilde{p}_1$ and $\tilde{p}_2$, equations (30) and (29) show that, if $s_2, t_1 > 0$, $\Omega_A > 0$, $\Omega_B > 1$, $\tilde{p}_1 > \tilde{p}_2$, and, thus, $dU/ds_2 < 0$. This proves proposition 4.

An additional result emerges if we compare the benchmark case with positive public spending with the previous case without any public spending. We phrase this in a corollary to proposition 4.

**Corollary to proposition 4:**

\[\text{Here, we make the assumption that the denominator of the expression for } \Omega_A \text{ in equation (30) is positive. This term is the derivative of labour income tax revenues to the labour income tax rate. Therefore, the assumption that the denominator of the expression for } \Omega_A \text{ is positive boils down to the assumption that the economy is on the left side of the relevant Laffer curve.}\]
A discrete degree of front-loading in pension contribution policies is more welfare-reducing in the benchmark case of positive public spending \((P > 0, \eta_1 = \eta_2, r = 0)\) than in the case without any public spending.

**Proof**

In case \(s_2 > 0\) and \(t_1 > 0\), \(\Omega_B\) in equation (29) is strictly higher than \((1 + \tilde{s}_2 \eta_2)/(1 - \tilde{t}_1 \eta_1)\) in equation (17). Hence, compared to the case of zero public spending, the first term at the RHS of equation (31) is now more negative. In addition, the second term at the RHS of equation (31) is now negative, whereas it is zero in case of zero public spending.

The explanation for the latter result can be found in the effect upon aggregate labour income,

\[
\left( \frac{d(q_1 l_1 + q_2 l_2)}{d s_2} \right)_S := -q_2 l_2 \left( \frac{\eta_1}{(1 + r)(1 - p - t_1)} \Omega_B - \frac{\eta_2}{1 - p + s_2} \right) \\
- \left( \frac{\eta_1 q_1 l_1}{1 - p - t_1} + \frac{\eta_2 q_2 l_2}{1 - p + s_2} \right) \left( \frac{d p}{d s_2} \right)_S
\]

(32)

where \((dp/ds_2)_S\) is given in equation (28).

Equation (32) shows that in the benchmark case of positive public spending \((P > 0, \eta_1 = \eta_2, r = 0)\) a marginal degree of front-loading has a zero effect on aggregate labour income (use equation (28) to show that \(dp/ds_2 = 0\) in case \(s_2 = t_1 = 0\)). A discrete degree of front-loading produces an unambiguous decline of aggregate labour income however. Indeed, both terms at the RHS of equation (32) turn negative if we move from a marginal to a discrete degree of front-loading (use equation (28) to show that \(dp/ds_2 > 0\) in case \(s_2, t_1 > 0\)). Through the government budget constraint, a decline of aggregate labour income necessitates an increase in the labour income tax rate. This adds to the welfare loss due to front-loading and explains why the welfare loss is now bigger than in the case of zero public spending.

**A more realistic case**

A rather different story emerges once we depart from the rather specific assumptions of the benchmark case. In case of a marginal reform, the welfare effect now reads as follows,

\[
\left( \frac{d U}{d s_2} \right)_{S,M} = -\frac{q_2 l_2}{(1 + \delta)^2} \left( \frac{q_1 l_1 (1 + r)^2 + q_2 l_2 (1 + r)}{((1 - \tilde{p} \eta_1) q_1 l_1 + (1 - \tilde{p} \eta_2) q_2 l_2)} \right) \beta \left( \frac{\eta_1}{1 + r} - \eta_2 \right)
\]

(33)
where we use index $M$ to denote that the effect refers to a marginal reform. Further, we use $\tilde{p}$ as a shortcut for $p/(1 - p)$.

The welfare effect in equation (33) is proportional with $\eta_1/(1 + r) - \eta_2$. Without any information on labour supply elasticities or the interest rate, it can be positive or negative. Empirical evidence is helpful to pin down the sign of $\eta_1/(1 + r) - \eta_2$, however.

Econometric studies find that labour supply elasticities increase strongly with age. In particular, French (2005) estimates that the labour supply elasticity is 0.3 for a 40-year old worker, but 1.1 for a worker of 60 years old. Fenge et al. (2006) finds that the labour supply elasticity rises with age: compensated elasticities are estimated at 0.010 for males aged 20-39 and 0.215 for males aged 40-59 (in contrast, for females the estimates do not differ much with respect to age). The figures reported in French and Jones (2012) are close to those in French (2005): the estimated labour supply elasticity is 0.17 for 40-year old workers and 1.17 for workers with age 60. Erosa et al. (2016) adopt a structural labour supply model that accounts for decision-making along the intensive and the extensive margin to calculate labour supply elasticities. They report that these elasticities are a U-function of age and highest for the oldest age group considered. For example, for high school individuals, they report elasticities of 2.01, 1.62, 1.90 and 2.74 for the 25-34, the 35-44, the 45-54 and the 55-61 age group respectively.

As regards the interest rate, historically real interest rates are generally positive. Although negative real interest rates cannot be excluded, they are a bit unlikely for the unit period of twenty years that we are concerned with here.

Hence, we define as a more realistic case the case $\eta_1/(1 + r) - \eta_2 < 0$. Equation (33) shows that in this case a marginal degree of front-loading is strictly welfare-increasing. We summarize this in proposition 5.

**Proposition 5:**

In the more realistic case of positive public spending ($P > 0$, $\eta_1/(1+r) - \eta_2 < 0$), a marginal degree of front-loading in pension contribution policies is strictly welfare-increasing.

**Proof**

This follows immediately [check! fn 5] from equation (33).

This effect reverses the effects above for the case of zero spending and the benchmark case of positive public spending. The reason for the first-order welfare loss of front-loading lies in the effect upon aggregate labour income.
To see this, we write down the effect upon aggregate labour income in case of a marginal degree of front-loading:

\[
\frac{d(q_1l_1 + q_2l_2)}{ds_2}_{S,M} = \left( -\frac{q_2l_2}{1 - \tilde{p}} \left( 1 + \frac{(\eta_2q_2l_2 + \eta_1q_1l_1)\tilde{p}}{(1 - \tilde{p}\eta_1)q_1l_1 + (1 - \tilde{p}\eta_2)q_2l_2} \right) \left( \frac{\eta_1}{(1 + r)} - \eta_2 \right) \right)
\]

As this equation shows, front-loading produces a first-order loss of aggregate labour income. Through the government budget constraint, it thus generates a first-order increase of the labour income tax rate. This fully explains the first-order effect of front-loading upon welfare, for, as we have seen in the previous section, the welfare effect of a marginal degree of front-loading that is due to the implicit pension tax and subsidy itself is zero.

An empirically very interesting case is the discrete equivalent of the case considered in proposition 5, i.e. with positive public spending and \(\eta_1/(1 + r) - \eta_2 < 0\). We may expect that increasing the degree of front-loading from zero to increasingly higher levels will reduce the welfare gain and may eventually turn the welfare gain into a welfare loss. For in the case of zero public spending and the benchmark case of positive public spending, moving from a marginal to a discrete reform turns a zero welfare effect into a welfare loss. Unfortunately, the expression in equation (31) is too complex to help us to verify this expectation. Therefore, we resort to numerical simulations for an illustration of the effects in this case.

The numerical calculations assume that \(r = 0.3478\). This is based on an annual interest rate of 1.5 percent and a unit period of 20 years. We assume the discount rate equal to the interest rate: \(\delta = r\). Furthermore, we take productivity to be a constant: \(q_1 = q_2 = 1\). For \(u_i = 1,2\) we adopt the specification \(-l_1^{1+1/\eta_i} / (1 + 1/\eta_i)\), which features constant labour supply elasticities. For \(\eta_l\) we use a value of 0.3. This is close to the value of 0.31 reported in Keane (2011) for a simple average of Hicks elasticities across male and female labour supply literatures.

The simulations highlight the role of \(\eta_2\), \(P\) and \(s_2\). For each of them, we adopt four different values. \(\eta_2\) is taken to be equal to \(\eta_1/(1 + r)\) (the benchmark case), \(\eta_1\), \(2\eta_1\) or \(3\eta_1\). These values for \(\eta_2\) are still at the lower end of estimates for \(\eta_2\) in empirical research (French, 2005; French and Jones, 2012). Taking even higher values would not change our results, as we will see below. For \(P\) we take a wide range: the values are 0, 0.2, 0.4 and 0.6. To choose values for \(s_2\) is somewhat more complicated as this implicit subsidy rate cannot be observed. We pursue as follows. We first note that
Table 1: The impact of front-loading on welfare

\[ \eta_2 = \eta_1 / (1 + r) = 0.2226 \]

\[
\begin{array}{|c|c|c|c|}
\hline
\downarrow s_2(\%) & P \rightarrow & 0.0 & 0.2 & 0.4 & 0.6 \\
3 & - & - & - & - \\
6 & - & - & - & - \\
9 & - & - & - & - \\
\hline
\end{array}
\]

\[ \eta_2 = \eta_1 = 0.3 \]

\[
\begin{array}{|c|c|c|c|}
\hline
\downarrow s_2(\%) & P \rightarrow & 0.0 & 0.2 & 0.4 & 0.6 \\
3 & - & + & + & + \\
6 & - & + & + & + \\
9 & - & - & + & + \\
\hline
\end{array}
\]

\[ \eta_2 = 1.5\eta_1 = 0.45 \]

\[
\begin{array}{|c|c|c|c|}
\hline
\downarrow s_2(\%) & P \rightarrow & 0.0 & 0.2 & 0.4 & 0.6 \\
3 & - & + & + & + \\
6 & - & + & + & + \\
9 & - & - & + & + \\
\hline
\end{array}
\]

\[ \eta_2 = 2\eta_1 = 0.6 \]

\[
\begin{array}{|c|c|c|c|}
\hline
\downarrow s_2(\%) & P \rightarrow & 0.0 & 0.2 & 0.4 & 0.6 \\
3 & - & + & + & + \\
6 & - & + & + & + \\
9 & - & + & + & + \\
\hline
\end{array}
\]

\( s_2 \) corresponds to the derivative of the implicit transfer due to front-loading policies with respect to labour income. This is measured by \( S_2/y_2 \) (see equation (8)). As we have seen, this variable is a function of \( r, g, \mu \) and \( \beta \). We now calculate the value of \( S_2/y_2 \) for a range of scenarios for \( g, r \) and \( \mu \): \( g \) and \( \mu \) take annual values of 0 or 1 percent and \( r \) values of 1, 2 or 3 percent. In all cases, we set \( \beta \), the pension benefit to labour income ratio, equal to
Excluding the scenarios that violate $r > \mu$, we then have 10 scenarios. Despite their variety, the outcomes for $S_2/y_2$ lie in a pretty small range. The minimum value is 5.34 percent and the maximum value 9.33 percent. Should we adopt a lower value for $\beta$, the outcomes would be proportionally lower. Obviously, this procedure can only give a very rough indication of the order of magnitude of the value for $S_2/y_2$ in real-world pension schemes, but is helpful in our case. Based on this, we choose values for $s_2$ in our simulations running from 0 to 9 percent.

Table 1 shows the results. It displays qualitative results, i.e. it displays whether a particular policy of front-loading (value of $s_2$) in a particular scenario (values of $\eta_2$ and $P$) implies higher welfare (+) or lower welfare (−) than the same scenario without front-loading ($s_2 = 0$). Appendix B contains the corresponding numerical results. The results in appendix B confirm the results derived above on the welfare losses of front-loading in the case of zero public spending and the benchmark case of positive public spending. In particular, in the case of zero public spending the welfare loss is larger, the higher the initial subsidy rate ($s_2$) and the more elastic the labour supply of older workers ($\eta_2$); in the benchmark case of positive public spending the welfare loss is larger, the higher the initial subsidy rate ($s_2$) and the larger is public spending ($P$).

Turning then to the qualitative results in table 1, we see that they confirm our earlier results and expectations. In the benchmark case where $\eta_1/(1 + r) - \eta_2 = 0$ (the upper block in table 1), front-loading implies a welfare loss, irrespective the values of $s_2$ and $P$. In case of a uniform elasticity, i.e. $\eta_1 = \eta_2$ (the second block in table 1), front-loading is also negative in most cases. Only if public spending is relatively large and the degree of front-loading relatively small, is front-loading welfare-increasing. The reason is clear: the cases in which front-loading implies a welfare gain combine a relative large initial distortion due to the labour income tax with a small distortion due to the implicit pension tax and subsidy. The case of a higher elasticity, i.e. $\eta_2 = 1.5\eta_1$ (the third block in table 1), is similar to the previous one. But given that this case features a higher elasticity for older workers than the previous one, most of the entries are now positive, indicating a welfare gain. The case of an elasticity for older workers that is double the elasticity for younger workers (the lower block in table 1) is even more outspoken. Now, all the entries are positive, except for the cases with zero government spending in which case there is no initial labour supply distortion.

There is a close connection between our results and those of the literature on optimal taxation. The latter states that optimal policies apply
relatively low tax rates on goods for which demand is relatively elastic. According to the empirical estimates referred to above, the demand for leisure by older workers is more elastic than that by younger workers. This suggests that younger workers should be taxed more heavily than older workers. The labour income tax in our model does not distinguish between young and old workers, however. A pension scheme with linear contribution and accrual rates can thus fill the gap that is left by tax policies. By implicitly differentiating tax rates with respect to age, these pension schemes can bring the economy closer to the first best and achieve a welfare gain. Be aware of the word ‘can’, however. Front-loading can also produce too much of a difference between tax rates on young and old workers and produce a welfare loss, as is nicely illustrated in table 1.8

5 The case of a collective pension scheme

Throughout the paper, we made an assumption that may have surprised one at first sight. We regard the pension scheme as an individual scheme, whereas front-loading policies are characteristic of collective schemes. Should the differences between individual and collective schemes affect any of our results, we would have to conclude that our approach was inappropriate. In this section, I will use the model without government spending to demonstrate that this is not the case: the welfare result derived for the simple case of an individual scheme simply carries over to the more complex case of a collective scheme.

Let us adopt then the model in section 3 but replace the individual scheme with a collective scheme. The analysis of the collective pension scheme itself is relegated to appendix C. There, it is shown that the transformation to a collective scheme implies a change in the relation between implicit tax revenues and implicit subsidies, \( T_1/S_2 \). In section 3, the ratio between the two variables equalled one. As appendix C shows, in the case of a collective scheme this ratio equals \( (1+r)/(1+g) \). In the case of a collective scheme, we need to impose dynamic efficiency in order to be able to account

\[\text{In order to find out what is the first-best amount of front-loading, one can rewrite the optimal tax problem as one with } \tau_1 \text{, defined as } p + t_1 \text{, and } \tau_2 \text{, defined as } p - s_2 \text{, as instruments. In order to be able to integrate the budget constraints of the pension fund and the government, we here assume } r = 0. \text{ The optimal tax rates } \tau_1 \text{ and } \tau_2 \text{ then follow from combining the Ramsey rule, } (\tau_1/(1 - \tau_1))/(\tau_2/(1 - \tau_2)) = \eta_2/\eta_1 \text{ with the corresponding budget constraint } \tau_1(1 - \tau_1)^{\eta_1} + \tau_2(1 - \tau_2)^{\eta_2} = P. \text{ Combining the optimal rates } \tau_1 \text{ and } \tau_2 \text{ with the labour income tax rate } p \text{ then gives the values of } t_1 \text{ and } s_2 \text{ that characterize the first-best amount of front-loading.} \]
for the interests of future generations, i.e. $r > g$. Hence, $T_1/S_2 > 1$ in case of a collective scheme. The balance between tax revenues and subsidies is less attractive for the workers that are covered all their working lives by a collective scheme. As appendix C clarifies, this reflects an implicit debt inherited from the time the scheme was started.

Hence, the pension fund constraint, which governs the relation between changes in the implicit tax rate $t_1$ and the implicit subsidy rate $s_2$, reads as follows in the case of a collective scheme,

$$\left(\frac{dt_1}{ds_2}\right)_F^C = - \left(\frac{T_1^C}{S_2^C}\right) \left(\frac{q_2 l_2 (1 + \tilde{s}_2 \eta_2)}{(1 + r) q_1 l_1 (1 - \tilde{t}_1 \eta_1)}\right) = \left(\frac{1 + r}{1 + g}\right) \left(\frac{q_2 l_2 (1 + \tilde{s}_2 \eta_2)}{(1 + r) q_1 l_1 (1 - \tilde{t}_1 \eta_1)}\right)$$

(35)

where we use superscript $C$ to refer to the collective scheme.

The corresponding expression for the welfare effect differs from the one in section 3 only in the factor $(1 + r)/(1 + g)$:

$$\left(\frac{dU_1}{ds_2}\right)_F^C = \left(\frac{(1 + r)}{(1 + \delta)^2} q_2 l_2 \left(1 - \left(\frac{1 + r}{1 + g}\right) \left\{\frac{1 + \tilde{s}_2 \eta_2}{1 - \tilde{t}_1 \eta_1}\right\}\right)\right)$$

(36)

Note that we now index $U$ to refer to the period in which the generation was born.

From the expression in equation (36) one might conclude that there is a difference between the collective scheme and the individual scheme. For in the present case, front-loading implies a first-order welfare loss (recall that $r > g$). But note that this expression refers to a steady-state generation, in particular the generation who is young when the scheme is started. In the case of a collective scheme this generation is not representative of the whole population.

This brings us to the second change. To do justice to the interests of all generations, we need to define a social welfare function that includes the interests of all generations involved, i.e. old, young and future generations:

$$W = (1 + r) U_0 + U_1 + \frac{1}{1 + r} U_2 + ...$$

(37)

This social welfare function uses the interest rate as discount rate. This is necessary if one wants to avoid that redistribution between generations will affect social welfare.
The welfare effect of a marginal increase in the degree of front-loading can be derived by elaborating $dW/ds_2$:

$$(dW/ds_2)_F = (1 + r)\left(\frac{dU_0}{ds_2}\right)_F + \frac{1}{1 + r} \left(\frac{dU_1}{ds_2}\right)_F + ...$$

$$= \left(\frac{1 + r}{1 + g}\right) \left(\frac{\partial U}{\partial s_2}\right) + \left(\frac{1 + r}{r - g}\right) \left(\frac{dU_1}{ds_2}\right)_F$$

(38)

The second line of equation (38) reflects that the welfare effect of the transition generation relates only to the change in $s_2$, not the change in $t_1$. This effect equals the partial effect $\partial U/\partial s_2$ that we already derived in section 3. Next, the second line reflects that the welfare effects of successive steady-state generations are identical, up to a proportionality factor $1 + g$.

Elaborating the expression by substituting the expressions for $\partial U/\partial s_2$ and $(dU_1/ds_2)_F$ yields the following:

$$(dW/ds_2)_F = \left(\frac{1 + r}{1 + g}\right) \left(\frac{1 + r}{1 + \delta}\right)^2 q_2 l_2$$

$$+ \left(\frac{1 + r}{r - g}\right) \left(\frac{1 + r}{1 + \delta}\right)^2 q_2 l_2 \left(1 - \frac{1 + r}{1 + g} \left\{\frac{1 + \tilde{s}_2 \eta_2}{1 - \tilde{t}_1 \eta_1}\right\}\right)$$

$$= \frac{(1 + r)}{(1 + \delta)^2} q_2 l_2 \left(\frac{1}{1 + g} + \frac{1}{r - g}\right) \left(1 - \left\{\frac{1 + \tilde{s}_2 \eta_2}{1 - \tilde{t}_1 \eta_1}\right\}\right)$$

(39)

Comparing this expression with the one derived for the case of an individual scheme in section 3, we see that the two are identical, except for a proportionality factor $1/(1 + g) + 1/(r - g)$, which reflects that the collective scheme accounts for the interests of all generations alive and those yet to be born.

This shows that front-loading policies work out the same for the case of a collective scheme and that of an individual scheme. The reason is that the only addition of a collective scheme to an individual scheme is a windfall gain enjoyed by the transition generation and paid for by the steady-state generations. In other words, the collective scheme adds redistribution between the transition generation and the steady-state generations. If the social discount rate equals the interest rate, this cannot change the welfare implications of front-loading policies.

6 Qualifications

As always, our analysis rests on assumptions of which some can be debated. One example is that workers are rational and forward-looking decision-makers who not only have access to all relevant information but also are
capable to interpret it. There is ample evidence that people are myopic (Van Rooij et al., 2012) and unable to interpret all relevant information (Lusardi and Mitchell, 2014), however. If we take this to the extreme, we could make the assumption that labour-supply decisions are unaffected by the implicit pension taxes and subsidies that we have explored. In this extreme case, our results would disappear.

We argue that this alternative assumption is too extreme, however. It conflicts with a bunch of empirical evidence that taxes affect labour supply behaviour. It also conflicts with evidence on retirement behaviour. For example, Kotlikoff and Wise (1988) and Samwick (1998) find that the accrual rate of retirement wealth affects the probability of retirement. Similarly, Chan and Stevens (2004), Asch et al. (2005), Coile and Gruber (2007), Mastrobuoni (2009), Euwals et al. (2010) and Hanel (2010) conclude that financial incentives determine the retirement behaviour of older workers. One way to reconcile the empirical evidence on retirement behaviour and that on the lack of rational, forward-looking behaviour is to assume that people heavily discount the future (Frederick et al., 2002), so that only the labour supply of older workers is affected by pension taxes. As one can easily verify, this would not change our basic results; it would make the case for a welfare gain of front-loading more likely.

Another debatable assumption is that the labour market is in equilibrium: the labour market accommodates any change in labour supply. In general, a model without unemployment is too simple to be true. It goes well beyond the scope of this paper to study the case of a disequilibrium labour market, but we can make a few observations. If demand and supply deviate, changes in the labour supply of younger and older workers would not necessarily imply changes in employment, but could translate into changes in unemployment. Hence, front-loading could imply lower unemployment for younger workers and higher unemployment for older workers. Still, essentially our results could remain unchanged. In particular, if some fraction of the labour supply is unemployed, as suggested by equilibrium unemployment theory (Pissarides, 2000), then changes in labour supply would continue to have an effect on employment and taxable income, although the effects would be smaller than we have calculated.

7 Concluding remarks

Our analysis has shown that the move from a scheme of linear accruals to one of degressive accruals that the Dutch government aims to make may be
welfare-decreasing. The reason is the loss of a policy instrument. Indeed, the reform implies that the taxation of labour income no longer differentiates with respect to age. This then also suggests a remedy: by differentiating the labour income tax with respect to age, the adverse effect studied here might be neutralized or, at least, mitigated. Such a reform of the tax scheme could actually improve welfare, as it would allow the government to fine its tax instruments to labour supply elasticities, which is much more difficult to organize through pension accruals.

References


Figure 1: Pension contributions and accruals over the working phase of the life cycle
Figure 2: Annual and cumulative implicit pension taxes over the working phase of the life cycle
Appendix A

Let start to repeat equation (20):

\[
\left( \frac{dU}{ds_2} \right)_{H1H2} = -\eta_1 \tilde{t}_1 l_1 \hat{q}_1 \left( \frac{q_2 l_2}{(1 + r)q_1 l_1} \right) \left( \frac{1 + \eta_2 \tilde{s}_2}{1 - \eta_1 \tilde{t}_1} \right) - \frac{\eta_2 \tilde{s}_2 l_2 \hat{q}_2}{1 + \delta}
\]

Given that we defined \( \hat{q}_1 \) as \( q_1((1 + r)/(1 + \delta))^2 \) and \( \hat{q}_2 \) as \( q_2((1 + r)/(1 + \delta)) \), we can rewrite this as follows:

\[
\left( \frac{dU}{ds_2} \right)_{H1H2} = \frac{(1 + r)}{(1 + \delta)^2} q_2 l_2 \left( -\eta_1 \tilde{t}_1 \left( \frac{1 + \eta_2 \tilde{s}_2}{1 - \eta_1 \tilde{t}_1} \right) - \eta_2 \tilde{s}_2 \right)
\]

\[
= \frac{(1 + r)}{(1 + \delta)^2} q_2 l_2 \left( 1 - (1 + \eta_2 \tilde{s}_2) - \frac{\eta_1 \tilde{t}_1}{1 - \eta_1 \tilde{t}_1}(1 + \eta_2 \tilde{s}_2) \right)
\]

\[
= \frac{(1 + r)}{(1 + \delta)^2} q_2 l_2 \left( 1 - \left( \frac{1 + \eta_2 \tilde{s}_2}{1 - \eta_1 \tilde{t}_1} \right) \right)
\]

The last line is the expression for the welfare effect of front-loading as derived in the main text (equation (17)).

Appendix B

This appendix gives the numerical results that correspond to the results displayed in table 1.

Appendix C

We adopt a model for a collective scheme that is identical to an individual scheme, except on one point: the pension contribution rate and accrual rate are the same for all participants. This implies that all working generations who entered the labour market before the start of the collective scheme, will enjoy a windfall gain as their life-cycle pension contributions will be lower than the pension rights they accrue. Let us denote these generations as transition generations. Figure 2 is illustrative in this respect. The windfall gain for a generation of age \( i \) equals the cumulation of implicit taxes at age \( i \), as entry at age \( i \) means that these implicit taxes will be foregone. These windfall gains have to be paid for by those generations that live all their life under the collective scheme, which we will denote as steady-state generations. The redistribution between transition generations and steady-state generations is identical with the redistribution that the start of a pay-as-you-go scheme implies. Such a scheme also carries an implicit debt that is to be paid for by workers till infinity (Sinn, 2000).
Table 2: The impact of front-loading on welfare

<table>
<thead>
<tr>
<th>$\eta_2 = \eta_1/(1 + r)$ = 0.2226</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\downarrow s_2(%)$</td>
<td>$P \rightarrow$</td>
</tr>
<tr>
<td>0</td>
<td>1.3761</td>
</tr>
<tr>
<td>3</td>
<td>1.3759</td>
</tr>
<tr>
<td>6</td>
<td>1.3755</td>
</tr>
<tr>
<td>9</td>
<td>1.3747</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\eta_2 = \eta_1 = 0.3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\downarrow s_2(%)$</td>
<td>$P \rightarrow$</td>
</tr>
<tr>
<td>0</td>
<td>1.3400</td>
</tr>
<tr>
<td>3</td>
<td>1.3398</td>
</tr>
<tr>
<td>6</td>
<td>1.3392</td>
</tr>
<tr>
<td>9</td>
<td>1.3383</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\eta_2 = 1.5\eta_1 = 0.45$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\downarrow s_2(%)$</td>
<td>$P \rightarrow$</td>
</tr>
<tr>
<td>0</td>
<td>1.2809</td>
</tr>
<tr>
<td>3</td>
<td>1.2807</td>
</tr>
<tr>
<td>6</td>
<td>1.2806</td>
</tr>
<tr>
<td>9</td>
<td>1.2788</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\eta_2 = 2\eta_1 = 0.6$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\downarrow s_2(%)$</td>
<td>$P \rightarrow$</td>
</tr>
<tr>
<td>0</td>
<td>1.2329</td>
</tr>
<tr>
<td>3</td>
<td>1.2327</td>
</tr>
<tr>
<td>6</td>
<td>1.2318</td>
</tr>
<tr>
<td>9</td>
<td>1.2304</td>
</tr>
</tbody>
</table>

It is easy to see that the collective scheme thus differs from the individual scheme in the pension fund constraint. We derive this pension fund
constraint as follows. Just like the individual scheme, aggregate pension contributions must equal aggregate accruals (and aggregate benefits), if we measure everything in euros at the same point in time. Hence, aggregate excess contributions of steady-state generations must finance the aggregate excess accruals of transition generations.

To make this concrete, let us start the scheme start at 1. Given that people spend $T_R$ years in the working phase, the transition phase will take $T_R - 1$ years. Year $T_R$ then marks the beginning of the steady-state phase of the scheme. The pension fund constraint now equates the excess accruals of the generations born in years 1 to $T_R - 1$ to the excess contributions of the generations born in year $T_R$ and thereafter, where all items are discounted to $T_R$,

$$\sum_{j=1}^{T_R-1} \left[ \sum_{i=1}^{j} (\alpha(1+\mu)^{j-i} - \pi(1+r)^{j-i})y_i \right] (1+r)^{T_R-j}$$

$$= \sum_{i=T_R}^{\infty} \left( \Pi_{T_R} - A_{T_R} \right) \left( \frac{1+g}{1+r} \right)^{-T_R+i}$$

(40)

and where $\Pi_{T_R}$ and $A_{T_R}$ are defined in equations (1) and (3). In order for the infinite sum of utilities of steady-state generations to converge, we need to assume that $r > g$, which implies dynamic efficiency (see section 2).

As mentioned above, we take the accrual rate in the collective scheme equal to that in the individual scheme. Hence, equation (40) can be elaborated to calculate the pension contribution rate in the collective scheme. Let us instead focus on the $T_R = 2$ case however, as this is all we need here.

Equation (40) simplifies considerably in the $T_R = 2$ case:

$$\alpha - \pi y_1 (1+r)$$

$$= \left( \alpha y_1 (1+\mu) + \alpha y_2 \right) \left( \frac{1+r}{r-g} \right)$$

(41)

Elaboration of equation (41) gives the following expression for the pension contribution rate,

$$\pi^C = \alpha \left( \frac{(r-g) + (2+\mu+g)}{(r-g) + (2+r+g)} \right)$$

(42)

where we have used index $C$ to refer to the collective scheme.

Equation (42) shows that the pension contribution rate is below the accrual rate, like in the individual scheme (recall that $r > \mu$ and $r > g$).
The front-loading of pension contributions is thus preserved when we move to the case of a collective scheme. But the contribution rate in the collective scheme is higher than the one in the individual scheme, reflecting the implicit debt that features the collective scheme. That this is so can be seen by a simple comparison. Rewrite the pension contribution rate in the individual scheme as specified in equation (4) for the case $T_R = 2$:

$$\pi = \alpha \left( \frac{2 + \mu + g}{2 + r + g} \right)$$  \hspace{1cm} (43)

Comparing the expression for $\pi$ in equation (43) with that for $\pi^C$ in equation (42) shows that $\pi^C > \pi$.

For implicit taxes and subsidies, we can derive the following expressions:

$$T^C_1 \equiv -(\partial Z_2 / \partial y_1)y_1$$
$$= -\alpha y_1 \left( (1 + \mu) - (1 + r) \left( \frac{(r - g) + (2 + \mu + g)}{(r - g) + (2 + r + g)} \right) \right)$$  \hspace{1cm} (44)

$$S^C_2 \equiv (\partial Z_2 / \partial y_2)y_2$$
$$= \alpha y_1 (1 + g) \left( 1 - \left( \frac{(r - g) + (2 + \mu + g)}{(r - g) + (2 + r + g)} \right) \right)$$  \hspace{1cm} (45)

It is easy to derive that $T^C_1$ is strictly larger and $S^C_2$ strictly smaller than $S_2$ in equation (8), as one would expect, given that $\pi^C > \pi$.

An expression for the ratio of the two cash flows is remarkably simple:

$$\frac{T^C_1}{S^C_2} = \left( \frac{1 + r}{1 + g} \right)$$  \hspace{1cm} (46)

Recalling $r > g$, this ratio is larger than one. Again, this reflects the implicit debt that features the collective scheme (the ratio equals one in the individual scheme).