Tightening Disability Screening Or Reducing Disability Benefits? Evidence and Welfare Implications*

Andreas Haller, University of Zurich
Stefan Staubli, University of Calgary, CEPR, IZA, and NBER
Josef Zweimüller, University of Zurich, CEPR, CESifo, and IZA

March 2018

Abstract

This paper provides a simple framework to evaluate the welfare effects of stricter disability insurace (DI) screening versus lower DI benefits. We develop sufficient-statistics formulas, capturing the insurance value and incentive costs of changes in screening stringency and benefit levels, and implement them for Austria and the United States. In Austria, we exploit exogenous variation in screening stringency and benefit levels arising from several reforms. We find that stricter screening significantly reduces DI inflow through both a mechanical effect, capturing that fewer applicants qualify for benefits under the stricter rules, and a behavioral effect, capturing that less people apply for benefits. We also find that a decrease in benefits is associated with a significant reduction in DI inflow. Our analysis suggests that DI screening among older workers in Austria has been too lenient, but benefit levels are optimal. For the United States, we use existing estimates from previous studies and find that both relaxing screening and increasing benefits would improve welfare, but the welfare gains from relaxing screening are greater.

Keywords: Disability insurance, screening, benefits, policy reform

JEL Codes: J14; J26; J65.

*Address: Andreas Haller, University of Zurich, CH-8000 Zürich, Switzerland; email *andreas.haller@uzh.ch*. Stefan Staubli, University of Calgary, Calgary, AB T2N 1N4, Canada; email *sstaubli@ucalgary.ca*. Josef Zweimüller, University of Zurich, CH-8000 Zurich, Switzerland; email *josef.zweimueller@econ.uzh.ch*.
We thank Richard Disney, Timothy Moore, Matthew Notowidigdo, Philippe Ruh, Florian Scheuer, Johannes Spinnewijn, Alexander Strand, and seminar participants at the University of Manitoba, the University of Melbourne, Universitat Pompeu Fabra, and the University of Zurich for helpful comments. This research was supported by the U.S. Social Security Administration through grant #1DRC12000002-04 to the National Bureau of Economic Research as part of the SSA DisabilityResearch Consortium. The findings and conclusions expressed are solely those of the author and do not represent the views of SSA, any agency of the Federal Government, or the NBER. All remaining errors are our own.

1 Introduction

In many countries, the share of individuals receiving Disability Insurance (DI) has increased significantly during the last two decades. For example, in the United States 2.6 percent of individuals in the age group of 20 to 64 were receiving DI benefits in 1992, but by 2012 this fraction had risen to 5.3 percent. The rapid expansion of the beneficiary population has generated substantial interest by policy makers and economists in measures that reduce growth in program caseloads and expenditures.

Two potential ways to slow program inflow are to tighten screening of DI applicants, for example by implementing stricter DI screening criteria, and to reduce benefit levels. Yet, little is known about the welfare effects of these measures. This paper helps to fill this gap by providing robust sufficient statistics formulas for welfare analysis that capture the insurance value and incentive costs of changes in screening stringency and benefit levels. These formulas are functions of high-level elasticities that can be estimated using design-based empirical methods.[1]

We implement these formulas empirically by estimating the relevant treatment effects directly for Austria. Studying the Austrian case has several advantages. *First*, we can use the Austrian Social Security Administration database (ASSD) which contains the complete labor market and earnings histories of all private-sector workers in Austria dating back to 1972. Additionally, we have detailed information on the various stages of the application process for all DI applications since 2004. *Second*, we are able to exploit exogenous variation in DI eligibility criteria and benefits which is generated by several policy reforms. The combination of detailed labor market and application data and quasi-experimental policy variation gives us the unique opportunity to study the impact of stricter screening and changes in benefit levels on DI inflow, DI applications, and labor force outcomes. *Third*, certain features of the Austrian DI- and social protection systems are similar to those of the United States. In particular, as described in more detail below, the Austrian reforms we are exploiting are comparable to reforms that have been proposed in the United States.

Our identification strategy to estimate the effects of stricter disability screening exploits variation in DI eligibility strictness that is generated by a policy reform. Prior to 2013 DI eligibility standards

[1] Recent applications of the sufficient statistic approach for optimal UI design include Shimer and Werning (2007), Chetty (2008), Kroft (2008), Landais et al. (2010), Kroft and Notowidigdo (2011), Schmieder et al. (2012), and Landais (2012). See the article by Chetty and Finkelstein (2013) for a detailed discussion of this literature.

were significantly relaxed for workers above age 57 relative to those below age 57. In 2013 the Austrian government increased the age threshold for relaxed DI access from age 57 to age 58, followed by further increases to age 59 in 2015 and age 60 in 2017. These step-wise increases imply that the strictness of DI eligiblity at a certain age varied by date of birth. On this basis, our estimation approach is a difference-in-differences design, comparing younger and older birth cohorts, who faced different DI eligiblity rules, over time. Our identification strategy to analyze the impact of benefit generosity exploits variation in DI benefits arising from a large pension reform. This reform changed DI benefit levels for indviduals with similar characteristics in different ways. This allows us to use a difference-in-differences approach that relates individuals' labor supply response to their differential change in benefit levels stemming from the policy reform.

The insights from our empirical analysis can be summarized by four broad conclusions. *First*, DI benefit receipt is responsive to changes in DI eligibility criteria. We estimate that relaxing DI eligibility standards at a certain age increases the receipt of DI benefits above that age by 5.8 percentage points (about 37 percent of the DI benefit receipt above the RSA). *Second*, using data on DI applications we can decompose the increase in DI benefit receipt into a behavioral effect, capturing that less people apply for benefits, and a mechanical effect, capturing that fewer applicants qualify for benefits under the stricter rules. We find that the bulk of the increase in benefit receipt is due to the mechanical effect while the behavioral effect is less important. *Third*, we find that relaxing DI access reduces employment, but also signficantly reduces the fraction of people receiving either unemployment or sickness insurance benefits. *Fourth*, we find that DI applications and DI claiming are also sensitive to the level of DI. Over our sample period, we estimate an elasticity of DI applications with respect to DI benefit levels of 0.7 for men and 0.9 for women. We estimate a smaller elasticity for DI claiming of 0.2 for both men and women.

Plugging the above estimates into our sufficient statistics formulas, we find that the DI eligiblity criteria for older workers in Austria were too generous before the reform. Thus, the increases in the age threshold for relaxed DI access improved overall welfare. On the other hand, we find that cutting DI benefits did likely reduce overall welfare. These conclusions are specific for Austria and cannot be generalized to other contexts. However, our sufficient statistics formulas can be used to make welfare statements in other contexts as well. To illustrate this point, we implement our formulas for the United States using empirical estimates from existing U.S. studies. This exercise illustrates that, in the U.S., relaxing DI eligiblity standards and increasing benefits would likely improve overall welfare.

There is a growing empirical literature studying the effects of DI on labor market outcomes (e.g. (Autor and Duggan, 2003; de Jong, Lindeboom, and van der Klaauw, 2011; Staubli, 2011; Maestas, Mullen, and Strand, 2013; Moore, 2015; Gelber, Moore, and Strand, 2017)) but empirical evidence on the effect of eligibility criteria on DI application behavior is scarce. Also, from a theoretical perspective relatively little is known about how imperfect information on disability status should be used to solve the incentive-insurance trade-off in the DI program. Diamond and Sheshinski (1995) and Parsons (1996) discuss medical screening in a static environment. More recently, Denk and Michau (2013) and Low and Pistaferri (2015) assess the optimal screening stringency in a dynamic environment and both conclude that screening stringency is too strict in the U.S. This paper builds on this literature and adds to it by exploring how changes in eligibility criteria and benefit levels affect DI application behavior and labor market outcomes of applicants. In particular, we are able to examine the relative impact of stricter eligibility criteria on DI enrollment due to more people being denied benefits under the stricter rules as opposed to more people self-screening, i.e. not applying for benefits.

The paper is organized as follows. In the next section presents a model of disability insurance and formulae for optimal disability screening and benefits. Section 3 describes the data and institutional backgrand in Austria. Sections 4 and 5 present the empirical results on stricter disability screening and changes in benefit levels, respectively. Section 6 measures the welfare consequences of stricter screening versus lower benefits in Austria and the United States. Section 7 summarizes the main results and draws policy conclusions.

2 Model

We start by analyzing a static model of disability insurance closely related to Diamond and Sheshinski (1995). We study the basic trade-offs when altering the key policy parameters of DI: strictness of screening and benefits.[2] The trade-offs are expressed as sufficient-statistics formulas

[2]In the spirit of Low and Pistaferri (2015), we define stricter screening as an increase in the disability standard set by the government. In Diamond and Sheshinski (1995) this is referred to as the "strictness of the disability standard". For brevity we use strictness of screening or stricter screening. The formal definition of strictness of screening is discussed in detail in section 2.1.

capturing the insurance value and incentive costs of DI. Moreover, we study the optimal policy mix, i.e. by how much should the strictness of screening be adjusted for every dollar change in benefits?

2.1 Agents and Planner's Problem

We adapt the DI model by Diamond and Sheshinski (1995) in two dimensions. First, by introducing DI application costs, we make the application decision dependent on strictness of screening. Second, as in Baily (1978) we split the model in two consecutive periods. In the first period, individuals work, earn a wage w, and pay lump-sum taxes τ to finance social insurance programs. Utility in the first period is given by $u(w-\tau)$ and individuals cannot make any choices. In the second period, disability shocks materialize and individuals can decide whether to work, apply for DI, or leave the labor force and receive welfare benefits.[3]

Disability is modeled as disutility of labor θ. We assume that θ is non-negative and distributed in the population with distribution $F(\theta)$.[4] To ease exposition, we assume that individuals only differ in θ. We show in Appendix E.1 that our results also apply in settings with other sources of heterogeneity. An individual with disutility level θ has utility $u(c_w)-\theta$ when working, where c_w denotes the consumption level. No taxes are levied in the second period, social insurance programs are purely financed through the first period's tax revenue. If unable or unwilling to work, an individual has access to two sources of benefits: disability and other benefits. The other benefits are unconditional and can be understood as summarizing all benefits other than DI such as unemployment insurance (UI) or retirement. In contrast, individuals need to qualify for DI benefits. Individuals have utility $v(c_r)$ if they receive other benefits and $v(c_d)$ if they are on the DI program.

The DI program is characterized by two policy parameters: The benefit level b_d and the strictness of screening θ^*. As in Diamond and Sheshinski (1995), the disutility level θ is only observed with noise, i.e. the planner observes $\theta_e = \theta + e(\theta)$ and this observed disutility has some conditional distribution $G(\theta_e|\theta)$. Strictness of screening θ^* is then modeled as a decision rule such that if the planner observes disutility levels above θ^* individuals are allowed on DI and otherwise rejected. Hence, the award probability of an individual with disutility θ at screening strictness θ^* is given by

[3] Splitting the model in two periods is convenient to rule out second order effects of changes in strictness of screening on behavior via changes in the tax rate. This is a standard assumption in the sufficient statistics literature, see e.g. Chetty and Finkelstein (2013).

[4] We assume $F(\theta)$ has a density $f(\theta)$ and the density is continuous.

$p(\theta;\theta^*) = 1 - G(\theta^*|\theta)$. A higher θ^* therefore implies a stricter screening rule with lower DI award probabilities.

Individuals can decide whether to apply for DI. If they apply for DI, they incur application costs ψ and are accepted with probability $p(\theta;\theta^*)$. If they are rejected, they can decide whether to return to work or to go on other benefits. We assume that $p(\theta;\theta^*)$ is weakly increasing in θ so that individuals with higher disutility levels have a weakly higher chance of being awarded DI benefits. Furthermore, we assume that DI benefits are higher than other benefits (otherwise there is no need for having a separate DI program) and that the marginal applicant returns to work if rejected[5]. Under these assumptions an individual will apply for DI iff

$$p(\theta;\theta^*)v(c_d) + (1 - p(\theta;\theta^*))(u(c_w) - \theta) - \psi \geq u(c_w) - \theta, \tag{1}$$

the marginal applicant is determined by

$$\theta^A = u(c_w) - v(c_d) + \frac{\psi}{p(\theta^A;\theta^*)}, \tag{2}$$

and the marginal type who decides to not work in case of DI rejection is given by

$$\theta^R = u(c_w) - v(c_r). \tag{3}$$

Figure 1 illustrates the choices in the second period.

Figure 1

The planner seeks to maximize welfare. Assuming no discounting, welfare is given by

$$\begin{aligned} W =& u(w-\tau) + \int_0^{\theta^A} u(c_w) - \theta dF(\theta) + \int_{\theta^A}^{\theta^R} (1 - p(\theta;\theta^*))(u(c_w) - \theta) dF(\theta) + \int_{\theta^A}^{\infty} p(\theta;\theta^*) v(c_d) dF(\theta) \\ &+ \int_{\theta^R}^{\infty} (1 - p(\theta;\theta^*)) v(c_r) dF(\theta) - \int_{\theta^A}^{\infty} \psi dF(\theta). \end{aligned} \tag{4}$$

[5] This is the empirically relevant case since we observe that rejected applicants return to work.

The planner is constrained by a balanced budget requirement

$$b_d \int_{\theta^A}^{\infty} p(\theta;\theta^*) dF(\theta) + b_r \int_{\theta^R}^{\infty} (1 - p(\theta;\theta^*)) dF(\theta) = \tau \tag{5}$$

and the behavioral responses of the agents

$$\theta^A = u(c_w) - v(c_d) + \frac{\psi}{p(\theta^A;\theta^*)}, \tag{6}$$

$$\theta^R = u(c_w) - v(c_r). \tag{7}$$

2.2 Welfare Analysis: Optimal Strictness of Screening

For simplicity we only discuss the intuition for stricter screening in this section, but our sufficient-statistics formulas apply for any direction of change in strictness of screening.[6] Setting a higher DI standard θ^* reduces DI inflow through a mechanical and a behavioral effect. To see this, we can write the DI inflow probability as $Pr(\text{Award}) = Pr(\text{Award}|\text{Apply}) * Pr(\text{Apply})$. Hence, stricter screening affects awards as follows

$$\frac{dPr(\text{Award})}{d\theta^*} = \underbrace{\frac{dPr(\text{Apply})}{d\theta^*} Pr(\text{Award}|\text{Apply})}_{\text{behavioral effect}} + \underbrace{\frac{dPr(\text{Award}|\text{Apply})}{d\theta^*} Pr(\text{Apply})}_{\text{mechanical effect}}. \tag{8}$$

The mechanical effect captures the change in DI inflow if only the award probability changed and the application probability remained unaffected. The behavioral effect is the change in DI inflow if only the application probability changed. Figure 2 illustrates these effects in the context of our model. Stricter screening shifts down the award probability curve and shifts the marginal applicant to the right (Panel a). The area between the two award probability curves is the mechanical effect. A fraction of rejected applicants due to mechanical effect goes back to work (gray area in Panel b). The other fraction has labor disutility too high to work and substitutes DI benefits with other benefits (blue area in Panel b). The change in the marginal applicant times the award probability of the marginal applicant is the behavioral effect (red area in Panel b). The same margins adjust when DI standards are lowered, but the adjustments are in the opposite direction. As we will show now, the relative size of these effects has important implications for welfare.

[6] The intuition for reducing strictness of screening is completely analogous, just in the opposite direction.

$\boxed{\textit{Figure 2}}$

Following from equations (4)-(7), the welfare effect of a budget-neutral change in strictness of screening is given by

$$\frac{dW}{d\theta^*} = -u'(w-\tau)\frac{d\tau}{d\theta^*} + \int_{\theta^A}^{\theta^R} \frac{dp(\theta;\theta^*)}{d\theta^*}\left[v(c_d) - [u(c_w) - \theta]\right] dF(\theta) + \int_{\theta^R}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*}\left[v(c_d) - v(c_r)\right] dF(\theta) \tag{9}$$

where

$$\frac{d\tau}{d\theta^*} = b_d \left(\underbrace{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)}_{\text{mechanical effect}} - \underbrace{\frac{d\theta^A}{d\theta^*} p(\theta^A) f(\theta^A)}_{\text{behavioral effect}} \right) - \underbrace{b_r \int_{\theta^R}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)}_{\text{benefit substitution}}. \tag{10}$$

A change in screening strictness has a fiscal effect (first term in equation 9) and changes the insurance value of DI (second and third term in equation 9). Equation (10) captures the fiscal effect. In particular, stricter screening relaxes the budget constraint through both the mechanical and the behavioral effect, but also tightens the budget constraint through the benefit substitution effect. The insurance value of DI changes because DI allowances become more selective. Some of the newly rejected applicants return to work and loose $v(c_d) - [u(c_w) - \theta]$ (second term in equation 9), while some substitute to other benefits and loose $v(c_d) - v(c_r)$ (third term in equation 9). Notably, the behavioral effect does not affect the insurance value. The behavioral effect is driven by marginal applicants deciding no longer to apply, and since they are indifferent between working and applying for DI, this has no first order welfare effect.

To ease interpretation, it is convenient to normalize the welfare gains of a change in screening by dividing through the mechanical effect. This normalization implies that we measure the welfare gain per rejected applicant. Moreover, to express the welfare gains as a money metric, we follow the standard approach in the sufficient statistics literature and divide the welfare gains by $u'(w-\tau)$. This normalization leads from (9) and (10) to

$$\frac{dW}{d\theta^*} \gtreqless 0 \iff \left(1 + \epsilon - \frac{b_r}{b_d} S\right) b_d \gtreqless \frac{v(c_d) - \left(v(c_r)S + (1-S)\left[u(c_w) - \widetilde{\theta}\right]\right)}{u'(w-\tau)} \tag{11}$$

where

$$S \equiv \frac{\int_{\theta^R}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)}{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)} = \frac{\text{benefit subsitution}}{\text{mechanical effect}}, \tag{12}$$

$$\epsilon \equiv -\frac{\frac{d\theta^A}{d\theta^*} p(\theta^A) f(\theta^A)}{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)} = \frac{\text{behavioral effect}}{\text{mechanical effect}} \tag{13}$$

and

$$\widetilde{\theta} = \frac{\int_{\theta^A}^{\theta^R} \frac{dp(\theta;\theta^*)}{d\theta^*} \theta dF(\theta)}{\int_{\theta^A}^{\theta^R} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)}. \tag{14}$$

The LHS of inequality (11) captures the budgetary effect. The term in the brackets measures the monetary multiplier of stricter screening per rejected applicant. On top of the direct effect, DI expenditures are also reduced because of fewer applications, measured by ϵ. The term $\frac{b_r}{b_d} S$ captures the fiscal externality of stricter screening, i.e. for every dollar saved in DI due to stricter screening, expenditures in other social insurance programs increase by $\frac{b_r}{b_d} S$ dollars. Therefore, $1 + \epsilon - \frac{b_r}{b_d} S$ is the monetary multiplier of saving one dollar in DI because of stricter screening. In our empirical setup, we can estimate all elements of the LHS of inequality (11) in a reduced form way.

The RHS of inequality (11) measures the monetary welfare loss from less generous insurance. In contrast to the sufficient statistics literature on UI, this loss is in absolute utility terms and not expressed as marginal utilities. This makes it more challenging to implement the RHS of inequality (11). The main challenge, however, is that the abstract quantity θ is showing up. Implementing the RHS directly would therefore require to know or assume the distribution of θ as well as the functional forms of the utility functions. To avoid this, we derive bounds for the RHS. These bounds only depend on risk aversion and replacement rates and are valid for a wide range of utility functions and any continuous distribution of θ. These bounds deliver sufficient conditions for welfare judgements. Appendix B contains the derivations for these bounds.

If we assume CRRA utility, then a sufficient condition for $\frac{dW}{d\theta^*} \geq 0$ is[7]

$$1 + \epsilon - \frac{r_r}{r_d} S \geq \frac{1}{1-\gamma} \left[r_d^{-\gamma} - \frac{r_r^{1-\gamma}}{r_d} \right] \tag{15}$$

where $r_r = \frac{b_r}{w-\tau}$ and $r_d = \frac{b_d}{w-\tau}$ are the replacement rates of other and disability benefits, respectively,

[7] see appendix B.1 for the derivations.

and γ is the coefficient of relative risk aversion. In Appendix B.1, we also derive a condition for general utility functions using a Taylor approximation. Analogously, we derive bounds to get a sufficient condition for $\frac{dW}{d\theta^*} \leq 0$. Appendix B.2 contains these sufficient conditions and derivations. The conditions we derive are sufficient but not necessary. The key question is therefore when these criteria are inconclusive. For given replacement rates and estimates of S and ϵ there is a range of risk aversion for which the conditions are indecisive. In our applications, however, these bounds are tight enough to make meaningful welfare statements. Appendix B.3 contains the details on the inconclusiveness of the bounds.

2.3 Welfare Analysis: Optimal DI Benefits

The second key policy parameter in DI is the benefit level. With the same model it is straightforward to derive a sufficient statistics formula for the optimality of DI benefits. Starting from equation (4) we get

$$\frac{dW}{db_d} = -u'(w-\tau)\frac{d\tau}{db_d} + \int_{\theta^A}^{\infty} p(\theta;\theta^*)v'(c_d)dF(\theta) \tag{16}$$

where

$$\frac{d\tau}{db_d} = -b_d\left(\frac{d\theta^A}{db_d}p(\theta^A)f(\theta^A)\right) + \int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta). \tag{17}$$

Hence,

$$\frac{dW}{db_d} \gtreqless 0 \iff \xi \lesseqgtr \frac{v'(c_d) - u'(w-\tau)}{u'(w-\tau)} \tag{18}$$

where $\xi := -\frac{\partial\theta^A}{\partial b_d}p(\theta^A)f(\theta^A)\frac{b_d}{\int_{\theta^A}^{\infty} p(\theta)dF(\theta)}$ is the elasticity of DI take-up wrt. DI benefits. This formula is analogous to the Bailey-Chetty formula of optimal UI. Intuitively, this captures the trade-off between moral hazard, higher DI take-up measured by ξ, and the insurance value of DI. To make the RHS of (18) operational we can again use the CRRA functional form to get [8]

$$\frac{dW}{db_d} \gtreqless 0 \iff \xi \lesseqgtr \left(\frac{c_d}{w-\tau}\right)^{-\gamma} - 1. \tag{19}$$

[8] Appendix C contains the formula using a Taylor approximation.

This provides again a local test whether benefits are too high / low for the given strictness of screening.

2.4 Welfare Analysis: Optimal Policy Mix

We derived ceteris paribus tests for optimal benefits and strictness of screening. The respective sufficient statistics formulas indicate in which direction strictness of screening and benefits should be adjusted. The next natural question concerns the optimal policy mix, i.e. how much should strictness of screening be adjusted for every dollar change in benefits? The sufficient statistics also shed light on this question. The gradient of the welfare function points in the direction of the greatest rate of increase of the function and therefore answers the question of the optimal policy mix. Figure 3 illustrates the idea. Following from equations (9) and (16), the gradient is given by

$$\nabla W = \begin{pmatrix} \frac{dW}{db} \\ \frac{dW}{d\theta^*} \end{pmatrix} = \begin{pmatrix} \alpha * \int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta) \\ \beta * \int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta) b_d \end{pmatrix} u'(w-\tau) \tag{20}$$

where

$$\begin{aligned} \alpha &= -\xi + \frac{v'(c_d) - u'(w-\tau)}{u'(w-\tau)}, \\ \beta &= -\left(1+\epsilon - \frac{b_r}{b_d}S\right) + \frac{v(c_d) - \left(v(c_r)S + (1-S)\left[u(c_w) - \widetilde{\theta}\right]\right)}{u'(w-\tau)b_d}. \end{aligned}$$

Therefore, the optimal direction of change in strictness of screening and benefits is given by

$$\frac{d\theta^*}{db_d}|_{optimal} = \frac{\beta}{\alpha} * \frac{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta) b_d}{\int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta)}. \tag{21}$$

α and β are the same formulas we implemented to determine the sign of the welfare derivatives with respect to benefits and strictness of screening in the previous sections. The sign of α determines the direction in which benefits should be adjusted, namely we have $\alpha \gtrless 0 \Leftrightarrow \frac{dW}{dbd} \gtrless 0 \Leftrightarrow db_d \gtrless 0$. Analogously, we have $\beta \gtrless 0 \Leftrightarrow \frac{dW}{d\theta^*} \lessgtr 0 \Leftrightarrow d\theta^* \lessgtr 0$. Hence, α and β determine the direction of adjustments in strictness of screening and benefits and the ratio of the two gives the relative direction. We can implement α and β empirically using the same bounds as before. Assuming CRRA utility we get

$$\alpha = -(1+\xi) + r_d^{-\gamma} \tag{22}$$

$$\beta < -\left(1+\epsilon - \frac{r_r}{r_d}S\right) + \frac{1}{1-\gamma}\left[r_d^{-\gamma} - \frac{r_r^{1-\gamma}}{r_d}\right] \tag{23}$$

$$\beta > -\left(1+\epsilon - \frac{r_r}{r_d}S\right) + \frac{S}{1-\gamma}\left[r_d^{-\gamma} - \frac{r_r^{1-\gamma}}{r_d}\right]. \tag{24}$$

Hence, for a given risk aversion this gives us a cone of optimal policy mixes. We can also vary γ, this gives us a wider bound of optimal policy mixes. This is illustrated in figure 3.

Figure 3

The optimal direction in (21) can be interpreted in two intuitive ways. First, it can be interpreted in terms of mechanical expenditure effects. A reduction of DI benefits by one dollar leads to a mechanical reduction in DI expenditures of $\int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta)$ dollars. Similarly, stricter screening translates to a mechanical reduction in DI expenditures of $\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*}dF(\theta)b_d$ dollars. Therefore, the optimal direction in (21) satisfies that for every dollar reduction of DI expenditures due to lower benefits, screening is adjusted such that DI expenditures mechanically change by $\frac{\beta}{\alpha}$ dollars. This is a simple rule. Suppose we have $\alpha < 0$ and $\beta < 0$, that is benefits should be reduced and screening should be stricter. Suppose DI benefits are reduced by one dollar and this translates to a mechanical reduction in DI expenditures of 1 Mio. dollars. Then the optimal direction rule implies that screening should be made stricter such that DI expenditures are mechanically reduced by $\frac{\beta}{\alpha}$ Mio. dollars. Hence, this optimal direction can be interpreted as a static scoring budgeting rule.

The second interpretation of the optimal direction is in terms of percent changes in benefits and DI award rate. The gradient can be rewritten as

$$\nabla W = \begin{pmatrix} \alpha * \frac{1}{b_d} \\ \beta * \frac{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*}dF(\theta)}{\int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta)} \end{pmatrix} u'(w-\tau)\int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta)b_d. \tag{25}$$

Note that a one dollar change in benefits translates to a $\frac{1}{b_d}$-percent change in benefits and that a change in screening stringency translates to a mechanical change of DI levels of $\frac{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)}{\int_{\theta^A}^{\infty} p(\theta;\theta^*) dF(\theta)}$ percent, which is equivalent to a change in the award rate since $\frac{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)}{\int_{\theta^A}^{\infty} p(\theta;\theta^*) dF(\theta)} = \frac{\Delta Pr[award|apply]}{Pr[award|apply]}$. Hence, the optimal direction can be seen in terms of these units. That is, the optimal direction can be interpreted as saying that for a 1-percent change in DI benefits, the award rate should be reduced by $\frac{\beta}{\alpha}$-percent.[9]

3 Institutional Background and Data

Institutional Background. Like in many developed countries, there are three programs in Austria that provide income replacement in the case of a separation from the labor market for economic or health reasons: disability insurance (DI), sickness insurance (SI), and unemployment insurance (UI). The DI program is financed by a payroll tax on earned income and provides partial earnings replacement to workers below the full retirement age who have accumulated at least 5 insurance years within the last 10 years. Insurance years include both contribution years (i.e., periods of employment, including sick leave) and non-contributory periods of labor force participation (e.g., unemployment). The required insurance years increase by one month for every two months above age 50 up to a maximum of 15 insurance years.[10]

To apply for DI benefits, an individual must submit an application to the local DI office. Employees at the DI office first check whether the applicant has not reached the full retirement age and meets the insurance years requirement. DI eligibility is not conditioned on earnings, so applicants are not required to stop working in order to apply for benefits. In a second step, a team of disability examiners and physicians assesses the severity of the medical impairment and the applicant's earnings capacity. An impairment is considered to be severe if it lasts at least six months and limits the applicant's mental or physical ability to engage in substantial gainful activity. The assessment of the applicant's *residual* earnings capacity depends on age and work experience. Unskilled applicants below age 60 are awarded benefits if the earnings capacity has been reduced to less than half of the earnings capacity of a healthy person in *any reasonable* occupation in the economy the individual

[9] A formal derivation of these two interpretations of the optimal policy mix can be found in Appendix D.
[10] The insurance years requirement does not apply if the disability is job-related; for each occupation there exists an explicit list of qualifying impairments.

could be expected to carry out.[11] For applicants who have worked in a *similar* occupation for 10 years in the last 15 years, eligibility criteria are substantially relaxed above the relaxed screening age (RSA) threshold by changing the comparison from a healthy worker performing any type of work in the economy to a healthy worker in a similar occupation. An occupation is considered similar if the following requirements are identical: manual and mental demands, amount of responsibility, posture, concentration, endurance, required care, and stress level (Wörister, 1999). Thus, older applicants are significantly more likely to be awarded benefits, as they are only compared to healthy workers in their occupation.[12] The RSA was 57 in 2004, but was increased in three one-year steps to 60 by 2016. As described in the next section, we exploit the variation in the RSA to identify the labor market effects of stricter disability screening. Once benefits are awarded, DI beneficiaries receive monthly payments until their return to work, medical recovery or death. DI benefits can be granted for a temporary period, but very few claimants (fewer than 4 percent) ever leave the DI rolls.

DI benefits are subject to income and payroll taxation and replace approximately 70 percent of pre-disability net earnings up to a maximum of about €4,500. The level of DI benefits is calculated by multiplying a pension coefficient, which varies by age and insurance years, with an assessment basis, which is average indexed capped earnings over a given period of time (e.g., the best 16 years in 2004 at the beginning of our study period). Younger applicants with limited work experience qualify for a special increment to supplement their benefits. DI beneficiaries may continue work, but those earning more than an exempt threshold lose up to 50 percent of their benefits, depending on their earnings.[13] A pension reform in 2004 gradually decreased pension levels for most workers, providing exogenous variation in benefit levels to identify the labor market effects of changes benefit generosity (section 5).

In case of a temporary illness, employers continue to pay 100% of earnings for up to 12 weeks. Once the right to full benefits paid by the employer has expired, individuals may claim benefits from the Austrian sickness insurance system. Sickness benefits replace approximately 65% of the last net

[11] Eligibility standards are less strict for semi-skilled and skilled applicants below age 60, whose set of reasonable occupations is more limited. To be classified as semi-skilled or skilled, an applicant must have worked in a semi-skilled or skilled occupation for 7.5 years or more in the most recent 15 years.

[12] Access to disability insurance is also relaxed in other countries at older ages, including Australia, Canada until 1995, Denmark, Sweden until 1997 (Karlström et al., 2008), and the United States (Chen and van der Klaauw, 2008).

[13] Ruh and Staubli (2016) show that this policy induces DI beneficiaries to keep their earnings below the exempt threshold in order to retain benefits.

wage up to the same maximum that applied to disability benefits. Continued wage payments and sickness benefits are both subject to income taxation. The benefit duration is 52 weeks for individuals who have worked at least 6 months in the previous 12 months, otherwise the duration is 26 weeks.

Regular UI benefits are a function of the wage on the last job and replace approximately 55 percent of the prior net wage subject to a minimum and maximum. The UI system is more generous for older workers. Specifically, while job losers below age 50 receive at most 39 weeks of regular UI benefits, job losers above age 50 can claim benefits for up to 52 weeks provided they have paid UI contributions for at least 9 years in the last 15 years. Job losers who exhaust the regular UI benefits can apply for "unemployment assistance" (UA). These means-tested transfers last for an indefinite period and are about 70 percent of regular UI benefits.

Data. We merge data from two administrative registers. First, the Austrian Social Security Database (ASSD) contains very detailed longitudinal information for the universe of workers in Austria since 1972. At the individual level the data include gender, nationality, month and year of birth, blue- or white-collar status, and labor market history. Labor histories are summarized in spells; all employment, unemployment, disability, sick leave, and retirement spells are recorded. Spells before 1972 are available for individuals who have claimed a public pension by the end of 2008. The ASSD also contains some firm-specific information: geographic region, industry affiliation, and firm identifiers that allow us to link both individuals and firms. See Zweimüller et al. (2009) for a detailed description of the data. Second, we use data on all DI applications since 2004 which contain detailed information on the date of the application, the date of the decision, the decision itself (i.e. reject or accept), the reported medical impairment of the applicant, and the stage of the application (first application, appeal etc.)

Starting from the population data set, we impose two restrictions. First, we exclude self-employed and civil service workers because they are covered by a different pension system than private-sector workers. Second, we exclude observations in which individuals are below age 54 or over 60, at which point many individuals become eligible for retirement benefits. We also exclude women with more than 38 insurance years, because they could claim an early retirement pension before age 60.[14] . Our sample covers more than three quarters of all active labor market participants in Austria. Since we can observe complete work histories, we can precisely calculate (1) how much DI benefits individuals would get at any point in time and (2) whether individuals have sufficient work experience to apply for DI benefits under the relaxed eligiblity criteria above the RSA.

Our empirical analysis focuses on the following key outcome variables. *DI benefits* is an indicator for whether an individual receives DI benefits.[15] *DI application* is an indicator for whether an individual has applied for DI benefit. The overarching goal of the different reforms we study was to foster employment among older workers. Therefore, the third outcome variable of interest is *employment*, which is an indicator for whether an individual has positive work days. The fourth outcome variable is *other benefits*, which measures whether individuals receive benefits from either the UI or the SI program. Estimating the magnitude of such benefit substitution effects is crucial to understand the fiscal and welfare effects of changes in the DI program.

Table 1 presents separate summary statistics for the "screening sample" used in our analysis for changes in DI screening. Summary statistics for the "benefit generosity sample" will be added.

We report separate statistics for men and women and whether they are eligible or not eligible for relaxed screening above the RSA threshold. Individuals are considered eligible for relaxed screening if they have at least 10 employment years in the past 15 years at any age between 56 and 59; otherwise they are considered ineligible for relaxed screening.[16]

Table 1

[14]Between 2004 and 2016, the Austrian government increased the earliest eligibility age for retirement pensions from age 57 to 60 for women, while leaving the earliest eligibility age unchanged at age 60 for men (see Staubli and Zweimüller, 2013). About 23 percent of women are eligible for an early retirement pension before age 60.

[15]DI spells are back-dated in the ASSD to the date the claim was filed, so an individual who applied for disability benefits late in the calendar year and was awarded benefits in the next calendar year is observed to claim disability benefits in the original calendar year.

[16]Note that only individuals who worked in a *similar occupation* for 10 of the last 15 years are eligible for relaxed screening, while our definition is based on whether somebody has worked in *any occupation* for 10 years of the last 15 years because we can only observe industry affiliation and not occupation. This implies that the eligible sample will include some individuals who are in fact not eligible for relaxed screening, but this number is likely small because what constitutes a similar occupation is defined very broadly.

4 Impacts of Stricter Disability Screening

4.1 The 2013 reform

In April 2012, the Austrian government announced the 2nd Stability Act (2. Stabilitätsgesetz), which became effective on January 1, 2013. The primary obejectives of this Act were to reduce expenditures in the public pension systems and to foster employment among older workers. The only change in the DI program was an increase in the RSA threshold by 3 years. This increase was phased-in gradually over time: the RSA was increased to age 58 in 2013, followed by further increases to age 59 in 2015 and age 60 in 2017. Individuals who had not worked in a similar occupation for 10 years in the last 15 years were not affected by this increase as they were not eligible for relaxed access to DI benefits above the RSA.

Figure 4

These step-wise increases implied that the RSA varied by date of birth. Figure 4 illustrates this variation in the RSA for the birth cohorts used in our subsequent analysis. For example, the RSA shifted from age 57 to age 58 for individuals born after November 1955. These individuals turned 57 in or after December 2012 and their application would be assessed under the new rules, effective January 1, 2013, because applications are assessed using the rules on the first of the month after filing. In contrast, individuals born in or before November 1955 could still apply for disability benefits under the old, less stringent rules. Thus, individuals born after November 1955 faced stricter DI eligibility criteria at age 57 compared to those born before December 1957. Similarly, the 2015 (2017) increases tightened disability eligibility criteria at age 58 (59) for individuals born in or after December 1956 (1957) relative to those born before.

In Figure 5 we provide descriptive evidence for the labor market effects of the RSA increses by plotting the percent of 50 to 59 year old men entering DI, applying for DI, working, and receving UI or SI benefits for different birth years.[17] Panel (a) shows a sharp increase in DI inflows at the age that coincides with a birth cohort's RSA; age 57 for individuals born in 1955 and age 58 for individuals born in 1956. The peak at age 58 is somewhat smaller in magintude than at age 57 .

[17]Since increases in the RSA affected individuals born in December and after, we define a birth year from December in the previous year to December this year. For example, the birth year 1955 refers to individuals born between December 1954 and December 1955.

Panel (b) shows that DI applications also exhibit a peak at their RSA, suggesting that individuals are more likely to apply for DI when eligibility criteria are relaxed. Panel (c) suggests that reaching the realxed screening for DI is associated with a reduction in employment as the employmnet rate drops significantly as soon as a birth cohort reaches its RSA. Panel (d) provides evidence for benefit substitution: the percent of individuals registered as unemployed or receiving sick leave benefits exhibits a permanent drop when a birth cohort reaches its RSA.

Figure 5

Figure 6 presents an analogous set of trends by birth year for women. Overall, we find similar patterns as for men, although the magnitude of the effects differs somewhat. For example, the peaks in DI inflow and DI applications at a cohort's RSA as well as the drops in employment and benefit substitution are generally smaller than for men. Employment also tends to be higher at any age for younger relative to older birth cohorts, perhaps driven by a general increase in female labor force participation. Similarly, DI inflow and DI applications tend to be lower at any age for younger relative to older birth cohorts. These figures are consistent with the hypothesis that a rise in the RSA reduces DI inflow and increases employment, but has also important spillover effects into other social insurance programs. In the next section, we describe our identification strategy to quantify the magnitudes of these effects empirically.

Figure 6

4.2 Estimation Strategy

Our research desgin exploits that the 2nd Stability Act created exogenous variation in the strictness of DI eligibility at certain ages by year of birth. On this basis, the primary estimation approach compares younger and older birth cohorts, who faced different DI eligiblity rules, over time. This comparison can be implemented by estimating regressions of the following type:

$$y_{it} = \alpha + \theta_a + \pi_c + \lambda_t + \sum_{k=-3\backslash -1.5}^{2.5} \beta_k I[age = RSA + k] + X_{it}'\delta + \varepsilon_{it}, \tag{26}$$

where i denotes individual, t year-quarter, c birth cohort; y_{ict} is the outcome variable of interest (such as an indicator for having applied for DI, an indicator for receiving DI benefits, and labor supply measures such an indicator for working), θ_i are dummies for age at a monthly frequency to control for age-specific levels in the outcome variable, π_c are dumies for birth cohort at a monthly frequency to capture time-constant differences across birth cohorts, λ_t are dummies for year and quarter to capture common time shocks and seasonal effects, and X_{ict} represent individual or region specific characteristics to control for any observable differences that might confound the analysis.

The key variables of interest are the indicators $I[age = RSA + k]$ which are equal to one if an individual's age, measured at a half-year frequency is equal to the interval $RSA + k$, where k runs from -3 to 2.5 using $k = -1.5$ as the reference categorie. Thus, each β_k-coefficient measures the average causal effect the effect of the RSA at the age$RSA + k$. For example, β_0 captures the effect of the RSA at the age where DI eligiblity standards are relaxed. Importantly, these indicators varies over time and across birth cohorts due to the RSA increase. For example, for individuals born before December 1955 $I[age = RSA+0]$ is one if the age is equal to 57, while for those born in or after December 1955 $I[age = RSA + 0]$ is only one if the age is equal to 58. To obtain the effect of the RSA over a wider age interval, we can simply sum up the different β_k-coefficients. Equation (26) is estimated separately for men and women using data in the age interval three years prior and post the RSA. Standard errors are clustered at the year-month of birth.

The main identification assumption is that, absent the increase in the RSA, the change in y_{it} at a certain age would have been comparable between birth cohorts not yet eligble for relaxed screening and those eligible after controlling for background characteristics. A potential concern of our estimation approach is that trends in the outcome variable at an age could be changing across birth cohorts over time for reasons unrelated to the RSA increase. Figures 5 and 6 show that there may be pre-existing trends in some outcome variables, less so for men than women. The estimated β_k-coefficients for $k < 0$ provide placebo checks for spurious trends, although they may capture possible anticipation effects to the RSA increase. In addition, we estimate equation (26) for the sample of individuals who are not eligible for relaxed DI screening because they have worked leass than 10 years in the past 15 years, and so we expect theβ_k-coefficients to be zero.

4.3 Empirical Results

Main Results. In Figures 7 and 8 we plot the estimated β_k-coefficients from equation (26) for men and women, respectively, for DI benefits, DI applications, employment, and other benefits (the

95 percent confidence interval is shown by the shaded area). In all four panels for both men and women the estimated coefficients for $k < 0$ are small and mostly statistically insignificant, providing evidence that our estimates are not cofounded by differential trends across birth cohorts.

The coefficients for DI benefits and applications turn significantly positive at $k = 0$, the age where eligibility criteria are relaxed. DI levels and applications increase continuously $k > 0$. The bottom left panel in Figures 7 and 8 shows that the increase in DI enrollment for $k \geq 0$ was accompanied by a substantial decreases in employment. However, we also find a siginficant decrease in other benefits.

Figure 7

Figure 8

In Figures 9 and 10 we plot the estimated β_k-coefficients from equation (26) for men and women, respctively, who are not eligible for relaxed DI screening above the RSA because they have worked less than 10 years in the past 15 years. In all four panels for both men and women the estimated coefficients are close to zero and almost always statistically insignificant, providing evidence that our estimates are not cofounded by differential trends across birth cohorts.

Figure 9

Figure 10

To get the overall effect of the RSA, we can simply sum up all β_k-coefficients for $k \geq 0$ (since point estimates are statistical insignificant for $k < 0$) and then divide by three, the number of years post-RSA. This estimate caputres the average effect of the RSA in the three-year window above the RSA. Table 2 reports these estimates for men (column 1) and women (column 2). The first row shows a sizeable increase in the percent of men receiving DI benefits when disability screening is relaxed. We estimate an increase in the DI recipiency rate for men of 5.77 percentage points, or about 37 percent of the total DI benefit receipt above the RSA (=5.77/15.8), and 8.6 percentage points (or 57 percent) for women. Row 2 shows a significant inscrease in the probability to apply for DI above the RSA.The increase is slightly smaller for men compared to women (3.02 percentage points compared to 4.07 percentage points).

One of the government's goals by tightening disability eligibility was to encourage employment among older workers. However, the DI program is only one of several transfer programs in Austria and individuals may have substituted towards these other programs rather than continued to work. Row 3 of Table 2 shows that the RSA leads to a significant decrease in employment of 2.63 percentage points among men and 3.07 percentage points among women. Row 4 shows that benefit substitution, i.e. whether soemebody is registered as unemployed or receives sick leave benefits, decreases by 3.09 percentage points for men and 4.92 percentage points for women above the RSA.

Figure 2

5 Impact of Benefit Generosity

The ideal experiment to analyze the impact of benefit generosity on DI claiming and applications would be to randomize the level of DI benefits across individuals. We emulate this ideal experiment with a quasi-experimental research design that exploits variation in DI benefits from a large pension reform. Our appraoch closely follows Mullen and Staubli (2016) who estimate the elasticity of DI claiming with respect to benefit generosity using variation in DI benefits in Austria from several reforms between 1987 and 2010. We differ from their study in two aspects: First, we update their estimates for a more recent time period (2004 to 2016). This period is characterized by lower replacement rates and stricter DI eligiblity compared to the 1980s and 1990s which could affect the responsiveness of DI claiming and applications to benefit levels. Second, we focus on older workers ages 50-59, while their sample also includes prime-age workers ages 35-49. This allows us to directly compare the benefit elasticity estimates with the screening elasticity estimates because they are estimated for the same age group and time period.

5.1 Estimation Strategy

In January 2004, the Austrian government implemented a series of changes to the calculation of DI benefits. These changes were part of a larger pension reform ("Pensionsreform 2003") and their effect was to reduce potential benefit levels for most individuals, although pension levels increased for some individuals with limited work experience. Specifically, before the reform applicants below age 56 with limited work experience qualified for a special increment to supplement their benefits. The reform gradually increased the age limit for the special increment to age 60 between 2004 and 2010. However, at the same time the reform phased in a reduction in the pension coefficient and increased the penalty for claiming benefits before the full retirement age (age 65 for men and age 60 for women).[18] In addition, the reform gradually increased the length of the assessment basis from 16 years to 40 years by 2028. As a result, the reform resulted in a large scale curtailment of benefits which was met with intense public criticism. Responding to the backlash, the Austrian government passed legislation in 2005 that reduced the maximum penalty for early claiming to 5-6.5 percent of the projected benefits under the pre-reform rules (with higher penalties phased in over time).

To estimate the elasticity of DI claiming and DI applications with respect to benefit generosity, we are interested in estimating regressions of the form:

$$y_{it} = \alpha + X_{it}'\beta + \gamma b_{it}(Z_{it}) + \lambda_t + \varepsilon_{it}, \tag{27}$$

where i denotes individual, t year-quarter, y_{it} is the outcome variable of interest such as applying for DI or claiming DI benefits, X_{it} is a vector of demographic and labor market characteristics, $b_t(Z_{it})$ are log potential DI benefits which are a function of labor market characteristics $Z_{it} \in X_{it}$ (e.g. age, insurance years, and assessment basis), λ_t are year-quarter fixed effects, and ε_{it} are any unobserved factors affecting DI claiming and applications such as tastes for work. The paramter of interest is γ which measures the aveage effect of a change in benefit levels on DI claiming and applications.

As discussed in Mullen and Staubli (2016), if b is a linear function of Z_{it}, we cannot separetly identify γ and β because no variation is left in b after controlling for Z_{it}. If γ is a non-linear function of Z_{it}, we can identify γ as long as sufficient residual variation is left in b after controlling for Z_{it}.[19] A drawback of this identification strategy is that it relies heavily on function form.

[18] Before the reform each insurance year increased the pension coefficient by 2 percentage points, while each year of claiming before the full retirement age reduced the pension coefficient by 3 percentage points (capped at a maximum of 10.5 percentage points or 15 percent of the pre-penalty pension coefficient, whichever is lower). The reform gradually reduced the pension coefficient adjustment for each insurance year from 2 to 1.78 percentage points between 2004 and 2009 and changed the penalty for each year of early claiming to 4.2 percent of the pension coefficient (capped at 15 percent of the full pension).

[19] For example, if we control for Z_{it} in a very flexible way by including polynomials or other transformations of Z_{it},γ may not be identified because benefits are collinear with Z_{it}.

This problem can be solved by exploiting the policy reform because it creates variation in b that is independent from Z_{it}. Intuitively, with the policy reform we observe individuals with similar Z_{it} but different benefits b. This approach is akin to a difference-in-differences estimation strategy, where identification is obatined by relating individuals' differential response to their differential change in benefit levels stemming from the policy reform.[20] Mullen and Staubli (2016) show that the policy-induced variation in b can be isolated by including the individual-specific (log) hypothetical benefits under each policy regime as additional controls in eqation (27). Due to the phased-in nature of the 2004 policy reform, we have 13 different hypothetical benefits for each year from 2004 to 2016. .

5.2 Empirical Results

Our main results are summarized in Table 3 with Panel A providing estimates of equation for (27) the entire age range spanned by our sample (ages 50-59) and Panel B displaying analogous estimates for the age group 55-59, which serve as inputs for the welfare analysis in the next section. Columns 1 and 2 indicate that an increase in benefit levels increases the propensity of applying for DI benefits. Specifically, for the age group 50-59 we find that a one percent increase in benefit levels raises the DI application rate by 3.28 percent among men and 3.54 percent among women. Dividing these estimates by the average DI application rate gives the implied application elasticity, which is higher among women (0.939) compared to men (0.657). For the age group 55-59, we find that application elasticities are smaller and similar for men (0.529) and women (0.575).

While the elasticity of DI applications is key to measure the behavioral response to changes in benefit levels, the elasticity of DI claiming with respect to benefits is important to caputre the fiscal effects. The ratio of the two elasticities is also informative on the award probability of applicants who are on the margin of applying for DI when benefits change. Columns 3 and 4 of Table 3 show that for the age group 50-59 the DI inflow elasticity varies beteen 0.225 to 0.249, about three to four times smaller than the DI application elasticity. This suggests that most marginal applicants get screened out by the application screening process. On the other hand, we find larger DI inflow elasticities for the age group 55-59 of 0.344 for women and 0.421, which is expected since most of these individuals are old enough to apply for DI under the relaxed eligiblity criteria.

[20]This approach has been used by Fevang et al. (2017) to estimate the effect of temporary disability insurance benefits on the duration of temporary disability insurance spells using policy variation in Norway and Nielsen et al. (2010) to estimate the response of college enrollment to changes in student aid using a Danish reform.

Table 3

6 Welfare Implications

6.1 Austria

Figures 7 and 8 show that the effects of the RSA increases differ by age relative to the RSA. To capture the total effect of the reform we aggregate the effects by taking the average of the estimates from RSA to RSA+3. We choose this period, because the estimates before the RSA are all insignificant and RSA+3 is the last period we observe. We therefore use the averaged estimates in our static sufficient statistics model.

For the welfare implications, it is important to decompose the DI effect into the behavioral and the mechanical effect. The behavioral effect is the change in the application probability multiplied by the conditional award probability (see equation 8). Our empirical analysis delivers a direct estimate of the change in the application probability. Estimating the conditional award probability is more tricky. Taking the average award probability before the reform is likely to overestimate the behavioral effect, because the individuals who react by no longer applying, i.e. those who drive the change in the application probability, tend to have a lower than average award probability. We therefore need a measure for the average award probability of the marginal applicants (those who react to the reform by no longer applying). We construct this award probability from our estimates of the benefit elasticity. The marginal applicants should be the same for changes in screening stringency and changes in benefits. If only benefits change we can observe the effect on applications and inflow, where the inflow effect is simply the application effect multiplied by the award rate of the marginal applicants. Therefore, to back out the award probability of marginal applicants, we simply divide the inflow effect by the application effect in table 3. This constructed award probability multiplied by the change in the application behavior delivers the behavioral effect. The mechanical effect is constructed as a residual by subtracting the behavioral effect from the DI effect. These estimates as well as the implied sufficient statistics $\epsilon = \frac{\text{behavioral effect}}{\text{mechanical effect}}$, $-\frac{b_r}{b_d}S = -\frac{b_r}{b_d}\frac{\text{benefit substitution}}{\text{mechanical effect}}$ and the monetary multiplier $1 + \epsilon - \frac{b_r}{b_d}S$ are reported in Table 4 for men and Table 5 for women.[21]

[21]The estimates in Tables are derived from the point estimates shown in Tables 6 and 7 in Appendix A. We set the replacement rates to $r_d = 0.7$ and $r_r = 0.5$.

Table 4

Table 5

Figure 11 plots the LHS and RHS of the sufficient statistics formula (11) for different values of risk aversion. We find that shifting the RSA by one year is welfare improving if risk aversion $\gamma < 1.88$ for men. For women, stricter screening is welfare improving if $\gamma < 1.7$ for women. For women stricter screening is not welfare improving if $\gamma > 2.4$, for values of risk aversion $1.7 \leq \gamma \leq 2.4$ our sufficient statistics are not informative for women. . Estimates from the literature suggest that the coefficient of relative risk aversion is below 2, Chetty (2006) finds an upper bound of $\gamma \leq 1.78$.

To test the benefit generosity, we plug the estimates from table 3 into the sufficient statistics formula (18). As illustrated in figure 12, we find that for a reasonable range of risk aversion around $\gamma = 1$ the sign of the welfare derivative changes. Hence, we cannot reject that the benefit generosity is optimal. Figure 13 plots the optimal policy mix for different values of risk aversion. For a risk aversion of $\gamma = 1.2$, the optimal direction implies that for increasing benefits by 1 percent, screening should be stricter such that the award rate falls by 4-5.8 percent.

Figure 11

Figure 12

Figure 13

6.2 United States

The U.S. DI eligibility criteria are also subject to vocational factors. This medical-vocational grid introduces sharp discontinuities in initial award rates by age. Chen and van der Klaauw (2008) use these discontinuities to estimate the labor supply effects of DI benefit receipt. We use their estimates for our sufficient statistics formula to discuss the welfare effects of abolishing/shifting these age cutoffs in the U.S. We find that in the U.S. abolishing/shifting these age cutoffs would be welfare reducing for reasonable values of risk aversion.

In contrast to Austria the U.S. age cutoffs do not seem to affect application behavior. There is no strategic bunching of applications at these ages, see Figure 6 in Chen and van der Klaauw (2008).

Chen and van der Klaauw (2008) argue that these rules are not well-known among DI applicants and therefore there is no systematic sorting around the age cutoffs. This has two implications. First, in absence of systematic selection around the age cutoffs, the RDD in Chen and van der Klaauw (2008) is valid. Second, there is no behavioral response with respect to these age cutoffs, i.e. in our sufficient statistics formula we have $\epsilon = 0$. According to the RDD estimates of Chen and van der Klaauw (2008), tightening DI eligibility criteria would increase labor force participation by 5-20 percentage points. Hence, if the medical-vocational grid was abolished, at least 80% of the rejected applicants would substitute to other benefits. In our sufficient statistics formula this corresponds to $S = 0.8$. This is consistent with Maestas et al. (2013). They find that for applicants on the margin of program entry, employment would have been 28 percentage points higher had they not received benefits. This would imply a lower bound of $S = 0.72$. However, this is the estimate for marginal applicants. Stricter screening is likely to affect also non-marginal applicants for whom we expect smaller employment effects and therefore higher benefit substitution. Therefore, setting $S = 0.8$. seems reasonable. Following Autor and Duggan (2003) we set the DI replacement rate to $r_d = 0.47$. Autor and Duggan (2003) (Table 1) report a DI replacement rate of 0.47 for the median earner of the age group 55-61 in the year 1999.[22] For the replacement rate of the welfare benefits we take the average replacement rates over 5 years of unemployment. We take this measure because in our model the welfare benefits reflect the benefits of a long term unemployed. Moreover, the five year period is close to our setup because we look at the effect of screening at age 55 and early retirement starts at age 62. According to the OECD tax and benefit models[23] the average replacement rate for individuals unemployed for 5 years is 29 percent in 2001 in the U.S..[24] This replacement rate also contains other benefits such as housing assistance and social assistance and therefore resembles the idea of other benefits in our model well. We therefore set $r_r = 0.29$.

These numbers imply that the monetary multiplier of stricter screening is 0.5, i.e. for every dollar saved due to stricter screening there are at least 50 cents of additional expenditures in other benefits. Plugging these numbers into our sufficient conditions for $\frac{dW}{d\theta^*} < 0$, equations (38) and (40)

[22] The replacement rates for below median earners are substantially higher. Taking higher DI replacement rates would only strengthen the results reported for the screening stringency. Higher replacement rates, however, might change the conclusions for the benefit generosity. Nevertheless, this does not change our conclusion that abolishing the vocational grid at age 55 is welfare reducing.

[23] see http://www.oecd.org/els/benefits-and-wages-statistics.htm

[24] 2001 is the earliest available year and closest to Chen and van der Klaauw (2008)'s time horizon 1990-1996.

in Appendix B.2, shows that abolishing/shifting these age cutoffs in the U.S. is welfare reducing if the coefficient of relative risk aversion, γ, is larger than 0.5. The sufficient statistics formula is plotted in figure 14 for different values of risk aversion. The value of γ is disputed. Chetty (2006) provides a range for γ from 0.15 to 1.78. The estimates of γ are highly context specific. Taking the methodology from Chetty (2006) and the income effect estimates from Gelber et al. (2017) we bound $\gamma \geq 1$. Following Chetty (2006) we have $\gamma = -\left(1+\frac{wl}{y}\right)\frac{\varepsilon_{l,y}}{\varepsilon_{l^c,w}} + \left(1+\frac{wl}{y}\right)\varepsilon_{u_{c,l}} \geq -\frac{\varepsilon_{l,y}}{\varepsilon_{l^c,w}}$, where $\varepsilon_{l,y}$ is the income elasticity of labor supply and $\varepsilon_{l^c,w}$ is the compensated wage elasticity of labor supply. Gelber et al. (2017) estimate the income effects for DI recipients with a Regression Kink Design and find $\varepsilon_{l,y} = 0.2$. According to their discussion in section V, income effects explain the largest part of the work disincentive effect of DI benefit receipt. This implies $-\frac{\varepsilon_{l,y}}{\varepsilon_{l^c,w}} > 1$ and with Chetty (2006)'s bounds it follows that $\gamma \geq 1$.[25] Therefore, making screening stricter at these age cutoffs is welfare reducing for reasonable values of risk aversion. Put differently, screening at ages below 55 is too strict.

To test the optimality of the benefit generosity we need an estimate of the DI take-up elasticity wrt. benefits. According to Low and Pistaferri (2015) (Table 7) empirical estimates of the application benefit elasticity range from 0.2 to 1.3. Low and Pistaferri (2015)'s model implies an application benefit elasticity of 0.62. For our sufficient-statistics formula we need the take-up elasticity. To obtain an upper bound of the take-up elasticity we multiply the application elasticity with the average award rate. This gives an upper bounds since the individuals who actually react to the benefits should have lower than average award rates. According to French and Song (2014) the award rate after 10 years from the initial application is 0.67. Hence, we get a take-up elasticity $\xi = 0.41$. With this take-up elasticity we reach the conclusion that benefits are too low for the given screening stringency if $\gamma > 0.4$ for CRRA and $\gamma > 0.75$ for the Taylor Approximation. Moreover, if we assume that $\gamma \approx 1$ then the take-up elasticity would need to be larger than 1 such that $\frac{dW}{db_d} > 0$ no longer holds. A DI take-up elasticity above one would imply an application elasticity of above 1.5, which is outside the range mentioned by Low and Pistaferri (2015). For reasonable values of risk aversion we therefore find that $\frac{dW}{d\theta^*} < 0$ and $\frac{dW}{db_d} > 0$.[26]

Our sufficient statistics formulas imply that benefit generosity should be increased and screening

[25] This approach bounds the coefficient of risk aversion for DI recipients. However, for our implementation of the sufficient statistics formula we have to assume that risk aversion is independent of the ability state anyway.

[26] Low and Pistaferri (2015) use $\gamma = 1.5$, hence for this level of risk aversion our results certainly hold.

should be more lenient in the U.S. for individuals older than 55. According to the discussion in section 2.4 we can also determine the optimal policy mix. Figure 15 plots the optimal range of policy mixes for different values of risk aversion.

Figure 14

Figure 15

Our findings are in line with the conclusions in Low and Pistaferri (2015). They conduct the same policy experiments, that we study with our sufficient-statistics model, in their structural model. They (i) change the generosity of DI benefits and (ii) make screening stricter. While they study the effects for the full population in a life-cycle model, our analysis focuses on the local effect at age 55. Nevertheless, we reach the same conclusions. Low and Pistaferri (2015) find that reforms, which increase benefit generosity or relax screening stringency, are welfare improving. We additionally study the optimal policy mix, this question is not studied in Low and Pistaferri (2015).

7 Conclusion

There are two ways to slow the rate at which workers exit the labor force and enter the DI program: tightening standards for DI eligiblity and lowering benefit levels. In this paper, we seek to understand the welfare effects of stricter DI eligibility criteria versus lower DI benefits by developing sufficient-statistics formulas, which capture the insurance value and incentive cost of these two measures. We implement the formulas by estimating all the relevant treatment effects empirically for Austria.

To estimate the effects of stricter disability screening, we exploit variation in DI eligibility strictness that is generated by a policy reform. Prior to 2013 DI eligibility standards were significantly relaxed for workers above age 57 relative to those below age 57. A 2013 pension reform increased the relaxed screening age (RSA) threshold from age 57 to age 58, followed by further increases to age 59 in 2015 and age 60 in 2017. These step-wise increases generate quasi-experimental variation in the strictness of DI eligiblity at a certain age by date of birth. Comparing younger and older birth cohorts over time, we find that being above RSA threshold increases the DI recipiency rate by 5.8 percentage points among men and 8.6 percentage points among women. The net effect of stricter eligibility criteria on DI awards can be decomposed into a behavioral effect, capturing that less people apply for benefits, and a mechanical effect, capturing that fewer applicants qualify for benefits under the stricter rules. We find that being above the RSA threshold increases the probability to apply for benefits by 3 percentage points for men and 4.1 percentage points for women. Together these estimates imply that the mechanical effect accounts for the majority of the increase in DI inflow (74-86 percent) above the RSA while the behavioral effect is less important (14-26 percent).

To examine the impacts of changes in DI benefit levels, we exploit a large pension reform that reduced potential benefit levels for most individuals, although pension levels increased for some individuals with limited work experience. Our estimates suggest that DI applications and DI inflow are sensitive to the level of DI benefits. More specifically, over our sample period we estimate an elasticity of DI applications with respect to benefit levels of 0.7 for men and 0.9 for women. We estimate a smaller elasticity for DI claiming of 0.2 for both men and women.

Plugging these estimates into our sufficient statistics formulas, we find that eligiblity criteria for older workers in Austria were too generous before the reform. Thus, the RSA increases did improve overall welfare. On the other hand, we find that cutting DI benefits likely reduced overall welfare. Importantly, these conclusions are specific for Austria and cannot be generalized to other contexts. However, our sufficient statistics formulas can be used to make welfare statements in other contexts as well. To illustrate this point, we implement our formulas for the United States using empirical estimates from existing U.S. studies. This exercise illustrates that in the U.S. context relaxing DI eligiblity standards and increasing benefits would likely improve welfare.

References

Autor, David H. and Mark G. Duggan. 2003. "The Rise in the Disability Rolls and the Decline in Unemployment." *Quarterly Journal of Economics*, *118* (1): 157–206.

Baily, Martin Neil. 1978. "Some aspects of optimal unemployment insurance." *Journal of Public Economics*, *10* (3): 379–402.

Chen, Susan and Wilbert van der Klaauw. 2008. "The Work Disincentive Effects of the Disability Insurance Program in the 1990s." *Journal of Econometrics*, *142* (2): 757–784.

Chetty, Raj. 2006. "A New Method of Estimating Risk Aversion." *American Economic Review*, *96* (5): 1821–1834.

Chetty, Raj. 2008. "Moral Hazard versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy*, *116* (2): 173–234.

Chetty, Raj and Amy Finkelstein. 2013. "Social Insurance: Connecting Theory to Data." In Alan J. Auerbach, Raj Chetty, Martin Feldstein, and Emmanuel Saez, eds., *Handbook of Public Economics: Volume 5*, chapter 3, Elsevier, pp. 111–193.

de Jong, Philip, Maarten Lindeboom, and Bas van der Klaauw. 2011. "Screening Disability Insurance Applications." *Journal of the European Economic Association*, *9* (1): 106–129.

Denk, Oliver and Jean-Baptiste Michau. 2013. "Optimal Social Security with Imperfect Tagging." Working Papers hal-00796521, HAL.

Diamond, Peter A. and Eytan Sheshinski. 1995. "Economic Aspects of Optimal Disability Benefits." *Journal of Public Economics*, *57* (1): 1–23.

Fevang, Elisabeth, Inés Hardoy, and Knut Rø ed. 2017. "Temporary Disability Insurance and Economic Disincentives." *The Economic Journal*, *127* (603): 1410–1432.

French, Eric and Jae Song. 2014. "The effect of disability insurance receipt on labor supply." *American Economic Journal: Economic Policy*, *6* (2): 291–337.

Gelber, Alexander, Timothy J. Moore, and Alexander Strand. 2017. "The Effect of Disability Insurance Payments on Beneficiaries' Earnings." *American Economic Journal: Economic Policy*, *9* (3): 229–61.

Karlström, Anders, Må rten Palme, and Ingemar Svensson. 2008. "The Employment Effect of Stricter Rules for Eligibility for DI: Evidence from a Natural Experiment in Sweden." *Journal of Public Economics*, *92* (10-11): 2071–2082.

Kroft, Kory. 2008. "Takeup, Social Multipliers and Optimal Social Insurance." *Journal of Public Economics*, *92* (3-4): 722–737.

Kroft, Kory and Matthew J. Notowidigdo. 2011. "Should Unemployment Insurance Vary With the Unemployment Rate? Theory and Evidence." *Mimeo*: 1–53.

Landais, Camille. 2012. "Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design." *Mimeo*, (July).

Landais, Camille, Pascal Michaillat, and Emmanuel Saez. 2010. "Optimal Unemployment Insurance Over the Business Cycle." *NBER Working Paper*, *16526*: 1–80.

Low, Hamish and Luigi Pistaferri. 2015. "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off." *American Economic Review*, *105* (10): 2986–3029.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand. 2013. "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt." *American Economic Review*, *103* (5): 1797–1829.

Moore, Timothy J. 2015. "The employment effect of terminating disability benefits." *Journal of Public Economics*, *124*: 30–43.

Mullen, Kathleen J. and Stefan Staubli. 2016. "Disability Benefit Generosity and Labor Force Withdrawl." *Journal of Public Economics*, *143*: 49–63.

Nielsen, Helena Skyt, Torben Sorensen, and Christopher Taber. 2010. "Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform." *American Economic Journal: Economic Policy*, *2*: 185–212.

Parsons, Donald O. 1996. “Imperfect "Tagging" in Social Insurance Programs.” *Journal of Public Economics*, *62* (1-2): 183–207.

Ruh, Philippe and Stefan Staubli. 2016. “Financial Incentives and Earnings of Disability Insurance Recipients: Evidence from a Notch Design.” *Working Paper University of Zurich*.

Schmieder, Johannes F., Till von Wachter, and Stefan Bender. 2012. “The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates Over 20 Years.” *Quarterly Journal of Economics*, *127* (2): 701–752.

Shimer, Robert and Iván Werning. 2007. “Reservation Wages and Unemployment Insurance.” *Quarterly Journal of Economics*, *122* (3): 1145–1185.

Staubli, Stefan. 2011. “The Impact of Stricter Criteria for Disability Insurance on Labor Force Participation.” *Journal of Public Economics*, *95* (9-10): 1223–1235.

Staubli, Stefan and Josef Zweimüller. 2013. “Does Raising the Early Retirement Age Increase Employment of Older Workers?” *Journal of Public Economics*, *108*: 17–32.

Wörister, Karl. 1999. “Länderbericht: Österreich.” In *Invalidenrenten: Europäische Trends und Politik*, Bern: Bundesamt für Statistik.

Zweimüller, Josef, Rudolf Winter-Ebmer, Rafael Lalive, Andreas Kuhn, Jean-Philipe Wuellrich, Oliver Ruf, and Simon Büchi. 2009. “Austrian Social Security Database.” *Working Paper University of Zurich*, *401*: 1–80.

Table 1: Summary statistics

	Men		Women	
	Eligible	Not eligible	Eligible	Not eligible
	(1)	(2)	(3)	(4)
Ever applied for DI (in %)	12.85	31.09	15.60	22.64
DI benefit receipt (in %)	9.26	31.67	9.82	22.38
Employment (in %)	81.35	17.66	79.47	15.71
Registered unemployment (in %)	5.92	18.04	6.66	8.79
Sick leave absence (in %)	1.43	1.73	1.37	0.92
Annual earnings	36,419	4,335	23,545	2,795
	(22,601)	(11,831)	(18,560)	(7,900)
Avg. earnings best 15 yrs.	40,339	24,714	23,806	15,912
	(10,568)	(15,139)	(11,132)	(10,557)
Any sick leave by age 50 (in %)	33.25	49.58	32.38	36.43
Unemployment years until age 50	0.53	2.03	0.65	1.24
	(1.07)	(2.82)	(1.20)	(2.21)
Employment years until age 50	13.87	5.53	11.58	3.65
	(2.02)	(5.09)	(3.29)	(4.46)
Insurance years until age 50	28.78	16.47	22.54	14.35
	(7.00)	(12.35)	(7.02)	(10.59)
No. Observations	2,481,863	1,039,503	1,153,363	1,317,823
No. Individuals	104,408	44,602	48,451	55,611

Notes: Table presents summary statistics.

Table 2: Average effect above RSA in percent

	Men		Women	
DI benefits	6.04***	(0.92)	8.24***	(0.90)
pre-reform mean	16.2		15.2	
DI applications	3.11***	(0.99)	3.98***	(0.92)
pre-reform mean	19.6		20.8	
Employment	-2.61***	(0.87)	-3.09***	(1.02)
pre-reform mean	74.2		72.7	
Other benefits	-3.56***	(0.62)	-4.81***	(1.05)
pre-reform mean	7.1		9.1	
No. Obs.	2,481,863		1,153,363	

Notes: Tabe presents average effect of the RSA for the ages above the RSA. The estimates are constructed by summing up all the β_k-coefficients from equation 26 for $k \geq 0$ and dividing by 3,the number of years above the RSA.

Table 3: Benefit elasticity estimates

	DI applications		DI inflow	
	Men (1)	Women (2)	Men (3)	Women (4)
A. Ages 50-59				
Log benefit	0.0328***	0.0354***	0.0062*	0.00387***
	(0.0065)	(0.0052)	(0.0035)	(0.0013)
F-test	57.14	65.11	10.96	25.69
Prob>F	<0.001	<0.001	0.001	<0.001
Observations	4,660,563	3,878,099	4,660,563	3,878,099
Avg. dependent variable	0.0500	0.0377	0.0274	0.0156
Implied elasticity	0.657***	0.939***	0.225*	0.249***
	(0.130)	(0.138)	(0.127)	(0.083)
B. Ages 55-59				
Log benefit	0.0402***	0.0267***	0.0198***	0.0078***
	(0.0080)	(0.0048)	(0.0046)	(0.0022)
F-test	54.36	45.57	31.48	18.44
Prob>F	<0.001	<0.001	<0.001	<0.001
Observations	1,992,981	1,411,616	1,992,981	1,411,616
Avg. dependent variable	0.0760	0.0467	0.0470	0.0231
Implied elasticity	0.529***	0.572***	0.421***	0.344***
	(0.106)	(0.102)	(0.099)	(0.097)

Notes: Tabe presents estimates of the application and benefit elasticity, respectively, using the empirical approach specified in Mullen and Staubli (2016).

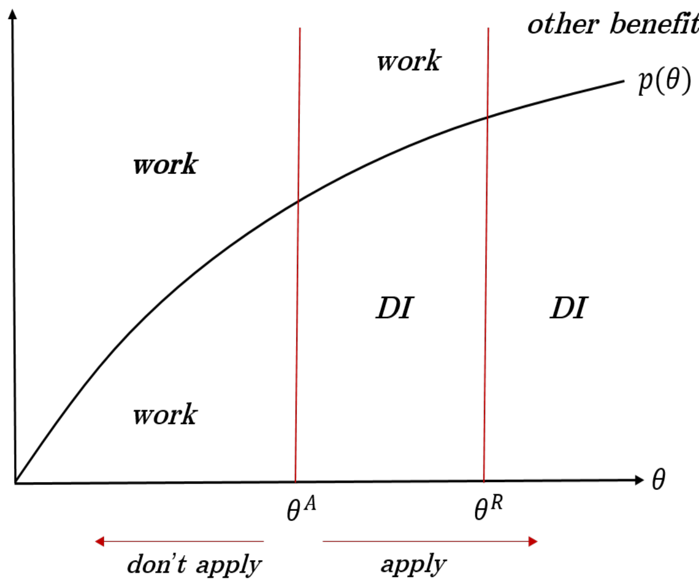
Table 4: Sufficient statistics estimates, men

	Inflow	Behavioral	Mechanical	ϵ	$-\frac{b_r}{b_d}S$	$(1+\epsilon-\frac{b_r}{b_d}S)$
Men	6.04	1.53	4.51	0.34	-0.56	0.77
Never on sick leave	2.49	1.29	1.20	1.08	-0.58	1.49
Was on sick leave	8.13	2.46	5.67	0.43	-0.61	0.83
Never unemployed	3.14	2.23	0.91	2.44	-0.56	2.87
Was unemployed	7.16	1.86	5.29	0.35	-0.61	0.74
Lifetime earnings Q1	8.37	2.27	6.10	0.37	-0.74	0.63
Lifetime earnings Q2	8.75	3.03	5.72	0.53	-0.32	1.21
Lifetime earnings Q3	6.90	2.78	4.12	0.67	-0.80	0.88
Lifetime earnings Q4	1.50	1.32	0.18	7.46	-1.33	7.13
DI replacement rate Q1	3.95	-0.34	4.29	-0.08	-0.85	0.07
DI replacement rate Q2	5.57	3.38	2.19	1.54	-0.78	1.76
DI replacement rate Q3	6.29	3.89	2.40	1.62	-0.32	2.30
DI replacement rate Q4	9.74	4.38	5.36	0.82	-0.48	1.33

Table 5: Sufficient statistics estimates, women

	Inflow	Behavioral	Mechanical	ϵ	$-\frac{b_r}{b_d}S$	$(1+\epsilon-\frac{b_r}{b_d}S)$
Women	8.24	1.20	7.03	0.17	-0.49	0.68
Never on sick leave	3.36	0.83	2.53	0.33	-1.23	0.10
Was on sick leave	13.37	4.39	8.98	0.49	-0.53	0.96
Never unemployed	2.94	-0.61	3.55	-0.17	-0.51	0.32
Was unemployed	11.45	4.55	6.90	0.66	-0.72	0.94
Lifetime earnings Q1	9.65	4.60	5.04	0.91	-0.89	1.02
Lifetime earnings Q2	6.55	-0.51	7.06	-0.07	-0.66	0.27
Lifetime earnings Q3	13.55	5.55	8.01	0.69	-0.58	1.11
Lifetime earnings Q4	5.63	1.68	3.94	0.43	-0.74	0.69
DI replacement rate Q1	5.38	1.17	4.21	0.28	-0.80	0.48
DI replacement rate Q2	5.66	1.79	3.88	0.46	-1.09	0.37
DI replacement rate Q3	8.49	1.82	6.67	0.27	-0.70	0.57
DI replacement rate Q4	16.94	6.85	10.09	0.68	-0.48	1.20

Figure 1: Illustration of model

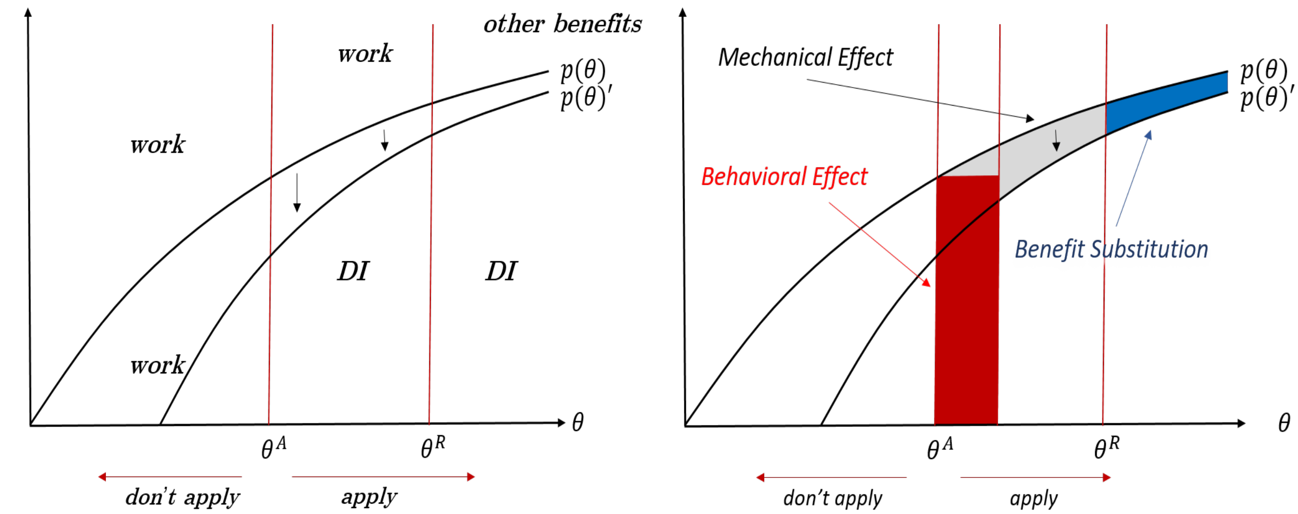


Note: This figure illustrates the basic setup. Individuals are characterized by disutility of labor θ and can choose whether to work, apply to DI or leave the labor force and consume other benefits. The award process to DI is noisy and individuals are awarded DI with probability $p(\theta)$. We assume that $p(\theta)$ is weakly increasing in θ. This captures that (i) it is difficult to assess the true disability level of an individual and (ii) the assessment contains nonetheless some valuable information on the true disutility. The marginal DI applicant is denoted byθ^A and individuals with $\theta \geq \theta^A$ apply to DI. The marginal other benefits type is denoted by θ^R and individuals with $\theta \geq \theta^R$ will go on other benefits if they are rejected.

Figure 2: Effects of stricter screening

(a) Shift of award probability curve and marginal applicant

(b) Different effects

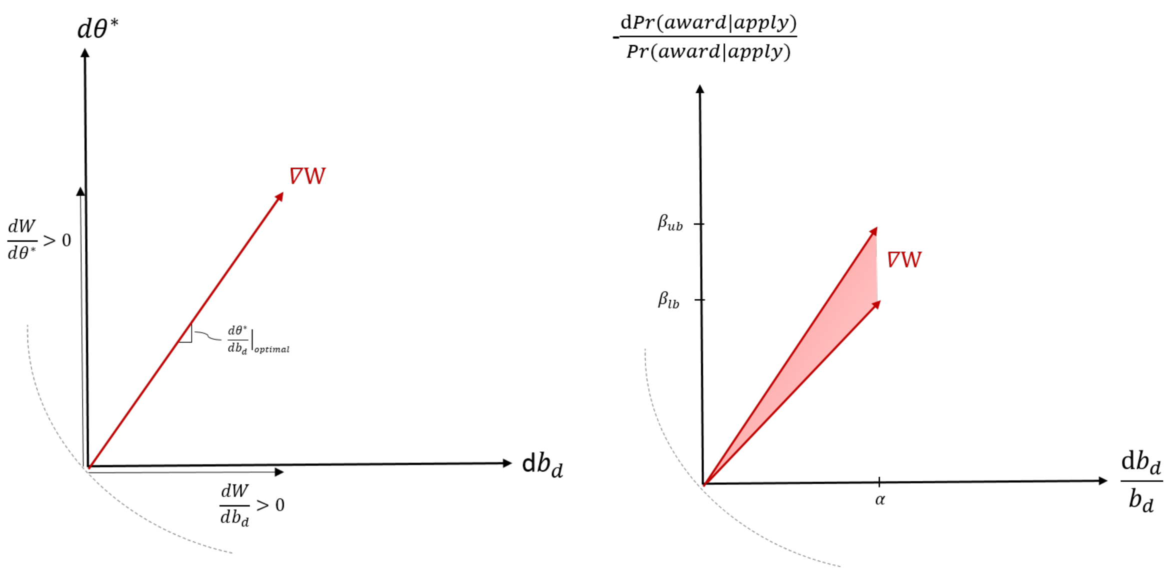


Notes: The figure illustrates the effects of stricter screening. Stricter screening shifts down the award probability curve and shifts the marginal applicant to the right (Panel a). The area between the two award probability curves is the mechanical effect. A fraction of rejected applicants due to the mechncial effect returns to work (gray area). The other fraction substitutes DI benefits with other benefits (blue area). The change in the marginal applicant times the award probability of the marginal applicant is the behavioral effect (red area).

Figure 3: Optimal policy mix

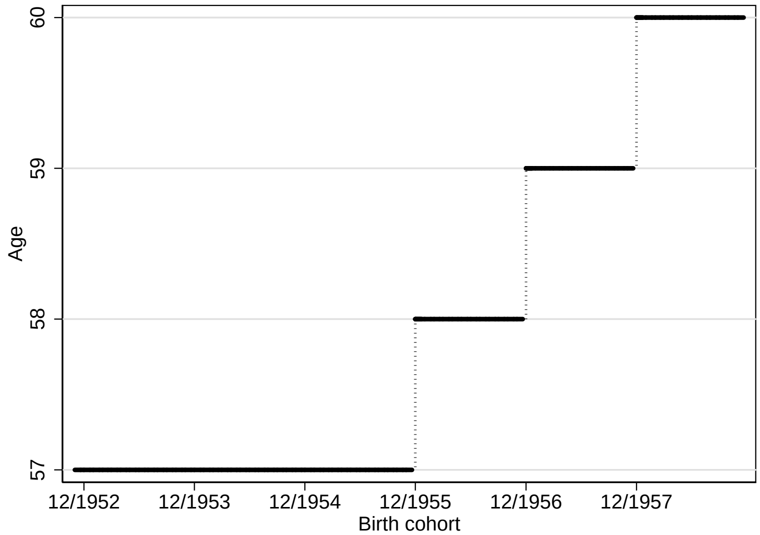
(a) Gradient of Welfare Function

(b) Empirical Implementation



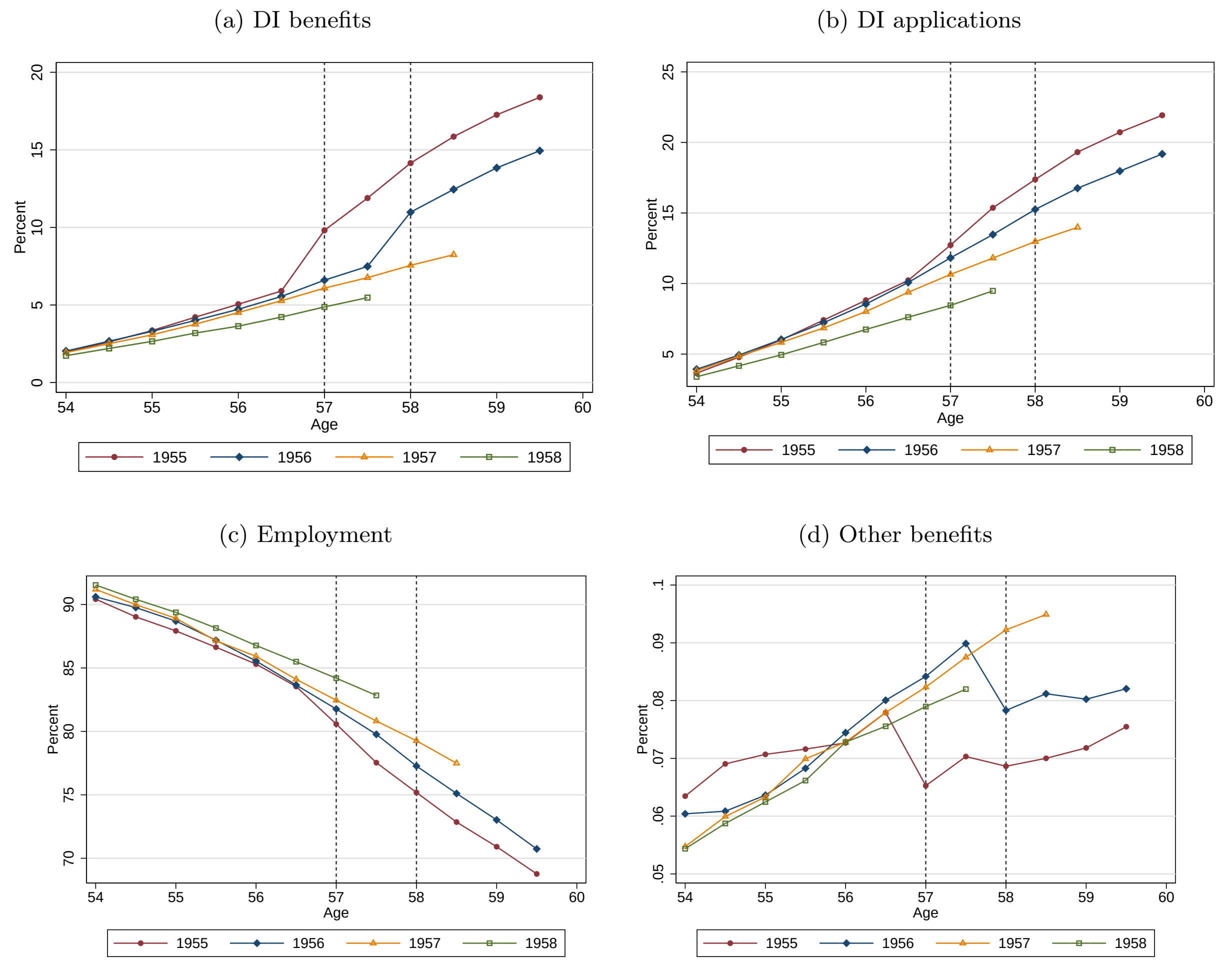
Notes: The figure illustrates the idea of the optimal policy mix. Panel a shows the gradient in case screening should be stricter and benefits should be more generous. The dotted line is the indifference curve of the welfare function of the current benefit level and strictness of screening. The gradient of the welfare function is orthogonal to the indifference curve and points in the direction of greates increase of the function. Panel b shows the empirical implementation of the gradient. Because we have to bound the welfare effects of stricter screening we get a cone of optimal directions for a given level of risk aversion.

Figure 4: Increase in the RSA



Notes: This figure displays all birth cohorts that are used in our subsequent analysis. The 2012 2nd Stability Act implemented an increase in the relaxed screening age (RSA) for DI benefits to age 60.
Source: Austrian federal law (Bundesgesetzblatt) no. 35/2012.

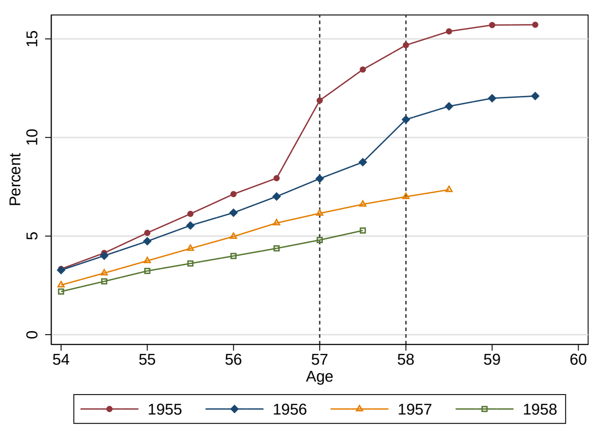
Figure 5: Trends by age and birth year for men



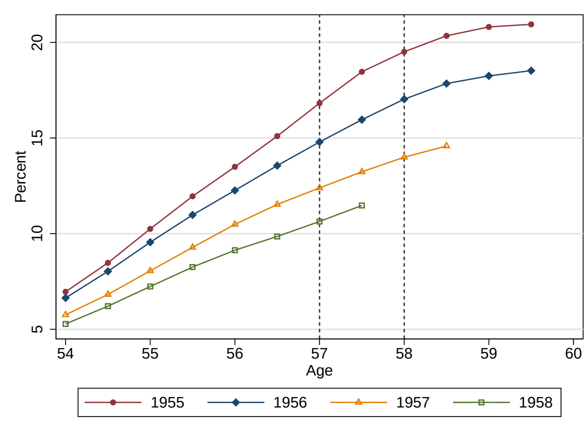
Notes: Figure shows DI inflow rates age and birth year for men and women.
Source: Own calculations, based on Austrian Social Security Data.

Figure 6: Trends by age and birth year for women

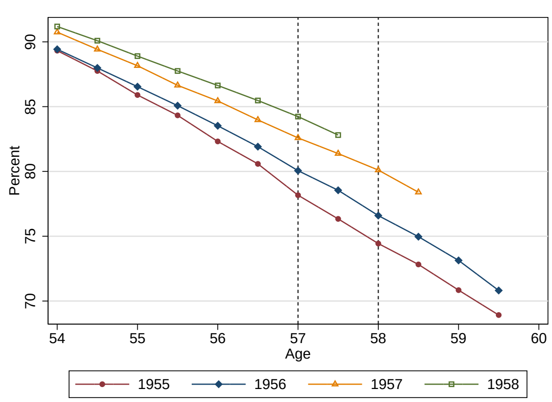
(a) DI benefits



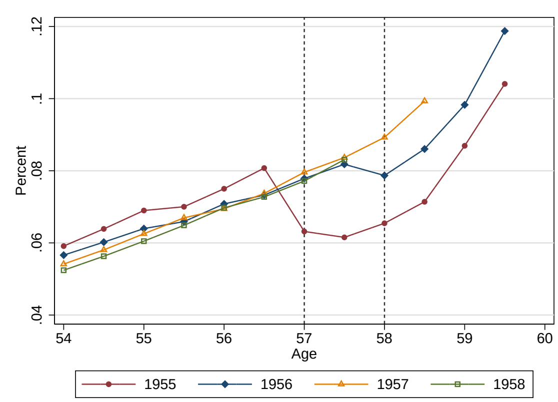
(b) DI applications



(c) Employment



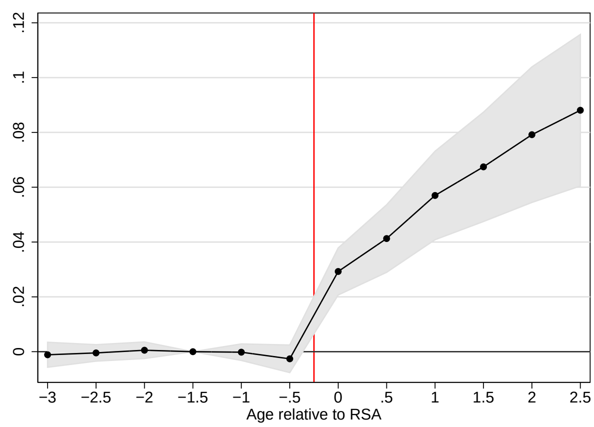
(d) Other benefits



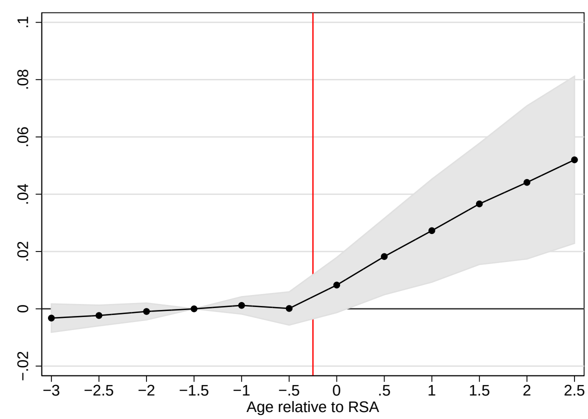
Notes: Figure shows DI inflow rates age and birth year for men and women.
Source: Own calculations, based on Austrian Social Security Data.

Figure 7: Estimates for eligible men

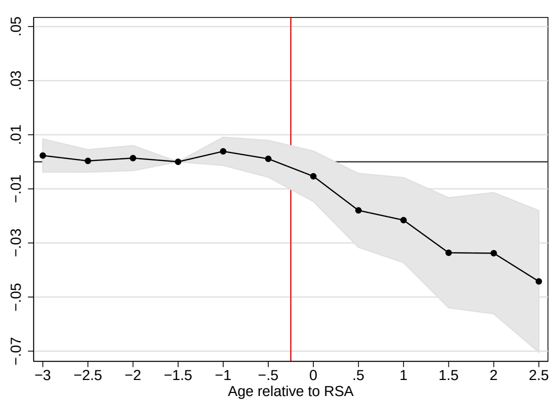
(a) DI benefits



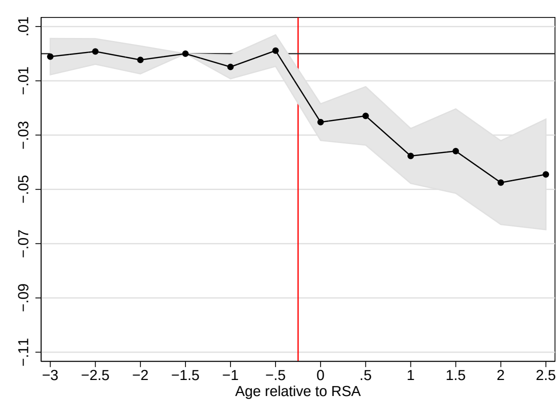
(b) DI applications



(c) Employment



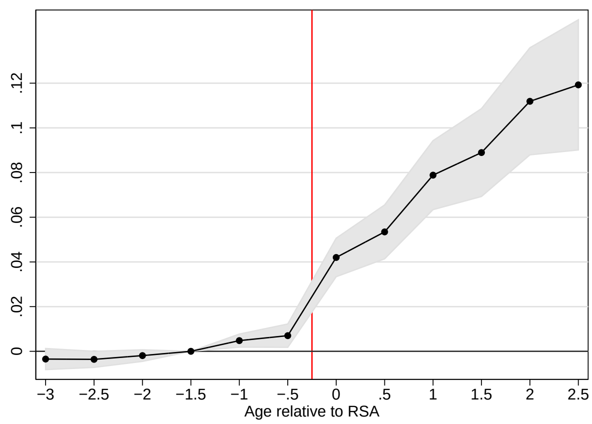
(d) Other benefits



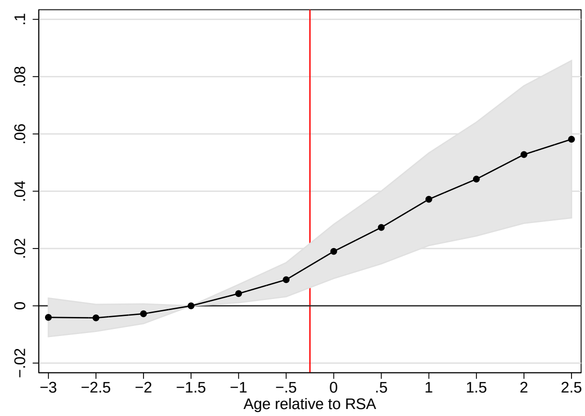
Notes:
Source: Own calculations, based on Austrian Social Security Data.

Figure 8: Estimates for eligible women

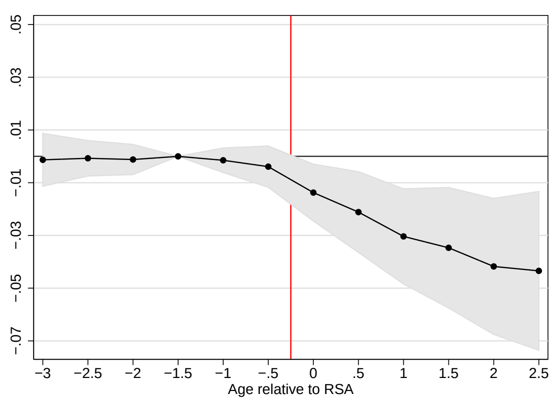
(a) DI level



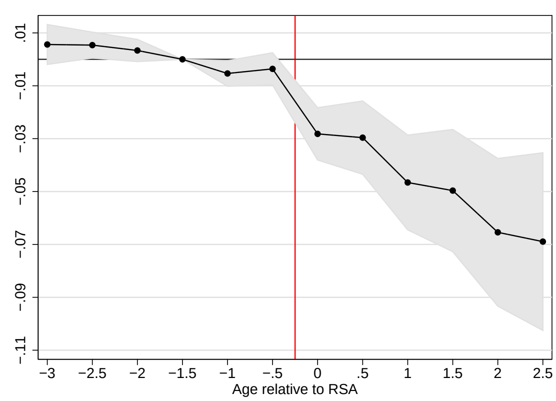
(b) DI applications



(c) Employment

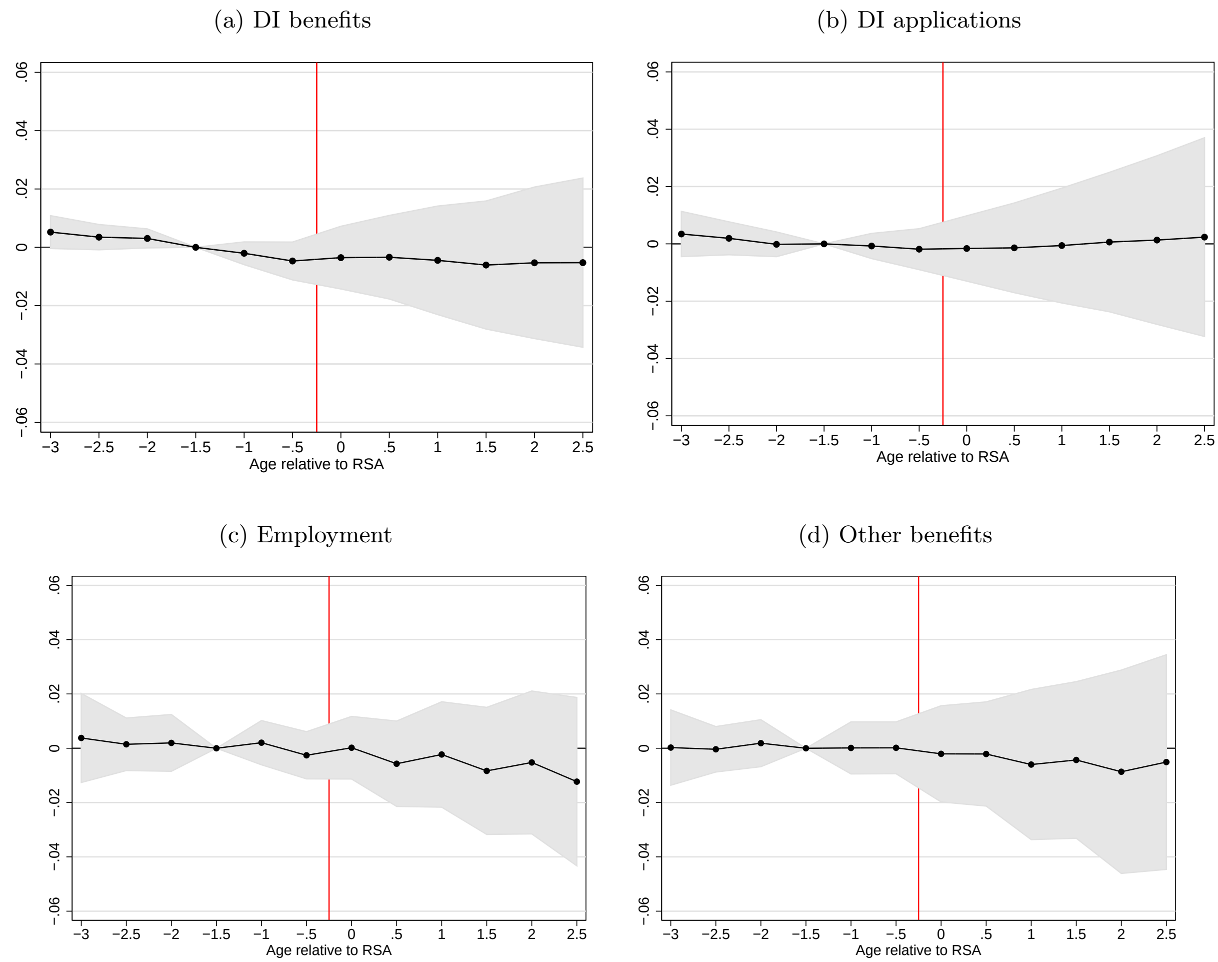


(d) Other benefits



Notes:
Source: Own calculations, based on Austrian Social Security Data.

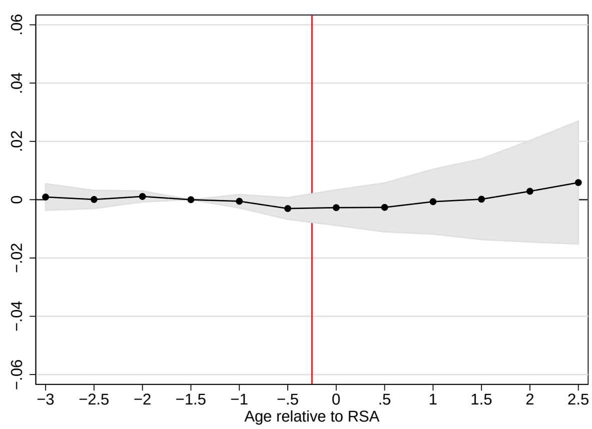
Figure 9: Estimates for ineligible men



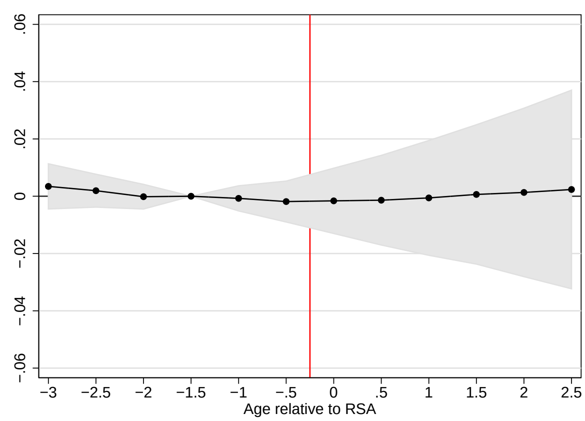
Notes:
Source: Own calculations, based on Austrian Social Security Data.

Figure 10: Estimates for ineligible women

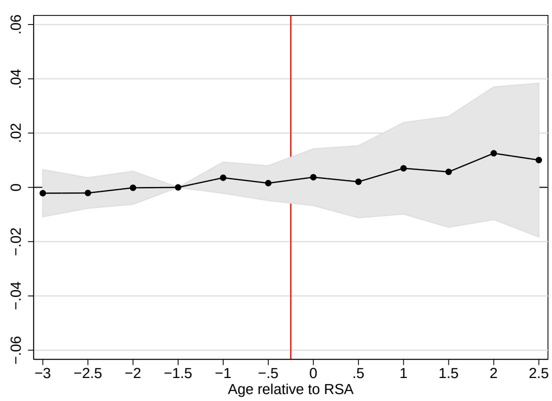
(a) DI level



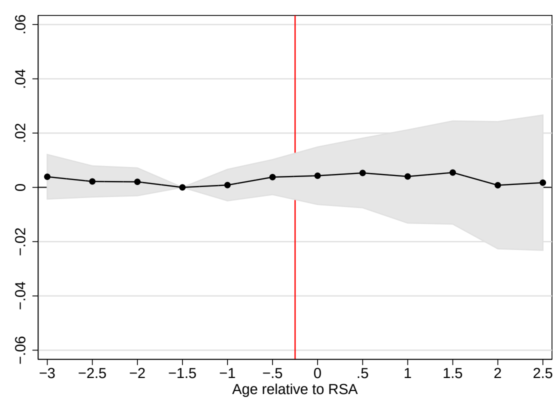
(b) DI applications



(c) Employment



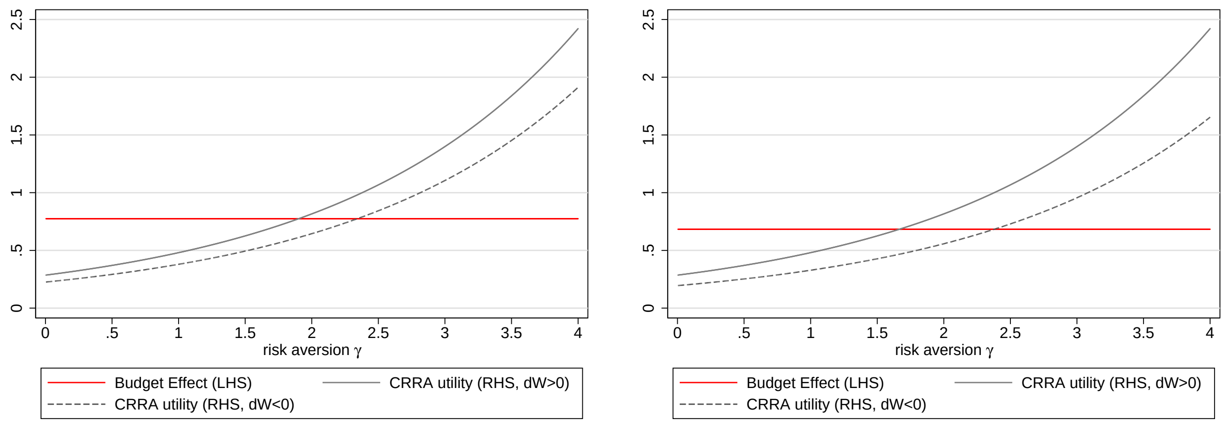
(d) Other benefits



Notes:
Source: Own calculations, based on Austrian Social Security Data.

Figure 11: Screening Stringency Austria

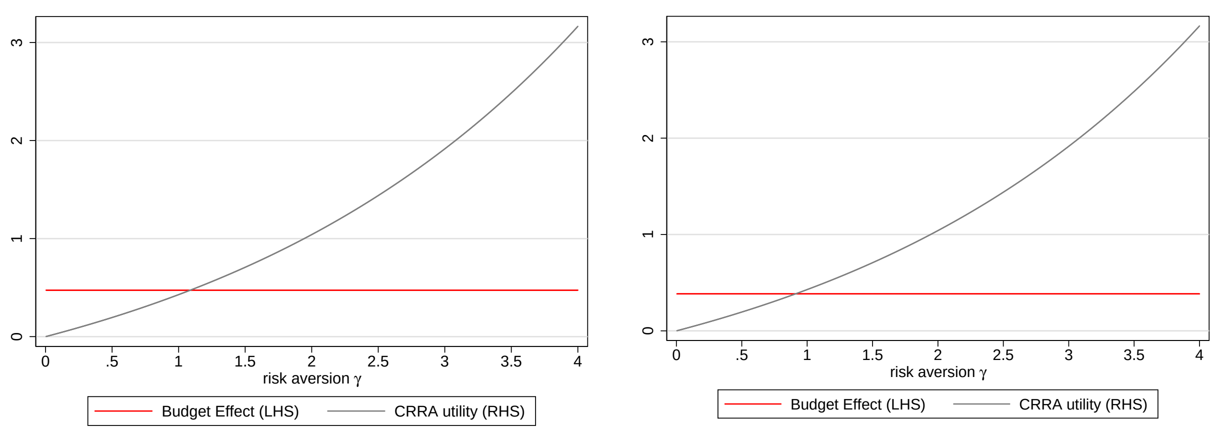
(a) Men (b) Women



Notes: Figure plots the LHS and the upper and lower bounds of the RHS of inequality (11) for men in panel (a) and women in panel (b) against different levels of risk aversion. If risk aversion is lower than the point where the solid grey line crosses the red line, then it is welfare improving to increase screening stringency. If risk aversion is higher than the point where the dashed grey line crosses the red line, then it is welfare improving to reduce screening stringency. For levels of risk aversion between these two points our sufficient statistics condition do not allow for a welfare statement.
Source: Own calculations.

Figure 12: Benefit Generosity Austria

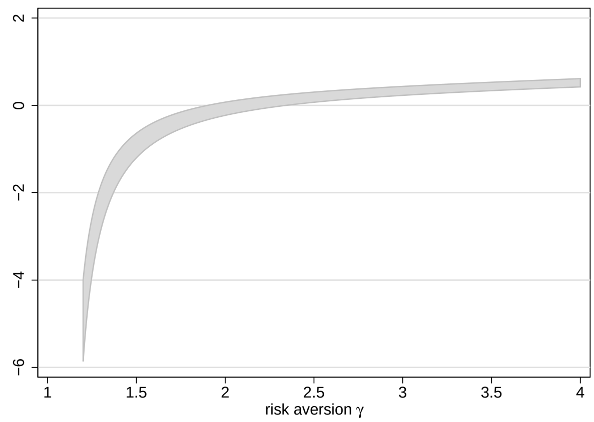
(a) Men (b) Women



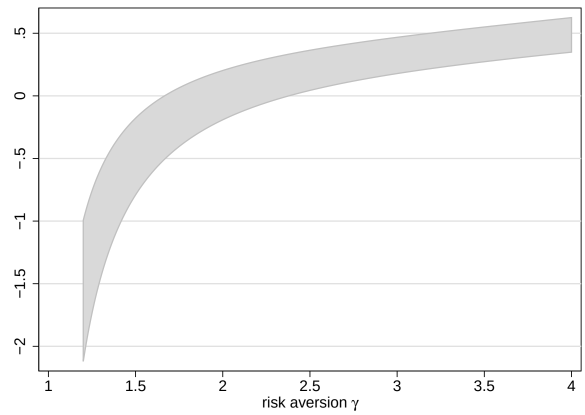
Notes: Figure plots the LHS and RHS of inequality (18) for men in panel (a) and women in panel (b) against different levels of risk aversion. If risk aversion is higher than the point where the grey line crosses the red line, then it is welfare improving to increase benefit generosity. If risk aversion is lower than this point, it is welfare improving to reduce benefit generosity.
Source: Own calculations.

Figure 13: Optimal Policy Mix Austria: Slope of Gradient

(a) Men



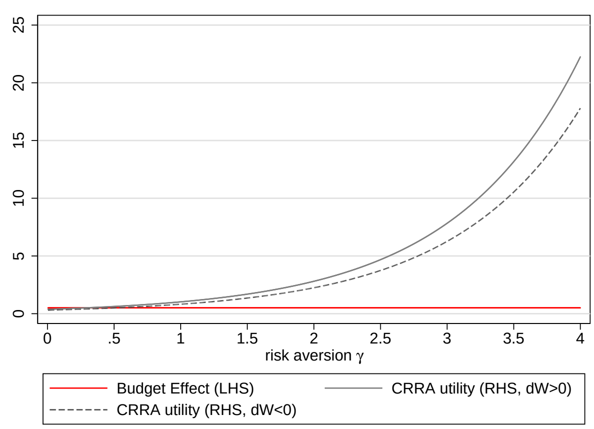
(b) Women



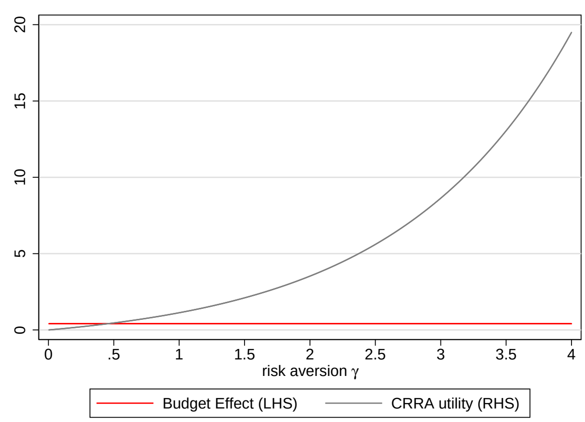
Notes: Figure plots the slope of the gradient β/α according to equation (21). The upper and lower bounds of this optimal direction are according to the implementation bounds in equations (22)-(24).
Source: Own calculations.

Figure 14: Screening Stringency and Benefit Generosity U.S.

(a) Screening Stringency

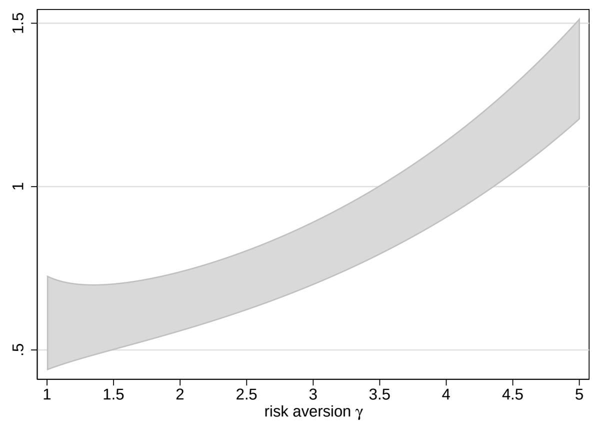


(b) Benefit Generosity



Notes: Figure plots the LHS and RHS of inequality (11) in panel A and inequality (18) in panel B against different levels of risk aversion.
Source: Own calculations, based on estimates from Autor and Duggan (2003), Chen and van der Klaauw (2008), Low and Pistaferri (2015) and OECD tax and benefit models as described in the main text.

Figure 15: Optimal Policy Mix U.S.: Slope of Gradient



Notes: Figure plots the slope of the gradient β/α according to equation (21). The upper and lower bounds of this optimal direction are according to the implementation bounds in equations (22)-(24). Interpretation: If the risk aversion was 1.5, the optimal direction rule implies that for a 1 percent increase in benefit generosity, screening stringency should be adjusted such that the award rate increases between 0.5 and 0.7 percent.

Table 6: Heterogeneity Effects Men

	Disability	Application	Employment	Other Benefits	No. Obs.
Full sample	6.04*** (0.92)	3.11*** (0.99)	-2.61*** (0.87)	-3.56*** (0.62)	2,481,863
Never on sick leave	2.49*** (0.86)	1.88** (0.85)	-1.51 (1.03)	-0.91 (0.67)	1,070,952
Was on sick leave	8.13*** (1.20)	3.58*** (1.23)	-3.31*** (1.16)	-4.74*** (0.87)	1,697,064
Never unemployed	3.14*** (0.83)	3.24*** (0.83)	-2.42** (1.19)	-0.33 (0.63)	1,105,104
Was unemployed	7.16*** (1.25)	2.71** (1.11)	-2.65** (1.16)	-4.74*** (0.87)	1,662,912
Lifetime earnings Q1	8.37*** (1.46)	3.30** (1.66)	-2.02 (1.89)	-5.80*** (1.25)	692,016
Lifetime earnings Q2	8.75*** (2.09)	4.41*** (1.38)	-6.21*** (1.96)	-3.47*** (1.28)	692,112
Lifetime earnings Q3	6.90*** (1.54)	4.03** (1.80)	-2.29 (1.58)	-4.60*** (0.89)	691,920
Lifetime earnings Q4	1.50 (0.95)	1.92** (0.98)	-1.17 (1.37)	0.56 (0.74)	691,968
DI replacement rate Q1	3.95*** (1.45)	-0.49 (1.55)	1.15 (1.49)	-5.92*** (1.39)	692,005
DI replacement rate Q1	5.57*** (1.55)	4.91*** (1.77)	-3.18* (1.68)	-2.99** (1.33)	692,013
DI replacement rate Q3	6.29*** (1.42)	5.63*** (1.54)	-5.22*** (1.55)	-1.64 (1.06)	692,009
DI replacement rate Q4	9.74*** (1.73)	6.36*** (1.83)	-6.11*** (1.58)	-1.72 (1.08)	691,989

A Additional Tables and Figures

B Bounds for Insurance Loss of Screening Stringency

Define the RHS in inequality (11) as

$$V_{\theta^*} := \frac{v(c_d) - \left(v(c_r)S + (1-S)\left[u(c_w) - \widetilde{\theta}\right]\right)}{u'(w-\tau)} \tag{28}$$

$$= \frac{u'(w-\tau)^{-1}}{-\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)} \left(-\int_{\theta^A}^{\theta^R} \frac{dp(\theta;\theta^*)}{d\theta^*} \left[v(c_d) - [u(c_w) - \theta]\right] dF(\theta) - \int_{\theta^R}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} \left[v(c_d) - v(c_r)\right] dF(\theta) \right)$$

We derive upper and lower bounds for V_{θ^*}. We then use these bounds to substitute for the RHS in (11). Using an upper bound of V_{θ^*} yields a sufficient condition for a welfare improvement of stricter screening, i.e. for $\frac{dW}{d\theta^*} \geq 0$. Using a lower bound for V_{θ^*}yields a sufficient condition for a

Table 7: Heterogeneity Effects Women

	Disability	Application	Employment	Other Benefits	No. Obs.
Full sample	8.24*** (0.90)	3.98*** (0.92)	-3.09*** (1.02)	-4.81*** (1.05)	1,153,363
Never on sick leave	3.36*** (1.07)	1.21 (1.07)	1.00 (1.36)	-3.75*** (1.10)	594,864
Was on sick leave	13.37*** (1.56)	6.40*** (1.32)	-6.68*** (1.57)	-6.12*** (1.46)	641,064
Never unemployed	2.94*** (1.07)	-0.89 (1.37)	-0.41 (1.48)	-1.85 (1.19)	411,720
Was unemployed	11.45*** (1.28)	6.65*** (1.19)	-4.49*** (1.36)	-6.33*** (1.25)	824,208
Lifetime earnings Q1	9.65*** (1.45)	6.72*** (2.03)	-3.38 (2.13)	-4.03** (1.60)	309,048
Lifetime earnings Q2	6.55*** (2.07)	-0.74 (1.80)	-0.04 (1.96)	-5.85*** (1.88)	308,952
Lifetime earnings Q3	13.55*** (1.84)	8.10*** (1.92)	-7.02*** (2.42)	-5.73*** (1.57)	308,952
Lifetime earnings Q4	5.63*** (1.69)	2.45 (1.64)	-1.56 (1.65)	-4.54*** (1.71)	308,976
DI replacement rate Q1	5.38*** (1.52)	1.72 (1.85)	-0.66 (2.15)	-4.02* (2.06)	308,978
DI replacement rate Q2	5.66*** (1.60)	2.61 (2.16)	0.27 (2.05)	-6.26*** (1.50)	308,968
DI replacement rate Q3	8.49*** (2.05)	2.66 (2.24)	-1.95 (2.51)	-6.13*** (1.39)	308,946
DI replacement rate Q4	16.94*** (2.27)	9.99*** (2.13)	-10.20*** (2.31)	-4.95*** (1.36)	309,036

welfare decrease of stricter screening, i.e. for $\frac{dW}{d\theta^*} \leq 0$. We derive the upper and lower bounds now separately.

B.1 Upper Bounds for Welfare Improvement

First, we use that $\forall \theta \in [\theta^A, \theta^R] \Rightarrow u(w) - \theta \geq v(c_r)$ and therefore

$$V_{\theta^*} \leq \frac{v(c_d) - v(c_r)}{u'(w - \tau)}. \tag{29}$$

Second, we replace consumption c_d and c_r by the benefits b_d and b_r to get an upper bound. Let consumption on benefits j be the sum of benefits b_j and other income/savings s_j, i.e. $c_j = b_j + s_j$. If $s_r \geq 0$ and $s_r \geq s_d$, concavity of $v(c)$ implies $v(c_d) - v(c_r) \leq v(b_d) - v(b_r)$. The condition $s_r \geq 0$ and $s_r \geq s_d$ can be expected to hold since welfare benefits are lower than DI benefits. Therefore, we have

$$V_{\theta^*} \leq \frac{v(b_d) - v(b_r)}{u'(w - \tau)}. \tag{30}$$

The problem of the loss of insurance being in absolute utility terms remains. We tackle this in two ways. First, we assume a functional form of the utility, specifically CRRA, to get

$$\frac{v(b_d) - v(b_r)}{b_d u'(w-\tau)} = \frac{1}{1-\gamma}\left[\frac{b_d^{1-\gamma} - b_r^{1-\gamma}}{b_d(w-\tau)^{-\gamma}}\right] = \frac{1}{1-\gamma}\left[r_d^{-\gamma} - \frac{r_r^{1-\gamma}}{r_d}\right] \tag{31}$$

where $r_r = \frac{b_r}{w-\tau}$ and $r_d = \frac{b_d}{w-\tau}$ are the replacement rates of welfare and disability benefits, respectively. This yields as a sufficient condition for $\frac{dW}{d\theta^*} \geq 0$

$$1 + \epsilon - \frac{r_r}{r_d} S \geq \frac{1}{1-\gamma}\left[r_d^{-\gamma} - \frac{r_r^{1-\gamma}}{r_d}\right]. \tag{32}$$

The second approach is to bound $v(b_d) - v(b_r)$ further by using $v(b_d) - v(b_r) \leq v'(b_r)(b_d - b_r)$, because $b_r < b_d$ and $v(c)$ is concave. This gives us a condition entirely expressed in terms of marginal utilities

$$V_{\theta^*} \leq \frac{v'(b_r)(b_d - b_r)}{u'(w-\tau)}. \tag{33}$$

We then use a Taylor approximation. Again we assume that the utility functions are the same whether on benefits or working, i.e. $v(c) = u(c)$, the Taylor approximation then yields

$$\begin{aligned} \frac{v'(b_r)(b_d - b_r)}{u'(w-\tau)} \approx& (b_d - b_r)\frac{u'(w-\tau) - u''(w-\tau)(w-\tau-b_r) + \frac{1}{2}u'''(w-\tau)(w-\tau-b_r)^2}{u'(w-\tau)} \\ =& (b_d - b_r)\left(1 + \gamma(1-r_r) + \frac{1}{2}\gamma\rho(1-r_r)^2\right) \end{aligned} \tag{34}$$

where $\gamma = -\frac{u''(c)}{u'(c)}c$ is the coefficient of relative risk aversion and $\rho = -\frac{u'''(c)}{u''(c)}c$ is the coefficient of relative prudence. Therefore, the sufficient welfare criteria becomes

$$1 + \epsilon - \frac{r_r}{r_d} S \geq \left(1 - \frac{r_r}{r_d}\right)\left(1 + \gamma(1-r_r) + \frac{1}{2}\gamma\rho(1-r_r)^2\right). \tag{35}$$

The assumption that the utility functions are the same whether on benefits or working, i.e. $v(c) = u(c)$, might not be innocent. It is not clear whether marginal utility of consumption is higher or lower on benefits or when working. Stories could be told for both $v'(c) > u'(c)$ and $v'(c) < u'(c)$. If marginal utility of consumption on benefits is smaller than while working, i.e.

$v'(c) < u'(c)$, then the assumption that $v'(c) = u'(c)$ is another upper bound and condition (35) is still valid. $v'(c) < u'(c)$ could be motivated by lower consumption needs of individuals on benefits because of more leisure time compared to working individuals.

B.2 Lower Bounds for Welfare Decrease

For the lower bound of V_{θ^*} we use that $\forall \theta \in [\theta^A, \theta^R] \Rightarrow u(w) - \theta < v(c_d)$ and therefore $V_{\theta^*} = v(c_d) - v(c_r)S - \frac{\int_{\theta^A}^{\theta^R} \frac{dp(\theta;\theta^*)}{d\theta^*}[u(w)-\theta]dF(\theta)}{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*}dF(\theta)} \geq v(c_d) - v(c_r)S - v(c_d)(1-S) = S\left(v(c_d) - v(c_r)\right).$ Hence,

$$V_{\theta^*} \geq S\frac{v(c_d) - v(c_r)}{u'(w-\tau)}. \tag{36}$$

As before we can either directly plug in CRRA utility or further bound $v(c_d) - v(c_r)$ and then use a Taylor approximation. Assuming CRRA utility we get

$$V_{\theta^*} \geq \frac{S}{1-\gamma}\left[\frac{c_d^{1-\gamma} - c_r^{1-\gamma}}{(w-\tau)^{-\gamma}}\right]. \tag{37}$$

For this inequality substituting benefits for consumption does not deliver a lower bounds. However, if consumption and benefit levels are close enough then the inequality will still be valid and we can express everything in replacement rates. If this were not the case then the replacement rates have to be understood as replacement rates of consumption levels. A sufficient condition for $\frac{dW}{d\theta^*} \leq 0$ is then

$$1 + \epsilon - \frac{r_r}{r_d}S < \frac{S}{1-\gamma}\left[r_d^{-\gamma} - \frac{r_r^{1-\gamma}}{r_d}\right]. \tag{38}$$

Alternatively, we can use as a lower bound $v(c_d) - v(c_r) \geq v'(c_d)(c_d - c_r)$, since $c_r < c_d$ and $v(c)$ being concave. To express everything as replacement rates we again assume that $v'(c_d)(c_d - c_r) \approx v'(b_d)(b_d - b_r)$ or interpret the replacement rates as replacement rates of consumption levels. The Taylor approximation is then

$$
\begin{aligned}
\frac{v'(b_d)(b_d-b_r)}{u'(w-\tau)} \approx & (b_d-b_r)\frac{u'(w-\tau)-u''(w-\tau)(w-\tau-b_d)+\frac{1}{2}u'''(w-\tau)(w-\tau-b_d)^2}{u'(w-\tau)} \\
= & (b_d-b_r)\left(1+\gamma(1-r_d)+\frac{1}{2}\gamma\rho(1-r_d)^2\right).
\end{aligned} \tag{39}
$$

A sufficient condition for $\frac{dW}{d\theta^*}<0$ is therefore

$$
1+\epsilon-\frac{r_r}{r_d}S < S\left(1-\frac{r_r}{r_d}\right)\left(1+\gamma(1-r_d)+\frac{1}{2}\gamma\rho(1-r_d)^2\right). \tag{40}
$$

B.3 When are the sufficient Conditions inconclusive?

The welfare conditions are only sufficient and not necessary. With these conditions we cannot make any welfare statements if

$$
\frac{v(c_d)-v(c_r)}{b_d u'(w-\tau)} > 1+\epsilon-\frac{b_r}{b_d}S > S\frac{v(c_d)-v(c_r)}{b_d u'(w-\tau)} \tag{41}
$$

or if we make the marginal utility approximation if

$$
\frac{v'(c_r)(c_d-c_r)}{b_d u'(w-\tau)} > 1+\epsilon-\frac{b_r}{b_d}S > S\frac{v'(c_d)(c_d-c_r)}{b_d u'(w-\tau)} \tag{42}
$$

or if we assume CRRA utility and do not use the marginal utility approximation

$$
\frac{1}{1-\gamma}\left[r_d^{-\gamma}-\frac{r_r^{1-\gamma}}{r_d}\right] > 1+\epsilon-\frac{r_r}{r_d}S > \frac{S}{1-\gamma}\left[r_d^{-\gamma}-\frac{r_r^{1-\gamma}}{r_d}\right]. \tag{43}
$$

holds.

For given replacement rates and estimates of S and ϵ we can determine a range of risk aversion γ for which the conditions are indecisive. In our applications, either the conditions are conclusive or this range of risk aversion coincides with reasonable values of risk aversion and hence could not be decided even if we had tighter bounds.

C Benefits

Instead of CRRA we can also use a Taylor approximation to get

$$\frac{dW}{db_d} \lesseqgtr 0 \iff \xi \lesseqgtr \gamma \left(1 - \frac{c_d}{w-\tau}\right) + \frac{1}{2}\gamma\rho\left(1 - \frac{c_d}{w-\tau}\right)^2. \tag{44}$$

As a sufficient condition for $\frac{dW}{db_d} \leq 0$ we can again plug in replacement rates instead of consumption, since $\left(1 - \frac{c_d}{w-\tau}\right) \leq \left(1 - \frac{b_d}{w-\tau}\right) = (1 - r_d)$.

D Policy Mix

The two interpretations of the optimal policy can be motivated formally by changes of basis vectors. The optimal direction in (21) is hard to interpret intuitively because of the abstract quantity θ^*. Therefore, we transform the coordinates by a change of basis to better interpretable units.

For the first interpretation in terms of static fiscal effects we note that a change in benefits translates to a mechanical expenditure effect of $\int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta)$. A change in strictness of screening translates to a mechanical expenditure effect of $\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*}dF(\theta)b_d$. Therefore, to tranlsate changes in benefits and strictness of screening to mechanical expenditure effects we have the following mapping $F_1 : (db_d, d\theta^*) \rightarrow M_1 * (db_d, d\theta^*)'$ where

$$M_1 = \begin{pmatrix} \int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta) & 0 \\ 0 & \int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*}dF(\theta)b_d \end{pmatrix}. \tag{45}$$

Therefore, to express the gradient in these units, i.e. in terms of the basis vectors $\left[\begin{pmatrix} \int_{\theta^A}^{\infty} p(\theta;\theta^*)dF(\theta) \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*}dF(\theta)b_d \end{pmatrix}\right]$ instead of the standard basis vectors $\left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right]$, we have

$$M_1^{-1}\nabla W = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} u'(w-\tau) \tag{46}$$

and the optimal direction in terms of these basis vectors is simply $\frac{\beta}{\alpha}$.

For the second interpretation in terms of percentage changes in benefit levels and the award rate we follow the same logic. Marginally increasing screening stringency θ^* translates to a mechanical reduction in DI recipients by $\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)$ or, measured in percent, by $\frac{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)}{\int_{\theta^A}^{\infty} p(\theta;\theta^*) dF(\theta)}$ percent. Similarly, reducing DI benefits by 1 dollar translates to a $\frac{1}{b_d}$ percent reduction in DI benefits. Hence, we have the mapping $F_2 : (db_d, d\theta^*) \to M_2 * (db_d, d\theta^*)'$ where

$$M_2 = \begin{pmatrix} \frac{1}{b_d} & 0 \\ 0 & \frac{\int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta)}{\int_{\theta^A}^{\infty} p(\theta;\theta^*) dF(\theta)} \end{pmatrix}. \tag{47}$$

The gradient in terms of this new basis vectors is given by

$$M_2^{-1} \nabla W = \begin{pmatrix} \alpha * \int_{\theta^A}^{\infty} p(\theta;\theta^*) dF(\theta) b_d \\ \beta * \int_{\theta^A}^{\infty} p(\theta;\theta^*) dF(\theta) b_d \end{pmatrix} u'(w-\tau). \tag{48}$$

and the optimal direction in terms of these basis vectors is again simply $\frac{\beta}{\alpha}$.

E Extensions

E.1 Extension 1: Other sources of heterogeneity (skill shocks)

Suppose individuals also differ in other dimension than disutility of work. Assume there are productivity differences $\chi \sim H(.)$. These affect the wage rate in the second period $w(\chi)$ and therefore also the application and retirement decision. Hence, the marginal applicants become dependent on χ. Welfare is therefore given by

$$\begin{aligned} W =& u(w-\tau) + \int \left(\int_0^{\theta^A(\chi)} u(c_w(\chi)) - \theta dF(\theta|\chi) + \int_{\theta^A(\chi)}^{\theta^R(\chi)} (1-p(\theta;\theta^*))(u(c_w(\chi)) - \theta) dF(\theta|\chi) \right) dH(\chi) \\ &+ \int \left(\int_{\theta^A(\chi)}^{\infty} p(\theta;\theta^*) v(c_d(\chi)) dF(\theta|\chi) + \int_{\theta^R(\chi)}^{\infty} (1-p(\theta;\theta^*)) v(c_r(\chi)) dF(\theta|\chi) - \int_{\theta^A(\chi)}^{\infty} \psi(\chi) dF(\theta|\chi) \right) dH(\chi). \end{aligned} \tag{49}$$

The planner is constrained by a balanced budget requirement

$$\int \left(b_d \int_{\theta^A(\chi)}^{\infty} p(\theta;\theta^*) dF(\theta|\chi) + b_r \int_{\theta^R(\chi)}^{\infty} (1 - p(\theta;\theta^*)) dF(\theta|\chi) \right) dH(\chi) = \tau \tag{50}$$

and the behavioral responses of the agents

$$\theta^A(\chi) = u(c_w(\chi)) - v(c_d(\chi)) + \frac{\psi(\chi)}{p(\theta^A;\theta^*)}, \tag{51}$$

$$\theta^R(\chi) = u(c_w(\chi)) - v(c_r(\chi)). \tag{52}$$

The change in welfare is then given by

$$\begin{aligned} \frac{dW}{d\theta^*} =& -u'(w-\tau)\frac{d\tau}{d\theta^*} + \int \left(\int_{\theta^A(\chi)}^{\theta^R} \frac{dp(\theta;\theta^*)}{d\theta^*} \left[v(c_d(\chi)) - [u(c_w(\chi)) - \theta] \right] dF(\theta|\chi) \right) dH(\chi) \\ &+ \int \left(\int_{\theta^R(\chi)}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} \left[v(c_d(\chi)) - v(c_r(\chi)) \right] dF(\theta|\chi) \right) dH(\chi). \end{aligned} \tag{53}$$

where

$$\frac{d\tau}{d\theta^*} = \int \left(b_d \left(\underbrace{\int_{\theta^A(\chi)}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta|\chi)}_{\text{mechanical effect}} - \underbrace{\frac{d\theta^A(\chi)}{d\theta^*} p(\theta^A(\chi)) f(\theta^A(\chi)|\chi)}_{\text{behavioral effect}} \right) - \underbrace{b_r \int_{\theta^R(\chi)}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta|\chi)}_{\text{benefit substitution}} \right) dH(\chi). \tag{54}$$

The interchangeability of the integration and differentiation follows from the assumptions on smooth utility functions. This makes the welfare function a smooth function and hence the Leibniz rule applies. Therefore, adding other non-health related shocks does not alter the basic result and intuition.

Moreover, to get a sufficient condition for $\frac{dW}{d\theta^*} \geq 0$ we can use the same bounding approach as before. First, use that $v(c_d(\chi)) - [u(c_w(\chi)) - \theta] \geq v(c_d(\chi)) - v(c_r(\chi)) \quad \forall \chi, \theta \in [\theta^A(\chi), \theta^R(\chi)]$. Second, by the concavity of the utility function it follows that $v(c_d(\chi)) - v(c_r(\chi)) \leq v(b_d) - v(b_r)$

because $c_i(\chi) = b_i + s_i(\chi)$, $s_r(\chi) \geq s_d(\chi)$ and $s_r(\chi) \geq 0$ as before. [27] Therefore, a sufficient condition for $\frac{dW}{d\theta^*} \geq 0$ is as before

$$1 + \epsilon - \frac{b_r}{b_d} S \geq \frac{v(b_d) - v(b_r)}{b_d u'(w - \tau)} \tag{55}$$

where

$$S \equiv \frac{\int \int_{\theta^R}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta|\chi) dH(\chi)}{\int \int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta|\chi) dH(\chi)} = \frac{\text{benefit subsitution}}{\text{mechanical effect}}, \tag{56}$$

$$\epsilon \equiv -\frac{\int \frac{d\theta^A}{d\theta^*} p(\theta^A) f(\theta^A|\chi) dH(\chi)}{\int \int_{\theta^A}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*} dF(\theta|\chi) dH(\chi)} = \frac{\text{behavioral effect}}{\text{mechanical effect}}. \tag{57}$$

From there on we can use the same bounding and approximation techniques as before.

To get a sufficient condition for $\frac{dW}{d\theta^*} \leq 0$ we use that $\forall \theta \in [\theta^A, \theta^R] \Rightarrow u(c_w(\chi)) - \theta < v(c_d(\chi))$ and by concavity we have $v(c_d(\chi)) - v(c_r(\chi)) \leq v'(b_d)(c_d(\chi) - c_r(\chi)) \leq v'(b_d)(b_d - b_r)$,where the inequality holds if savings of rejected applicants are weakly larger than non-rejected applicants. With these bounds we arrive at the same condition as in Appendix B.2.

E.2 Extension 2: One Period Model

With just one period welfare is given by

$$\begin{aligned} W = &\int_0^{\theta^A} u(w - \tau) - \theta dF(\theta) + \int_{\theta^A}^{\theta^R} (1 - p(\theta;\theta^*))(u(w - \tau) - \theta) dF(\theta) + \int_{\theta^A}^{\infty} p(\theta;\theta^*) v(c_d) dF(\theta) \\ &+ \int_{\theta^R}^{\infty} (1 - p(\theta;\theta^*)) v(c_r) dF(\theta) - \int_{\theta^A}^{\infty} \psi dF(\theta). \end{aligned} \tag{58}$$

The planner is constrained by a balanced budget requirement

$$b_d \int_{\theta^A}^{\infty} p(\theta;\theta^*) dF(\theta) + b_r \int_{\theta^R}^{\infty} (1 - p(\theta;\theta^*)) dF(\theta) = \tau \left(F(\theta^R) - \int_{\theta^A}^{\theta^R} p(\theta;\theta^*) dF(\theta) \right) \tag{59}$$

[27] We assume b_i does not dependent on χ since the wage in the first period is the same for all individuals and hence with replacement rates depending on the previous wage the benefits do not depend on χ or θ.

and the behavioral responses of the agents

$$\theta^A = u(w-\tau) - v(c_d) + \frac{\psi}{p(\theta^A;\theta^*)}, \tag{60}$$

$$\theta^R = u(w-\tau) - v(c_r). \tag{61}$$

The Welfare derivative is then given by

$$\begin{aligned} \frac{dW}{d\theta^*} &= -u'(w-\tau)\frac{d\tau}{d\theta^*}\left(F(\theta^R) - \int_{\theta^A}^{\theta^R} p(\theta;\theta^*)dF(\theta)\right) \\ &+\int_{\theta^A}^{\theta^R} \frac{dp(\theta;\theta^*)}{d\theta^*}\left[v(c_d) - [u(c_w)-\theta]\right]dF(\theta) + \int_{\theta^R}^{\infty} \frac{dp(\theta;\theta^*)}{d\theta^*}\left[v(c_d) - v(c_r)\right]dF(\theta) \end{aligned} \tag{62}$$

where

$$\frac{d\tau}{d\theta^*} = \frac{1}{F(\theta^R) - \int_{\theta^A}^{\theta^R} p(\theta;\theta^*)dF(\theta)}\left[(b_d+\tau)\left(\underbrace{\int_{\theta^A}^{\infty}\frac{dp(\theta;\theta^*)}{d\theta^*}dF(\theta)}_{\text{mechanical effect}} - \underbrace{\frac{d\theta^A}{d\theta^*}p(\theta^A)f(\theta^A)}_{\text{behavioral effect}}\right) - (b_r+\tau)\underbrace{\int_{\theta^R}^{\infty}\frac{dp(\theta;\theta^*)}{d\theta^*}dF(\theta)}_{\text{benefit substitution}}\right] \tag{63}$$

This assumes that $\frac{\partial\theta^A}{\partial\tau} = \frac{\partial\theta^R}{\partial\tau} = 0$, i.e. we ignore second order effects of changes in screening stringency via taxes on the application and retirement decision. The only difference compared to the two period model is that there are larger budgetary effects of individuals returning to work by saving the DI benefits and the increased tax returns. Hence, we get

$$\frac{dW}{d\theta^*} \gtreqless 0 \iff \left(1+\epsilon - \frac{b_r+\tau}{b_d+\tau}S\right)(b_d+\tau) \gtreqless \frac{v(c_d) - \left(v(c_r)S + (1-S)\left[u(w-\tau)-\widetilde{\theta}\right]\right)}{u'(w-\tau)}. \tag{64}$$

From there on we can follow the same steps as in the two period model.