Pension Adequacy Standards: Empirical Estimates for the United States and Germany

Julian Schmied
(Max-Planck-Institute for Demographic Research)

Christian Dudel
(Max-Planck-Institute for Demographic Research)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 2D-2: Retirement and Aging

Time: Tuesday, August 21, 2018 [14:00-17:30]

Pension adequacy standards: Empirical estimates for the United States and Germany*

Christian Dudel Julian Schmied

July 10, 2018

Abstract

Due to increasing life expectancy pension provisions need to last longer and longer, casting doubt on the financial security of pensioners. Surprisingly, it is unclear what pension level can be considered to be adequate. In this paper, we propose a general framework for the estimation of pension adequacy standards. Applying a range of econometric estimation techniques to data from the U.S. and Germany, we find that a net pension income around 100% of the last net working life income can be considered adequate, give or take 10 percentage points. Extensive sensitivity checks suggest that this finding is robust.

Keywords: Replacement rate; Pension adequacy; Retirement income; Nonparametric estimation;

JEL Classification: C14; D19; H55; J26; J32;

*Financial support by the Deutsche Forschungsgemeinschaft (DFG), grant DU 1566/1-1, is gratefully acknowledged. We thank Martin Werding for useful comments and suggestions. We also would like to thank seminar participants at the Institute for Fiscal Studies, at the Center for Population Change, and at the Max Planck Institute for Demographic Research. The usual disclaimer applies.

1 Introduction

This paper presents empirical standards for pension adequacy for the U.S. and Germany. Pension adequacy is defined as the level of the replacement rate which during retirement allows the same standard of living as during working life. We propose a general framework for the estimation of such pension adequacy standards, and describe parametric, semiparametric, and nonparametric estimation procedures. For both countries we find that a net pension income equal to 100% of the last net working life income is needed to maintain the standard of living.

In many high-income countries populations are aging, and the population aged 65 and older is expected to increase substantially due to increasing life expectancy and low fertility (United Nations, 2015). For instance, the U.S. Census Bureau predicts that the share of the population aged 65 and older will increase from 15% in 2014 to 24% in 2060 (Colby and Ortman, 2015). Population aging puts a strain on public finances, and calls their sustainability into question, as spending on pension provisions increases (Attanasio et al., 2007; Meier and Werding, 2010). As a reaction, in many countries pension systems have been reformed or are currently undergoing changes. In the U.S. the Social Security retirement age has been increased from 65 to 66 for the cohorts of 1943 to 1954, and it will increase further for later born cohorts (Behagel and Blau, 2012). The aim is to increase the length of working life, and through this to increase sustainability. In Germany the statutory retirement age also is increasing (OECD, 2013).

In this context of population aging, concerns have been raised about the financial security of retirees and whether the pensions they receive are adequate. One commonly used indicator for the measurement of pension adequacy are replacement rates, which are defined as post-retirement income relative to pre-retirement income (Boskin and Shoven, 1984); e.g., a replacement rate of 80% would imply that post-retirement income amounts to 80% of pre-retirement income.

Surprisingly, no clear benchmark exists for assessing what pension level is adequate. Often, benchmarks levels are set heuristically, with values between 70% and 100% being common choices (e.g., Haveman et al., 2007). A data driven approach is based on calculating the smallest possible replacement rate needed to not live in poverty (Love et al., 2008), but it only allows to derive a minimum level. The life cycle model and its extensions also have been used to derive benchmarks. Somewhat simplified, they are based on assuming that individuals smooth marginal utility of consumption over the life course, and calibrated to empirical data they imply a certain level of the replacement rate (e.g., Mitchell and Moore, 1998). While life cycle models have proven to be valuable for assessing saving behavior, it is less clear how optimality relates to adequacy – Crawford and O'Dea (2012) raise the point that optimal saving might not lead to an adequate pension level.

In this paper, we study an empirical approach which allows to derive the replacement rate needed to maintain the living standard achieved during working life. We propose to use this replacement rate as a benchmark for pension adequacy. While we do not claim that this is the only valid yardstick for adequacy, it is attractive for several reasons. First, our benchmark is easily interpretable. Second, the conceptual framework we present is straightforward and rather general, but it is still rich in the sense that it structures and informs empirical assessments. Third, the data demands of our approach are modest and due to its generality it is compatible with many different types of data. Fourth, as our empirical results show, our benchmark seems to be close to what individuals expect to be an adequate standard (Binswanger and Schunk, 2012). Fifth, historically in some countries pension systems were expected to provide a constant living standard, and policy debates are still framed in this way (e.g., Wilke, 2014).

From a methodological perspective, we present a general identification strategy based on expenditure data, but it is also compatible with other welfare indicators. We apply parametric, semiparametric, and nonparametric methods for estimation. The parametric approach and the semiparametric approach are were originally devised to estimate equivalence scales (Deaton and Muellbauer, 1986; Pendakur, 1999), while the nonparametric approach is adopting recent results on partial identification (Fan et al., 2017). All approaches are informed by the literature on life cycle models. While the literature on equivalence scales and the literature on life cycle models study questions different from ours, similar problems need to be solved, and building on them helps avoiding potential issues.

The parametric estimation approach and the semiparametric estimation approach require that replacement rates do not depend on the pre-retirement income. That is, an individual with a low income during working life needs the same replacement rate to keep the living standard constant as an individual with a high income. The nonparametric approach does not require this assumption, but only yields set estimates, i.e., partially identified estimates in the sense of Manski (2003). We use a parametric test and a semiparametric test of income-dependence which are taken from the equivalence scale literature, and we propose and apply a simple nonparametric approach.

We also conduct extensive sensitivity checks to assess how robust our approach is. Among other things, we check how sensitive our findings are with respect to the choice of welfare indicator and whether endogeneity of retirement status biases our results. In combination with the three estimation procedures and the accompanying tests described above the sensitivity analyses allow us to assess how reliable our results are and whether the assumptions they are based on are reasonable. As it turns out, our main findings are rather robust.

We study two countries: the United States and Germany. They have very different pension systems. Germany is usually considered to be the archetype of the Bismarckian model, relying heavily on social security contributions and until recently with low importance of private savings. The U.S., on the other hand, is a Beveridge system based on taxes, with a high importance of private provisions. While we do not expect that adequate replacement rates differ by country, even finding the same replacement rate would have different implications in these two countries. Moreover, incentives to stay in the labor force until reaching the statutory retirement age differ, which might affect endogeneity of the retirement decision.

For the U.S., our findings are based on data of the most recent wave of the Health and Retirement Study from 2014; and for Germany we use the most recent wave of the Income and Expenditure Survey (Einkommens- und Verbrauchsstichprobe) from 2013. Applying our methods to the U.S. data and the German data, our main finding is that adequate net replacement rates are around 100% in both countries, consistent across methods and robust to many sensitivity checks. Results for income independence are less clear cut: for Germany, income independence cannot be rejected, while for the U.S. the results depend on the type of test and the choice of welfare indicator.

Comparing our results to actual replacement rates U.S. retirees received, as reported by Love et al. (2008), we find that on average the gap is small. Still, as a sizable part of the older U.S. population has no or only little provisions for retirement (Lusardi and Mitchell, 2007), there is a large group of retirees who will achieve pension incomes considerably lower than our standard. For Germany there currently is a gap of 20 to 30 percentage points by which actual replacement rates provided by the statutory pension insurance are lower than the adequate rates we find. According to stylized projections by the OECD (2013), this gap might increase substantially in the future, showing the need to intensify efforts for work-related pension schemes and private provisions.

In summary, we contribute to the literature by establishing a general framework for the estimation of adequate replacement rates, and by discussing and comparing estimation approaches with different levels of econometric sophistication and underlying assumptions. We also assess whether adequate replacement rates depend on pre-retirement income and apply econometric tests, including a new nonparametric test. We provide comparable benchmarks for pension adequacy in the U.S. and Germany. Our benchmark helps individuals to make informed decisions on pension saving, which in light of longer lives and decreasing public pensions are becoming more important. Our benchmark helps policy makers to assess the well-being of the retiree population; and it helps pension providers and insurance companies to provide individuals with the pension packages they need.

The remainder of this paper is structured as follows. In section 2 we discuss the related literature. Our economic framework is described in section 3 and our identification strategy is discussed in 4. The data we use is presented in section 5. We present our main findings in section 6, and additional findings and sensitivity analyses in section 7. Section 8 concludes.

2 Related work

At first sight, it might appear that a level of the replacement rate of 100% should allow a constant living standard, at least if the time shortly before and early after retirement is considered. But departures from this are possible for several reasons. On the one hand, values below 100% seem possible, as retired individuals will have no work-related expenses (e.g., commuting); will not save for retirement anymore; and will have more time for household production (Aguiar and Hurst, 2005; Love et al., 2008). On the other hand, values of the replacement rate above 100% could occur because health-related expenses might increase with age or might be anticipated and lead to precautionary saving (Blundell et al., 2016); and more free time could also imply more pastime expenditures (Crawford and O'Dea, 2012).

Whether differences in taxation of retirees and non-retirees play a role for replacement rates depends on whether gross replacement rates (GRRs) or net replacement rates (NRRs) are considered. In case of GRRs taxation will usually be an argument for replacement rates below 100%, as retirees will often be taxed less. In case of NRRs taxation does not matter. In this paper, we will focus on the estimation of NRRs, as these are are more readily comparable across countries. In contrast, because of differences in taxation the same GRR could imply quite different living standards in different countries. In the literature, both GRRs and NRRs can be found, and in our supplementary materials we also supply estimates of GRRs to ease comparison to other studies.

In the literature on pension savings and incomes, often heuristic benchmarks are used for pension adequacy. For instance, Haveman et al. (2007) assume for the

U.S. that a replacement rate of 70% is adequate, while Schulz and Carrin (1972) use a value of 80%. According to Love et al. (2008) values between 70% and 100% are common. Similar values can be found in the literature for Germany. Often, the chosen value is not further justified, perhaps except mentioning that it is in the established range. For Germany, a value of 70% of the (net) replacement rate is sometimes justified by arguing that it was the highest value which was ever provided by the German public pension system (Schnabel, 2003).

A more data-based approach allows to derive the minimum level of the replacement rate required to not live in poverty (e.g., Love et al., 2008). While this approach does not establish an adequate level of the replacement rate, it sets a lower bound. To do so, a poverty threshold is calculated. This can be, for instance, the threshold suggested by the OECD, calculated as 50% of the median equivalized disposable income (Knoef et al., 2016). Given this threshold, it is possible to calculate the replacement rate an average household would require to be above the poverty threshold. A related approach was suggested by VanDerhei and Copeland (2010), who used expenditure data and assessed the minimum income required to reach a certain expenditure level. Again, only a lower bound for pension adequacy is established.

A theoretically-grounded approach which allows the derivation of replacement rates is based on the life cycle model, originating from work by Modigliani and Brumberg (1954) and Friedman (1957). The life cycle model implies smoothing of the marginal utility of consumption over the life course and, at least in simple variants, also smoothing of consumption itself. This implies that pre-retirement income and post-retirement income should not differ too much to achieve similar consumption levels, except perhaps because of work-related expenses and in case of gross replacement rates differences in taxation (Wolfson, 2011). A similar reasoning without recourse to the life cycle model was proposed by Henle (1972), requiring equal disposable pre-retirement income and post-retirement income.

The life cycle model essentially is a model of optimal saving and consumption, and given a functional form of the utility function and further assumptions like, e.g., interest rates, can be calibrated to empirical data. It has been found that at retirement consumption drops occur in a way that is not compatible with consumption smoothing (Attanasio and Weber, 2010). This lead to discussions whether life cycle models were missing important factors (Bernheim et al., 2001), or whether the discrepancies between model and data are due to measurement errors (Aguiar and Hurst, 2005, 2007). More recent takes on the life cycle model can predict the drop in consumption (Pagel, 2017).

While much of the literature focuses on optimal saving behavior, life cycle model estimates also imply replacement rates. For the U.S., these have often been found to be between 80% and 90% (Hamermesh, 1984; Bernheim, 1992; Mitchell and Moore, 1998), but other values also have been reported. For instance, the results published by Scholz et al. (2006) imply a replacement rate of around 66%. These results show that the life cycle model does not necessarily imply constant consumption and thus (net) replacement rates around 100%. While replacement rates derived from life cycle models are optimal given some model assumptions and optimization constraints, they are not necessarily adequate, as optimality of saving behavior does not imply adequacy (Crawford and O'Dea, 2012): a replacement rate derived from optimal behavior might not be adequate, and an adequate replacement rate might not be optimal.

While life cycle models are usually calibrated to expenditure data, another approach uses data on subjective assessments. Binswanger and Schunk (2012) asked individuals in the U.S. and in the Netherlands about their prefered retirement incomes. More specifically, based on a respondents' current income, several pairs of pre-retirement income and post-retirement income were presented, each representing different choices about saving for retirement and resulting replacement rates; e.g., high disposable pre-retirement representing little saving and a correspondingly low retirement income, or low disposable pre-retirement income and high post-retirement income and replacement rate. Their results show that both Americans and Dutch prefer replacement rates between 80% and 100%. As many of the surveyed individuals were below retirement age (with a median age between 51 and 52), the results of Binswanger and Schunk (2012) are partly based on expectations individuals have about their needs during retirement.

Dudel et al. (2016) also used subjective assessments and chose an approach close to the one presented later in this paper. Based on individual satisfaction with household income and applying the equivalence scale framework, they calculated the replacement rate needed to keep the standard of living constant. Using German panel data, they find that the level of the replacement rates needs to be 87%.

3 Conceptual framework

In this section we develop the economic framework we use and discuss its implications. Let $V(\mathbf{z}, x, \mathbf{p})$ be the indirect utility function of an individual with characteristics $\mathbf{z}$, net income y, and facing prices $\mathbf{p}$. Using $V(\cdot)$, we can define an income function $I(\mathbf{z}, u, \mathbf{p}) = \min_y[y|V(\mathbf{z}, y, \mathbf{p}) = u]$, which gives the minimum income an individual with characteristics $\mathbf{z}$ and facing prices $\mathbf{p}$ needs to achieve welfare level u. Moreover, let d be a binary variable capturing whether an individual is retired ($d = 1$) or not ($d = 0$), and let $\mathbf{z}_d = (d, \mathbf{z})$.

Using this notation, we define the replacement rate which keeps the living standard constant as

$$R(\mathbf{z}'_0, \mathbf{z}''_1, u, \mathbf{p}) = I(\mathbf{z}''_1, u, \mathbf{p})/I(\mathbf{z}'_0, u, \mathbf{p}). \tag{3.1}$$

Thus, $R(\mathbf{z}'_0, \mathbf{z}''_1, u, \mathbf{p})$ is the income a retired individual needs to attain welfare level u relative to the income a non-retired individual needs to achieve u. Except from the retirement status, we will mostly assume that the retiree and the non-retiree are otherwise similar, e.g., $\mathbf{z}' = \mathbf{z}''$, although this is not required and they could differ with respect to, for instance, age.

This setup varies from the one proposed by Boskin and Shoven (1984) and the one usually considered in life cycle models. Specifically, we only assess welfare at a single point in time. This way, we avoid assumptions on separability of utility over time and assumptions on discounting, which are necessary for life cycle models (Attanasio and Weber, 2010). Also, changes in household composition during the life course like, e.g., children entering and leaving households will not complicate our analyses (Boskin and Shoven, 1984).

Equation (3.1) is similar in structure to the definition of equivalence scales (Lewbel and Pendakur, 2008). Equivalence scales are used to compare households of different composition, e.g., couples with a child and couples without children. For this, the retirement indicator d is replaced with an indicator of household composition. There is a rich literature on theoretical issues and the empirical estimation of equivalence scales which we use to inform our approach.

An important decision which we derive from the literature on equivalence scales is that we only consider (indirect) utility functions of individuals, in contrast to the literature on equivalence scales, which often starts from household utility functions (this holds also for the life cycle literature; see Attanasio and Weber, 2010). Household utility functions require strong assumptions to be meaningful, and ignore decision making processes and the allocation of resources within the household (Chiappori, 2016). This is also relevant for the elderly, as shown empirically by Lundberg et al. (2003) and Cherchye et al. (2012). As the data we use is only at the household level and does not allow to identify individual consumption, we chose to restrict ourselves to single-person households. This also simplifies analyses and identification of households of retirees, which for two-person households might be less straightforward (Moreau and Stancanelli, 2015).

Nevertheless, we conduct sensitivity checks in which two-person households are included in the analyses.

Even restricting ourselves to single-person households equation (3.1) is not easily identified (Blundell and Lewbel, 1991). The approach we will follow is related to the reasoning of Engel as discussed by Deaton and Muellbauer (1986) and on the reasoning outlined in van Praag (1991). Essentially, we assume that an indicator variable is available which measures the welfare level u. For this, we focus on expenditure data, using several different indicators and conducting robustness checks with respect to the choice of indicator (see section 5.2). Note that our approach only requires the welfare indicator to be comparable across individuals in the sense that a specific value u' means the same for all individuals; i.e., all individuals with a specific value u' have the same welfare level; and that higher (lower) values than u' mean that individuals are better (worse) off.

As we use cross-sectional data we assume that prices are fixed and the same for all households; this allows us to ignore price dependence, i.e., $R(\mathbf{z}_0', \mathbf{z}_1'', u, \mathbf{p}) = R(\mathbf{z}_0', \mathbf{z}_1'', u)$. Including prices would allow to assess the elasticity of replacement rates with respect to price changes for specific commodities. Price changes have been shown to be relevant for equivalence scales (Pendakur, 2002). Unfortunately, this is not possible with the data we use in this paper.

A related issue is so called income-independence, or independence of base (Lewbel, 1989). Income-independence means that the level of the replacement rate needed to maintain a constand living standard does not depend on the welfare level u, i.e., $R(\mathbf{z}_0', \mathbf{z}_1'', u) = R(\mathbf{z}_0', \mathbf{z}_1'')$. This means that, for instance, the level of the adequate replacement rate is the same for a person with high income during working-life and a person with low income during working life. For equivalence scales, independence of base is usually rejected in empirical studies (Donaldson and Pendakur, 2006; Biewen and Juhasz, 2017).

While there are no similar empirical results for replacement rates yet, the findings on expected replacement rates by Binswanger and Schunk (2012) suggest that income independence might be violated. In their survey on prefered replacement rates, individuals with low income tend to prefer higher replacement rates than individuals with higher income. In the next section we discuss several tests to assess income dependence, including a new one.

4 Identification strategy

4.1 Basic estimation problem

To discuss identification and estimation of replacement rates as defined through equation (3.1) we make use of the potential outcomes framework (e.g., Imbens, 2004). As above, let d denote an indicator variable which captures whether individuals are retired or not. Let W_d denote the welfare level. $Y_d(w)$ is the income individuals need to achieve welfare level w given their retirement status. In what follows we assume that observed income is equal to $Y_d(W_d)$ and thus equal to the income function $I(\cdot)$ defined in the previous section, or at least a close approximation. As already mentioned in section 2 we will assume that income is the net income after taxes and transfers, but everything described here works with gross income as well.

Each individual is either observed being retired or not retired, and never both at the same time. Thus, we either observe $(W_0, Y_0(W_0))$ for the non-retired, or $(W_1, Y_1(W_1))$ for the retired. As it turns out, this is not enough to estimate equation (3.1) without strong assumptions. Essentially, replacement rates as defined by equation (3.1) are given by $Y_1(W_0)/Y_0(W_0)$. Formulated as an econometric estimation problem, we have

$$\mathrm{E}\left[\mathrm{E}\left(\frac{Y_1(W_0)}{Y_0(W_0)}\middle| \mathbf{z}\right)\right], \tag{4.1}$$

where $\mathbf{z}$ again is a vector of covariates. With some simple algebra, this can be shown to be

$$\mathrm{E}\left[\frac{E(Y_1(W_0|\mathbf{z})}{E(Y_0(W_0|\mathbf{z})}\right] - \mathrm{E}\left[\frac{1}{\mathrm{E}(Y_0(W_0|\mathbf{z})}\mathrm{Cov}\left(\frac{Y_1(W_0)}{Y_0(W_0)}, Y_0(W_0)\middle| \mathbf{z}\right)\right]. \tag{4.2}$$

As explained in more detail below, the first term in this equation can be identified using standard assumptions, while the second term cannot. Essentially, the first term only requires the marginal distributions of Y_1 and Y_0 to be identified, which can be achieved using standard assumptions. The second term, on the other hand, requires the joint distribution of Y_1 and Y_0, for which no point identification is possible in the setup we assume (Abbring and Heckman, 2007).

Assuming that the pair (Y_0, Y_1) is independent from d conditional on W and $\mathbf{z}$, the first term in equation (4.2) can be identified. This is the so called unconfoundedness assumption (Imbens, 2004). Essentially, unconfoundedness means that there is no unobserved selection into retirement. This allows for differences between retirees and non-retirees, as long as they can be captured by $\mathbf{z}$. In addition, a further assumption is required. The overlap condition requires that $0 < \Pr(d = 1|\mathbf{z}) < 1$. Somewhat simplified it means that for all values and combinations of $\mathbf{z}$ there are both retirees and non-retirees. Moreover, the stable unit treatment value assignment is invoked, implying that whether one individual is retired or not does not depend on the retirement status of other individuals.

In contrast, the second term in equation (4.2) is not identified using standard assumptions (Abbring and Heckman, 2007). Specifically, this is due to the covariance of $Y_1(W_0)/Y_0(W_0)$ and $Y_0(W_0)$. This covariance term captures to which extent the replacement rate an individual needs to keep a constant living standard depends on the income and living standard before retirement. It is directly related to income independence of the adequate replacement rate: If the covariance is zero, then the replacement rate does not depend on the baseline income; otherwise it does. If the covariance is negative, then higher baseline incomes go along with lower replacement rates; and if the covariance is positive is positive, then higher incomes imply higher replacement rates.

In what follows, we outline three different approaches to estimate equation (4.2), inspired by methods for estimation of equivalence scales. These approaches go along with different tests to empirically assess income independence, and they differ with respect to their complexity and underlying assumptions. We discuss a fully parametric approach that can be implemented very easily via standard linear regression, but it requires the strong assumption that the relationship between income and the welfare indicator is (log) linear. It also assumes income independence. The semiparametric approach we borrow from Pendakur (1999) and Stengos et al. (2006) is more complicated to use, but does not require linearity; it still requires income independence, though. The third approach we study is based on a general approach to identification developed by Fan et al. (2017), which we apply to the estimation problem presented here. It does neither require linearity nor income independence, but it only yields identification bounds on replacement rates and no point estimates.

Note that none of the three approaches just described requires a specific sign of the effect of income on the welfare indicator; that is, the indicator might both be increasing but also decreasing in income; what is important is that similar values of the welfare indicator imply the same level of the welfare function.

For all approaches and tests we provide functions for the statistical software R (R Core Team, 2017) which allow easy use of our framework. For the implementation of the semiparametric and nonparametric approach we use the `np` package provided by Hayfield and Racine (2008).

4.2 Parametric approach

A classical approach for estimation of equivalence scales attributed to Engel works as follows (Deaton and Muellbauer, 1986). W is considered to be a function of income Y, retirement status d, and some other covariates $\mathbf{z}$, i.e., $W(Y, d, \mathbf{z})$. This Engel curve can be estimated empirically, allowing to use its inverse, $W^{-1}(w, d, \mathbf{z})$. Given a welfare level w', the adequate replacement rate can then be calculated as $W^{-1}(w', 1, \mathbf{z})/W^{-1}(w', 0, \mathbf{z})$.

A common implementation of the Engel approach builds on the model specification proposed by Working (1943) and Leser (1963),

$$W_d = a + \log Y b_Y + d b_d + \mathbf{z}'\mathbf{b}_z + \epsilon, \tag{4.3}$$

where a, b_Y, b_d, and $\mathbf{b}_z$ are regression coefficients, and ϵ is a well behaved error term. Given parameter estimates, which can easily be calculated using least squares, consider equation (4.3) for a retiree and a non-retiree which are similar with respect to $\mathbf{z}$; assume that these have the same welfare level; and equate both variants of the equation and solve for Y_1/Y_0. This yields

$$\mathrm{E}\left(\frac{Y_1}{Y_0}\right) = \exp\left(-\frac{\hat{b}_d}{\hat{b}_Y}\right). \tag{4.4}$$

In some sense the regression of Y on W is 'reversed', and this is used as W^{-1}. From the perspective of the potential outcome framework this amounts to imputing Y_1 and Y_0 through the parameter estimates. Note that (4.4) does neither depend on Y_0, nor on W, and it implies that the covariance term in equation (4.2) is zero.

The fully parametric framework of equation (4.3) allows to assess income dependence by adding additional, nonlinear terms. For instance, a quadratic term of log income can be added. As discussed for equivalence scales by Lancaster and Ray (1998), this yields a formula for the income ratio which is more complicated than equation (4.4), and which depends on the baseline income. Pendakur (1999) suggested to test whether the interaction of the treatment indicator d and the quadratic term is statistically significant, a proposal we will follow in our analyses. This is not a direct test of income independence, but of shape invariance, which in our case means that the Engel curves of retirees and non-retirees have the same shape. Shape invariance is sufficient for income independence, but not necessary. This means that failing to reject shape invariance also means that income independence cannot be rejected. In contrast, rejecting shape invariance does not imply that income independence can be rejected.

4.3 Semiparametric approach

The log-linear functional form used by the Working-Leser model might not hold empirically depending on the welfare indicator (Banks et al., 1997). Pendakur (1999) proposed a semiparametric approach for the estimation of equivalence scales, which does not require strong assumptions on the functional form of the relationship between log income and the welfare indicator, except for some smoothness restrictions. The welfare indicator is assumed to be a nonlinear function of income, which depends on retirement status: $W(\log Y - d\alpha, d) + d\mu$, where α is the adequate replacement rate, and μ is an additional elasticity parameter. Note that there are no other covariates $\mathbf{z}$ included. Given a welfare level w' and inverting W, this setup leads to $W^{-1}(w', 1)/W^{-1}(w', 0) = \alpha$, irrespective of the level of w'.

Practically, $W(\log Y, 0)$ and $W(\log Y, 1)$ are estimated separately using non-parametric regression techniques. Specifically, we use kernel regression, with

$$W(\log Y, D) = \frac{\sum\limits_{i:d_i=D} K_h(\log Y - \log y_i) w_i}{\sum\limits_{i:d_i=D} K_h(\log Y - \log y_i)},$$

where K_h is a kernel function with bandwidth h. Essentially, the kernel function $K(\cdot)$ gives observations with values of y_i close to Y a high weight, while values of y_i far from Y get a low weight. This way, the estimate of W at Y is based not only on observations with $y_i = Y$, but also observations close to Y. In a next step, we follow Stengos et al. (2006) and estimate α and μ such that

$$L(\alpha, \mu) = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{W}(\log y_i - (1-d)\alpha, 1) + \mu - \hat{W}(\log y_i + d\alpha, 0) \right)^2 \tag{4.5}$$

is minimized. This is achieved through grid search. Somewhat simplified, the nonparametric estimates $\hat{W}(\log Y, 0)$ and $\hat{W}(\log Y, 1)$ are shifted by α and μ in such a way that they are as close as possible. See the supplementary materials for details of our implementation.

Rephrasing the approach in terms of the potential outcomes framework, for each individual i we observe $y_i = y_i^0(1-d) + y_i^1 d$ and $w_i = w_i^0(1-d) + w_i^1 d$. If $d_i = 0$, y_i^1 is imputed as $y_i^0/\exp(\alpha)$. If $d_i = 1$, then $\hat{y}_i^0 = y_i^1 \exp(\alpha)$. Given estimates of $E(W_1(Y))$ and $E(W_0(Y))$, imputed values of y imply values for w. α is chosen such that the imputed welfare level, $\hat{w}_i$, is as close as possible to the observed welfare level, i.e., $\hat{w}_i \approx w_i$.

This semiparametric approach also requires that the covariance between the replacement rate and baseline income is zero. It further requires shape invariance of Engel curves, i.e., that the relation between (log) income and the welfare indicator has the same shape for both retirees and non-retirees, except shifts of the curve by α and μ (Pendakur, 1999). As already noted shape invariance is sufficient for income independence, but not necessary. Testing shape invariance with the semiparametric approach can proceed using simulations, based on comparing the value of $L(\alpha, \mu)$ as defined in equation (4.5) above calculated from empirical data to simulated values which are generated assuming shape invariance (see supplementary materials for implementation).

4.4 Nonparametric approach

The parametric approach and the semiparametric approach rely on the strong assumption that the covariance term in equation (4.2) is zero; i.e., that there is no income dependence of the adequate replacement rate. They also require shape invariance of Engel curves. Both the parametric approach and the semiparametric approach go along with tests to check this assumption; but if it is rejected, it is not clear whether income independence can be rejected, and the point estimates of the parametric and semiparametric are potentially biased.

Dropping the assumption of income dependence and shape invariance, we use a nonparametric approach, building on recent results by Fan et al. (2017). This approach does not point identify the replacement rate, and gives set estimates in the sense of Manski (2003); i.e., it can only be established that the replacement rate is in a specific closed interval for which the endpoints can be estimated from the data.

Fan et al. (2017) build on work by Cambanis et al. (1976) for Frechét-Hoeffding bounds to show how bounds on expectations of the form $\mathrm{E}(k(Y_1, Y_0)|\mathbf{X})$ can be derived, where $k(\cdot)$ is a strictly subadditive function. The lower bound can be calculated as

$$\mathrm{E}^L(k(Y_1, Y_0)|\mathbf{X}) = \int_0^1 k\left(F_1^{-1}(t|\mathbf{X}), F_0^{-1}(t|\mathbf{X})\right) dt, \tag{4.6}$$

and the upper bound is given by

$$\mathrm{E}^U(k(Y_1, Y_0)|\mathbf{X}) = \int_0^1 k\left(F_1^{-1}(t|\mathbf{X}), F_0^{-1}(1-t|\mathbf{X})\right) dt, \tag{4.7}$$

where $F_1^{-1}(u|\mathbf{X})$ and $F_0^{-1}(u|\mathbf{X})$ are the quantile functions of the conditional marginal distributions of Y_1 and Y_0, respectively. These quantile function can be estimated fully nonparametrically (see appendix).

Assuming that Y_1 and Y_0 will always be strictly positive, i.e., $Y_1 > 0$ and $Y_0 > 0$, $k(Y_1, Y_0) = Y_1/Y_0$ is strictly subadditive and the approach of Fan et al. (2017) allows to calculate bounds for $\mathrm{E}(Y_1/Y_0|\mathbf{X})$. This can be used to derive bounds on the covariance term in equation (4.2), as it is bounded by

$$\begin{aligned} &\mathrm{E}(Y_1|\mathbf{X}) - \mathrm{E}^U(Y_1/Y_0|\mathbf{X})\mathrm{E}(Y_0|\mathbf{X}) \leq \\ &\qquad \mathrm{Cov}(Y_1/Y_0, Y_0|\mathbf{X}) \leq \mathrm{E}(Y_1|\mathbf{X}) - \mathrm{E}^L(Y_1/Y_0|\mathbf{X})\mathrm{E}(Y_0|\mathbf{X}), \end{aligned} \tag{4.8}$$

where $\mathrm{E}^U(Y_1/Y_0|\mathbf{X})$ is the upper bound for $\mathrm{E}(Y_1/Y_0|\mathbf{X})$ and $\mathrm{E}^L(Y_1/Y_0|\mathbf{X})$ is the lower bound (also see Dudel, 2015). As Fan et al. (2017) show that the bounds on $\mathrm{E}(Y_1/Y_0|\mathbf{X})$ are sharp, it follows that the bounds on the covariance given by equation (4.8) are also sharp.

Note that this approach does not use the logarithm of income, in contrast to the parametric and the semiparametric approach. The reason is that neither the expectation of $\log Y_1/\log Y_0$ nor $\log(Y_1/Y_0)$ are very helpful. The expectation of the latter is equivalent to the geometric mean of Y_1/Y_0 (Szulc, 2009), which has the undesirable property that it moves further away from the mean the higher the variance is (Cartwright and Field, 1978). The former is not a meaningful quantity.

While this approach does not require the income independence assumption, the bounds still relate to a single quantity. This quantity is the population average of the adequate replacement rate, where there are no restrictions on replacement rates of individuals, while the parametric and semiparametric approach force the replacement to be equal across individuals.

The nonparametric procedure also implies a test of income independence. Assuming that income independence holds true, the identification bounds of the covariance as given by equation (4.8) include zero. It is not possible that both bounds are equal to zero, as this would imply point identification, which is only possible if at least one of the marginal distributions of Y_0 or Y_1 is degenerate (Fan et al., 2017). Based on bootstrap resampling, the probability that the identification bounds of the covariance include zero can be calculated. We propose to use this probability to test income independence. If it is lower as conventional significance levels, then income independence can be rejected (see appendix for implementation).

5 Data

5.1 Datasets: HRS and EVS

We use two different datasets for our analyses. For the U.S., we employ data of the Health and Retirement Study (HRS) 2014, while for Germany we use data of the Income and Expenditure Survey (Einkommens- und Verbrauchsstichprobe; EVS) 2013.

The HRS is a panel study focusing on Americans aged 50 or older that has been running since 1992 (Juster and Suzman, 1995). It is conducted by the Survey Research Center of the Institute for Social Research of the University of Michigan, and is supported by the National Institute on Aging (NIA) and the Social Security Administration (SSA). The HRS covers a broad range of questions, including questions on employment, pensions, housing, and assets. Respondents are interviewed every two years, and the data for the year 2014 is the last wave available to us. As our identification strategy does not rely on longitudinal data, we focus on this most recent wave.

The EVS is a cross-sectional household survey conducted every 5 years by the German Federal Statistical Office. The last available wave is for 2013. For each household, detailed information on the households income, expenditures, and savings is collected for one quarter of the year. In addition, socio-demographic variables are included, like the educational attainment of household members.

For our analyses, we restricted both the HRS sample and the EVS sample in two ways. The first restriction is with respect to household size, where we only look at households of single persons. This avoids the need to introduce household utility functions (see section 3). Moreover, the resulting samples are more homogeneous, because, for instance, single adults might differ from couples with respect to bequest motives (De Nardi et al., 2010; Hann and Prowse, 2014). Also, for households of couples it is not always clear cut whether they should count as households of retirees, as one partner might be retired and the other not, complicating analyses (Moreau and Stancanelli, 2015). Finally, we also exclude individuals living alone who report to be married, but this is only a small number of observations.

The second restriction we apply is with respect to the age range: we restrict our samples to respondents aged between 60 and 69. Restricting the sample in this way means that we are essentially comparing non-retirees shortly before retirement and retirees shortly after retirement, as in Germany the statutory retirement age for most people in the EVS was 65 or slightly more, while for the respondents of the HRS 2014 in this age range the earliest retirement age was 62 and the full retirement age was around 66. This restriction helps to guarantee that for all ages under consideration there are both retirees and non-retirees, as required for identification per the overlap assumption (see section 4); for older ages there might only be retirees or very few non-retirees, while for younger ages it is the other way round.

5.2 Main variables: Welfare indicator, income, and retirement status

Three ingredients are required to implement the approaches described in the previous section: a welfare indicator, income, and retirement status.

As the main indicator for the welfare level we use the the income share of expenditures for food at home, which can be calculated for both the HRS and EVS data. This share has long been used for the estimation of equivalence scales, building on Engel's law which states that the share of food expenditures decreases as (log) income increases (Deaton and Muellbauer, 1986). Food expenditures are also commonly used as an indicator of household consumption in the life cycle literature (Browning and Crossley, 2001).

As the main income concept we will use net household income plus annuitized wealth including housing. This makes the replacement rates we calculate net replacement rates, which are usually higher than gross replacement rates. While the EVS covers net income the HRS only covers gross income. To calculate net income for the US, we make use of the tax simulations provided by RAND and follow the methodology of Pantoja et al. (2017) and Blundell et al. (2016). We annuitize non-housing wealth with 2.5% annually and housing wealth with 1.25% annually (Crawford and O'Dea, 2012).

To remove outliers with respect to income from from our analyses, we drop households with net income below zero. For the HRS this can occur because taxes are only simulated. In a next step, the income distribution is trimmed and the top and bottom 2.5% incomes are dropped from analyses. In the EVS incomes below zero are rare but also possible and due to the way the Federal Statistical Office calculates household income. By design, no households with incomes above $12,000$ Euro per month are included in the EVS, because of which there is no need to drop high-income outliers. As a threshold below which we drop households from our analyses we use the level of the federal welfare benefits to which all individuals with income below the threshold are entitled.

Assessing whether an individual is retired or not is straightforward for the case of the EVS. This is because in Germany retirement is a rather clear-cut transition, at least compared to other countries. Specifically, we use an indicator readily included in the data, which captures the labor force state of the respondent and whether she is retired or not. In addition, we make use of information on working hours and assume that all individuals with 20 or more hours worked per week are not retired. There are only very few individuals whose labor force state is retired and who reported to work more than 20 hours, though.

For the HRS, it is more complicated to assign the retirement status, and several definitions of the retirement status can be used (e.g., Behagel and Blau, 2012). Our main analyses use the labor force status provided by RAND, and we conduct several sensitivity analyses using varying definitions of the retirement status. The labor force status provided by RAND is based on several variables, including the number of hours worked and whether the respondent considers herself retired. As in case of Germany, we assume that individuals who work more than 20 hours per week are not retired.

For all three main variables both the HRS and EVS offer many alternatives. For instance, instead of food at home, nondurable expenditure could be used, or satisfaction with income. Because of this, we conducted extensive sensitivity analyses, using a large number of different variables. An overview of findings is given in section 7, while a detailed description is given in the supplementary materials. Moreover, the supplementary materials also include results using gross income.

5.3 Other variables

Both the parametric and the nonparametric approach allow the inclusion of control variables, controlling for potential heterogeneity between retirees and non-retirees. For instance, an obvious difference will be with respect to age. While there is a lot of overlap between both groups regarding this variable, retirees on average might be older than non-retirees.

In addition to age, we use the following control variables: education; gender; whether the individual is divorced or not, as this might go along with alimony payments; whether the individual owns the home she is living in; and whether the county an individual is residing in is rural, in a metropolitan area, or in between.

Table 1: Main results for the U.S. (HRS data) and Germany (EVS data), including the net replacement rate (NRR), its standard error (SE), and 95% confidence intervals (CI) by method. Also shown are the p-values for the tests of income independence and the number of observations. Source: Own calculations based on HRS and EVS.

USA	NRR	SE	95% CI	p(ESE)	n
Parametric	0.912	0.069	[0.777,1.047]	0.068	798
Semiparametric	1.052	0.112	[0.670,1.170]	0.373	878
Nonparametric	[0.888,1.194]	0.011;0.012	[0.868,1.212]	[-8515, 582]	798
Germany					
Parametric	0.968	0.035	[0.899,1.037]	0.011	2310
Semiparametric	1.076	0.067	[0.870,1.150]	0.383	2310
Nonparametric	[0.863,1.061]	0.003;0.004	[0.857,1.067]	[-1263,17]	2310

For the EVS data, we also include whether an individual resides in eastern or western Germany, and the quarter the data was collected in. For the HRS data, we add race/ethnicity with a dummy variable indicating whether an individual is non-white or not; while the latter is a rather heterogeneous group, we cannot distinguish further, because of sample size.

5.4 Descriptive findings

6 Main results

Our main results for the US and Germany are displayed in Table 1. It shows the estimates of the net replacement rate needed to achieve a constant living standard, as well as their standard errors and the tests for shape invariance or income independence. For the nonparametric approach, the identification interval of the replacement rate is given, as well as the standard error of both the lower and the upper bound of this interval. For the tests of shape invariance or income independence values below 0.05 indicate that shape invariance or income independence can be rejected, respectively. In case of the nonparametric approach, the confidence interval of the covariance is shown.

The parametric point estimate for Germany amounts to 97%. This means that the net retirement income roughly needs to be at the same level as the net pre-retirement income to avoid changes in the living standard. The semiparametric point estimates is roughly 10 percentage points higher. The estimates for the US are only a few percentage points different from the German results, and confidence intervals mostly overlap. For the US data standard errors are generally higher than for the German data, due to smaller sample size.

The identification intervals of the nonparametric approach are rather wide: The difference between the upper and the lower bound amounts to 31 percentage points in case of the US, and 20 percentage points for Germany. While one can argue that the upper bound especially for the US is not very helpful, the lower bounds of the intervals are informative and rule out some of the heuristic values found in the literature like, for instance, the NRR level of 70% often quoted for Germany in the literature. The results of the parametric method and the semiparametric approach lie within the identification bounds, except for the semiparametric approach applied to German data.

For both countries there is a clear pattern with the bounds of the nonparametric approach having the smallest standard errors and the semiparametric approach having the highest. The latter might be due to the fact that the semiparametric approach does not include control variables, and thus is affected more by changes in sample composition than the other approaches.

Constructing confidence intervals for the nonparametric approach is not as straightforward as for the parametric and semiparametric approach. Here, we use confidence intervals which cover the entire identification interval with a fixed probability, roughly similar to Horowitz and Manski (2000) with some modifications (see appendix). We implement this using the bootstrap and we choose the smallest interval which covers the identification intervals of 95% of the bootstrap resamples. For the US, this confidence interval amounts to 87% and 121%, and for Germany it ranges from 86% to 107%. While these intervals showing both sampling variability and uncertainty due to identification assumptions are rather wide they are still informative.

The results of the tests of shape invariance and income independence are consistent with income independent replacement rates. For the US, none of the three tests rejects income independence. In case of Germany, the parametric test rejects shape invariance, but shape invariance is only a sufficient condition for income independence and not necessary.

In summary, the point estimates for both countries point to an adequate replacement rate around 100%. While the nonparametric approach does lead to wide identification intervals pointing to some uncertainty, it is mostly consistent with the point estimates. There is no evidence for income dependence of replacement rates.

Table 2: Results comparing non-retired individuals 60 to 69 with retirees aged 70 to 79; for the U.S. (HRS data) and Germany (EVS data), including the net replacement rate (NRR), its standard error (SE), and 95% confidence intervals (CI) by method. Also shown are the p-values for the tests of income independence and the number of observations. Source: Own calculations based on HRS and EVS.

USA	NRR (70-79)	SE	95% CI	p(ESE)	n
Parametric	1.003	0.118	[0.773,1.234]	0.831	1059
Semiparametric	1.054	0.072	[0.990,1.200]	0.414	1172
Nonparametric	[0.939,1.169]	0.010;0.012	[0.920,1.192]	[-7357,141]	1059
Germany					
Parametric	0.977	0.036	[0.906,1.048]	0.052	3009
Semiparametric	—	—	—	—	—
Nonparametric	[0.951,1.149]	0.004;0.005	[0.942,1.160]	[-1188,-16]	3009

7 Further findings and robustness checks

7.1 Adequate replacement rates by age

Our main analyses are restricted to individuals aged 60 to 69. This was motivated by the overlap condition, which requires that individuals of all studied ages are in both the control group (non-retirees) and the treatment group. If we used, say, ages 60 to 89, then there might not be any non-retirees above a certain age, or only very few.

Still, it has been argued that replacement rates might depend on age (Knoef et al., 2016; Dudel et al., 2016). On the one hand, older individuals might on average be less active and spend less for leisure or transportation, leading to replacement rates declining with age. On the other hand, health expenditures might increase with age and thus the income required to keep the living standard constant. The latter might be more important for the U.S. than Germany, where health insurance coverage is comparatively generous.

Here, we present results on adequate replacement rates by age. We use non-retired individuals aged 60 to 69 as the control group, and compare them to retired individuals aged 70 to 79; otherwise we follow our main analyses. This procedure violates the overlap condition with respect to age, and results are essentially based on extrapolation. While the findings might give an indication whether age plays a role for adequate replacement rates, they should be viewed with care.

Estimates are shown in table 2. The semiparametric approach did not converge using the German data and thus no results are shown for it. For both the US and Germany the identification bounds are shifted upwards compared to the main results. The identification regions of the main results and the results for the older age group overlap, though, and it cannot be ruled out that the true value is identical. Point estimates are also close to the main findings. Income independence is rejected for Germany, while this is not the case for the US. These findings imply that for Germany replacement rates might become income-dependent with increasing age, even though they do not directly depend on age. While for the US there also is only very weak evidence of age-dependent replacement rates the analyses presented here are only exploratory, and further research might be needed to better understand how needs change with age.

7.2 Adding single-earner couples

The main findings presented in the previous section are restricted to households of single persons, and thus are missing couples which are a large part of the population around retirement age. Other household constellations (e.g., parents and children) occur, but are not very common in the age range we study.

There are two motivations for this restriction. First, including couples within the framework described in section 3 would require to assume that all functions relate to households; e.g., household utility functions. To be meaningful and indicative about the level of welfare of individuals within the household, household utility functions require strong assumptions. Assuming utility functions for individuals instead requires to look at decision making processes and resource sharing within households (Chiappori, 2016). The data we use unfortunately does not allow us to identify individual consumption within households.

Second, even for single individuals it is not always clear-cut whether they can be counted as retired or not. In case of couples this is more complicated, as one household member might retired and the other not; or one household member might retire while the other one starts to work again after a phase of retirement.

Nevertheless, one-person households are only a subset of the population in the age range we study, and they might be a selected group in one way or another. Couples will have different consumption patterns than single persons because of economies of scale and economies of scope; and they might also have different preferences, e.g., with respect to bequest motives. Our main findings might thus not be representative of the total population around retirement age.

Here, we present additional analyses which also include households of couples in addition to one-person households. For households of non-retirees we only include couples for which one partner is working and the other is a homemaker. The

Table 3: Results including both single person households and households of couples with a single-earner, for the U.S. (HRS data) and Germany (EVS data), including the net replacement rate (NRR), its standard error (SE), and 95% confidence intervals (CI) by method. Also shown are the p-values for the tests of income independence and the number of observations. Source: Own calculations based on HRS and EVS.

USA	NRR (70-79)	SE	95% CI	p(ESE)	n
Parametric	0.860	0.040	[0.781,0.938]	0.089	1321
Semiparametric	1.137	0.069	[0.950,1.200]	0.477	1433
Nonparametric	[0.880,1.226]	0.007;0.010	[0.866,1.245]	[-15287,240]	1321
Germany					
Parametric	0.956	0.023	[0.912,1.001]	0.004	4935
Semiparametric	1.112	0.067	[0.930,1.119]	0.390	4935
Nonparametric	[0.902,1.088]	0.002;0.002	[0.898,1.092]	[-1643,4]	4935

working partner is required to be in the age range 60 to 69, the other is not. In many cases this will imply that the retirement of the working partner means that all household members are off the labor market. This is, of course, an extreme simplification and will not be true for all observations, but allows to define the control group in a straightforward way.

For retirees, we include all couples with both partners retired and the older one in the age range 60 to 69. Note that this group also includes couples who were once dual-earners and who might not perfectly match the couples in the control group, but for many cases in the HRS and for all observations in the EVS we are not able to control for previous labor force states of both partners.

Results are shown in table 3. These should be viewed with care for the reasons discussed above.

Overall, the order of magnitude of the estimates is roughly comparable to our main results shown in table 1. For both the US and Germany, the results for the parametric approach are a few percentage points below the main estimates, while in case of the semiparametric approach estimates are somewhat higher. The bounds of the identification intervals are rather similar. Results for income dependence are more mixed in the case of Germany, which might be due to the more heterogeneous sample.

7.3 Welfare indicators, income concept, and retirement status

We conducted several robustness checks with respect to the welfare indicator, the income concept, and the definition of the retirement status. For the U.S., we use four different income concepts, five welfare indicators, and four definitions of the retirement status. We run all possible combinations of these variables with both the parametric and the semiparametric approach, giving 160 estimates in total. A few of these are dropped, as the semiparametric approach does show convergence issues, leading to 152 estimates in total. For Germany, we also use four different income concepts, but only four welfare indicators, and two definitions of the retirement status. In combination with the parametric and semiparametric approach this yields 58 models when dropping two semiparametric estimates due to convergence issues.

For the US, the overall range of point estimates is rather wide, but roughly 78% of all point estimates arising from the sensitivity analyses fall within the identification bounds of our main results. For Germany, it is 93%. The analyses leading to estimates deviating from our main findings mostly are based on quite different welfare indicators: satisfaction with household income, which leads to lower results than expenditure-based estimates and is responsible for the larger share of estimates for the US which lie outside of the identification bounds; and nondurable expenditure including expenditure for transportation covering gas, repairs of vehicles, and public transportation.

For equivalence scales satisfaction-based estimates are known to be often lower than expenditure-based scales. This is likely due to the fact that satisfaction measures are influenced by other things than consumption, e.g., comparison of one own's situation to others (Ferrer-i-Carbonell, 2005). In case of nondurable expenditure, expenses for transportation can be rather high and strongly dependent on the need to commute; with retirement there is a big drop in these expenses, which is not reflective of a drop in the welfare level. Thus, using this welfare indicator likely overstates the replacement rate needed to maintain a constant standard of living.

More detailed results are presented in the appendix.

7.4 Instrumenting retirement status

For our analyses it is crucial that the effect of retirement on the welfare indicator is estimated correctly. This requires the retirement decision to be exogenous.

This might not be true, and individuals might select into retirement, based on expectations of their welfare level in retirement and incentives to not retire. For instance, in Germany retiring before the nominal retirement age with, say, age 63 leads to reduced pension benefits. Individuals who are well of and have large savings might not be concerned about this reduction, while individuals with no savings and expected low pensions might be and decide against retiring.

A solution to this problem is to exploit exogenous variability in pension eligibility with a (fuzzy) regression discontinuity approach (Imbens and Lemieux, 2008). For instance, in the U.S. the first claiming age is 62 and the nominal retirement age is 66, with both thresholds being set by policy makers and not influenced by individual retirement aspirations. This leads to exogenous discontinuities in the age-specific probability of retirement at the age thresholds. Regression discontinuity designs exploiting pension eligibility have been used succesfully using both U.S. and German data (Eibich, 2015). In the life cycle literature this has been used to study consumption smoothing (Battistin et al., 2009).

Here, we use linear probability models to predict the probability of retirement conditional on age, which then is used as an instrument for retirement status (Eibich, 2015). We implement this using the retirement status as described in section 5. For the U.S., we use data on individuals aged 55 to 75, and exploit the age thresholds at 62 and the nominal retirement age by cohort, which for some of the cohorts in our data is below 66. For Germany, we also use observations aged 55 to 75 to estimate the probability of retirement, using discontinuities at ages 60 and 65. While for Germany the retirement age is increasing cohort by cohort, we are not able to exploit the variation introduced through this, as it would require information on the exact day of birth, and we only have access to the year of birth. The German system allows retirement before age 65 without penalties given that a minimum number of contribution years is reached, but this does not induce a clear discontinuity (Eibich, 2015). More details on the implementation and diagnostics can be found in the appendix.

The probability of retirement is a quantitative variable, which can be easily used in combination with the parametric approach, but not with the semiparametric approach and the nonparametric approach, which require a binary retirement status. Because of this, we only can present results for the parametric approach. Moreover, the treatment effect estimated using a regression discontinuity design is a local treatment effect, which in our case means that it is essentially the effect very shortly before and after retirement. In this sense, it might be less representative of the pension needs of the retirees population than our (already restricted) main

Table 4: Results of the regression discontinuity design for the US (HRS) and Germany (EVS).

	NRR	SE	p(ESE)	n
USA	0.534	0.460	0.082	787
Germany	0.978	0.137	0.002	1608

findings.

Results of the second state of estimation are shown in table 4. For the US, the point estimate is considerably lower than for our main parametric result (53% vs 91%), but is estimated very imprecisely and is not statistically significantly different. For Germany the point estimate of the RDD is very close to the main parametric estimate. For both the US and Germany the tests of income dependence are consistent with our main results. Thus, overall endogeneity does not seem to pose a threat to our results.

8 Summary and concluding remarks

In this paper we propose a framework to assess the retirement income needed to maintain the standard of living achieved during working life. We report net replacement rates, which are defined as the net retirement income needed to maintain the living standard relative to the net pre-retirement income. Applying parametric, semiparametric, and nonparametric estimation approaches to U.S. data and German data, we find that replacement rates are slightly below 100%; i.e., the income during retirement roughly needs to be at the same levels as during working life to avoid a drop in the welfare level. We conduct extensive sensitivity checks and show that our main findings are robust with respect to the indicators of welfare, income, and retirement used, and endogeneity, among other things. Moreover, for Germany we show that replacement rates are likely flat across the whole income range, and the level of the replacement rate does not depend on pre-retirement income. For the U.S. results on income independence of replacement rates are less clear cut.

For the US, the pension provisions recent retirees receive should be enough to reach the level we find in our analyses (Love et al., 2008). For Germany, our findings point to a substantial gap in the pensions individuals receive and the pensions they need. According to Kluth and Gasche (2015), recent retirees receive a factual net replacement rate of around 70% from the public pension system, which for many people is the main source of pension income (Bönke et al., 2010).

Projections by the OECD indicate that the replacement rate provided by the public system will decrease until 2060 to around 55% (OECD, 2015). Comparing these recent and projected numbers to our main findings yields a gap in pension provisions of around 25% or more percent, which will increase to up to 40% in the future.

Given the projections of the OCED, a replacement rate close to 100% seems hard to reach in the future, especially since pension provisions will need to last longer due to increasing life expectancy. For policy makers, a replacement rate of 100% might conflict with other policy goals, such as sustainability, and likely is only possible at a cost, such as delayed retirement (Kitao, 2014). For individuals, this might require high saving efforts. Especially in the German context with little private savings for retirement today (Bucher-Koehnen and Lusardi, 2011), it seems unlikely that individuals will reach this level on their own. Informing individuals better about the pensions they will receive, the pension they need, and how they can reach a desired pension level might enable them to reach more adequate pension levels (Bernheim and Garret, 2003; Dolls et al., 2018).

One important caveat of our findings is that the replacement rates we report are population averages, potentially ignoring heterogeneity in replacement rates. Our results showed no clear indication of income dependence, though. Nevertheless, this is not the only potential source of heterogeneity, and replacement rates could depend, for instance, on health. While we provided results by age, further disentangling heterogeneity in replacement rates potentially is possible with the methods presented here, and is a potential avenue for future study.

The methods presented here can easily be adopted to other welfare measures, data sets, and countries. As our findings show, the parametric approach seems to yield reasonable estimates, and it is simple to understand and has low data demands. With respect to income independence, results were not clear cut for the US, and the nonparametric approach seems more trustworthy, though. This shows that the more complex approaches also have their use, especially as a check for simpler methods.

References

Abbring, J. H. and Heckman, J. J. (2007). Econometric evaluation of social programs, Part III. In *Handbook of Econometrics*, volume 6B, pages 5146–5303. Elsevier.

Aguiar, M. and Hurst, E. (2005). Consumption versus expenditure. *Journal of Political Economy*, 113:919–948.

Aguiar, M. and Hurst, E. (2007). Life-cycle prices and production. *American Economic Review*, 97:1533–1559.

Angelini, V., Brugiavini, A., and Weber, G. (2014). The dynamics of homeownership among the 50+ in Europe. *Journal of Population Economics*, 27:797–823.

Attanasio, O. P., Kitao, S., and Violante, G. L. (2007). Global demographic trends and social security reform. *Journal of Monetary Economics*, 54:144–198.

Attanasio, O. P. and Weber, G. (2010). Consumption and saving: Models of intertemporal allocation and their implications for public policy. *Journal of Economic Literature*, 48:693–751.

Banks, J., Blundell, R., and Lewbel, A. (1997). Quadratic engel curves and consumer demand. *Review of Economics and Statistics*, 79:527–539.

Battistin, E., Brugiavini, A., Rettore, E., and Weber, G. (2009). The retirement consumption puzzle: Evidence from a regression discontinuity approach. *American Economic Review*, 99:2209–2226.

Behagel, L. and Blau, D. M. (2012). Framing social security reform: Behavioral responses to changes in the full retirement age. *American Economic Journal: Economic Policy*, 4:41–67.

Bernheim, B. (1992). Is the baby boom generation preparing adequately for retirement? Technical Report, Princeton NJ, Merrill Lynch.

Bernheim, B. D., Skinner, J., and Weinberg, S. (2001). What accounts for the variation in retirement wealth among U.S. households? *American Economic Review*, 91:832–857.

Bernheim, D. A. and Garret, D. M. (2003). The effects of financial education in the workplace: evidence from a survey of households. *Journal of Public Economics*, 87:1487–1519.

Biewen, M. and Juhasz, A. (2017). Direct estimation of equivalence scales and more evidence on independence of base. *Oxford Bulletin of Economics and Statistics*, 79:875–905.

Binswanger, J. and Schunk, D. (2012). What is an adequate standard of living during retirement? *Journal of Pension Economics and Finance*, 11:203–222.

Blundell, R., Crawford, R., French, E., and Tetlow, G. (2016). Comparing retirement wealth trajectories on both sides of the pond. *Fiscal Studies*, 37:105–130.

Blundell, R. and Lewbel, A. (1991). The information content of equivalence scales. *Journal of Econometrics*, 50:49–68.

Bönke, T., Schröder, C., and Schulte, K. (2010). Incomes and inequality in the long run: The case of German elderly. *German Economic Review*, 11:487–510.

Boskin, M. J. and Shoven, J. B. (1984). Concepts and measures of earnings replacement during retirement. NBER Working Paper No. 1360.

Browning, M. and Crossley, T. F. (2001). The life-cycle model of consumption and saving. *Journal of Economic Perspectives*, 15:3–22.

Bucher-Koehnen, T. and Lusardi, A. (2011). Financial literacy and retirement planning in germany. *Journal of Pension Economics and Finance*, 10:565–584.

Cambanis, S., Simons, G., and Stout, W. (1976). Inequalitites for $ek(x, y)$ when the marginals are fixed. *Probability Theory and Related Fields*, 36:285–294.

Cartwright, D. I. and Field, M. J. (1978). A refinement of the arithmetic mean-geometric mean inequality. *Proceedings of the American Mathematical Society*, 71:36–38.

Cherchye, L., De Rock, B., and Vermeulen, F. (2012). Economic well-being and poverty among the elderly: An analysis based on a collective consumption model. *European Economic Review*, 56:985–1000.

Chiappori, P.-A. (2016). Equivalence versus indifference scales. *Economic Journal*, 126:523–545.

Colby, S. L. and Ortman, J. M. (2015). Projections of the size and composition of the U.S. population: 2014 to 2060. Washington, D.C.: U.S. Census Bureau.

Crawford, R. and O'Dea, C. (2012). The adequacy of wealth among those approaching retirement. Institute for Fiscal Studies Report No. R72.

De Nardi, M., French, E., and Jones, J. B. (2010). Why do the elderly save? the role of medical expenses. *Journal of Political Economy*, 118:39–75.

Deaton, A. and Muellbauer, J. (1986). On measuring child costs: With application to poor countries. *Journal of Political Economy*, 94:720–744.

Dolls, M., Doerrenberg, P., Peichl, A., and Stichnoth, H. (2018). Do retirement savings increase in response to information about retirement and expected pensions? CESifo Working Paper 6842.

Donaldson, D. and Pendakur, K. (2006). The identification of fixed costs from consumer behaviour. *Journal of Business and Economic Statistics*, 24:255–265.

Dudel, C. (2015). Nonparametric bounds on equivalence scales. *Economics Bulletin*, 35:2161–2165.

Dudel, C., Ott, N., and Werding, M. (2016). Maintaining ones living standard at old age: What does that mean? *Empirical Economics*, 51:1261–1279.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of Health Economics*, 43:1–12.

Fan, Y., Guerre, E., and Zhu, D. (2017). Partial identification of functionals of the joint distribution of "potential outcomes". *Journal of Econometrics*, 197:42–59.

Ferrer-i-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics*, 89:997–1019.

Ferrer-i-Carbonell, A. and Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *Economic Journal*, 114:641–659.

Friedman, M. (1957). *A Theory of the Consumption Function*. Princeton University Press.

Hall, P., Racine, J. S., and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, 99:1015–1026.

Hamermesh, D. (1984). Consumption during retirement: The missing link in the life cycle. *Review of Economics and Statistics*, 66:1–7.

Hann, P. and Prowse, V. (2014). Longevity, life-cycle behavior and pension reform. *Journal of Econometrics*, 178:582–601.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, 21:1926–1947.

Haveman, R., Holden, K., Romanov, A., and Wolfe, B. (2007). Assessing the maintenance of savings sufficiency over the first decade of retirement. *International Tax and Public Finance*, 14:481–502.

Hayfield, T. and Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27:1–32.

Henle, P. (1972). Recent trends in retirement benefits related to earnings. *Monthly Labor Review*, 95:12–20.

Horowitz, J. and Manski, C. (2000). Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, 95:77–84.

Imbens, G. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142:615–635.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86:4–29.

Juster, F. T. and Suzman, R. (1995). An overview of the Health and Retirement Study. *Journal of Human Resources*, 30:S7–S56.

Kitao, S. (2014). Sustainable social security: Four options. *Review of Economic Dynamics*, 17:756–779.

Kluth, S. and Gasche, M. (2015). Ersatzraten in der gesetzlichen Rentenversicherung. *Journal of Economics and Statistics*, 235:553–583.

Knoef, M., Been, J., Alessie, R., Caminada, K., Goudswaard, K., and Kalwij, A. (2016). Measuring retirement savings adequacy: developing a multi-pillar approach in the Netherlands. *Journal of Pension Economics and Finance*, 15:55–89.

Lancaster, G. and Ray, R. (1998). Comparison of alternative models of household equivalence scales: The Australian evidence on unit record data. *Economic Record*, 74:1–14.

Leser, C. E. V. (1963). Forms of engel functions. *Econometrica*, 31:694–703.

Lewbel, A. (1989). Household equivalence scales and welfare comparisons. *Journal of Public Economics*, 39:377–391.

Lewbel, A. and Pendakur, K. (2008). Equivalence scales. In Durlauf, S. N. and Blume, L. E., editors, *New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2nd edition.

Love, D. A., Smith, P. A., and McNair, L. C. (2008). A new look at the wealth adequacy of older U.S. households. *Review of Income and Wealth*, 54:616–642.

Luengo-Prado, M. J. and Sevilla, A. (2013). Time to cook: Expenditure at retirement in spain. *Economic Journal*, 123:764–789.

Lundberg, S., Startz, R., and Stillman, S. (2003). The retirement-consumption puzzle: a marital bargaining approach. *Journal of Public Economics*, 87:1199–1218.

Lusardi, A. and Mitchell, O. S. (2007). Baby Boomer retirement security: The roles of planning, financial literacy, and housing wealth. *Journal of Monetary Economics*, 54:205–224.

Manski, C. (2003). *Partial Identification of Probability Distributions*. Springer.

Meier, V. and Werding, M. (2010). Ageing and the welfare state: securing sustainability. *Oxford Review of Economic Policy*, 26:655–673.

Mitchell, O. and Moore, J. (1998). Can Americans afford to retire? New evidence on retirement savings adequacy. *The Journal of Risk and Insurance*, 65:371–400.

Modigliani, F. and Brumberg, R. (1954). Utility analysis and the consumption function: An interpretation of cross-section data. In Kurihara, K., editor, *Post-Keynesian Economics*, pages 388–436. Rutgers University Press, New Brunswick, NJ.

Moreau, N. and Stancanelli, E. (2015). Household consumption at retirement: A regression discontinuity study on French data. *Annals of Economics and Statistics*, 117/118:253–276.

Munell, A. H. and Soto, M. (2005). What replacement rates do households actually experience in retirement? CRR Working Paper 2005-10.

OECD (2013). Pensions at a glance 2013. Available online at http://www.oecd.org/pensions/public-pensions/OECDPensionsAtAGlance2013.pdf.

OECD (2015). Pensions at a glance 2015. Available online at http://www.oecd-ilibrary.org/social-issues-migration-health/pensions-at-a-glance-2015_pension_glance-2015-en.

Pagel, M. (2017). Expectations-based reference-dependent life-cycle consumption. *Review of Economic Studies*, 84:885–934.

Pantoja, P., Hurd, M., Martin, C., Meijer, E., Rohwedder, S., and St Clair, P. (2017). Tax calculations for hrs 2000 and 2014. RAND Documentation.

Pendakur, K. (1999). Semiparametric estimates and tests of base-independent equivalence scales. *Journal of Econometrics*, 88:1–40.

Pendakur, K. (2002). Taking prices seriously in the measurement of inequality. *Journal of Public Economics*, 86(1):47–69.

R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria.

Racine, J. S. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119:99–130.

Schnabel, R. (2003). *Die neue Rentenreform: Die Nettorenten sinken.* Deutsches Institut für Altersvorsorge, Köln.

Scholz, J., Seshadri, A., and Khitatrakun, S. (2006). Are Americans saving 'optimally' for retirement? *Journal of Political Economy*, 114:607–643.

Schulz, J. and Carrin, G. (1972). The role of savings and pension systems in maintaining living standards in retirement. *Journal of Human Resources*, 7:343–365.

Stengos, T., Sun, Y., and Wang, D. (2006). Estimates of semiparametric equivalence scales. *Journal of Applied Econometrics*, 21:629–639.

Szulc, A. (2009). A matching estimator of household equivalence scales. *Economics Letters*, 103:81–83.

United Nations (2015). World population ageing 2015, highlights. Department of Economic and Social Affairs, http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015.

van Praag, B. M. S. (1991). Ordinal and cardinal utility. An integration of the two dimensions of the welfare concept. *Journal of Econometrics*, 50:69–89.

VanDerhei, J. and Copeland, C. (2010). The EBRI retirement readiness rating: Retirement income preparation and future prospects. EBRI Issue Brief 344.

Wilke, F. (2014). Abschied von der Lebensstandardsicherung. Altersvorsorge im

Spannungsfeld zwischen Unsicherheit und langfristiger Zielsetzung. *Sozialer Fortschritt*, 3/2014:58–65.

Wolfson, M. C. (2011). Projecting the adequacy of Canadians' retirement income. IRPP Study No. 17.

Working, H. (1943). Statistical laws of family expenditure. *Journal of the American Statistical Association*, 38:43–56.

Table 5: Results comparing estimates based on net income and gross income, for the U.S. (HRS data) and Germany (EVS data). Source: Own calculations based on HRS and EVS.

USA	NRR	GRR
Parametric	0.912	1.156
Semiparametric	1.052	0.905
Nonparametric	[0.888,1.194]	[0.812;1.154]
Germany		
Parametric	0.968	0.881
Semiparametric	1.076	—
Nonparametric	[0.863,1.061]	[0.657;0.856]

A Replacement rates using gross income

Results for replacement rates based on gross income (gross replacement rates; GRR) are shown in table 5. NRRs are shown for comparison. Overall, GRR estimates seem to be less stable than NRR estimates. For the US identification bounds shift only slightly as compared with NRR results, while for Germany identification bounds even do not overlap. For both the US and Germany the parametric estimates are outside of the identification bounds. The semiparametric approach is doing better in case of the US, but did not converge for Germany.

B Sensitivity analyses for the main variables: Income, welfare indicator, and retirement status

B.1 Variables: Welfare indicators

For the main results presented in the text, we use the expenditure share of food at home as the welfare indicator. Expenditure for food at home is known to be only an approximate measure of household consumption (Bernheim et al., 2001). Apart from not capturing food expenditure away from home at restaurants, at work, etc. this indicator also might mismeasure consumption as expenditure might be substituted with time for home production, while consumption stays constant (Aguiar and Hurst, 2005; Luengo-Prado and Sevilla, 2013).

Because of this, we use several other indicators to assess the robustness of our results. This includes the income share of expenditure for food at home and away from home; the absolute level of expenditure for nondurable goods defined in several ways; and satisfaction with household income. Expenditure for nondurable goods has been argued to give a more complete picture of consumption (Battistin et al., 2009). Satisfaction with household income differs from the expenditure based indicators in that it is a subjective measure, assumed to capture a subjective assessment of household welfare (van Praag, 1991). While satisfaction with household income is a qualitative variable, we follow earlier literature (Ferrer-i-Carbonell and Frijters, 2004; Dudel et al., 2016) and model it similar to the other welfare indicators and thus as a metric variable.

While the expenditure based measures are readily available for the EVS, this is not the case for the HRS. Information on food away from home is collected, but excludes expenditures for food at work or school, making this variable potentially incomplete and not directly comparable to the EVS data. Nondurable expenditures are not covered in the HRS itself, but in the Consumption and Activities Mail Survey (CAMS), which was collected in 2015 after the 2014 HRS wave. This data is only available for a part of the HRS sample, leading to a small number of observations. Satisfaction with household income was covered as part of the HRS 2014 survey, but only in a leave-behind questionnaire administered to half of the respondents. This leave-behind questionnaire was not answered by all respondents, also leading to a small number of observations.

For all expenditure-based welfare indicators we removed some outliers. For both the EVS and HRS we excluded households from the analyses if the income share of food was above 70%, or if the amount spent on nondurables amounted to more than 80% of household income.

B.2 Variables: Income concepts

For our main analyes, we used the net income plus annuitized wealth and annuitized housing wealth. While this income concept can easily be implemented for both the German and the U.S. data, other income concepts could be used instead (for a discussion see Munell and Soto, 2005; Crawford and O'Dea, 2012). For instance, housing wealth might not generate steady income, and could be illiquid Angelini et al. (2014).

To assess the sensitivity of our results with respect to the income concept use, we conduct robustness checks using net income excluding wealth; net income plus annuitized non-housing wealth, i.e., excluding housing wealth; and net income plus annuitized non-housing wealth plus imputed rent instead of housing wealth. Imputed rent is readily available for the EVS data as calculated by the Federal

Statistical office, while for the HRS we generate imputed rent assuming a 5% rental yield (Munell and Soto, 2005; Crawford and O'Dea, 2012).

B.3 Variables: Retirement status

For our main analyses, for the US we used the labor force status as defined by RAND in combination with the number of weekly hours worked. In the sensitivity analyses, we use three variants: only using the RAND indicator, without dropping individuals who work more than 20 hours per week from the group of retirees; setting retirement status based on receipt of any pension income, including pension plans and annuities; and self-reported retirement status, i.e., whether respondents consider themselves to be retired or not.

For Germany, in the main analyses we used the labor force state as defined by the Federal Statistical Office and the number of hours worked to define whether an individual is retired. In the sensitivity analyses, we drop the hours worked per week.

B.4 Results and discussion

The robustness checks are summarized in Figure 1. It shows kernel densities based on 157 net replacement rate estimates for the US and 58 estimates for Germany, all either based on the parametric or semiparametric approach. Each of these models is based on a different combination of welfare indicator, income concept, and retirement status. For some, the semiparametric approach exhibited convergence issues (3 for the US, 2 for Germany). The point estimates of the parametric and semiparametric approach of our main results are shown as solid lines, and the identification bounds as dashed lines.

For the US, 78% of all estimates fall within the identification bounds of our main results. The exceptions to this form two groups. The first group are occasional outliers, mostly occurring for the semiparametric approach. Generally, the semiparametric approach produces results which are more spread out than the results of the parametric approach. For instance, using the CAMS food data and the RAND labor force status, the parametric estimates for the four different income concepts range from 102% to 111%, while for the semiparametric approach they range from 88% to 107%. This could partly be due to the small sample size, as the number of observations we can use for our analyses is only 220 for the CAMS sample. Still, the results are almost all in the identification bounds and centered around our main estimates.

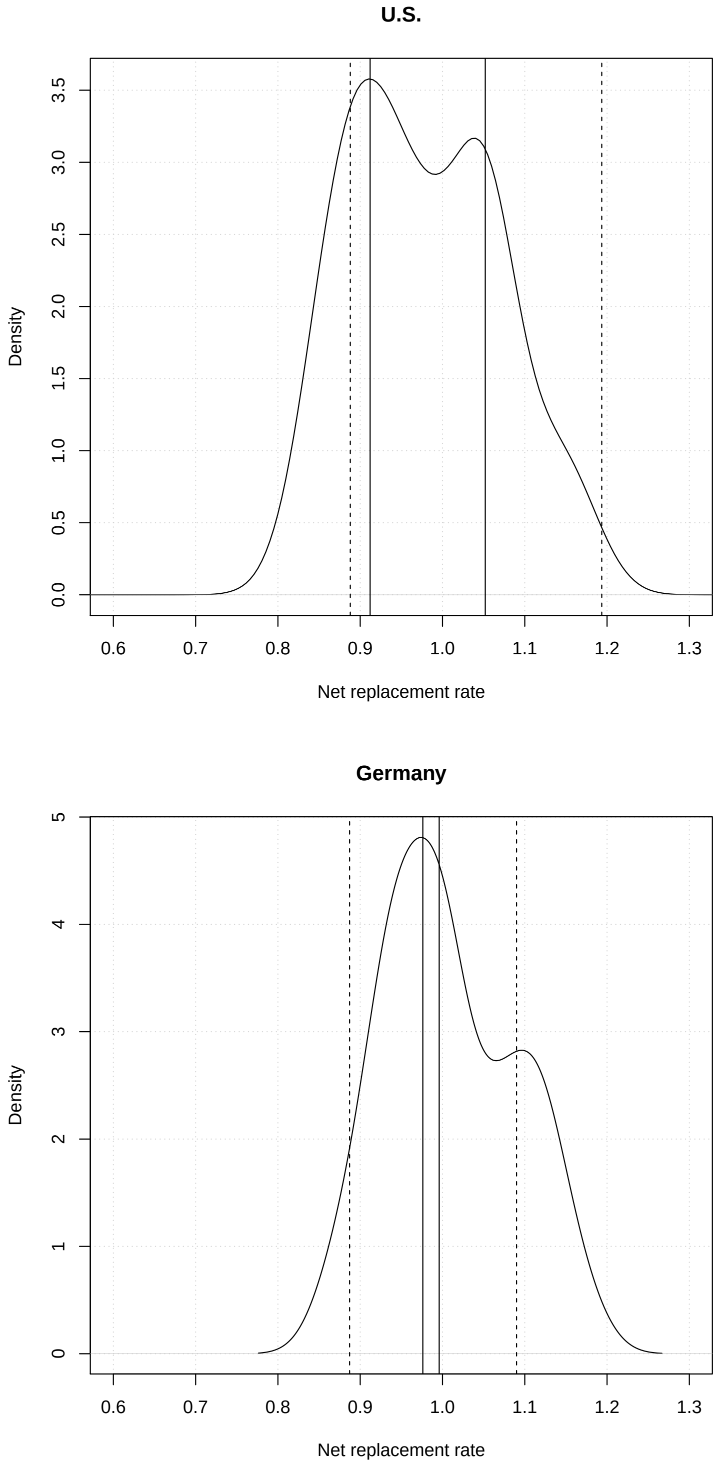

Figure 1: Kernel density estimates of the net replacement rates resulting from different welfare indicators, income concepts, and definitions of the retirement status. For the US results are based on 157 estimates of net replacement rates; for Germany on 58 estimates. Source: Own calculations using HRS and EVS.

The second group are some of the estimates based on expenditures for food, which are below the lower identification bound of 89%. All of these are estimates 83% or more, and thus not far away from the lower bound.

For Germany, the results of the robustness checks are insofar different compared with the US in that the overall range of estimates is lower, and there are more estimates above the identification bounds, while there are no estimates below the bounds. This is due to the different welfare indicators used: no satisfaction based measure is available in the EVS, but different definitions for nondurables can be applied, while still maintaining a large sample size (in contrast to the HRS and CAMS). High estimates of the replacement rate around 110% and above mostly result when using a rather wide concept of nondurables which includes, among other things, nondurable expenditure for transportation covering gas, repairs of vehicles, and public transportation. These expenses can be rather high and strongly dependent on the need to commute; with retirement there is a big drop in these expenses, which is not reflective of a drop in the welfare level. Thus, using this welfare indicator likely overstates the replacement rate needed to maintain a constant standard of living.

C Additional results on the RDD

C.1 Implementation

To estimate the probability of retirement, we use a linear probability model. d_i denotes the binary retirement status (1=retired,0=not retired). Let x_i be individual age, and let $\mathbb{I}(\cdot)$ be the indicator function. Roughly following Eibich (2015), we specify the model for Germany as

$$\begin{aligned} d_i &= a + b_x x_i + b_{60}\mathbb{I}(65 > x_i \geq 60) + b_{65}\mathbb{I}(x_i \geq 65) \\ &+ b_{x60} x_i \mathbb{I}(65 > x_i \geq 60) + b_{y65} x_i \mathbb{I}(x_i \geq 65) + \epsilon_i, \end{aligned}$$

where ϵ_i is an error term, and b_x, b_{60}, b_{65}, b_{x60}, and b_{x65} are the coefficients to be estimated. Coefficient estimates are then used to predict $\mathrm{E}(d_i|x_i)$. These predicted values are then used to instrument retirement status in equation (4.3). For the U.S. we use 62 instead of 60, and 65 is replaced with the retirement age r_i applying to each individual.

C.2 Diagnostics

D Details on the semiparametric approach and the nonparametric approach

The semiparametric approach and the nonparametric approach make use of nonparametric kernel regression and nonparametric estimates of quantile functions. We use the implementation in the `np` package for R, developed by Hayfield and Racine (2008). This package implements generalized product kernels, which allow mixing of continuous and categorical explanatory variables (Racine and Li, 2004). These kernels are defined as

$$K(\mathbf{x}_i - \mathbf{x}) = \prod_{k=1}^{m_1} \frac{1}{h_k} K_k^{(c)}(x_{ki} - x_k) \prod_{l=m_1+1}^{m} K_l^{(d)}(x_{ki} - x_k),$$

where $\mathbf{x}$ is a vector of explanatory variables with elements x_k. m is the number of elements of $\mathbf{x}$, with the first m_1 elements being continuous and the other $m - m_1$ elements being categorical. h_k is the bandwidth for variable x_k. The type of the kernel function depends on the type of variable, where $K^{(c)}$ indicates the kernel function for continuous variables, and $K^{(d)}$ the kernel function for categorical variables. For continuous variables, we used a second-order Gaussian Kernel and for the categorical case the kernel function proposed by Aitchison and Aitken was utilized (Hayfield and Racine, 2008). For bandwidth selection, see Hall et al. (2004).

For statistical inference, for the semiparametric estimation approach we use a residual bootstrap as proposed by Pendakur (1999). For the nonparametric estimation approach, we use the resampling bootstrap. This procedure could potentially be conservative (see Härdle and Mammen, 1993), but residual-based approaches are not well defined for estimates of conditional quantile functions. For both the semiparametric estimation approach and the nonparametric estimation approach standard errors and confidence intervals are based on 1000 bootstrap replications.

Confidence intervals are based on percentiles of the bootstrap replications. In case of the nonparametric identification region, we construct an interval which covers the complete identification region with a fixed probability (95%), roughly similar to Horowitz and Manski (2000). If (L_s, U_s) denote the bounds resulting for the sth bootstrap replication, then we choose the bounds l and u of the confidence interval such that $l < L_s$ and $U_s < u$ for 95% of the bootstrap replications. As l

and u will usually not be unique, we choose the values of l and u for which the interval width, $u - l$, is smallest.

To calculate the smallest interval, an iterative procedure is used and re-run several times. Each run takes the 5% percentile of the lower bound L and the 95% percentile of the upper bound U as starting values, perturbed with noise $e_L \sim \mathcal{N}(0, \mathrm{sd}(L))$ or $e_U \sim \mathcal{N}(0, \mathrm{sd}(U))$, respectively. Let the resulting bounds be denoted by $L^{(0)}$ and $U^{(0)}$. The coverage achieved with these values is equal to $\rho^{(0)}$. If $\rho^{(0)}$ is smaller than $1-\alpha$, $L^{(0)}$ and $U^{(0)}$ are decreased and increased, respectively, by a stepsize $\lambda_L = 0.1\mathrm{sd}(L)$ or $\lambda_U = 0.1\mathrm{sd}(U)$ to get new values: $L^{(1)} = L^{(0)} - \lambda\epsilon_L^{(0)}$ and $U^{(1)} = U^{(0)} + \lambda\epsilon_U^{(0)}$, where $\epsilon_U^{(0)}$ and $\epsilon_L^{(0)}$ follow a uniform distribution. If $\rho^{(0)}$ is larger than $1-\alpha$ the signs for λ are changed to instead decrease the interval width. $\rho^{(1)}$ is the coverage achieved after these adjustments. Depending on whether it is above or below $1-\alpha$, the adjustments are applied to get updated values $L^{(2)}$ and $U^{(2)}$; $\rho^{(2)}$ is checked against $1-\alpha$ again etc. until $\rho^{(k)} = 1-\alpha$. This procedure is re-run for 100 different starting values and the interval with the smallest width is reported.

The semiparametric approach makes use of simulation-based inference to assess income independence, where we follow Pendakur (1999). Income independence is assessed by comparing the empirical value of $L(\alpha, \mu)$, as defined in equation (4.5), to simulated values of L arising from assuming income independence. These are generated by using the residual bootstrap, where predicted values for retirees are generated using the kernel regression function of non-retirees shifted by the estimates of α and μ. Using this bootstrap sample, the semiparametric approach is fitted again, yielding a value of L^* based on shape invariance and thus income independence. The distribution of simulated values L^* is used to assess the probability of the empirical value of L.

E R functions