

Max-Planck-Institut für demografische Forschung
Max Planck Institute for Demographic Research

Konrad-Zuse-Strasse 1 • D-18057 Rostock • Germany • Tel +49 (0) 3 81 20 81 - 0 • Fax +49 (0) 3 81 20 81 - 202 • www.demogr.mpg.de

MPIDR Working Paper WP 2019-003 | January 2019

**Pension adequacy standards:
An empirical estimation strategy
and results for the United States and
Germany**

Christian Dudel | dudel@demogr.mpg.de
Julian Schmied | schmied@demogr.mpg.de

This working paper has been approved for release by: Peter Eibich (eibich@demogr.mpg.de),
Deputy Head of the Research Group: Labor Demography.

© **Copyright is held by the authors.**

Working papers of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in working papers are attributable to the authors and do not necessarily reflect those of the Institute.

Pension adequacy standards: An empirical estimation strategy and results for the United States and Germany*

Christian Dudel[†]

Julian Schmied[‡]

Abstract

Due to increasing life expectancy pension benefits and savings need to last longer, which casts doubt on the financial security of retirees. Surprisingly, what pension benefit level can be considered adequate remains unclear. In this paper, we propose an identification strategy for the estimation of pension adequacy standards. Applying a range of econometric techniques to data from the U.S. and Germany, we find that a net pension income of around 100% of the last net working life income, plus or minus 10 percentage points, can be considered adequate.

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

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

*Financial support provided by the Deutsche Forschungsgemeinschaft (DFG), grant number 277165179, is gratefully acknowledged. We thank Thomas Crossley and Martin Werding for their useful comments and suggestions. We also thank seminar participants at the Institute for Fiscal Studies, at the Centre for Population Change, at the Free University of Berlin, and at the Max Planck Institute for Demographic Research; and conference participants at the 35th IARIW General Conference in Copenhagen, 2018, and at the annual meeting of the German Statistical Society in Rostock, 2017. The usual disclaimer applies.

[†]Corresponding author; Max Planck Institute for Demographic Research, Konrad-Zuse-Str. 1, 18057 Rostock, Germany; email: dudel@demogr.mpg.de

[‡]Max Planck Institute for Demographic Research; Freie Universität Berlin; email: schmied@demogr.mpg.de

1 Introduction

In many high-income countries the population is aging, and the share of the population aged 65 or 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 or 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 the sustainability of government programs into question, as spending on pensions increases (Attanasio et al., 2007). In response to these trends, the pension systems in many countries 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 born between 1943 and 1954, and it will increase further for the cohorts born later (Behagel and Blau, 2012). The aim of raising the retirement age is to increase the length of working life, and to thereby ensure the sustainability of the Social Security program. In Germany, the statutory retirement age is also increasing (OECD, 2013), and private pension savings are playing an increasingly important role.

In this context of population aging, concerns have been raised about the financial security of retirees, and about whether the pension incomes they receive are adequate. One indicator that is commonly used to measure pension adequacy is the replacement rate, which is defined as post-retirement income relative to pre-retirement income (Boskin and Shoven, 1984). For example, a replacement rate of 80% would imply that an individual's post-retirement income is equivalent to 80% of her pre-retirement income. Surprisingly, no clear benchmark exists for assessing what level of the replacement rate can be considered adequate. In the literature, a wide range of values of between roughly 60% and 100% can be found (Love et al., 2008).

In this paper, we present an approach to derive the replacement rate an individual would need to maintain the living standard achieved by the end of working life. Specifically, our approach is based on keeping the individual's welfare level constant shortly before and after retirement. We propose using this replacement rate as a benchmark for pension adequacy, and call it the adequate replacement rate. We discuss econometric identification of the adequate replacement rate, and we show several methods for estimating it. These methods are applied to data from the U.S. and Germany. Potential limitations of our approach are examined through sensitivity checks.

The adequate replacement rate is an attractive measure of pension adequacy for several reasons. First, our benchmark is easily interpretable, as a constant

living standard is easy to understand. This makes it a useful policy indicator. Second, and related to the first point, the pension systems in some countries have been expected to provide a constant living standard, and policy debates in these countries are sometimes still framed with this goal in mind (e.g., Wilke, 2014). Third, while the conceptual framework we present is rather general, it is easy to apply, and it has modest data demands.

While other methods have been used to estimate replacement rates, they do not imply adequacy. Often, the benchmark levels are set heuristically, and the value chosen is justified as being a common choice in the literature (e.g., Haveman et al., 2007). Another approach is based on calculating the lowest possible replacement rate needed to avoid living in poverty (Love et al., 2008). While this rate represents the minimum level necessary to meet the goal of adequacy, it is likely not sufficient. The life-cycle model and its extensions have also been used to derive benchmarks. A somewhat simplified description of the life-cycle model is that the model is based on the assumption that individuals smooth the marginal utility of consumption over the life course. Our approach, by contrast, is based on the goal of keeping constant the welfare of individuals around the time of retirement. Calibrated to empirical data life-cycle models imply a certain level of the replacement rate (e.g., Mitchell and Moore, 1998). While life-cycle models have been proven to be valuable for assessing saving behavior, it is less clear how optimality relates to adequacy. Crawford and O’Dea (2012) have made the point that optimal saving behavior may not lead to an adequate pension level.

From a methodological perspective, we present a general identification strategy for determining adequate replacement rates based on the potential outcomes framework (e.g., Imbens, 2004), and that is compatible with any indicator of welfare. We focus on expenditure data, and use subjective measures of welfare to perform robustness checks. We apply parametric, semiparametric, and nonparametric methods for estimation. The parametric method and the semiparametric method were originally devised to estimate equivalence scales (Deaton and Muellbauer, 1986; Pendakur, 1999), while the nonparametric method is based on recent results on partial identification by Fan et al. (2017).

Applying all of the procedures we discuss in this paper provides us with easily interpretable results from the simpler approaches, while enabling us to test the assumptions and the robustness of these results using the more sophisticated methods. A requirement of the parametric estimation approach and the semiparametric estimation approach is that the replacement rates do not depend on the pre-retirement income. It is thus assumed that an individual with a low income

during working life needs the same replacement rate as an individual with a high income to maintain a constant living standard. The nonparametric approach is not based on this assumption, and yields set estimates only; i.e., partially identified estimates, as described in Manski (2003). In our analyses, we use a parametric test and a semiparametric test drawn from the literature on equivalence scales to assess income independence, and we propose and apply a simple nonparametric test. We also conduct additional checks to assess how sensitive our results are with respect to endogeneity and other potential issues. The outcomes of these checks indicate that our main findings are rather robust.

We study two countries: the United States and Germany. For the U.S., our analysis is based on data from the most recent wave of the Health and Retirement Study, which was conducted in 2014. For Germany, we use the most recent wave of the Income and Expenditure Survey (Einkommens- und Verbrauchsstichprobe), which was conducted in 2013. The U.S. and Germany have very different pension systems. The German retirement system is usually considered to be the archetype of the Bismarckian model, as it relies heavily on social security contributions. Until recently, the role of private savings in the German system was small. By contrast, the U.S. retirement system is a Beveridge system based on taxes and private savings play a large role. While we do not expect that the adequate replacement rates differ by country, finding the same replacement rate would have different implications in each of these two countries, as the absolute pension levels differ between the U.S. and Germany. Thus, even if the two countries had the same adequate replacement rate, there could be a gap between the actual and the required pension levels in one country, but not in the other.

In summary, we contribute to the literature in several ways. First, in this paper we provide a blueprint for the empirical estimation of pension standards by establishing a conceptual framework for the estimation of adequate replacement rates, and by discussing and comparing estimation approaches with different levels of econometric sophistication and different underlying assumptions. Second, we investigate the question of whether adequate replacement rates depend on pre-retirement income, and we apply econometric tests, including a new nonparametric test. Third, we provide comparable benchmarks for pension adequacy in the U.S. and Germany. Fourth, our benchmark can be used to help individuals make informed decisions about saving for retirement, which are becoming more important as longevity increases and public pension benefits decline. In addition, our benchmark can help policy-makers assess the well-being of the retiree population, and can help pension providers and insurance companies ensure that

individuals have access to the pension plans they need. Fifth, together with this paper, we provide functions for the statistical software R that readily implement our methods, making them easily applicable to other data sets and countries (<https://osf.io/hde4y/>).

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 explained in section 4. The data we use is described in section 5. We present our main findings in section 6, and additional findings and sensitivity checks in section 7. Section 8 concludes.

2 Related work

Intuitively, it might appear that a replacement rate of 100% would allow a retiree to maintain a constant living standard, at least if the time shortly before and after retirement is considered. It may therefore be assumed that if having a certain income level enabled an individual to achieve a certain living standard before retirement, then having the same income should be sufficient to maintain this living standard after retiring.

In the literature, however, several reasons for why a replacement rate of 100% may be either above or below the adequate level have been put forward. On the one hand, values below 100% may be considered adequate given that retired individuals have no work-related expenses (e.g., commuting), are unlikely to have children living in the household, no longer have to save for retirement, and have more time for household production (Aguiar and Hurst, 2005; Love et al., 2008). On the other hand, replacement rate values above 100% may be required because retired individuals could find that their health-related expenses are increasing with age, and that precautionary saving is therefore necessary (Blundell et al., 2016). In addition, because retirees have more free time, they may wish to spend more money on leisure activities (Crawford and O’Dea, 2012).

Taxes might also play a role. To what extent differences in the taxation levels of retirees and non-retirees affect replacement rates depends on whether gross replacement rates or net replacement rates are considered; i.e., whether replacement rates are based on gross income before and after retirement, or on net income before and after retirement. If gross replacement rates are considered, it may be argued that replacement rates below 100% are adequate, given that retirees are usually taxed at lower rates than income earners. If net replacement rates are

considered, the differences in the taxation levels of retirees and non-retirees should not matter.

In this paper, we will focus on the estimation of net replacement rates, as these rates are more readily comparable across countries. In the literature, both gross and net replacement rates can be found. In the appendix, we also supply estimates of gross replacement rates to make it easier to compare our results with those of other studies.

In the literature on pension savings and incomes, heuristic benchmarks for pension adequacy are often used. For instance, Haveman et al. (2007) assumed for the U.S. that a net replacement rate of 70% is adequate, while Schulz and Carrin (1972) used a value of 80%. According to Love et al. (2008), replacement rates of between 70% and 100% are common. Similar values can be found in the literature for Germany. In many of these studies, the authors do not justify the chosen value, other than by declaring that the replacement rate is in the established range. In research on Germany, the authors sometimes justify using a net replacement rate value of 70% by arguing that 70% is the highest value that has ever been provided by the German public pension system (Schnabel, 2003).

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

A theoretically grounded approach is based on the life-cycle model, which was introduced by Modigliani and Brumberg (1954) and Friedman (1957). The life-cycle model assumes that the marginal utility of consumption is smoothed over the life course, and that – at least in simple variants – consumption itself is also smoothed. This implies that for an individual to maintain a consistent consumption level, pre-retirement income and post-retirement income should not differ too much, except perhaps after taking changes in work-related expenses or taxation levels into account (Wolfson, 2011). A similar reasoning without recourse to the life-cycle model was proposed by Henle (1972), who argued that equal levels of disposable pre-retirement income and post-retirement income are needed.

While much of the literature on life-cycle models focuses on optimal saving behavior, life-cycle model estimates also imply replacement rates. For the U.S., these rates 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, results by Scholz et al. (2006) imply a replacement rate of around 66%. These results showed that the life-cycle model does not necessarily imply constant consumption, and thus no replacement rates around 100%.

While replacement rates derived from life-cycle models are optimal given the model assumptions and optimization constraints, they are not necessarily adequate (Crawford and O’Dea, 2012), as a replacement rate based on optimal saving might not be adequate, and an adequate replacement rate might not be optimal. For instance, whereas an individual with a low income might save optimally but still not receive an adequate pension income after retiring, an individual with a high income might not save optimally but still have an adequate income after retiring.

In contrast to the approaches that rely on expenditure data, Binswanger and Schunk (2012) used empirical data on subjective assessments to estimate pension adequacy. They asked individuals in the U.S. and in the Netherlands about their preferred retirement incomes. More specifically, based on the respondents’ current income levels, they presented several pairs of pre-retirement income and post-retirement income levels, each of which represented different retirement saving choices and resulting replacement rates. For example, a respondent was given the choice between having a high disposable pre-retirement income with a low savings level and a correspondingly low retirement income; or a low disposable pre-retirement income and a high post-retirement income and replacement rate. Their results showed that both the American and the Dutch respondents preferred net replacement rates of between 80% and 100%. As many of the surveyed individuals were below retirement age (with a median age of between 51 and 52), the results of Binswanger and Schunk (2012) are partly based on the expectations individuals have about their needs in retirement.

Dudel et al. (2016) also looked at subjective assessments using an approach close to the one presented later in this paper. Based on data on individual satisfaction with household income and applying the equivalence scale framework, they calculated the replacement rate needed to keep an individual’s standard of living constant. Using German panel data, they estimated that a replacement rate of between 82% and 90% is adequate.

3 Conceptual framework

In this section, we develop the conceptual framework we use, and discuss its implications. Let $V(\mathbf{z}, y, \mathbf{p})$ 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 ($d = 0$). The vector $\mathbf{z}_d = (d, \mathbf{z})$ consists of all characteristics including retirement status.

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

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

where \mathbf{z}'_0 captures the covariate values before retirement; and, similarly, \mathbf{z}''_1 includes the values after retirement. 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 in terms of their retirement status, we will mostly assume that the retirees and the non-retirees are similar, unless otherwise indicated; i.e., $\mathbf{z}' = \mathbf{z}''$. This is, however, not a requirement of our approach, and we will also present some results for which \mathbf{z}' and \mathbf{z}'' differ; e.g., results by age. As we use cross-sectional data, we also assume that prices are fixed and the same for all households; i.e., $R(\mathbf{z}'_0, \mathbf{z}''_1, u, \mathbf{p}) = R(\mathbf{z}'_0, \mathbf{z}''_1, u)$.

This setup varies from the one proposed by Boskin and Shoven (1984) for replacement rates, and the one usually considered in life-cycle models. Specifically, while we assess welfare at a single point in time only and require that the utility level is constant, the other approaches require that marginal utility is smoothed over time. In our approach, we avoid assumptions about the separability of utility over time and discounting, which are necessary for life-cycle models (Attanasio and Weber, 2010). Moreover, changes in the household composition over the life course, such as children entering or leaving the household, will not complicate our analysis (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 with different compositions; e.g., couples with children and couples without children. In these comparisons, the retirement indicator d is replaced with an indicator of household composition. Whereas the literature on equivalence scales often starts

from household utility functions (this is also the case for the life-cycle literature; see Attanasio and Weber, 2010), we consider the (indirect) utility functions of individuals. Household utility functions require that strong assumptions are met, and ignore decision-making processes and the allocation of resources within the household (Chiappori, 2016). This is also the case for the elderly, as was shown empirically by Lundberg et al. (2003) and Cherchye et al. (2012). As the data we use is at the household level only, and therefore does not allow us to examine the welfare of each individual in the household, we have chosen to restrict ourselves to studying single-person households. We also conduct sensitivity checks in which two-person households are included in the analysis.

Even after restricting our analysis 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 that measures the welfare level u . We focus on expenditure data and use several different indicators. We conduct robustness checks with respect to the choice of indicator, including subjective measures of welfare (see section 5.2). Our approach only requires that the welfare indicator is comparable across individuals in the sense that a specific value u' has the same meaning for all individuals; i.e., that 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.

An important issue that arises in this context is income independence, or independence of base (Lewbel, 1989). Income independence means that the replacement rate needed to maintain a constant living standard does not depend on income y , and, in turn, on the welfare level u ; i.e., $R(\mathbf{z}'_0, \mathbf{z}''_1, u) = R(\mathbf{z}'_0, \mathbf{z}''_1)$. This means, for instance, that the adequate replacement rate is the same for a person with a high income during working-life as it is for a person with a 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, as yet, no similar empirical results for replacement rates, the findings on expected replacement rates from Binswanger and Schunk (2012) suggest that income independence might be violated. In their survey on preferred replacement rates, they found that individuals with low incomes tend to prefer higher replacement rates than individuals with higher incomes. In the next section,

we discuss several tests that can be used to assess income independence, including a new one.

4 Identification strategy

4.1 Basic estimation problem

To discuss the identification and the 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 that captures whether individuals are retired or not. Let W_d denote the welfare level, measured through a welfare indicator based on, for example, expenditure data or subjective measures. $Y_d(w)$ is the income individuals need to achieve welfare level w given their retirement status. In the following, we assume that observed income is equal to $Y_d(W_d)$, and is thus equal to the income function $I(\mathbf{z}_d, u, \mathbf{p})$ defined in the previous section. As we mentioned in section 2, we will assume that income is the net income after taxes and transfers, but all of the calculations described here work with gross income as well.

Each individual is observed as being either retired or not retired, and never as being both at the same time. Thus, we observe $(W_0, Y_0(W_0))$ for the non-retired, and $(W_1, Y_1(W_1))$ 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 $Y_1(W_0)/Y_0(W_0)$. Taking expectations we have

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

where \mathbf{z} is a vector of covariates, assuming that the covariate values are the same before and after retirement. With some simple algebra, this can be shown to be

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

As we explain 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 while the second term requires the joint distribution of Y_1 and Y_0 (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 differences between retirees and non-retirees, as long as they can be captured by \mathbf{z} . In addition, a further assumption is needed. The overlap condition requires that $0 < \Pr(d = 1|\mathbf{z}) < 1$. Somewhat simplified, this means that for all values of \mathbf{z} , there are both retirees and non-retirees. Moreover, the stable unit treatment value assignment is invoked, which implies that whether one individual is or is not retired does not depend on the retirement status of other individuals.

The second term in equation (4.2) is not identified using these assumptions (Abbring and Heckman, 2007). Specifically, this is because of the covariance of $Y_1(W_0)/Y_0(W_0)$ and $Y_0(W_0)$, which requires knowledge of the joint distribution of Y_1 and Y_0 . This covariance term captures to what extent the replacement rate depends on the income before retirement. It thus captures the income independence of the adequate replacement rate: the replacement rate does not depend on the baseline income if the covariance is zero, but otherwise it does. If the covariance is negative, then higher baseline incomes are accompanied by lower replacement rates; and if the covariance is positive, then higher incomes are accompanied by higher replacement rates.

In the following, we outline three different approaches for estimating equation (4.2): a parametric approach and a semiparametric approach based on the equivalence scale literature, and a nonparametric method. Each of these approaches go along with different tests to empirically assess income independence, and they differ in terms of their complexity and underlying assumptions. Applying all of the estimation procedures and tests we discuss enables us to provide easily interpretable results for the simpler approaches, while testing the assumptions and the robustness of these results using the more sophisticated approaches.

We discuss a fully parametric approach that can be implemented very easily via standard linear regression, but that requires the strong assumption that the relationship between income and the welfare indicator is log-linear. It also assumes income independence. While the semiparametric approach we borrow from Pendakur (1999) and Stengos et al. (2006) does not require log-linearity, it still requires income independence. 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. While it requires neither log-linearity

nor income independence, it yields only identification bounds on replacement rates, and no point estimates.

For all of these approaches and tests, we provide functions for the statistical software R (R Core Team, 2017) that make using our framework easy (see <https://osf.io/hde4y/>). For the implementation of the semiparametric and nonparametric approaches, we use the `np` package provided by Hayfield and Racine (2008).

4.2 Parametric approach

A classical approach for the estimation of equivalence scales attributed to Engel can be adapted to our estimation problem 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 us 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),

$$W_d = a + \log Y b_Y + db_d + \mathbf{z}'\mathbf{b}_z + \epsilon, \quad (4.3)$$

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 who are similar with respect to \mathbf{z} ; assume that they have the same welfare level; and equate both variants of the equation and solve for Y_1/Y_0 . This yields

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

Thus, in a 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 not depend on Y_0 or on W , and it implies that the covariance term in equation (4.2) is zero.

The fully parametric framework of equation (4.3) allows us to assess income independence by adding nonlinear terms. For instance, a quadratic term of log income can be added. As was discussed for equivalence scales by Lancaster and Ray (1998), this yields a formula for the income ratio that is more complicated

than equation (4.4), and that depends on the baseline income. If the quadratic term is statistically significant, income independence can be rejected. We use this as a test for income independence.

4.3 Semiparametric approach

The log-linear functional form of the Working-Leser model might not hold empirically, depending on the welfare indicator used (Banks et al., 1997). Pendakur (1999) proposed a semiparametric approach for the estimation of equivalence scales that does not require strong assumptions about the functional form of the relationship between the 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 no other covariates \mathbf{z} are 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, as suggested by Pendakur (1999), we use kernel regression, with

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

where K_h is a kernel function with bandwidth h . 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 have a low weight. Thus, the estimate of W at Y is based not only on observations with $y_i = Y$, but also on observations close to Y . In a next step, we follow Stengos et al. (2006) and estimate α and μ such that

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

is minimized, where n is the sample size. Put simply, 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 appendix for details of our implementation.

Rephrasing the approach in terms of the potential outcomes framework, for each individual i we observe $y_i = y_i^0(1-d) + y_i^1d$ and $w_i = w_i^0(1-d) + w_i^1d$. 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 . Equation (4.5) means that α is chosen so that the imputed welfare level, \hat{w}_i , is as close as possible to the observed welfare level; i.e., $\hat{w}_i \approx w_i$.

The semiparametric approach requires that the covariance between the replacement rate and baseline income is zero. It further requires shape invariance of the Engel curves. Shape invariance means that the relationship between the log income and the welfare indicator has the same shape for both retirees and non-retirees, except for shifts of the curve by α and μ (Pendakur, 1999). Shape invariance is thus more general than the assumption of log-linearity employed for the parametric approach. It is sufficient, but not necessary, for income independence.

The testing of shape invariance with the semiparametric approach proceeds using simulations based on a comparison of the value of $L(\alpha, \mu)$ that is calculated from empirical data to simulated values that are generated assuming shape invariance (see the appendix for details). The distribution arising from the simulated values is used to assess the probability of the empirically observed value given shape invariance. If this probability is below the conventional 5%, threshold shape invariance is rejected.

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 the adequate replacement rate is income independent. They also require shape invariance or linearity of Engel curves.

Dropping the assumptions of income independence, shape invariance, and linearity, 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) built on work by Cambanis et al. (1976) on Frechét-Hoeffding bounds to show how bounds on expectations of the form $E(k(Y_1, Y_0)|\mathbf{X})$ can be derived, where $k(\cdot)$ is a strictly subadditive function. The lower bound can be calculated as

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

and the upper bound is given by

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

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 functions 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 us to calculate bounds for $E(Y_1/Y_0|\mathbf{X})$. This can be used to derive bounds on the covariance term in equation (4.2), which is bounded by

$$E(Y_1|\mathbf{X}) - E^U(Y_1/Y_0|\mathbf{X})E(Y_0|\mathbf{X}) \leq \text{Cov}(Y_1/Y_0, Y_0|\mathbf{X}) \leq E(Y_1|\mathbf{X}) - E^L(Y_1/Y_0|\mathbf{X})E(Y_0|\mathbf{X}), \quad (4.8)$$

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

The bounds of the covariance can be used to test income independence. Assuming that income independence holds, the identification bounds of the covariance as given by equation (4.8) include zero. We propose using this to test income independence. If the confidence interval of the identification region of the covariance includes zero, then the data is consistent with income independence, and income independence cannot be rejected. Specifically, we construct confidence intervals that cover the entire identification interval with a fixed probability; roughly in line with the approach used in Horowitz and Manski (2000), albeit with some modifications (see the appendix). We implement this approach using the bootstrap, and we choose the smallest interval that covers the identification intervals of 95% of the bootstrap resamples.

5 Data

5.1 Data sets: HRS and EVS

We use two different data sets for our analysis. For the U.S., we employ data of the Health and Retirement Study (HRS) from 2014, while for Germany we use data

of the Income and Expenditure Survey (Einkommens- und Verbrauchsstichprobe; EVS) from 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 five years by the German Federal Statistical Office. The last available wave was conducted in 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 analysis, we restricted both the HRS sample and the EVS sample in two ways. The first restriction is with respect to household size, as we include only single-person households. This was done to avoid the need to introduce household utility functions (see section 3). Moreover, the resulting samples are more homogeneous, as it is, for instance, likely that the bequest motives of single adults differ from those of couples (De Nardi et al., 2010; Hann and Prowse, 2014). Moreover, it is not always clear whether households of couples should count as households of retirees, as one partner might be retired while the other is not, which would complicate the analysis (Moreau and Stancanelli, 2015). Finally, we also excluded individuals living alone who reported being married, but the number of such cases was very small number.

The second restriction is of the age range: we restricted our samples to respondents aged 60 to 69. Restricting the samples to these ages allows us to compare non-retirees who were relatively close to retirement and retirees who had retired relatively recently; as for most of the EVS respondents in this age range the statutory retirement age was 65 or slightly older, while for the HRS 2014 respondents in this age range the lowest retirement age was 62 and the full retirement age was around 66. This restriction helps to guarantee that for all of the ages under consideration there are both retirees and non-retirees in the samples, as is required for identification in the overlap assumption (see section 4). If the samples had covered older ages, they might have consisted mainly of retirees

or very few non-retirees; whereas if the samples had covered younger ages, they might have consisted mainly of non-retirees.

For the HRS, these restrictions leave us with 878 observations for which the main variables are available, and 798 observations for which we have complete information on the main variables and all control variables (see below). For the EVS, there are 2310 observations with complete information on the main variables and the control variables.

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

Three ingredients are required in applying our approach: a welfare indicator, income, and retirement status.

As the main indicator for the welfare level we use the income share of expenditures for food at home. We chose this as the welfare indicator as it can be calculated for both the HRS and the EVS data. It has long been used for the estimation of equivalence scales, building on Engel’s observation that the share of income devoted to food expenditures decreases as (log) income increases (Deaton and Muellbauer, 1986). Spending on food is also commonly used as an indicator of household consumption in the life-cycle literature (e.g., Browning and Crossley, 2001; Smith, 2006). Moreover, this indicator is appealing because it is available in many other data sources, and not just in the EVS and the HRS. Still, as we are aware of the potential issues that can arise when using this indicator (e.g., Aguiar and Hurst, 2005), we have conducted extensive sensitivity checks using alternative welfare indicators. These checks generated qualitatively similar results (see below and the appendix for sensitivity checks).

As the main income concept we use net household income plus annuitized wealth including housing. This means that the replacement rates we calculate are net replacement rates. While the EVS measures net income, the HRS only captures gross income. To calculate net income for the U.S., we make use of the tax simulations provided by RAND, and follow the methodology of Pantoja et al. (2017) and Blundell et al. (2016). For both the HRS and EVS we annuitize non-housing wealth with 2.5% annually, and housing wealth with 1.25% annually (similar to Crawford and O’Dea, 2012). For non-retired individuals, we exclude all components of wealth that are related to pension savings; for Germany, this includes life insurance. For some asset types, such as stocks or cash savings, it is not clear whether they are intended as retirement savings. These assets are always included in wealth.

To remove outliers with respect to income, we drop households with net incomes below zero. For households in 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 our analysis. Among the households in the EVS, incomes below zero are rare, but are possible because of the way the Federal Statistical Office calculates household income. By design, no households with incomes above 12,000 euros per month are included in the EVS. Thus, there is no need to drop the high-income outliers. As the threshold below which we drop households from our analysis, we use the German welfare benefit level to which all individuals with incomes below this threshold are entitled.

Determining whether an individual is retired is rather straightforward for the EVS sample, as retirement is a comparatively clear-cut transition in Germany. Specifically, we use an indicator readily included in the data that captures the labor force state of the respondent, and whether she was retired. In addition, we make use of information on working hours, and assume that all individuals who were working 20 or more hours per week were not retired. There are, however, very few individuals whose labor force status was retired and who reported working more than 20 hours per week.

For the HRS sample, assigning the retirement status is more complicated, as several definitions of retirement status can be used (e.g., Behagel and Blau, 2012). Our main analysis uses the labor force status provided by RAND, and we conduct several sensitivity checks 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 were working more than 20 hours per week were not retired.

For all three main variables (welfare, income, retirement) both the HRS and EVS offer alternatives. For instance, instead of data on expenditures on food at home, we could use data on non-durable expenditures, which are available in both the HRS and EVS. Another alternative indicator is satisfaction with income, which is available in the HRS only. We have we conducted extensive sensitivity checks using these and other alternative indicator. An overview of these checks is given in section 7, while a detailed description is given in the appendix. Moreover, the appendix also includes results using gross income.

5.3 Other variables

Both the parametric approach and the nonparametric approach allow for the inclusion of covariates, controlling for potential heterogeneity between retirees and non-retirees. For instance, there are obvious differences with respect to age. While there is considerable overlap in this variable between the two groups, the retirees in our samples were, on average, older than the non-retirees (see the appendix for descriptive results).

In addition to age, we use the following control variables: education; gender; whether the individual was divorced, as this might indicate alimony payments; whether the individual reported owning the home she was living in; and whether the county where the individual was living was in a rural area, a metropolitan area, or an area with a medium population density. In the EVS data, we also include whether the individual was living in eastern or western Germany, and the quarter in which the data were collected. In the HRS data, we have added race/ethnicity with a dummy variable indicating whether an individual was white or non-white. While the non-white category is rather heterogeneous, we cannot break it down further because of our sample size.

A descriptive overview of all of the variables we use can be found in the appendix.

6 Main results

Our main results for the U.S. and Germany are displayed in Table 1, which 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, together with the standard errors of the lower bound and the upper bound of this interval. All of the underlying results (e.g., regression coefficients) are available upon request. For the tests of shape invariance or income independence, values below 0.05 indicate that shape invariance or income independence can be rejected. For the nonparametric approach, the 95% confidence interval of the covariance is shown. If it includes zero, then income independence cannot be rejected; if, on the other hand, zero is not in the identification set, then income independence can be rejected.

The parametric point estimate for Germany is 98%. This means that the net retirement income needs to be at roughly the same level as the net pre-retirement income to avoid changes in the living standard. The semiparametric point estimate

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 (for the nonparametric approach the confidence interval of the covariance), and the number of observations. Source: Own calculations based on HRS and EVS.

USA	NRR	SE	95% CI	p(ESE)	n
Parametric	0.949	0.073	[0.805,1.093]	0.085	798
Semiparametric	1.052	0.118	[0.650,1.160]	0.371	878
Nonparametric	[0.867,1.156]	0.009;0.011	[0.848,1.176]	[-8305, 545]	798
Germany					
Parametric	0.976	0.035	[0.907,1.045]	0.001	2310
Semiparametric	0.996	0.066	[0.850,1.150]	0.371	2310
Nonparametric	[0.887,1.090]	0.004;0.005	[0.878,1.100]	[-1254,18]	2310

is only slightly higher. The parametric estimate for the U.S. is about 95%, and is thus relatively similar to the parametric estimate for Germany; whereas the semiparametric estimate for the U.S. is 10 percentage points higher. However, the confidence intervals overlap. The standard errors are generally higher for the U.S. data than for the German data because of the smaller sample size. In the U.S. data, there are several observations with missing values for the control variables. This explains why the number of observations is lower in the parametric approach (and in the nonparametric approach) than in the semiparametric approach, which does not include these control variables. In the German data, all of the observations that fulfill the inclusion criteria are complete. Thus, the sample size is the same for all three approaches.

The identification intervals of the nonparametric approach are rather wide: the difference between the upper and the lower bound is 33 percentage points for the U.S. and is 22 percentage points for Germany. The lower bounds of the identification intervals are informative in the sense that they allow us to rule out some of the heuristic values found in the literature, like, for instance, the net replacement rate of 70% that is 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 results of the semiparametric approach applied to German data, which are slightly above these bounds. For both countries, we see a clear pattern in which the bounds of the nonparametric approach have the smallest standard errors the semiparametric approach has the largest standard

errors. The latter finding might be attributable to the fact that the semiparametric approach does not include control variables.

The results regarding income independence are mixed. For both the U.S. and Germany, the semiparametric test does not reject shape invariance; and the nonparametric test does not reject income independence. The parametric approach, on the other hand, indicates a rejection of income independence in the case of Germany. For the U.S., income independence can be rejected at the 10% level. Overall, these results overall provide weak evidence for income independence, as the semiparametric test and the nonparametric test rely on assumptions that are not as strong as those of the parametric test, and should therefore be more reliable. Still, heterogeneity in adequate replacement rates might require further study.

In summary, the point estimates for both countries suggest that the adequate replacement rate is around 100%. While the wide identification intervals found when applying the nonparametric approach indicate some degree of uncertainty, they are mostly consistent with the point estimates. The tests of income independence produced mixed results, but overall provided evidence in favor of income independence.

7 Further findings and robustness checks

7.1 Adequate replacement rates by age

Our main analysis is restricted to individuals aged 60 to 69. This restriction was motivated by the overlap condition, which requires that individuals of all studied ages are in both the non-retiree group and the retiree group. If we had used, say, ages 60 to 89, then there would have been only a few non-retirees above a certain age.

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 on goods like 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 analysis. This procedure violates the overlap condition with respect to age, and the results are essentially

Table 2: Results comparing non-retired individuals aged 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]	0.001	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	0.920	0.037	[0.849,0.993]	0.001	3302
Semiparametric	—	—	—	—	—
Nonparametric	[0.951,1.221]	0.004;0.005	[0.942,1.160]	[-1188,-16]	3302

based on extrapolation. While the findings might give some indication of whether age plays a role in adequate replacement rates, they should be viewed with care.

Estimates are shown in Table 2. As the semiparametric approach did not converge using the German data, no results are shown for it (see the appendix for details).

For both the U.S. and Germany, the identification bounds are shifted upwards compared to the main results. However, the identification regions of the main results and the results for the older age group overlap, and it cannot be ruled out that the true value is identical. For the U.S., the point estimates are also close to the main findings. For Germany, the parametric point estimate is outside of the identification region, possibly because the income independence assumption is violated (see below).

With respect to income independence, the results are mixed for the U.S., as the parametric test rejects income independence, while the semiparametric test and the nonparametric test do not. The results for Germany are clearer, as both the parametric and the nonparametric approach reject income independence, which might explain why the parametric point estimate is outside of the identification region of the nonparametric approach. These findings imply that for Germany the replacement rates might become income-dependent with increasing age, even though they do not directly depend on age. For the U.S. there is only very weak evidence of age-dependent replacement rates. Further research might be required to gain a better understanding of how income needs change with age.

7.2 Adding couples

As our main findings are restricted to single-person households, couples are missing. However, large shares of the population around retirement age are married or cohabiting. For instance, in the EVS data the majority of individuals aged 60 to 69 were living with a partner. Other household constellations (e.g., parents and children) occur, but are not very common in the age range we are studying.

There are two reasons for this restriction. First, including couples within the framework described in section 3 would require us to assume that all functions are related to households; e.g., household utility functions. When using household utility functions, strong assumptions are needed to be meaningful and indicative of the level of welfare of the individuals within the household. Using utility functions for individuals instead requires us to look at the decision-making processes and resource-sharing within households (Chiappori, 2016). Unfortunately, the data we are using either do not allow us to assess the welfare of individuals within households (HRS), or do so to a very limited extent only (EVS). Second, even for single individuals it is not always clear whether they can be counted as retired. In case of couples this is more complicated, as one household member might have retired while the other has not; or one household member might have retired while the other one has started working again after a period of retirement.

Nevertheless, one-person households are only a subset of the population in the age range we are studying, and they might be a selected group in one way or another. Couples tend to have different consumption patterns than single people because of economies of scale and economies of scope. Moreover, these two groups might also have different preferences, such as different bequest motives. It is therefore possible that our main findings are not representative of the total population around the retirement age.

Here, we present an additional analysis that also includes households of couples as well as one-person households. For households of non-retirees, we only include couples in which both partners were working, and one of the partners was in the 60-69 age range. For retirees, we include all couples in which both partners were retired and the older one was in the 60-69 age range. This group also includes couples who had been in single-earner households, and who might not perfectly match the dual-earner couples in the control group. But for many of the households in the HRS and for all of the households in the EVS, we are not able to control for the previous labor force status of both partners, as this information is not available.

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	SE	95% CI	p(ESE)	n
Parametric	0.860	0.040	[0.781,0.938]	0.001	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

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

The order of magnitude of the estimates is roughly comparable to our main results shown in Table 1. For both the U.S. and Germany, the results for the parametric approach are a few percentage points below the main estimates from Table 1, while for the semiparametric approach, the estimates are somewhat higher. The bounds of the identification intervals are rather similar, but some of the point estimates are outside of the bounds. The results of the tests of income independence are consistent across countries and are comparable to the main results; i.e., the semiparametric tests show evidence of shape invariance, and the nonparametric test does not reject income independence, while the parametric test does. As for the main results, we interpret this finding as being slightly in favor of income independence.

7.3 Welfare indicator, income concept, and retirement status

We have 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 (e.g., net income with and without wealth), five different welfare indicators (e.g., expenditures on durable goods or satisfaction with household income), and four definitions of the retirement status. For a complete overview of the variables, see the appendix. We have run all possible

combinations of these variables with both the parametric and the semiparametric approach. As the semiparametric approach shows convergence issues in some cases, we have 147 estimates in total. For Germany, we also use four different income concepts, but only four welfare indicators are available, and two definitions of the retirement status. In combination with the parametric and semiparametric approach, this yields 58 models when six semiparametric estimates are dropped due to convergence issues.

Detailed results are presented in the appendix, and here we provide an overview.

For the U.S., roughly 81% of all point estimates arising from the sensitivity check fall within the identification bounds of our main results shown in Table 1. For Germany, this figure is 72%. The analysis leading to estimates that deviate from our main findings is mostly based on quite different welfare indicators: on satisfaction with household income, which leads to lower results than the expenditure-based estimates; and on nondurable expenditures, that in some cases lead to estimates that are higher than those based on food expenditures.

For equivalence scales, satisfaction-based estimates tend to be lower than expenditure-based scales. This is likely because satisfaction measures are influenced by factors other than consumption, such as a comparison of one's own situation with that of others (Ferrer-i-Carbonell, 2005). Of the nondurable expenditures, expenses for transportation tend to be especially high and strongly dependent on the need to commute. Thus, there is a big drop in these expenses at retirement that is not reflective of a decline in welfare. It therefore appears that using this welfare indicator would overstate the replacement rate needed to maintain a constant standard of living.

7.4 Instrumenting retirement status

For our analysis, it is crucial that the effect of retirement on the welfare indicator is estimated correctly. This means that the retirement decision must be exogenous. This might not be the case, as individuals might select into retirement based on their expectations of their welfare level in retirement, and their incentives to retire or to continue working (Samwick, 1998). For instance, in Germany, retiring before the nominal retirement age, such as at age 63, leads to reduced pension benefits. Individuals who are well off and have substantial savings might not be concerned about this reduction, whereas individuals who have no savings and expect to receive low pension benefits might decide against retiring at a younger age.

A potential solution to this problem is to exploit the exogenous variability in pension eligibility with a (fuzzy) regression discontinuity approach (Imbens

and Lemieux, 2008). For instance, in the U.S., the lowest age at which Social Security benefits can be claimed is age 62, and the normal retirement age (NRA) is currently age 66. Both of these thresholds have been set by policy-makers, and are 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 applied to U.S. and German data (Eibich, 2015; Kämpfen and Maurer, 2016). In the life-cycle literature, this approach 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 is then used as an instrument for retirement status (Eibich, 2015). We implement this approach 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 of 62 and the NRA, which for some of the cohorts in our data is below age 66. For Germany, we also use observations at ages 55 to 75 to estimate the probability of retirement, as well as discontinuities at ages 60 and 65. While the retirement age in Germany is increasing cohort by cohort, we are not able to exploit the variation introduced through this trend, as in order to do so we would need information on the exact day of birth. Unfortunately, our data provide information on the year of birth only. The German system allows individuals to retire before reaching age 65 without penalties provided they have a minimum number of contribution years, but the availability of this option 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 that 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. For this reason, we can present results for the parametric approach only. 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 immediately before and after retirement. Thus, these estimates might be less representative of the pension needs of the retiree population than our main findings.

The results of the second stage of estimation are shown in Table 4. For the U.S., the point estimate is considerably lower than our main parametric result (53% vs 95%). However, this estimate is very imprecise, and it is not statistically significantly different given the relatively small sample size. For Germany, the point estimate of the regression discontinuity approach is very close to the main

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

	NRR	SE	p(ESE)	n
USA	0.532	0.449	0.082	787
Germany	0.982	0.056	0.001	1608

parametric estimate. For both the U.S. and Germany, the results of the tests of income independence are consistent with our main results. Thus, the overall endogeneity does not seem to pose a threat to our results. The estimation issues for the U.S. that lead to the large standard error of the estimate in Table 4 are discussed in the appendix.

8 Summary and concluding remarks

In this paper, we have described a framework for assessing the retirement income individuals need to maintain the standard of living they achieved while working, and we have proposed using this framework as a measure of pension adequacy. We estimated net replacement rates, which are defined as the net retirement income levels individuals need to maintain their living standard relative to their net pre-retirement income. Applying parametric, semiparametric, and nonparametric estimation approaches to U.S. data and German data, we found adequate net replacement rates of about 100%. Thus, our results suggest that to avoid a decline in welfare in retirement, people need to have roughly the same income levels during retirement that they had while working.

We found weak evidence to support the claim that the adequate replacement rate is flat across the whole income range; i.e., that the level of the replacement rate does not depend on pre-retirement income. This finding is based on several tests of income independence, most of which did not reject it. But as some exceptions were found, additional evidence is needed to arrive at a more definite conclusion. We also conducted extensive sensitivity checks, which showed that our main findings are robust to endogeneity, and with respect to the indicators of welfare, the definition of income, and the definition of retirement status used.

The methods presented in this paper can be easily adapted for use with other welfare measures, data sets, and countries; as we supply the R code that implements our methods (<https://osf.io/hde4y/>). Thus, our paper can serve as a blueprint for the empirical estimation of pension standards. As our findings show, the

parametric approach yields reasonable estimates, is simple to understand, and has low data demands. However, the results for income independence were not clear. This means that the more complex approaches are needed to check the results of the simpler methods. Thus, ideally, the whole range of methods we presented here should be applied.

In the literature, replacement rate values of between roughly 60% and 100% have been deemed adequate, or at least as within a reasonable range. Our results clearly point to the upper part of this range, and are thus comparatively high. Our estimates are compatible with the results presented by Binswanger and Schunk (2012), which indicated what replacement rates individuals consider adequate and would prefer. However, our finding that the adequate replacement rate is high leads us to ask whether the retirement income levels people actually achieve are high enough to reach it.

For the U.S., the pension incomes and savings recent retirees have accumulated are, on average, roughly equal to the level we find in our analysis (Love et al., 2008). Still, as sizable shares of the older U.S. population have no or relatively little retirement savings (Lusardi and Mitchell, 2007), a large group of retirees will likely have pension incomes that are considerably lower than our standard. For Germany, our findings point to a 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 to around 55% by 2060 (OECD, 2015). Comparing these recent and projected numbers to our main findings yields a gap in pension of around 25% or more, a figure that is expected rise as high as 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 benefits 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 is likely is only possible with certain trade-offs, such as a delay in retirement (Kitao, 2014). For individuals, this might mean that they will have to step up their saving efforts. Especially in the German context, where private retirement savings are currently quite low (Bucher-Koehnen and Lusardi, 2011), it seems unlikely that individuals will reach this level on their own. Providing individuals with better information about the pension benefits they will receive, the pension benefits they will need, and how they can reach their desired level of pension benefits might help them

achieve greater financial security in retirement (Bernheim and Garret, 2003; Dolls et al., 2018).

One important caveat of our findings is that the replacement rates we have reported are averages, and we mostly ignore heterogeneity in replacement rates, except in the tests for income independence and the results by age. Income and age might not be the only potential source of heterogeneity. For instance, replacement rates might depend on health and health-related expenditures, as medical expenses have been shown to account for large shares of household consumption in the U.S. (Banks et al., 2018; Finkelstein et al., 2013). While we provided results by age, further disentangling the heterogeneity in replacement rates is possible with the methods presented in this paper, and is a potential avenue for future study.

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.
- 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., Levell, P., and Smith, J. P. (2018). Life-cycle consumption patterns at older ages in the US and the UK: Can medical expenditures explain the difference? *American Economic Journal: Economic Policy*, forthcoming.
- 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.

- Bellemare, C., Melenberg, B., and van Soest, A. (2002). Semi-parametric models for satisfaction with income. *Portuguese Economic Journal*, 1:181–203.
- 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.
- Börsch-Supan, A. (2000). Incentive effects of social security on labor force participation: evidence in germany and across europe. *Journal of Public Economics*, 78(1):25 – 49.
- 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). Inequalities for $ek(x, y)$ when the marginals are fixed. *Probability Theory and Related Fields*, 36:285–294.
- 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.
- Finkelstein, A., Luttmer, E. F. P., and Notowidigdo, M. J. (2013). What good is wealth without health? The effect of health on the marginal utility of consumption. *Journal of the European Economic Association*, 11:221–258.
- 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.
- Kämpfen, F. and Maurer, J. (2016). Time to burn (calories)? The impact of retirement on physical activity among mature americans. *Journal of Health Economics*, 45:91 – 102.
- Kitao, S. (2014). Sustainable social security: Four options. *Review of Economic Dynamics*, 17:756–779.
- 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.
- Lee, D. S. and Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, 142(2):655 – 674. The regression discontinuity design: Theory and applications.
- 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.
- 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.
- Munnell, 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.
- 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.

- 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.
- Samwick, A. A. (1998). New evidence on pensions, social security, and the timing of retirement. *Journal of Public Economics*, 70:207–236.
- 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.
- Smith, S. (2006). The retirement-consumption puzzle and involuntary early retirement: Evidence from the British Household Panel Survey. *Economic Journal*, 116:C130–C148.
- Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.
- Stengos, T., Sun, Y., and Wang, D. (2006). Estimates of semiparametric equivalence scales. *Journal of Applied Econometrics*, 21:629–639.
- 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.

A Descriptive statistics

Table 5: The Health and Retirement Study 2014, summary sample statistics by retirement status. Source: Own calculations based on the HRS.

	Pre-retirees	Retirees
<i>Retirement status defined by:</i>		
1) Labor force status from RAND, n	423	710
2) as 1) but obs. working >20hrs are treated as pre-retired, n	449	684
3) Whether or not receive any pension income, n	828	291
4) Self-reported retirement status, n	370	734
<i>...applying definition 2)</i>		
<i>Welfare indicators</i>		
Food share (at home; in %), mean	14.3	17.4
Food share (at home and away; in %), mean	14.0	16.6
Monthly expenditures for nondurables (in 1000 dollars), mean	17.7	11.6
Income satisfaction (1 to 5), mean	3.1	3.2
<i>Income concepts</i>		
Monthly net income (in 1000 dollars), mean	2.5	2.0
Monthly net income (in 1000 dollars), min	0.1	0.1
Monthly net income (in 1000 dollars), max	6.8	7.2
w/o annuities w/o housing (in 1000 dollars), mean	2.3	1.7
w/o housing (in 1000 dollars), mean	2.4	1.9
w/t imputed rent (in 1000 dollars), mean	2.8	2.2
<i>Covariates</i>		
Age (years), mean	63.2	64.6
Share of males (%)	34.5	34.6
Share of homeowner (%)	54.8	49.4
Share of highly educated (%)	32.1	27.9
Share of non-whites (%)	41.2	43.1

Table 6: The German Income and Expenditure Survey 2013, summary sample statistics by retirement status. Source: Own calculations based on the EVS.

	Pre-retirees	Retirees
<i>Retirement status defined by:</i>		
1) Labor force status, n	807	1510
2) as 1) but obs. working >20hrs are treated as pre-retired, n	816	1501
<i>... applying definition 2)</i>		
<i>Welfare indicators</i>		
Food share (at home; in %), mean	12.2	14.6
Food share (at home and away; in %), mean	15.8	18.2
Monthly exp. for nondurables narrow (in 1000 euros), mean	0.5	0.5
Monthly exp. for nondurables wide (in 1000 euros), mean	0.7	0.6
<i>Income concepts</i>		
Monthly net income (in 1000 euros), mean	2.4	1.8
Monthly net income (in 1000 euros), min	0.2	0.3
Monthly net income (in 1000 euros), max	11.5	10.8
w/o annuities w/o housing (in 1000 euros), mean	2.3	1.6
w/o housing (in 1000 euros), mean	2.3	1.7
w/t imputed rent (in 1000 euros), mean	2.6	2.0
<i>Covariates</i>		
Age, mean	61.9	65.6
Share of males (%)	31.5	29.5
Share of former GDR (%)	21.8	25.9
Share of homeowner (%)	45.5	43.4
Share of highly educated (%)	33.9	24.5

B Replacement rates using gross income

As was discussed in the main text, the replacement rates based on gross income may differ from those based on net income. To determine whether this is the case, the results for replacement rates based on gross income (gross replacement rates; GRR) are shown in Table 7. The net replacement rates (NRRs) are shown for comparison. Gross income is defined here as monthly income from work and pensions before taxes, plus annuitized wealth and housing wealth, as described in the main text for net income. The analysis is based on the samples of single-person households in the age range 60-69.

Table 7: Results comparing replacement rate estimates based on net income (NRR) and gross income (GRR), for the U.S. (HRS data) and Germany (EVS data). Source: Own calculations based on HRS and EVS.

USA	NRR	GRR
Parametric	0.949	0.924
Semiparametric	1.052	1.014
Nonparametric	[0.867,1.156]	[0.837;1.169]
Germany		
Parametric	0.968	0.980
Semiparametric	1.076	1.001
Nonparametric	[0.863,1.061]	[0.827;1.037]

The differences between GRR and NRR are, overall, not large; and the GRR estimates are rather close to our main findings. The results on income independence also do not differ (details available upon request from the authors). A potential explanation for these findings is that both gross income and net income account for annuitized wealth and housing wealth in a similar way. Moreover, given the same gross income level, the differences in the taxation of retirees and non-retirees are not very large, at least for our sample.

C Sensitivity analysis for the main variables: Income, welfare indicator, and retirement status

C.1 Variables: Welfare indicators

For the main results presented in the text, we use the share of total expenditures allocated to food at home as the welfare indicator. The share of expenditures on food at home is known to be only an approximate measure of household consumption (Bernheim et al., 2001). In addition to not capturing food expenditures away from home at locations like restaurants and workplaces, this indicator might mismeasure consumption, as expenditures might be substituted with time for home production, while consumption stays constant (Aguiar and Hurst, 2005; Luengo-Prado and Sevilla, 2013).

For this reason, we use several other indicators to assess the robustness of our results, including the income share of total expenditures devoted to food at home and away from home; the absolute level of expenditures on nondurable goods defined in several ways; and satisfaction with household income. It has been argued that expenditures on nondurable goods 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 that is 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 in a similar manner as the other welfare indicators; and thus as a quantitative variable.

In the EVS data, no subjective measures are available. While the HRS collects information on expenditures on food away from home, because it excludes expenditures on food at work or at school, this variable may be incomplete, and is 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. As these data are only available for a subset of the HRS sample, the number of observations is small. 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. As this leave-behind questionnaire was not answered by all respondents, the number of observations is again small.

Keeping these restrictions in mind, we use five different welfare indicators for the HRS data: food at home, as in the main analysis; food at home and outside of home, excluding work; food at home and dining (CAMS); nondurable expenditures (CAMS); and satisfaction with income (leave-behind questionnaire). For the EVS data, we use four indicators: food at home, as in the main analysis; food at home and outside of home; nondurable expenditures without transportation; and nondurable expenditures including transportation.

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

C.2 Variables: Income concepts

For our main analysis, we used the net income plus annuitized wealth and annuitized housing wealth. While this income concept can be easily implemented for both the German and the U.S. data, other income concepts could be used instead (for a discussion, see Munnell 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 used, 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 data, we generate imputed rent assuming a 5% rental yield (Munnell and Soto, 2005; Crawford and O’Dea, 2012).

C.3 Variables: Retirement status

For the U.S., we use in our main analysis the labor force status, as defined by RAND, and the number of hours worked per week to define whether an individual was retired. In the sensitivity checks, we use three different approaches: using only the RAND indicator, without dropping individuals who were working more than 20 hours per week from the group of retirees; defining the retirement status based on receipt of any pension income, including from pension plans and annuities; and using the self-reported retirement status, i.e., whether the respondent reported being retired.

For Germany, we use in the main analysis the labor force status, as defined by the Federal Statistical Office, and the number of hours worked per week to define whether an individual was retired. In the sensitivity analysis, we drop the criterion for the hours worked per week.

C.4 Results and discussion

The results of the robustness checks are summarized in Figure 1. The results show kernel densities based on 147 net replacement rate estimates for the U.S. and 58 estimates for Germany, all of which are based on either the parametric or the semiparametric approach. Each of these models is based on a different combination of welfare indicators, income concepts, and retirement statuses. In some of these models, the semiparametric approach exhibits convergence issues (14 for the U.S., mostly based on CAMS data with small sample size; six for Germany). The point estimates of the parametric and the semiparametric approach in our main results are shown as solid vertical lines, and the identification bounds are shown as dashed vertical lines.

For the U.S., 81% of all estimates fall within the identification bounds of our main results. The exceptions to this general pattern form two groups. In the first group are estimates based on nondurable expenditures, which are sometimes below and sometimes above the identification bounds. In the second group are estimates based on satisfaction with household income, which in some cases are lower than the expenditure-based estimates. The latter finding might be attributable to the fact that satisfaction measures may be influenced by factors other than consumption, such as a comparison of one's own situation with that of others (Ferrer-i-Carbonell, 2005). Somewhat surprising is our finding that in these robustness checks the parametric and the semiparametric estimates based on satisfaction can differ, as the contrary outcome has been reported for equivalence scales (Bellemare et al., 2002). The finding that nondurable expenditures are lower than they were in our main findings is difficult to explain, whereas the finding that nondurable expenditures are higher than they were in our main findings might be explained as follows. Nondurable expenditures include expenses that can be rather high, and that are strongly dependent on the need to commute. Thus, these expenses may fall sharply in retirement without being accompanied by a decline in the welfare level. Using this welfare indicator likely overstates the replacement rate needed to maintain a constant standard of living, which explains why these estimates are higher than our main results.

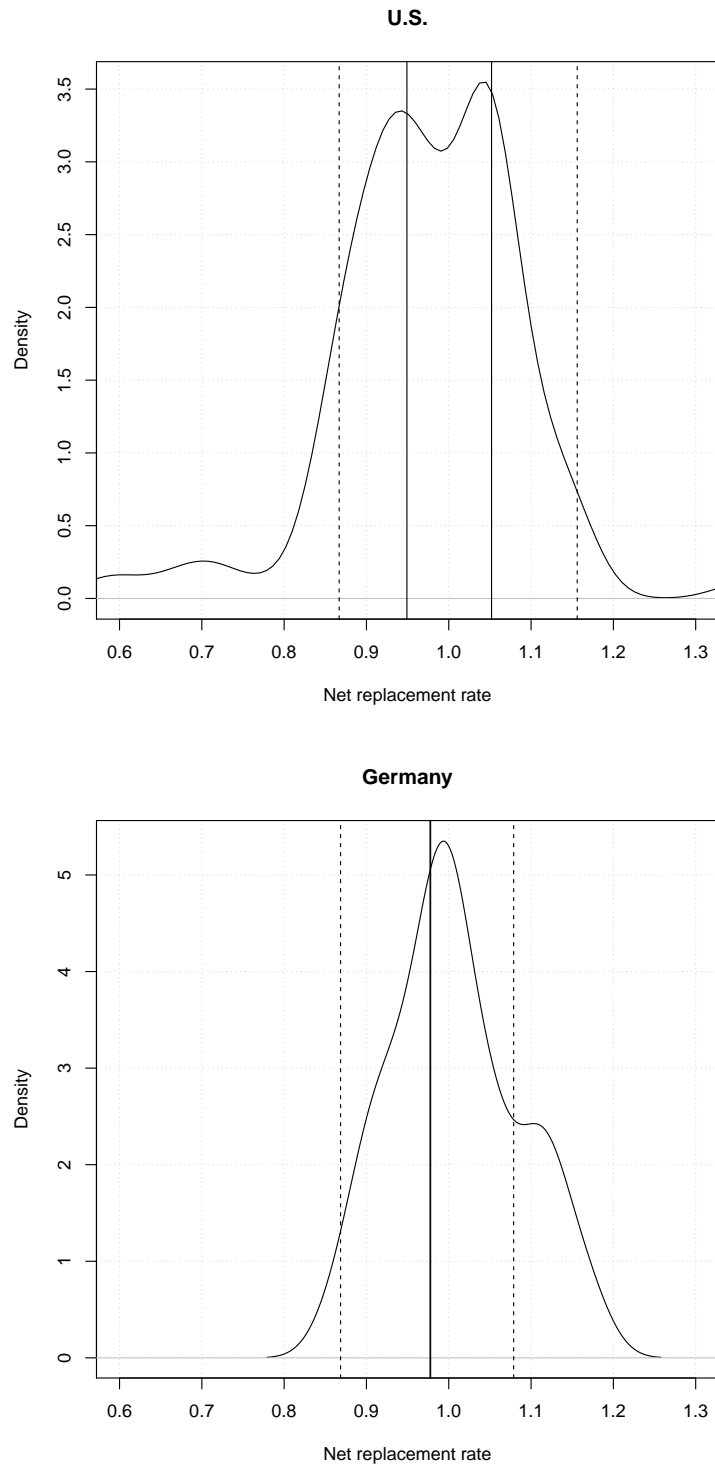


Figure 1: Kernel density estimates of the net replacement rates resulting from different welfare indicators, income concepts, and definitions of the retirement status. For the US results are based on 157 estimates of net replacement rates; for Germany on 58 estimates. Source: Own calculations using HRS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

The results of the robustness checks for Germany differ from those for the U.S. in several ways: i.e., the overall range of estimates is lower; and while there are more estimates above the identification bounds, there are no estimates below the bounds. These differences in the findings can be explained by the different welfare indicators used: no satisfaction-based measure is available in the EVS data, but different definitions of nondurables can be applied, while still maintaining a large sample size (unlike in the HRS and the CAMS data). High estimates of the replacement rate, of around 110% or higher, can result when a rather broad definition of nondurables is used. As described above, the use of such a definition might lead to the overestimation of the replacement rate needed to maintain a constant standard of living.

The sensitivity checks also included several different retirement status definitions and income concepts. The variation in the results for these variables is less systematic than it was for the welfare indicators. Given the same welfare indicator and using the same econometric approach, the effects on the results of the income concept and the retirement status used on the results are small for Germany. For the U.S., where we explore more definitions of the retirement status, the variability is larger. Still, except for some outliers that are found when retirement is defined only through the receipt of pension benefits, the overall results are largely consistent.

D Additional information on the RDD

D.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 $E(d_i|x_i)$. These predicted values are then used to instrument retirement status in equation (4.3). For the U.S. we use age 62 instead of age 60, and age 65 is replaced with the NRA that applies to each individual.

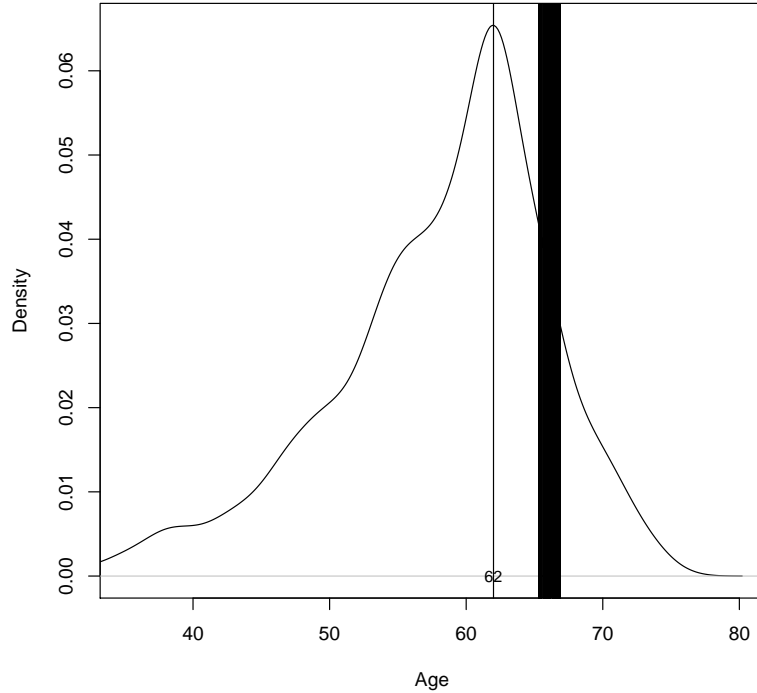


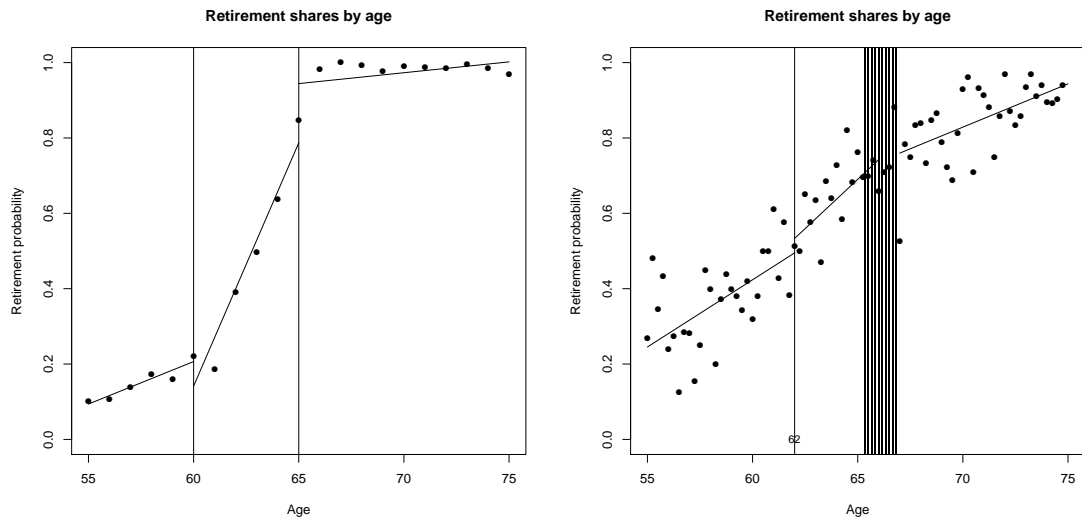
Figure 2: Distribution of self-reported age of retirement taking into account cohorts born in 1939-1959, HRS. Note: the vertical lines represent eligibility ages at age 62 and the normal retirement age (NRA) depending on the year of birth.

The German data we use include only the year of birth, and not the exact age of each respondent. It is therefore not possible to determine exactly how close the respondents were to retirement age. Still, as the German system provides strong financial incentives to retire at certain thresholds, the cutoffs at ages 60 and 65 are relatively clear (Eibich, 2015). Thus, the instrument is rather strong (Börsch-Supan, 2000). The HRS data, by contrast, provide information on the exact date of birth.

D.2 Diagnostics: Discontinuities

Following Imbens and Lemieux (2008), we test the validity of the RDD mostly graphically. As noted above, the ages of the respondents in Germany can only be calculated based on their years of birth. This means that the running variable has to be treated as discrete, and computing averages very close to the cutoff point is, therefore, not possible (Lee and Card, 2008).

The first question we want to assess graphically is whether there is a clear cutoff in the probability of retiring in the running variable; i.e., age. The left



(a) GERMANY in year bins

(b) The U.S. in quarter bins

Figure 3: Age specific retirement probabilities and retirement eligibility

panel of Figure 3 demonstrates that in the German EVS sample, the share of retiring individuals rises sharply at age 60 and at age 65, which indicates that these retirement eligibility ages are important thresholds. Although the system offers incentives to retire early at the age of 63 (Börsch-Supan, 2000), the graph suggests that this threshold is less relevant here.

For the U.S., a small discontinuity is recorded at age 62, which is the youngest age at which Americans become eligible for retirement benefits (Kämpfen and Maurer, 2016). This finding is confirmed by Figure 2, which shows that most Americans retire at this age. The NRA, which differs depending on the year of birth, is less clear-cut. Whether that age can serve as a relevant instrument must therefore be questioned. This pattern likely explains the high standard error we find when applying the RDD approach to the HRS data, as reported in the main text.

To test the relevance of the instruments, we calculate the F-statistic for excluding the retirement eligibility ages when estimating the probability of retiring (See C.1). For both age 62 and the NRA combined, we find that the F-value is 5.3, and is thus below 10, a benchmark often used in this literature (Staiger and Stock, 1997). For comparison, the F-statistic for the joint instruments of age 60 and age 65 in the EVS sample amounts to 253.

D.3 Diagnostics: Further checks

We checked several other assumptions underlying the RDD approach. We briefly summarize the findings here. Overall, these findings confirm that the RDD approach should work well for the German data, but less well for the U.S. data. Detailed results are available upon request.

First, for the RDD to be valid, respondents are not allowed to have influence on the running variable. In our case, age is the running variable. Other than by misreporting their birthday or year of birth, the respondents cannot influence this variable. While age might be misreported in the data we use, it is likely that misreporting is not a big issue. Second, the outcome variable (welfare indicator) must be a smooth function of the assignment variable. A visual assessment shows that this is the case for Germany, while the results are less conclusive for the U.S. Third, we assess the sensitivity of the RDD results with respect to the choice of bandwidth to the polynomial degree, and to the choice of cutoff points. Again, the results show no issues for Germany. But since the results for the U.S. are found to be rather sensitive to these choices, we again conclude that the RDD approach may not be valid for the subset of HRS data we use.

E 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 for mixing of continuous and categorical explanatory variables (Racine and Li, 2004). Generalized product kernels are used to include control variables in the nonparametric approach. 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 kernel function depends on the type of variable, where $K^{(c)}$ indicates the kernel function for continuous variables, and $K^{(d)}$ indicates the kernel function for categorical

variables. For continuous variables, we use a second-order Gaussian kernel; and for the categorical case, we use the kernel function proposed by Aitchison and Aitken (Hayfield and Racine, 2008). For bandwidth selection, see Hall et al. (2004).

For statistical inference, we use for the semiparametric estimation approach a residual bootstrap, as proposed by Pendakur (1999). For the nonparametric estimation approach, we use the resampling bootstrap. This procedure could lead to conservative inference (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, the standard errors and the confidence intervals are based on 1000 bootstrap replications.

Confidence intervals are based on percentiles of the bootstrap replications. For the nonparametric identification region, we have constructed an interval that covers the complete identification region with a fixed probability (95%); roughly similar to the approach used by Horowitz and Manski (2000). If (L_s, U_s) denote the bounds resulting for the s th 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 $e_L \sim \mathcal{N}(0, \text{sd}(L))$ or $e_U \sim \mathcal{N}(0, \text{sd}(U))$, 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 step size $\lambda_L = 0.1\text{sd}(L)$ or $\lambda_U = 0.1\text{sd}(U)$ 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.

To find the values of α and μ that minimize equation (4.5) for the semiparametric approach, we insert values between 0.4 and 1.2 for α , starting from 0.4 and using increments of 0.001. Conditional on α , μ is identified and can be easily calculated, allowing us in turn to calculate (4.5). α is then chosen as the value in the interval

(0.4, 1.2) which yields the lowest value of equation (4.5). If the lowest value of α is either 0.4 or 1.2, this indicates that (4.5) is likely not a convex function over $[0.4, 1.2]$, and we therefore say that the semiparametric approach did not converge.

The semiparametric approach makes use of simulation-based inference to assess income independence, following 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 values of L 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 on income independence. The distribution of simulated values L^* is used to assess the probability of the empirical value of L .