Retirement, pension eligibility and home production*

Emanuele Ciani†[a,b]

[a]Bank of Italy, Regional Economic Research Division, Florence Branch, Italy

[b]Center for the Analysis of Public Policies, University of Modena and Reggio Emilia

April 29, 2014

Abstract

I study the change in home production at retirement. Descriptive evidence from the 2007 Italian Survey on Income and Living Conditions shows that retirees spend much more time than workers on household chores, shopping and caring, even when the comparison is made for individuals of a given age. To account for the endogeneity of retirement, I exploit the discontinuity in pension eligibility generated by the Italian Social Security system. Estimates show that women increase time spent on household production at retirement by more than 400 minutes per week. No evidence of an equally large change is found for men.

JEL: J22, J26, D1

Keywords: retirement, house work, regression discontinuity

*I wish to thank Marco Francesconi, Elena Stancanelli, Marcello Morciano, Carlo Mazzaferro, Patrick Nolen, Mark Bryan, Roberto Nisticò, Jonathan James, Claudio Deiana, Ludovica Giua, Guglielmo Weber, Matthias Parey, Maurizio Lisciandra, Tindara Addabbo, Paolo Sestito and participants at seminars at Essex, the SIE 2013, EALE 2013, and RES 2014 conferences for helpful suggestions. This paper was part of my PhD dissertation at the University of Essex, during which I received funding from the Economic and Social Research Council and from the Royal Economic Society Junior Fellowship, that are gratefully acknowledged. The views expressed in this paper are those of the author and do not necessarily reflect those of the Bank of Italy. SILC microdata can be requested from ISTAT (Italian National Statistical Office), while programs calculating the estimates from raw data are available from the author. All estimates and figures are obtained using Stata™12, plus programs `ivreg2` (Baum et al., 2007) and `esttab` (Jann, 2007).

†Contacts: emanuele.ciani@bancaditalia.it

1 Introduction

The evidence of a drop in consumption at retirement spurred a large stream of research which tried to reconcile it with the permanent income hypothesis. In his summary of the literature, Hurst (2008) argues that this reduction can be explained by unexpected retirement due to deteriorating health, by a reduction in work-related expenses, and by an increase in home production. In this paper I focus on the latter and I provide new evidence about the change in time spent on producing household goods and services at retirement, using data from the 2007 Italian Survey on Income and Living Conditions (SILC).[1]

As argued by Rogerson and Wallenius (2012), the comparison between employed and retired individuals at any given age can provide a biased estimate for the quantity of interest, because retirees may have different preferences for leisure and house work. To manage this problem, I use the fuzzy Regression Discontinuity Design (RDD) outlined in Battistin et al. (2009), which exploits the discontinuities in pension eligibility induced by the Italian Social Security system. While they employed it to estimate the drop in consumption at retirement, I focus on time spent on house work, for which no information was available in their dataset.

Previous empirical research providing evidence of an increase in home production at retirement can be found in Aguiar and Hurst (2005), Szinovacz and Harpster (1994), Szinovacz (2000) and Hurd and Rohwedder (2005, 2006) for the U.S.; Schwerdt (2005) and Luhrmann (2010) for Germany; and Luengo-Prado and Sevilla (2013) for Spain. To the best of my knowledge, only Stancanelli and van Soest (2012) used (fuzzy) RDD to address this question. They exploited the discontinuity in retirement at age 60 induced by the French system to estimate the causal effect of either partners' retirement on house work in couples. An advantage in studying the Italian setting is that eligibility depends on both age and years of contributions, generating discontinuities in retirement even when keeping one or the other fixed. Furthermore, the system has

[1] At the moment of writing, I am aware of only one economic related study using SILC data on home production. Addabbo et al. (2011) studied time allocation within working age couples, but they did not analyse retirement.

been the subject of several reforms in the last two decades, so that different rules apply to individuals who retired in different years. The nature of the home production information in SILC is also different from Stancanelli and van Soest (2012). While their data are collected from a single day diary, in SILC respondents are asked about time spent in house work during an average week.

Finally, the Italian case is interesting in itself, because comparative international evidence shows that gender differences are stronger than in other countries, with Italian men spending much less time on household production (Burda et al., 2006, p. 16-19). I present separate estimates for men and women, in order to understand whether retirement has an equalising effect.

The main results from my RDD estimates show that women increase home production by more than 400 minutes per week on average, while for men there is no evidence of an equally large increase. This gender difference has no parallel in studies from Germany, France, Spain and the US. Results can therefore justify a drop in consumption associated with women's retirement, while they do not seem to be sufficient to explain it for the case of men leaving their job at pensionable age.

Section 2 presents the identification strategy, while section 3 introduces the dataset. The main results are reported in section 4. The final section concludes.

2 Identification strategy

I follow the identification strategy outlined by Battistin et al. (2009), which exploits the discontinuity in retirement behaviour with respect to time to/from eligibility for a pension. As they noticed, if I define an individual as retired only when s/he does not work and s/he is recipient of a retirement pension, I should not observe anybody in this state before meeting the requirements. Restricting the sample to individuals who are currently employed or retired from work, I observe a sensible increase in the proportion of retired individuals between one year before eligibility and one year after. This motivates a RDD.

Define S_i as time to/from eligibility, $D_i \equiv \mathbf{1}[S_i \geq 0]$ as the dummy for being eligible, R_i as a dummy for being retired from work. Individuals are indexed by $i = 1, ..., N$. Let Y_{1i} be the time spent on home production if i was retired, while Y_{0i} if s/he was still employed. For each single individual, I actually observe only one or the other, so that the observed outcome is (Hahn et al., 2001)

$$Y_i = \delta_i R_i + \epsilon_i, \tag{1}$$

$$\epsilon_i \equiv Y_{0i},\ \delta_i \equiv Y_{1i} - Y_{0i}. \tag{2}$$

In order to exploit the RDD to identify the average of the causal effect δ_i, I need a discontinuity in retirement:

$$\text{(A1).}\ R_i = \gamma_D D_i + h_R(S_i) + \xi_i$$

$$\text{with } \gamma_D \neq 0;\ h_R(S_i = s) \text{ continuous at } s = 0;\ E(\xi_i|S_i) = 0.$$

Given that the majority of retirement benefits in Italy come from state-managed funds, the eligibility rules are expected to have a strong effect on retirement behaviour. This prior is corroborated by previous results from Battistin et al. (2009), who found a 43.5 percentage points increase in the proportion of retired household heads at $s = 0$.

In order to exploit this discontinuity, the potential time spent on house work without retirement must not change discontinuously at eligibility:

$$\text{(A2).}\ E[\epsilon_i|S_i = s] = h_Y(S_i = s),\ h_Y \text{ continuous at } s = 0$$

so that

$$Y_i = \delta_i R_i + h_Y(S_i) + \eta_i, \tag{3}$$

$$\eta_i \equiv \epsilon_i - h_Y(S_i) \tag{4}$$

However, there might be age-specific effects that force individuals to exit the labour market and spend more time on home production. For instance, their partners' health may deteriorate, demanding a considerable amount of caregiving. The probability of such an event is quite likely to be a function of age and seniority, but there is no particular reason to believe it to be discontinuous at the specific and rather arbitrary point of eligibility. Workers are also hardly able to manipulate S in order to become eligible, given that the National Social Security Institutions keeps track of each worker's contribution history. Furthermore, given that requirements have been subject to several reforms since 1992, individuals were not able to exactly predict the timing of their eligibility in advance.

Under assumptions (A1) and (A2), the average causal effect is equivalent to the ratio of the discontinuities in the reduced forms $E\left[Y_i|S_i = s\right]$ and $E\left[R_i|S_i = s\right]$ at $s = 0$, because any change in household production at eligibility can be attributed to retirement. However, identification is complicated by the fact that S is not directly observed. Instead, I recovered it using information on current age, age at first job, years spent in paid work, years of social contributions and job description. This introduces three additional problems.

First of all, in SILC I can measure time/to from eligibility only in discrete units (years). As argued by Lee and Card (2008), this forces us to choose a parametric specification, which can be used as an approximation of the correct model. Define a vector P_i containing a polynomial in S_i, possibly interacted with D_i. Assuming for the moment a constant treatment effect, the model can be rewritten as

$$Y_i = \alpha_0 + \delta R_i + P_i\alpha + \eta_i^* + \eta_i, \tag{5}$$

$$R_i = \gamma_0 + \gamma_D D_i + P_i\gamma + \xi_i^* + \xi_i, \tag{6}$$

where $\eta_i^* \equiv h_Y\left(S_i\right) - P_i\alpha$ and $\xi_i^* \equiv h_R\left(S_i\right) - P_i\gamma$ can be interpreted as the residuals from the Best Linear Projection ($BLP\left[\cdot\right]$) of the true functions h_R and h_Y on the vector P_i. Note that this implies that errors are clustered on S, so that standard inference may lead to wrong conclusions. Having kept the same polynomial in both eq. (5) and (6), the causal effect δ can be recovered using 2SLS and instrumenting R_i with D_i. For it to be consistent, apart from assumptions (A1) and (A2) it must be that

$$\text{(A3).}\ BLP\left[\eta_i^*|D_i, P_i\right] = BLP\left[\eta_i^*|P_i\right] \equiv 0.$$

This implies that the approximation does not introduce any discontinuity in the main equation of interest (5), so that D_i can be excluded from it. If $BLP\left[\xi_i^*|D_i, P_i\right] = BLP\left[\xi_i^*|P_i\right] \equiv 0$, then the discontinuity in the BLP of R_i on (D_i, P_i), call it γ_D^*, is also equal to the true jump in retirement (γ_D). However, the equation for retirement is only a first stage, and therefore we only need it to be the best linear projection.[2] With this caveat in mind, the main estimates will employ a simple 2SLS strategy, choosing the polynomial that provides the best fit in the reduced form for Y. One might prefer to look at the two reduced forms $E\left[Y_i|S_i\right]$ and $E\left[R_i|S_i\right]$ separately and then estimate δ as the ratio of the two discontinuities. In this parametric setting, however, using 2SLS has the advantage of being clearer, given that it is equivalent to an instrumental variable approach.

The second problem of identification is caused by the fact that S is discrete because it is rounded in years. Dong (2014) shows that the OLS estimator for the discontinuity in Y at eligibility is biased. Nevertheless, she showed that, under certain conditions, the bias can be recovered if the marginal distribution of the true continuous distance is known. In particular, one must assume that the moments of the rounding error are independent from S and that the true functional forms for $E\left[Y_i|S_i\right]$ and $E\left[R_i|S_i\right]$ are polynomials of possibly unknown order.[3] Unfortunately, at the moment I do not have access to any additional archive that I can use to observe S in smaller intervals of time. Nevertheless, I calculated the bias-corrected estimates assuming a uniform

[2]The reason is that, under assumption (A3), the discontinuity in the BLP of Y_i on (D_i, P_i) would be equal to $\delta\gamma_D^*$, so that 2SLS is still consistent. Caution should be applied, because if the equation for R_i is only a BLP, then testing for a discontinuity in it is not equivalent to testing the presence of a discontinuity in the true retirement equation. Therefore I may be using a discontinuity in retirement that does not exist, for instance confounding a jump with a kink.

[3]See Dong (2014) for the other assumptions. It must be added that the current literature does not discuss the potential problems arising from the presence of both rounding and misspecification. Note, however, that in the main results I always fail to reject the null of correct specification for the reduced form of Y.

distribution inside each year interval. This seems to be at least a good approximation, given that eligibility depends on a mixture of years of contribution and age, so that it is equivalent to assume that individuals started to work and were born more or less uniformly during the year.[4]

The last problem, discussed in Battistin et al. (2009), is that the process of recovering S from other survey information introduces measurement error, which smooths the discontinuity in R at $s = 0$. In particular, if S was correctly measured I should not observe anyone in the retirement status before being eligible, that is when $S < 0$. The reason is that, following their strategy, I defined individuals as retired only if they received a pension. As they argued, 2SLS is consistent as long as the measurement error process is statistically independent from (Y, R) conditional on the true value of distance to/from eligibility. One concern is that S is necessarily calculated differently for workers and retired. In particular, the need to determine the year in which the individual has gone into retirement introduces an additional source of measurement error that has no counterpart for workers. For women, whose retirement behaviour is influenced more by the National Retirement Age (NRA), I also estimated the effect on household production using only age as running variable. Results are broadly in line with the conclusions discussed here, and the point estimate is quite similar when I use a linear polynomial and covariates are introduced. However, results are less clear, mainly because a large proportion of women go into retirement as soon as eligible, which is generally earlier than the NRA. A full discussion can be found in Appendix C.

Finally, if there are heterogeneous treatment effects, then I can still interpret the 2SLS coefficient as a Local Average Treatment Effect for those who retire as soon as

[4] I estimated the distribution of date of birth within a year using data from the Italian administrative records (http://demo.istat.it/altridati/IscrittiNascita/, last access: 06/03/13). Unfortunately, they are available only for recent years, between 2001 and 2011. The first four empirical moments (0.507, 0.339, 0.255, 0.203) are similar to the theoretical ones from a uniform distribution (0.500, 0.333, 0.250, 0.200; see Dong, 2014, for a similar comment on the US). I also used data on the month of hire for employees, years 2009-2011 (*Comunicazioni Obbligatorie*, available only for some regions at http://www.venetolavoro.it/servlet/dispatcherServlet?load=/osservatorio/seco/SeCo_04_12.xls, last access: 09/03/13). Although there are downs in December and August, followed by picks in September and January, the first four empirical moments (0.492, 0.338, 0.259, 0.211) are not too far from the theoretical ones with equiprobability of being hired in each month (0.500, 0.348, 0.273, 0.228).

eligible. In this case I also need R_i as a deterministic function of S_i to be monotonic *near* $s = 0$, while δ_i and $R_i(S_i)$ must be jointly independent of S_i (see Hahn et al., 2001). This can be defended using the same arguments advanced for assumptions (A1) and (A2). Despite its local properties, the LATE at eligibility is of interest for a policy maker who is planning to strengthen the seniority and age requirements.

3 Data

The Italian component of the European Union Survey of Income and Living Conditions is a stratified sample of the households' population conducted by the Italian National Statistical Office (ISTAT) every year since 2004.

Here I discuss only the main steps I followed in generating the estimation sample, while details are provided in Appendix A. I identified retired individuals as those who reported not to be working in the week prior to the interview because they were "*in pensione da lavoro",* literally "in work-related pension". Conversely, I defined workers as individuals with "employed" as self-reported employment status, excluding those who have not worked in the week prior to the interview because of being temporarily unemployed or under a temporary layoff public scheme called *cassa integrazione*.

Distance to/from eligibility S is calculated as age at interview minus age at eligibility. Firstly, age at retirement is recovered as age at first regular job, plus years spent in paid job, plus one. The final correction is taken to account for rounding.[5] Secondly, the age at eligibility is then recovered simulating the rules that applied in the year in which the individual went into retirement, calculated as year of birth plus age at retirement, or plus the current one for workers. To summarise, eligibility depended on a combination of different rules based on age and on the number of years the individual had contributed to social security. Different requirements applied to different categories (self-employed, public sector or private employees) and to women. Last but not least, rules were more generous in the past and they have been changed almost

[5]A full discussion of the reasons underlying this choice are contained in the online Appendix A, together with results without this correction, which show that the estimates of the effect on Y are in line with those discussed here, although slightly smaller for women.

every year since 1992 (see Brugiavini and Peracchi, 2004; Morciano, 2007; Intorcia, 2011, for details).

Table 1 reports sample selection by gender. I kept only workers or retirees, for two main reasons. Firstly, I am not interested in comparing them to housewives or other inactive individuals. Secondly, S is not defined for those who have never worked in a paid job. I also excluded all proxy interviews, which is the case when another household member provides the information on an individual who is not available at the time of interview.[6] The reason is that they are likely to increase measurement error and not to be particularly reliable for Y. There are few missing values for house work.[7]

As in Battistin et al. (2009), I kept only the window $S \in [-10, 10]$, in order to limit the influence of observations far away from the eligibility threshold, and I excluded observations with $S_i = 0$. The fact that contributions, age at first job and time spent in paid work are measured in years implies that the observed S is obtained by rounding either up or down, so that $S_i = 0$ includes both cases at the left and at the right of eligibility. One simple solution, suggested by Dong (2014), is to discard observations with $S_i = 0$.[8]

I did not use sample weights, because they were designed for the original sample and it is not clear whether they would be appropriate in the selected one. Nevertheless, in section 4.4 I discuss what happens when I include stratification variables in the regression or I employ sample weights.

The main variable of interest was collected from the question *"On average, how much time per week do you spend on domestic and family-related work (household chores, shopping, caregiving), in hours and minutes?"*.[9] Hereafter, Y_i is equal to the

[6] Including proxy interviews the graphical evidence is less clear for women, but all main estimates lead to the same conclusions. See appendix A for full results.

[7] The other variables employed here do not contain any missing for the sole reason of having been imputed by the ISTAT using multivariate methods. While for income data an imputation factor is available, no such information is reported for qualitative variables. Although this standard practice is debatable, ISTAT does not release the original raw data and therefore I cannot provide details.

[8] Including the zeros, the main estimates lead to the similar conclusion (see Appendix A for full results).

[9] The question was also asked in the following year. However, the 2008 cross-section contains a large number of missing values (18.05%) which casts doubts on its validity.

individual answer to this question, measured in minutes per week.

To better understand the content of this information, I compared it with the Italian Time Use Survey 2008-2009 (TUS), where "family related" work consists of cooking, doing the dishes, cleaning the house, doing the laundry, sewing, knitting, shopping, and general administrative work. It also includes gardening, taking care of pets, maintenance of the house and vehicles. Lastly, it accounts for time spent on caring for children or adults. Unfortunately, the TUS does not collect information on years of contribution, so that it is not possible to replicate the RDD.

From Table 2, it can be noticed that on average time spent on house work is lower in SILC with respect to "family related" work from TUS (column TUS (A)). The difference is proportionally larger for men. After age 65, both samples display a drop in participation and average minutes per day for women. However, the decrease is larger in SILC. For men I observe an increase in average minutes using both datasets, but SILC shows a drop in participation rate against an increase in TUS. Comparing retired and employed individuals, in both samples retirees spend more time on house work, but the difference is larger in the Time Use survey. Moreover, participation slightly drops for women in SILC while it increases in the other dataset.

One might conclude that there is a substantial under-reporting in SILC. However, the difference with TUS data, which is stronger among the elderly, is more likely to be related to a different definition. The general question posed in SILC might exclude some activities. While caring and shopping are explicitly mentioned, "household chores" is likely to be associated with cooking and "core" household work, as defined by Stancanelli and van Soest (2012, pg. 7): "cleaning, doing the laundry, ironing, cleaning the dishes, setting the table, and doing administrative paper work for the household". However, it might exclude "semi-leisure" chores, such as gardening. To provide indirect evidence in favour of this hypothesis, in columns labeled TUS (B) I redefined the variable in the Time Use Dataset, keeping only shopping, cooking, caring and "core" household work. As expected, the averages for men are generally closer to SILC, in particular for those aged 65 or over and for retirees.

4 The change in house work at retirement

4.1 Graphical analysis

Figure 1 draws attention to individuals aged 50-70. It shows the average time spent on household production (Y) at any age, by gender and employment status. There are two main stylised facts that can be drawn from it. The first is that, at any age, the average Y is larger for retirees than for workers. Secondly, not only men spend much less time on house work than women, but also for females the difference between retirees and workers is almost double than the one for males.

As discussed in the introduction, simply comparing workers and retirees may lead to biased estimates. Figure 2 instead focuses on the pattern of retirement and household production with respect to the distance S to/from eligibility from a pension. For both genders I observe a small proportion of individuals who retired before meeting the eligibility criteria. Between $S = -1$ and $S = 1$ there is a large step-up in the fraction of retirees, which continues at a declining rate until reaching 90% or more at $S = 10$.

Time spent on house work is slightly increasing before eligibility is met. After the average Y for men progressively increases, but there is no clear evidence of a discontinuity. I observe an increase at $S = 0$ around 50 minutes/week, but it is followed by alternate falls and rises. For women, time spent in home production is quite constant before eligibility. I then observe a jump at $S = 0$ by nearly 160 minutes/week, followed by an increase. A linear polynomial predicts a discontinuity. A quadratic does not, but it is important to note that it seems to overfit the mean for Y at $S = 0$, predicting a lower value. The comparison of predicted values with the sample average at eligibility is useful in evaluating the polynomial fit, because I am not using observations with $S_i = 0$ in estimating the regressions.

4.2 Estimates of the jump in retirement at eligibility

To test for the presence of a discontinuity in retirement at eligibility, table 3 shows the results of regressions of R_i on the eligibility dummy D_i, a polynomial in S_i and their interactions.[10] I focus on regressions up to the 3rd order because graphical evidence, available on request, shows that 4th order polynomials tend to overfit at $S_i = 0$. For model selection, I focus on minimizing the Akaike (AIC) and Bayesian (BIC) information criteria. The first is suggested by Lee and Lemieux (2010), while the second is useful in this context as it puts more weight on the number of parameters to be estimated. I also discuss Ramsey's RESET test of correct specification, obtained testing the significance of the square and cube of fitted values as additional covariates. Lastly, I test whether the constraints imposed by the polynomial specification are rejected, using Lee and Card (2008) G statistic. It compares the regression with an unrestricted one that includes a dummy for each value of S. In order to be conservative I computed the version valid under homoskedasticity. Using the heteroskedastic-robust version leads to larger p-values in all the models shown in the tables.

For men (columns (1)-(3) in Table 3), both a cubic and quadratic polynomial estimate a jump in retirement (γ_D) around 30 percentage points at eligibility, this being statistically significant at the 1% level.[11] The Akaike criterion favours the highest order, though all 3rd order terms are not statistically significant at conventional levels and the Bayesian criterion is minimized with the linear specification. Ramsey's RESET test does not reject the null of correct specification at the 5% level. Differently, the G test strongly rejects the constraints imposed by the polynomials. Lee and Card (2008) argued that this is not a problem, as far as the best linear projections of the specification error does not bias the estimator of the discontinuity. In this case, they proposed to correct the standard errors by clustering on S. The p-values for the test of $\gamma_D = 0$ is still less than 1%. Lastly, Dong's (2012) corrected estimates are smaller for the 3rd order polynomial, with a p-value 0.055, but they do not differ much in the other two cases.

[10]Results with no interactions, available on request, are stronger for the retirement discontinuity and more precise for the effect on Y.

[11]I can also compare $\widehat{\gamma_D}$ with results from Battistin et al. (2009), who estimated an increase in the proportion of retired male heads at $s = 0$ by 0.435 (s.e. 0.038), using a quadratic polynomial with no interactions. If I run the same regression on SILC, I obtain $\widehat{\gamma_D} = 0.398$ (s.e. 0.027). A t-test for equality fails to reject the null with p-value 0.427. If instead I use a quadratic polynomial with interactions on their dataset, $\widehat{\gamma_D}$ is 0.252 (s.e. 0.069), closer to the equivalent result in SILC (0.313, s.e. 0.048). I used Battistin et al. (2009) files available on the American Economic Review website.

For women (columns (4)-(6) in Table 3), the estimated discontinuity in R at $S = 0$ is small and not statistically significant using the 3rd order polynomial. However, with a quadratic it is around 24 percentage points and statistically significant with either robust or clustered standard errors. The statistical tests do not give a clear indication. The G test is passed at the 5% level with the cubic and not with the quadratic, but the RESET test gives the opposite result. The Akaike information criteria leads us to choose the cubic regression, but the Bayesian is minimized for the second order, and it should be noted that the R^2 does not change up until the third decimal place between the two models. Given the strong institutional reasons for expecting a jump at eligibility, I find it reasonable to focus on the quadratic specification and take it as supporting evidence in favour of the presence of a discontinuity. Dong's correction suggests a smaller jump (0.182), but still statistically significant at conventional levels.[12]

4.3 The effect of retirement on home production

Given the evidence of a jump in retirement at eligibility, I expect that, in the presence of an effect on household production, I should also observe a discontinuity in Y around $S = 0$. In Table 4 I show regressions of Y on D a quadratic or linear polynomial in S. I do not consider higher orders, given that information criteria invariably lead us to prefer the simplest specification and graphical analysis did not show large differences.[13]

Despite the strong evidence of a jump in retirement at eligibility for men, none of the estimated models show a parallel discontinuity in the average time spent on home production (Table 4, columns (1)-(3)). Regression analysis is therefore in line with the intuitions resulting from graphical inspection. To recover the causal effect δ of retirement on house work, I use 2SLS, instrumenting R with D. The highest estimate (Table 5, column (3)) is 73 minutes/week, obtained including only S. It is around 25%

[12] Dong's corrected estimate is similar to the one I obtain by keeping a quadratic polynomial at the right of the discontinuity and a cubic at the left (point estimate 0.181, p-value 0.009).

[13] In the case of the linear polynomial with no interactions, Dong's correction is zero. The reason is that the bias is due to the presence of a kink at eligibility, but using only S there is no change in the slope at $S = 0$.

of the relative OLS estimate (see the last row in Table 5), and it is not statistically significant at conventional levels.[14]

To understand whether results differ sensibly across different groups, Table 7 shows 2SLS estimates splitting by education, area and category. I do it separately, because of sample size.[15] The estimated effect is economically significant for college graduate (176 minutes/week) and in the North (148 minutes/week), but not statistically significant. For private and public employees it is larger than for self-employed (113 and 105 minutes/week against -21), but not far from the one estimated for the whole sample. The only estimate which is statistically significant, even if only at the 10% level, is the one for men living in densely populated areas (34% of the sample), which is approximately 225 minutes/week, similar to the OLS results.

From a theoretical point of view, it is strange that the effect for men is, at least overall, quite small and not sensibly different from zero. Given the strong increase in available time associated with retirement, I should expect at least a partial increase in time spent on home production.[16] There are two possible reasons. The first is that men, at retirement, usually put the most of their effort on "semi-leisure" chores, such as gardening or house-repairing. Indeed, Stancanelli and van Soest (2012) found that men's increase in time spent on home production was mostly in this category. Furthermore, there seems to be some effect for men in densely populated areas, where probably there is less scope for these activities. Another explanation is that, within couples, the unequal division of household chores by gender is not levelled-off at retirement. To provide some evidence, I also split the sample between those living with a partner and those who do not. Among the former, I also distinguished those who are

[14] From graphical inspection it seems that there is a kink in house work at eligibility. I tried to exploit it instrumenting retirement with both D and $D \times S$, where the latter captures the kink (see Dong, 2010; Card et al., 2009). The only exogenous regressor included in the equation of interest is S. Although $\widehat{\delta}$ becomes 135.5, with p-value 0.063, it is not stable to the inclusion of covariates, where it drops down to 90.7 (p-value 0.329). Adding S^2 as an additional covariate leads to the same conclusions.

[15] Results are obtained with no other covariates from X. However, adding the covariates not used in each split sample lead to similar conclusions, with few differences (see Appendix A).

[16] To get a magnitude of the increase in available time, I can use 2SLS with time spent on a paid job as dependent variable. Including $(1 - D) \times S$, $D \times S$, the estimated drop in working time is 2489 minutes/week (s.e. 77), almost equal to the average time spent by workers (2523 minutes/week). Results with a quadratic polynomial are quantitatively similar.

married and the few cases in which they only cohabit. The interesting result is that the change is very close to zero for married men, while it is large for those who are not living with a partner (413 minutes/week), although statistically significant only at the 10% level (p-value 0.069).[17] Those who are not married but they cohabit show quite a large increase. One may speculate that less traditional families have a different distribution of household chores, but the number of observations is far too limited to draw any conclusion.

Results in columns (4)-(6) of Table 4 provide evidence on the presence of a discontinuity in Y at $S = 0$ for women, around 222 minutes/week using a linear polynomial without interactions. Although a second order polynomial shows no discontinuity, the information criteria indicate a preference for the simpler polynomials, for which both the G and the RESET tests fail to reject the null of correct specification.[18] Dong's correction does not lead to different conclusions.

Using the simplest linear specification, and instrumenting R with D, the 2SLS estimate for δ (Table 5, column (6)) is 435 minutes/week, statistically significant at the 5% level. Compared to the equivalent OLS regression, it is 32% smaller. If I use a linear polynomial with interactions (Table 5, column (5)), the estimated effect is quite similar.[19]

While women with a high school or lower degree exhibit estimates for δ similar to that obtained in the main 2SLS regressions, the change in time spent on home production is negative for college graduates (Table 7). The magnitude is very large (289 minutes/week), but it is probably driven by the weakness of the instrument and by the small sample size. The effect is stronger in the North (569 minutes/week). In the Centre and in the South it is still economically relevant (347 and 204 minutes/week

[17]Among married men living with a partner, there are 14 who actually report to be *de facto* separated from their spouse, so that I can infer that they are cohabiting with a different person. Removing them has a very small effect on the estimates. Similarly for women, though there are only 3 cases.

[18]A very similar estimate (212 minutes/week, p-value 0.018) is obtained by a regression of Y on D, S and S^2, with no interactions as in Battistin et al. (2009).

[19]It might be that using a linear polynomial we are confounding a kink with a jump. An alternative would be to use a linear specification (including S), and instrument R with D and $D \times S$, where the latter picks up the kink. The estimated δ would be 412 (s.e. 146), and it is robust to the inclusion of covariates. Adding S^2 as an additional covariate leads to the same conclusions.

respectively), but statistically imprecise. It is also stronger in densely populated areas and in intermediate ones (more than 600 minutes/week), while it is negative, but not statistically significant in thinly populated areas. With respect to job type, the increase for public sector employees (349 minutes/week) is smaller than for other categories, probably because their contracts already allow them to take paid and unpaid days off if they have family needs, such as an elderly parent with impaired health. Differently from men, married women living with their partner show an increase (around 400 minutes/week), though this is smaller than for those not living with a partner (730 minutes/week, p-value 0.062). Estimates for women living with their partner, but not married, are quite imprecise due to a very low first stage.

Results might be driven by the choice of the window size. I checked how they change when it is decreased, using 2SLS regressions including $(1 - D) \times S$, $D \times S$ as covariates, and using D as an instrument for R. Appendix A includes a graph that depicts the different results. The estimates for men oscillate around zero and they are never statistically significant at the 10% level.

For women, $\widehat{\delta}$ is quite stable for to $|S| \geq 5$. However, the 95% confidence interval becomes larger and includes zero. At size 4, the estimate is almost zero, while for size 3 and 2 the first stage F is very small. One reason is that four points are probably not enough to obtain precise estimates of the linear fit with interactions. Another is that, given that measurement error smooths down the discontinuity in retirement, I need other points away from $S = 0$ to partially correct for it. Nevertheless, I tried to exploit only the data point close to eligibility, limiting the sample to $S \in \{-1, 1\}$ and using a simple Wald estimator with R instrumented by D. In this case $\widehat{\delta}$ is equal to 423 minutes/week, very similar to the main results, but much less precise (s.e. 366).

4.4 Discontinuities in other covariates

One way to check the plausibility of the continuity assumption (A2) is to inspect whether some baseline characteristics exhibits discontinuities at eligibility. I focus on three sets of variables:[20]

1. Geographical dummies for area of residence and population density (which were used for stratification);

2. Dummies for highest educational achievement;

3. Variables used to build S.

Geographical dummies are relatively smooth (Table 8, for graphs see Appendix B). We observe an increase at eligibility in the proportion of men residing in the Centre (p-value 0.052) and in the proportion of women in densely populated area (p-value 0.042). However, a test for joint significance of all the discontinuities in geographical dummies fails to reject the null at conventional levels for both genders.[21]

Educational dummies are fairly smooth for women, while for men we observe an increase in the proportion of high school graduates at eligibility and a decrease in those who only completed the middle school degree. This discontinuity is a problem if it is evidence of endogenous sorting of individuals. In the present context, one possibility is that they were able to exploit rules related to their educational level: in Italy university graduates are allowed to pay-back social contributions to cover the years of higher education and become eligible earlier. But in this case I should have also found an increase in university graduates at eligibility, while I found no evidence of such a discontinuity. Another problem could be the 1963 educational reform, that had an effect on cohorts from 1949 (see Brunello et al., 2012, p. 19). It seems that this is a minor issue in this context. Firstly, by construction $S = 0$ does not include a single cohort: the proportion of cohort 1949 at $S = -1$ is 0.725, quite close to the proportion at $S = 1$ (0.621). Secondly, if this was the problem, we should expect a decrease in the educational level at eligibility, given that those at $S \geq 0$ are older individuals.

[20] Other changes may be due to retirement itself, such as a reduction in household size or an improvement in health (Battistin et al., 2009; Coe and Zamarro, 2011). In Appendix D I show that this does not occur in this case.

[21] The same applies if we separately test the joint significance of area dummies and of population density dummies

To further inspect the change in overall educational level, I calculated the total years of schooling by attributing the official length to each degree. This allows to account for some shorter vocational training degrees that are included in "high school" dummy. As shown in Table 8, there is no evidence of a discontinuity for both genders. I also calculated the difference between age 6 and the age at which the individual has taken his/her highest degree. This is larger than years of education, both because of grade retention and individuals taking degrees later in life. This variable seems to show a drop in the "age at highest degree - 6" variable, not necessarily in line with an increase in the educational level. It is not statistically significant, although the joint test for the discontinuities in both additional educational variables has p-value 0.046.

Among variables used to build S, age, years of contribution and age at retirement are fairly smooth. Differently, for men we observe a decrease at eligibility in the proportion of public employees, compensated by an increase in self-employed. This is related to an increase in years spent in paid job, which makes sense given that some self-employed individuals may have not contributed for some years to the retirement scheme, because they were included in the national insurance only at the end of the fifties. For women we observe a decrease in private sector employees at eligibility, though statistically significant only at the 10%, mostly compensated by an increase in self-employed (not statistically significant). We also observe an increase in years spent in a paid job, though less precise (p-value 0.091), and an increase in the age at first job of around one year, statistically significant at the 5%.

To summarize, there seem to be discontinuities mainly related to some of the variables that enter in the definition of eligibility. Having excluded the possibility of sorting related to the educational qualification, one alternative explanation is that the retirement reforms created some discontinuities across workers with different employment histories. The source of these differential treatments does not seem to be precisely manipulable by the single individual, given that the repeated changes in the rules between 1992 and 2007 were hardly predictable at the time s/he started his/her career. However, the resulting discontinuities make individuals across eligibility not completely comparable.

One possible solution is to state all assumptions (A1)-(A3) conditional on the different covariates and employ the RDD on cells defined by employment category. Although this is not feasible given the sample size, I have already discussed how estimates differ when the sample is split according to baseline characteristics (taking one variable at a time).

Another solution would be to adopt a parametric framework, where the counterfactual $\epsilon_i \equiv Y_{0i}$ depends linearly on these additional variables, which, therefore, enter all regressions as a vector of covariates X (see Frölich, 2007, for a non-parametric alternative). To understand how the introduction of X affects the estimates I plotted against S the fitted values for R and Y obtained from a regression on dummies for education, geographical area and employment category, plus age, years of social contribution, years spent in paid work and age at which the respondent began the first regular job (the graph is reported in Appendix B for space constraints).[22] For both genders there is a small drop in fitted retirement probability at eligibility. Also fitted hours of domestic work show a small drop at $S = 0$ for males (18 minutes/week if estimated using a linear polynomial) and a larger decrease for women (54 minutes/week).

Indeed, when using covariates, estimates for the discontinuity in retirement are basically unchanged for men, while they are slightly smaller, but still statistically significant for women. For both men and women the estimated discontinuity in Y is larger. For men, the highest estimate for δ is 89 minutes/week (Table 6, column (6)), but not statistically significant at conventional levels and still far from the OLS results. Also for women the estimates with covariates are larger: using a linear polynomial with interactions the results is 528 minutes/week, while it is 483 including only S.

The discontinuities in covariates may also be due to decision not to use sample weights. For men, using them I still find the discontinuities in geographical area and educational dummies, while those for employment category have similar size but are not statistically significant (full results available on request). 2SLS estimates for δ are

[22] Age at retirement is not included because it is a nonlinear function of the other variables (see Appendix A). 2SLS results are basically unchanged when adding it.

therefore larger, with a maximum of 139 minutes/week using covariates, but never statistically significant. For women the discontinuities in baseline covariates become all statistically not significant when using weights, although they do not change much in size. Estimates for δ are smaller than those presented in the main text, but still larger than 400 minutes/week. In a nutshell, overall conclusions are confirmed using sample weights.

Finally, I know from McCrary (2008) that a discontinuity in the density function at eligibility might be a sign of individuals sorting around the threshold, even if a continuous density function is neither a sufficient nor a necessary condition for identification. Density plots are reported in Appendix B. I observe no change in the density at $S = 0$ for men. For women I observe a drop of around 1 percentage point, if estimated with a linear fit. However, if individuals were able to manipulate their distance to/from eligibility, there would be no reason to expect them to misreport it in order to become ineligible. Given that retirement is not generally compulsory at $S = 0$ according to Italian rules, most individuals have an incentive to manipulate S_i in the opposite direction, so that I should find an increase in density at eligibility. Hence I do not take the observed drop as evidence of sorting.

5 Discussion

I used a RDD that exploits the discontinuity in retirement behaviour induced by the Italian Social Security System. Although the proportion of men leaving employment at eligibility is quite large, the strong discontinuity in retirement is not associated with a jump in time spent on home production. Conversely, for women I observe an increase in both retirement and house work at eligibility. The resulting estimate for the causal effect of retirement on house work is between 430 and 530 minutes per week (nearly an hour per day), depending on the introduction of covariates and on whether or not we interact S with D.

The strong gender difference found for Italy seems to have no parallel in the US, France, Germany and Spain. Hurd and Rohwedder (2005, 2006), using data from the Health and Retirement Study, showed that women who retired between 2001 and 2003 increased by 309 minutes per week the time spent in activities with close market substitutes.[23] However, they found a sensible increase for men as well, of around 361 minutes/week. The gerontology literature provides similar evidence (Szinovacz and Harpster, 1994). Szinovacz (2000), using US panel data, found that husbands' increase their relative contribution not only in "male tasks (outdoor tasks, repairs, paying bills)", but also in "female tasks (preparing meals, doing the dishes, cleaning house, laundry)" (Szinovacz, 2000, p. 82). For France, Stancanelli and van Soest (2012) estimated that at retirement wives spend 2 hours 40 minutes per weekday more on house work, but they found that husbands also increased house work by around 3 hours per weekday. Furthermore, there is evidence for Germany (Schwerdt, 2005; Luhrmann, 2010) of an increase in housework at the retirement of the household heads, who are mostly men.[24] Lastly, Luengo-Prado and Sevilla (2013) provided evidence that in Spain the retirement of the household head causes a reallocation of household duties, with men increasing their involvement in shopping and cooking. They also suggested that this equalizing effect is the result of a move towards more egalitarian social norms.

One explanation for the different result in Italy is that, after retirement, men mostly focus on "semi-leisure activities", such as gardening, which are not included in the SILC definition of home production. This argument is supported by the descriptive comparison with Time Use Data and by results from Stancanelli and van Soest (2012), who showed that (in France) the increase for men is concentrated in these activities. Furthermore, it must be noted that some weak evidence of an increase is found for men residing in densely populated areas, who are probably less likely to specialise in

[23]Activities included house cleaning, yard work/gardening, food preparation, home improvements, washing/ironing, shopping and finances related.

[24]Schwerdt (2005), analysing data from the German SOEP (1994-2000), studied home production (errand, housework and yardwork) around a window of 2 years before and after retirement. Distinguishing low and high income replacement household heads, he estimated that the former spent around 714 minutes more per week on house work after retiring, while the latter 504 minutes. Similarly, Luhrmann (2010) found an increase of home production (cooking, preparing meals, paperwork and gardening) in households with a retired head by about 574 minutes per week, using German Times Use surveys 1991/92 and 2001/02.

these "semi-leisure activities". Another explanation is that married men living with their spouse do not increase their participation in household chores or caregiving at the moment of retirement, leaving them to their wives. Indeed, when I focus on this group the estimate is very small, while it is around 400 minutes/week for those living without a partner, even if statistically significant only at the 10% level. Differently, for women the estimate is positive and statistically significant in both cases. Clearly, although these sample split exercises are suggestive, they cannot prove whether this explanation is ultimately correct. Furthermore, one can note that also for women the estimate is larger for singles, and therefore interactions within the household are worth to be further studied.[25] Indeed, a natural extension of this work would be to look at the interrelations between partners around retirement, in the spirit of Stancanelli and van Soest (2012), but this analysis goes beyond the purpose of the present paper and I postpone it to future research.

Is the increase in house work sufficient to explain the change in consumption at retirement? The literature from Italy provided evidence of a drop in expenditure at the time when household heads (usually men) leave work. In line with research from other countries (Hurst, 2008), this decrease is mainly on food and work-related expenses. Battistin et al. (2009) found that retirement of male household heads was associated with a drop in expenditure on non-durable goods by 9.8% and on food (including meals out) by 14.1%. Part of this change was explained by adult children leaving the household, so that focusing on equivalised expenditure the reduction was only 4.1% for non-durable goods and 8.4% on food, only the latter statistically significant.[26] Miniaci et al. (2010), using a cohort analysis on data from the 1985-1996 Italian Survey of Family Budgets, found a drop in total consumption expenditure at retirement by 5.4%. Their evidence suggested that this fall could be explained by increased home

[25]Similarly to the result for men, the effect for women not living with a partner is statistically significant only at the 10% level (p-value 0.062), probably because of reduced sample size. It must be added that a statistical test for the equality of the effect for married (and living with a partner) men and for those not living with a partner rejects the null, although still only at the 10% level (p-value 0.094). Differently, for women the test fails to reject the null (with p-value 0.443).

[26]In Appendix D I show that in my sample there is no evidence of a reduction in household size at retirement.

production of meals and decreased purchase of work-related goods.

If we focus on female retirement, my estimates could explain a drop in expenditure on consumption. On the one hand, they are sufficient to justify a significant increase in the home production of goods, such as food, which could substitute for products purchased on the market. On the other hand, increased effort on shopping may also reduce expenditure. Aguiar and Hurst (2007) found evidence that, in the US, a good's price decreases by 7 to 10 percent when shopping frequency doubles. For instance, from the TUS I estimate that employed women aged 55-59 spend on average 189 minutes/week on shopping, so that doubling this time input is compatible with the estimated increase in household work at retirement.

However, results from Battistin et al. (2009) refer to the retirement of male household heads. But for the case of men my estimates do not seem to indicate a large change in home production, and therefore they are less likely to explain a drop in consumption. On the one hand, the increase for them can be positive, but small in absolute value and therefore difficult to detect, given also that the estimate is quite imprecise. On the other hand, the estimated effect is very close to zero for married men, who represent the majority of the individuals in the sample. Unless both partners' are employed and leave work in the same year, the drop in consumption associated with the retirement of men is more likely to be explained by other factors discussed in the literature, such as a decrease in work-related expenses.

Tables and Figures

Table 1: Sample selection

	Male		Female	
	obs	% change	obs	% change
Raw 2007 SILC data	21,522		23,611	
Worker or Retired	16,958	-21%	12,162	-48%
Non Proxy	13,979	-18%	10,856	-11%
Missing house work	13,437	-4%	10,546	-3%
$S \in [-10, 10]$	4,139	-69%	2,795	-73%
$S \neq 0$	3,970	-4%	2,700	-3%

Table 2: Participation and average minutes per day spent in house work, by gender and characteristics, SILC 2007 and TUS 2008-2009, all individuals aged 15+.

	Women						Men					
	Avg minutes/day			% participants			Avg minutes/day			% participants		
	SILC	TUS (A)	TUS (B)	SILC	TUS (A)	TUS (B)	SILC	TUS (A)	TUS (B)	SILC	TUS (A)	TUS (B)
TOTAL	238	285	271	0.93	0.93	0.92	64	96	70	0.63	0.68	0.64
AGE												
15-24	92	94	91	0.75	0.72	0.71	23	27	23	0.36	0.37	0.35
25-44	245	292	285	0.96	0.94	0.94	60	75	65	0.66	0.66	0.64
45-64	284	332	315	0.97	0.98	0.98	66	112	79	0.67	0.73	0.68
65+	219	306	278	0.89	0.93	0.92	80	150	97	0.63	0.84	0.76
SELF DEFINED EMPLOYMENT STATUS												
Employed	194	219	212	0.96	0.94	0.93	56	73	60	0.65	0.64	0.61
Unemployed	252	296	285	0.94	0.92	0.92	59	94	76	0.51	0.63	0.62
Housewife	338	389	370	0.97	0.97	0.97						
Student	62	70	68	0.70	0.68	0.67	25	23	20	0.40	0.39	0.37
Retired	236	325	296	0.92	0.96	0.95	85	168	108	0.67	0.87	0.78
Other	214	271	253	0.84	0.87	0.87	77	107	81	0.57	0.64	0.61

Note: estimated on original microdata using sample weights. In SILC I excluded missing values in house work. TUS (A) refers to total "family related" work, while column (B) contains only shopping, cooking, caring and "core" household work. To calculate average minutes per day in SILC, I divided Y by 7. TUS data refer to an average day calculated from averaging diaries collected in different days of the week.

Table 3: First stage OLS regressions for retirement status, SILC 2007 (selected sample)

Dep var R	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	0.270***	0.313***	0.398***	0.104	0.236***	0.471***
	(0.082)	(0.048)	(0.027)	(0.097)	(0.059)	(0.034)
$(1-D)\times S$	-0.004	0.021***	0.011***	0.081***	0.034***	0.008***
	(0.024)	(0.007)	(0.001)	(0.028)	(0.009)	(0.002)
$(1-D)\times S^2$	-0.004	0.001		0.012**	0.002***	
	(0.005)	(0.001)		(0.005)	(0.001)	
$(1-D)\times S^3$	-0.000			0.001**		
	(0.000)			(0.000)		
$D\times S$	0.138***	0.075***	0.043***	0.206***	0.137***	0.048***
	(0.052)	(0.017)	(0.004)	(0.059)	(0.019)	(0.004)
$D\times S^2$	-0.016	-0.003**		-0.023**	-0.008***	
	(0.010)	(0.001)		(0.011)	(0.002)	
$D\times S^3$	0.001			0.001		
	(0.001)			(0.001)		
Constant	0.089**	0.117***	0.097***	0.188***	0.134***	0.078***
	(0.036)	(0.021)	(0.012)	(0.046)	(0.028)	(0.014)
Observations	3970	3970	3970	2700	2700	2700
R^2	0.570	0.569	0.568	0.703	0.703	0.696
$H_0:\gamma_D=0$ (p-val)	0.001	0.000	0.000	0.286	0.000	0.000
- (p-val cluster)	0.004	0.001	0.000	0.168	0.003	0.000
Dong's $\widehat{\gamma_D}$	0.197	0.285	0.382	0.036	0.182	0.451
Dong's $\widehat{\gamma_D}$ (p-val)	0.055	0.000	0.000	0.767	0.005	0.000
RESET2	0.064	0.129	0.048	0.021	0.109	0.000
RESET23	0.138	0.055	0.139	0.070	0.244	0.000
G (p-value)	0.005	0.003	0.001	0.093	0.032	0.000
AIC	1949.554	1949.817	1954.425	380.124	382.628	441.741
BIC	1999.846	1987.536	1979.571	427.332	418.034	465.345

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. γ_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table 4: Reduced form OLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	-20.802	21.280	30.915	16.112	209.189**	221.914**
	(64.606)	(38.201)	(37.069)	(149.527)	(89.015)	(86.830)
$(1-D) \times S$	-0.456	3.916		21.163	8.689	
	(15.672)	(3.411)		(36.313)	(8.334)	
$(1-D) \times S^2$	-0.384			1.097		
	(1.341)			(3.139)		
$D \times S$	42.465*	17.065***		104.541**	21.533*	
	(23.501)	(5.474)		(50.732)	(11.424)	
$D \times S^2$	-2.307			-7.502*		
	(2.153)			(4.547)		
S			9.391***			13.881**
			(3.029)			(6.783)
Constant	408.109***	417.312***	449.276***	1648.099***	1621.634***	1652.413***
	(38.964)	(22.567)	(20.526)	(91.999)	(54.469)	(46.421)
Observations	3970	3970	3970	2700	2700	2700
R^2	0.018	0.017	0.016	0.036	0.035	0.034
$H_0 : \beta_D = 0$ (p-val)	0.747	0.578	0.404	0.914	0.019	0.011
- (p-val cluster)	0.661	0.547	0.463	0.799	0.018	0.017
Dong's $\widehat{\beta_D}$	-42.583	14.705		-27.010	202.767	
Dong's $\widehat{\beta_D}$ (p-val)	0.546	0.708		0.867	0.026	
RESET2	0.799	0.278	0.069	0.539	0.117	0.414
RESET23	0.756	0.554	0.130	0.723	0.259	0.259
G (p-value)	0.166	0.200	0.101	0.760	0.649	0.654
AIC	61615.827	61613.328	61615.615	45048.293	45047.607	45046.465
BIC	61653.546	61638.474	61634.475	45083.699	45071.211	45064.168

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. β_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table 5: 2SLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	-66.374	53.429	73.225	68.362	444.351**	434.811***
	(208.601)	(95.124)	(86.878)	(628.995)	(183.315)	(165.434)
$(1-D) \times S$	0.909	3.339		18.825	5.186	
	(18.423)	(4.037)		(51.406)	(9.113)	
$(1-D) \times S^2$	-0.327			0.939		
	(1.441)			(4.054)		
$D \times S$	47.462	14.745*		95.166	0.271	
	(36.742)	(8.809)		(126.991)	(18.234)	
$D \times S^2$	-2.499			-6.950		
	(2.603)			(8.670)		
S			7.606			3.428
			(4.946)			(10.308)
Constant	415.898***	412.140***	436.378***	1638.943***	1586.944***	1576.825***
	(57.339)	(28.883)	(34.626)	(158.098)	(63.618)	(72.433)
Observations	3970	3970	3970	2700	2700	2700
$H_0 : \delta = 0$ (p-val)	0.750	0.574	0.399	0.913	0.015	0.009
- (p-val cluster)	0.672	0.505	0.384	0.790	0.001	0.000
First Stage F	42.196	216.080	265.562	15.930	197.380	255.250
OLS est. for δ	298.222***	285.657***	287.267***	677.528***	657.555***	638.151***
	(29.401)	(28.573)	(27.504)	(75.147)	(71.879)	(68.582)

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The p-val cluster is calculated by clustering on S.

Table 6: 2SLS regressions for time spent on house work (in minutes per week), including covariates, SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	-96.849	81.312	89.194	389.471	527.824***	483.067***
	(220.787)	(95.816)	(93.195)	(653.412)	(188.988)	(179.980)
$(1-D) \times S$	22.190	11.839		19.743	20.639	
	(28.001)	(10.382)		(74.186)	(21.139)	
$(1-D) \times S^2$	-0.034			-0.709		
	(1.440)			(3.756)		
$D \times S$	62.865	17.065		43.092	-5.571	
	(42.160)	(10.595)		(143.451)	(24.246)	
$D \times S^2$	-3.180			-3.543		
	(2.722)			(8.735)		
S			14.668			8.237
			(9.743)			(21.288)
Constant	1354.507**	913.408**	978.413**	3088.969*	2773.672***	2528.197***
	(656.012)	(410.811)	(392.830)	(1853.701)	(928.576)	(933.595)
Observations	3970	3970	3970	2700	2700	2700
$H_0 : \delta = 0$ (p-val)	0.661	0.396	0.339	0.551	0.005	0.007
- (p-val cluster)	0.524	0.287	0.236	0.239	0.000	0.000
First Stage F	43.794	247.004	269.478	15.431	192.845	222.697
OLS est. for δ	292.556***	281.519***	282.089***	610.274***	605.257***	586.585***
	(32.982)	(31.026)	(30.970)	(78.814)	(73.651)	(72.597)

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The p-val cluster is calculated by clustering on S. Covariates include age, age at first job, years of contributions, years spent in a paid job, plus geographic area, population density, education and employment category dummies. Coefficients are available on request.

Table 7: 2SLS estimates splitting the sample by education, area, employment category and marital status, SILC 2007 (selected sample)

		Men				Women			
		$\widehat{\delta}$	p-value	First stage F	obs	$\widehat{\delta}$	p-value	First stage F	obs
By education:	Middle school or less	72.183	0.579	138	2020	468.949	0.049	141	1353
	High school	28.098	0.830	101	1492	573.677	0.018	116	1006
	College	176.195	0.448	30	458	-288.646	0.620	16	341
By area:	North	148.417	0.176	157	1953	568.688	0.007	148	1369
	Centre	86.169	0.619	61	966	346.976	0.302	61	717
	South	-105.052	0.635	50	1051	203.989	0.598	51	614
By degree of urbanization:	Densely populated	225.247	0.067	108	1360	641.716	0.002	98	985
	Intermediate area	-28.386	0.855	97	1689	690.372	0.013	107	1109
	Thinly populated	-24.008	0.896	63	921	-197.694	0.608	53	606
By category:	Private employee	112.762	0.222	268	2080	482.389	0.028	207	1073
	Public employee	104.895	0.690	34	761	349.389	0.226	77	980
	Self-employed	-21.271	0.934	25	1129	569.424	0.182	33	647
By marital status:	Living with partner, married	0.899	0.992	211	3132	399.771	0.026	236	1827
	Living with partner, not married	679.165	0.077	11	105	-346.473	0.840	2	55
	Not living with partner	413.374	0.069	46	733	729.858	0.062	38	818

Note: all estimates include only a constant and S, while R is instrumented by D. The p-value and the first stage F are calculated using robust standard errors.

Table 8: Regressions for different socio-economic variables, SILC 2007 (selected sample)

	Men			Women		
Dep. var.	$\widehat{\gamma}_D$	p-value	G (p-value)	$\widehat{\gamma}_D$	p-value	G (p-value)
Geographical dummies						
North	-0.046	0.196	0.249	-0.035	0.430	0.975
Centre	0.059*	0.052	0.508	0.013	0.743	0.060
South	-0.014	0.656	0.042	0.022	0.561	0.373
Densely pop area	-0.030	0.368	0.680	0.087**	0.042	0.284
Intermediate area	0.026	0.458	0.048	-0.068	0.120	0.790
Thinly pop area	0.004	0.881	0.229	-0.019	0.590	0.279
Test for joint significance		0.368			0.245	
Educational dummies						
College	-0.015	0.522	0.627	0.016	0.608	0.893
High school	0.084**	0.013	0.051	0.005	0.904	0.318
Middle school	-0.065**	0.041	0.048	0.004	0.921	0.970
Primary sch.	-0.004	0.889	0.460	-0.025	0.523	0.370
Test for joint significance		0.064			0.909	
Additional educational variables						
Years of schooling	0.282	0.318	0.769	0.508	0.170	0.828
Age highest edu - 6	-1.180	0.102	0.318	0.736	0.366	0.007
Test for joint significance		0.046			0.373	
Variables used in building S						
Age	-0.123	0.589	0.036	-0.131	0.516	0.670
Y. of contribution	0.175	0.592	0.284	-0.220	0.732	0.125
Age at retirement	0.064	0.828	0.064	-0.087	0.855	0.285
Age at first job	0.060	0.860	0.798	1.175**	0.032	0.725
Years spent in a paid job	0.976**	0.018	0.304	1.186*	0.092	0.325
Private employee	-0.013	0.704	0.020	-0.071*	0.091	0.485
Public employee	-0.055**	0.044	0.469	0.024	0.579	0.194
Self-employed	0.068**	0.031	0.005	0.048	0.212	0.210
Test for joint significance		0.001			0.000	

* p<.10 ** p<.05 *** p<.01. The regressions include $(1 - D) \times S$, $D \times S$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. The null hypothesis for the test for joint significance is that there is no discontinuity in all variables of each group, and it is run by using Stata command `suest` with robust standard errors. In the case of mutually exclusive dummies (for instance North-Centre-South), one constraint is removed, but the result does not depend on which one is chosen.

Figure 1: Average minutes/week of house work by employment status (circles for workers, triangles for retirees) and age (in years), SILC 2007, only age-employment cells with at least 10 obs. Lines are fit from a 2nd order polynomial (with 95% c.i.).

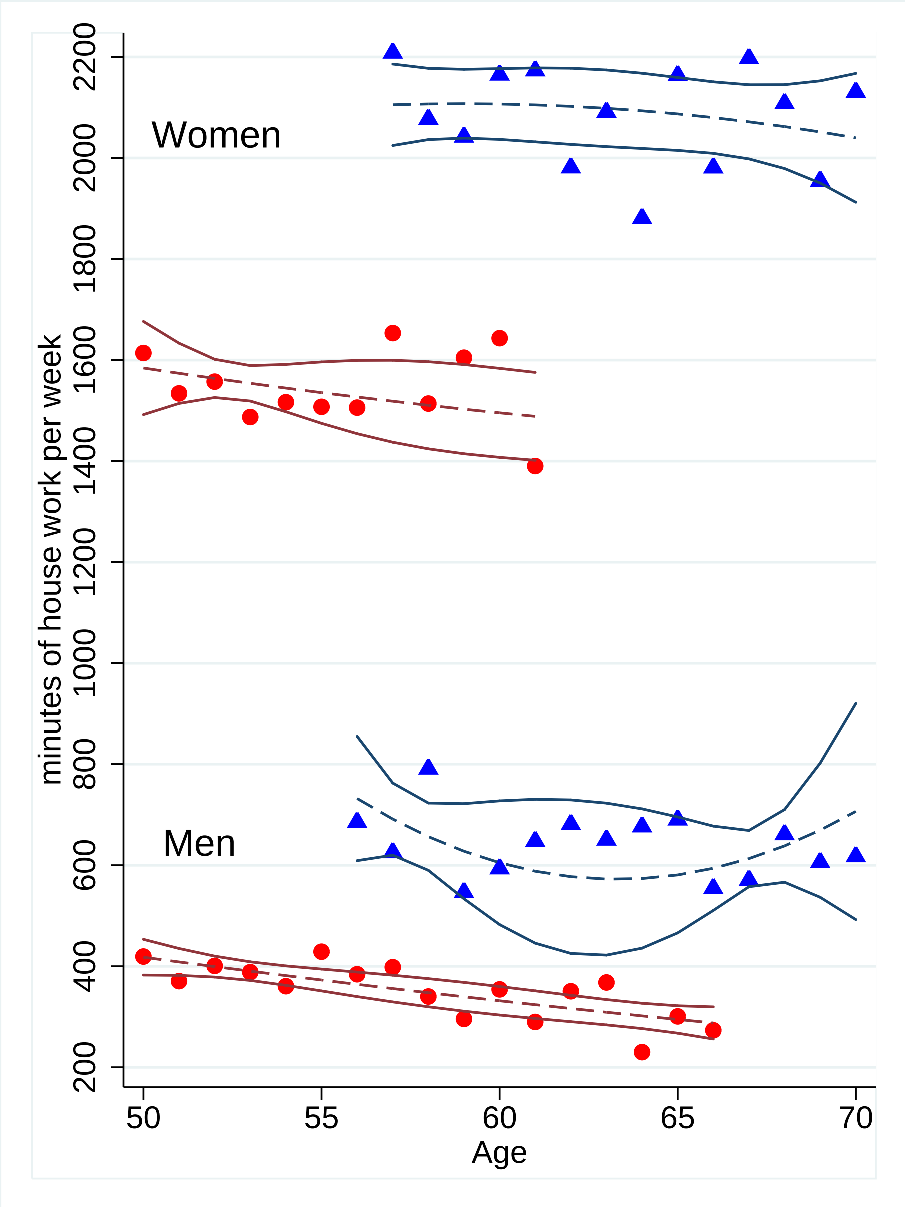
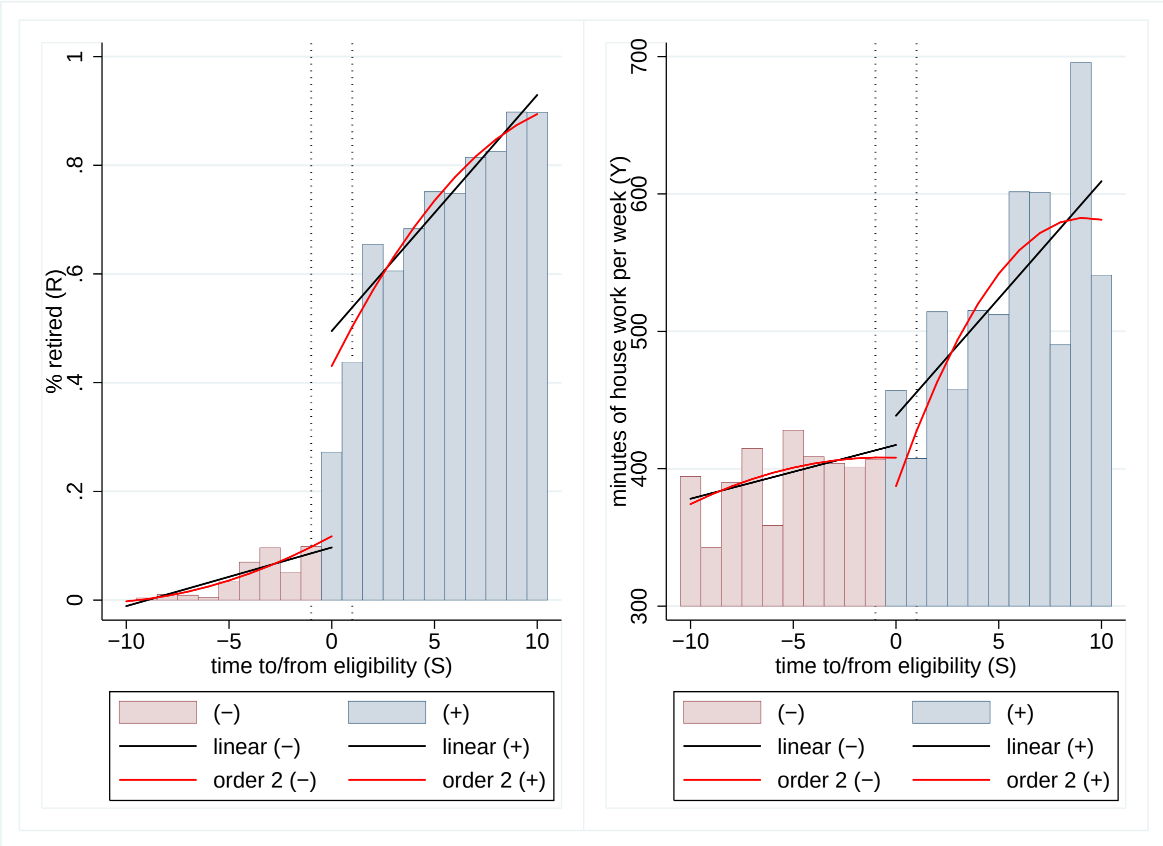
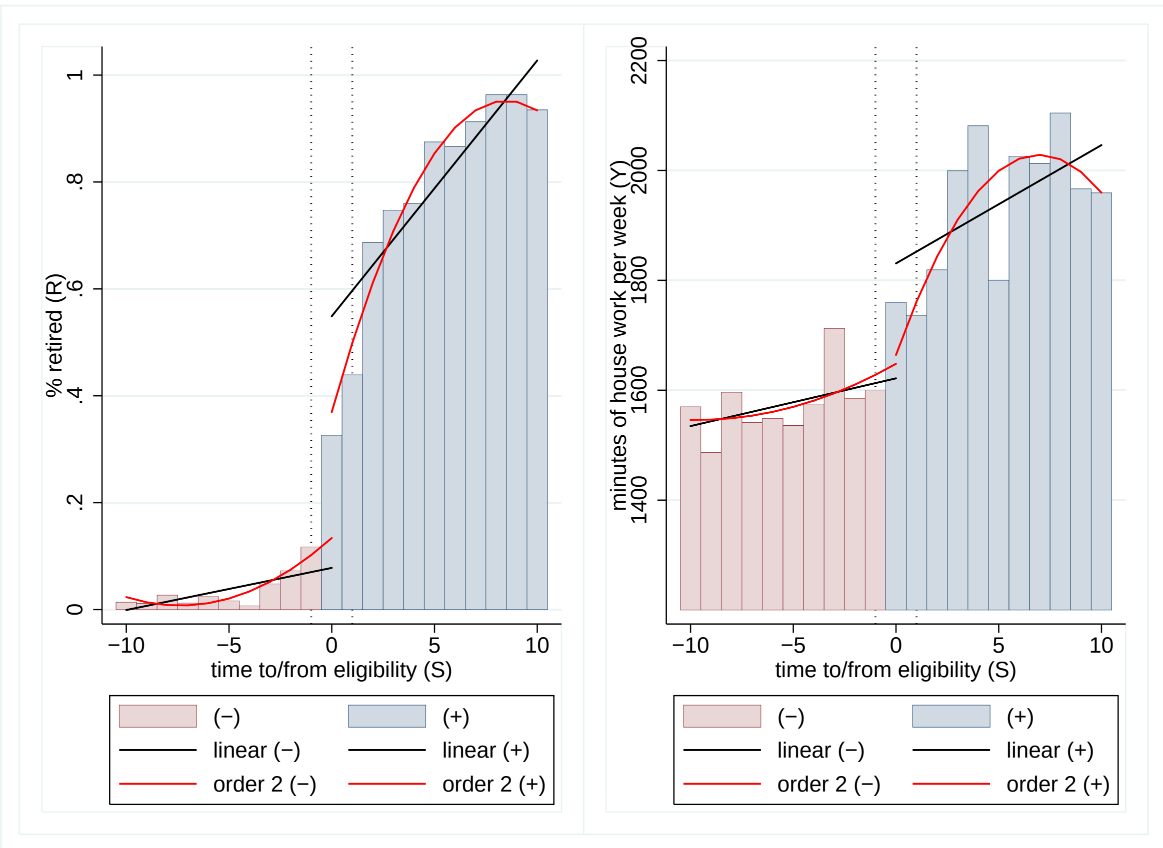


Figure 2: Retirement and house work with respect to S, SILC 2007 (selected sample).



(a) Men



(b) Women

References

Addabbo, T., Caiumi, A., Maccagnan, A., 2011. Allocation of Time within Italian Couples: Exploring the Role of Institutional Factors and their Effects on Household's Wellbeing. CAPPaper No. 85.

Aguiar, M., Hurst, E., 2005. Consumption versus Expenditure. Journal of Political Economy 113(5), 919–948.

Aguiar, M., Hurst, E., 2007. Life-cycle Prices and Production. The American Economic Review 97(5), 1533–1559.

Battistin, E., Brugiavini, A., Rettore, E., Weber, G., 2009. The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach. American Economic Review 99(5), 2209–2226.

Baum, C.F., Schaffer, M.E., Stillman, S., 2007. Enhanced routines for instrumental variables/GMM estimation and testing. Boston College Economics Working Paper No. 667.

Brugiavini, A., Peracchi, F., 2004. Micro-Modeling of Retirement Behavior in Italy. in J. Gruber and D. Wise (eds), Social Security Programs and Retirement around the World: Micro-Estimation, University of Chicago Press .

Brunello, G., Weber, G., Weiss, C.T., 2012. Books Are Forever: Early Life Conditions, Education and Lifetime Income. IZA Discussion Paper No. 6386.

Burda, M.C., Hamermesh, D.S., Weil, P., 2006. The Distribution of Total Work in the EU and US. IZA Discussion Paper No. 2270.

Card, D., Lee, D.S., Pei, Z., 2009. Quasi-Experimental Identification and Estimation in the Regression Kink Design. Princeton University - Industrial Relations Section Working Paper No. 553.

Coe, N.B., Zamarro, G., 2011. Retirement effects of health in Europe. Journal of Health Economics 30, 77–86.

Dong, Y., 2010. Jumpy or Kinky? Regression Discontinuity without the Discontinuity. MPRA Paper 25461.

Dong, Y., 2014. Regression Discontinuity Applications with Rounding Errors in the Running Variable. Journal of Applied Econometrics DOI: 10.1002/jae.2369.

Frölich, M., 2007. Regression Discontinuity Design with Covariates. IZA Discussion Paper 3024.

Hahn, J., Todd, P.E., der Klaauw, W.V., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69(1), 201–209.

Hurd, M.D., Rohwedder, S., 2005. Changes in Consumption and Activities in Retirement. Michigan Retirement Research Center Working Paper 2005-096.

Hurd, M.D., Rohwedder, S., 2006. Some answers to the retirement-consumption puzzle. NBER Working Paper 12057.

Hurst, E., 2008. The Retirement of a Consumption Puzzle. NBER Working Paper 13789.

Intorcia, M., 2011. L'ordinamento pensionistico (quarta edizione). INPS.

Jann, B., 2007. Making regression tables simplified. The Stata Journal 7(2), 227–244.

Lee, D.S., Card, D., 2008. Regression discontinuity inference with specification error. Journal of Econometrics 142, 655–674.

Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. Journal of Economic Literature 48, 281–355.

Luengo-Prado, M.J., Sevilla, A., 2013. Time to cook: expenditure at retirement in Spain. The Economic Journal 123, 764–789.

Luhrmann, M., 2010. Consumer Expenditures and Home Production at Retirement - New Evidence from Germany. German Economic Review 11(2), 225–245.

McCrary, J., 2008. Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics 142, 698–714.

Miniaci, R., Monfardini, C., Weber, G., 2010. How does consumption change upon retirement? Empirical Economics 38, 257–280.

Morciano, M., 2007. Un modello di microsimulazione a popolazione dinamica per l'analisi del sistema di protezione sociale italiano. PhD dissertation, available at http://amsdottorato.cib.unibo.it/147/1/TESI_MORCIANO_MARCELLO.pdf (last access: 22/11/2012) .

Rogerson, R., Wallenius, J., 2012. Retirement, Home Production and Labor Supply Elasticities. Society for Economic Dynamics 2012 Meeting Papers 41.

Schwerdt, G., 2005. Why does consumption fall at retirement? Evidence from Germany. Economics Letters 89, 300–305.

Stancanelli, E., van Soest, A., 2012. Retirement and Home Production: A Regression Discontinuity Approach. American Economic Review, Papers and Proceedings 102(3), 600–605.

Szinovacz, M., Harpster, P., 1994. Couples' Employment/Retirement Status and the Division of Household Tasks. Journal of Gerontology 49(3), S125–S136.

Szinovacz, M.E., 2000. Changes in Housework After Retirement: A Panel Analysis. Journal of Marriage and the Family 62, 78–92.

Appendices for "Retirement, pension eligibility and home production"
(ONLINE SUPPLEMENTARY MATERIAL)

Contents

Appendix A: additional data description and robustness checks

Additional data description

Definition of retirement

I identify retired individuals as those who report not to be working in the week previous to the interview because they were "*in pensione da lavoro*", literally "in work-related pension". Battistin et al. (2009) used a similar definition, but controlling for whether or not they were actually recipients of a pension. This cannot be done using SILC cross-sectional data, because income information refers to the calendar year previous to the one of interview. It should be added that under the current rules it is possible to retire and receive the first payment only a few months later.

The definition of retirement implies that $R = 1$ corresponds to zero hours of paid work by construction. Stancanelli and van Soest (2012) used instead the self-defined economic status, so that some of the retired may be working for some time during the week. However, in SILC 2007, among those whose self-reported status is "retired from work", only 1.15% of men and 0.58% of women have worked at least one hour in the week previous to the interview. If they did not work, though, 6.26% of men and 10.47% of women report not to have searched for a job for reasons different from being "in work-related pension". For instance, around 1% have health related problems and another 1% is taking care of relatives, while 2.97% of men and 5.97% report not to search for a job for age-related reasons (see Table A1). Furthermore, among those reporting to be "in work-related pension", 6% of the women have "housewives" as occupational status, so that it is possible that individuals receive a pension but do not report to be "retired from work".[1]

Differently, I define workers as individuals with "employed" as self-reported em-

[1] I also estimated the main regressions setting $R = 1$ if the individual's self-defined occupational status is "retired from work". Results for men are almost unchanged. The estimated effect for women including only S as covariate is 360 minutes per week (p-value 0.046), smaller than the one reported here.

Table A1: Reasons why the respondent did not search for a job; only individuals who did not work in the week previous to the interview and whose self-reported occupational status is "retired from work"; SILC 2007.

	Men	Women	Total
Has found a job starting in less than 3 months	0.05%	0.00%	0.03%
Has found a job starting in more than three months	0.04%	0.02%	0.03%
Health related problems	0.88%	0.93%	0.90%
To look after children or relatives	0.71%	1.38%	1.01%
In retirement from work	93.74%	89.53%	91.85%
S/he believes s/he cannot find a job	0.16%	0.22%	0.18%
Age-related reasons	2.89%	5.97%	4.27%
S/he does not need	0.21%	0.43%	0.31%
Is waiting results from job applications	0.04%	0.04%	0.04%
Is unable to work	0.55%	0.50%	0.52%
Other reasons	0.74%	0.97%	0.85%
Total	100.00%	100.00%	100.00%

ployment status, excluding those who have not worked in the week previous to the interview because temporarily unemployed or under a job maintaining scheme called *Cassa integrazione*.[2]

Construction of the time to/from eligibility

The key variable is age at eligibility to retirement. In 2007, individuals could retire if they met one of three alternative criteria. The first required a minimum of 35 years of social contribution and 57 years of age for employees, or 58 for self-employed. The second demanded at least 39 years of contribution (40 for self-employed). To meet the third criterion, men had to be aged 65 or more, while women 60 or older.

Before 2007, the requirements were lower and had been gradually increased since 1992 (see Table 2 in Battistin et al., 2009, for details). Therefore I need to understand the year in which the individual went into retirement, in order to know the correct rules. For retired individuals, I calculated the age at retirement as the age in which

[2]In the selected sample used for estimates, 6.3% of those self-defined as workers report to have received a work-related pension in the previous year, while among those classified as retired 12.9% report that they have not received a pension. I do not correct their status because this is likely to correct measurement error in R only at some positive distance from eligibility, while leaving the same situation at $S = 0$. The reason is that for someone who retired in the current year we do not know whether s/he is in receipt or not of the pension. If I drop these cases, estimates of the effect of retirement on household production are smaller for women, but still close to 400 minutes/week and statistically significant, while larger (around 120 minutes/week depending on the specification) but not statistically significant at conventional levels for men.

the respondent began the first regular job plus the number of years spent in paid job, plus one.[3] I added one year because it seems that respondents do not report the last year of work if it consisted of less than 12 months. To understand this step, define T as the difference between current age and age at retirement. Among retired, I expect to observe almost nobody with $T < 0$, then a positive jump in the frequency at $T = 0$, and a gradual decrease toward zero for larger T. However, if I do not add one year, there are very few retired with $T = 0$, and the discontinuity is at $T = 1$, implying that almost everybody retired at least one year before. Furthermore, if I build T for workers as well, I would expect the mode of the distribution to be $T = 0$. However, if I do not add one the mode of the distribution is at $T = 1$, with a frequency of only 1.21% at $T = 0$.

With this caveat in mind, I recovered the year of retirement as year of birth plus age at retirement. This strategy clearly introduced measurement error, because of the possible presence of gaps, of the retrospective nature of this information, and of rounding errors.

I then calculated the minimum age at which the individual could have retired (or should retire). For this purpose I used the information on the years of social contribution, available for all respondents who had worked at some point in their life. To distinguish self-employed and employees, I exploited information on the current job for workers and on the last job for retired. For some of the years before 2004, rules were somewhat more favourable to employees in the public sector. However, I have this information only for those currently working. Therefore I use the Statistical Classification of Economic Activities in the European Union (NACE code) for both workers and retired.[4] I define as employees in the public sector those working in "Public administration and defence, compulsory and social security", "Education" or "Health and social work". Among workers in 2007, only the 18.91% of those belonging to these three groups report to work for the private sector, and together these three categories

[3] I also corrected the age at retirement to be equal to the current age for 0.30% (29 obs) of the retired for whom the first was larger than the second.

[4] See http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/NACE_backgrounds for details (last access: 12/07/2012).

accounted for the 84.2% of total public sector employees. One might argue that, given the availability of the public/private information for those currently working, I should use the NACE code only for retired individuals. Given that in 2007 rules for employees are independent from the sector of activity, it would make no difference.

Lastly, time to eligibility is calculated as age at interview minus age at eligibility. Current age is equal to the year of interview minus year of birth. If the quarter of interview is one or more quarters before the quarter of birth, I reduced age by one.[5] Although quarter of birth is available, I keep the minimum unit of measurement equal to one year, because the other variables (age at first job, years in paid employment and years of social contributions) are measured in years and not in quarters. Therefore it does not make much sense to mix the two.[6]

Descriptive data on covariates

[5] Among workers and retired, the quarter of interview is missing for 1.73% of the observations (504). In 310 cases, I replace it with the non-missing quarter of household interview. For the rest of the cases, I assume it took place in the fourth quarter, which is the designed quarter for interviews, and the most likely (92.49% of the cases).

[6] Indeed, doing so I observe peaks in the distribution of time to eligibility around each first quarter of the year.

Table A2: Descriptive statistics of covariates, Men, SILC 2007 (selected sample)

	mean	median	sd	min	max	count
Y	452.6914	300	571.778	0	5400	3970
Bad health	.0639798	0	.2447479	0	1	3970
Disabled	.0403023	0	.196692	0	1	3970
Missing health	.0244332	0	.1544094	0	1	3970
hsize	2.950126	3	1.191653	1	7	3970
S	-1.039295	-2	6.266241	-10	10	3970
Centre	.2433249	0	.4291437	0	1	3970
South	.2647355	0	.4412479	0	1	3970
intermediate area	.4254408	0	.494472	0	1	3970
thinly populated area	.2319899	0	.4221558	0	1	3970
Married	.8183879	1	.3855731	0	1	3970
Separated	.0277078	0	.164155	0	1	3970
Widowed	.0244332	0	.1544094	0	1	3970
Divorced	.0302267	0	.1712321	0	1	3970
College	.1153652	0	.3195024	0	1	3970
High school	.3758186	0	.4843946	0	1	3970
Middle school	.2979849	0	.4574304	0	1	3970
age	56.89421	57	6.128908	44	75	3970
ycontrib	32.5602	33	5.74599	5	50	3970
age first job	18.5204	17	4.959113	8	50	3970
years paid job	32.92393	33.5	7.500483	4	60	3970
employee public	.1916877	0	.3936782	0	1	3970
self-employed	.2843829	0	.4511768	0	1	3970
Observations	3970					

Table A3: Descriptive statistics of covariates, Women, SILC 2007 (selected sample)

	mean	median	sd	min	max	count
Y	1722.239	1500	1032.371	0	6000	2700
Bad health	.062963	0	.2429413	0	1	2700
Disabled	.0362963	0	.187061	0	1	2700
Missing health	.0285185	0	.1664797	0	1	2700
hsize	2.732963	3	1.135579	1	9	2700
S	-1.317037	-3	6.366438	-10	10	2700
Centre	.2655556	0	.4417104	0	1	2700
South	.2274074	0	.4192355	0	1	2700
intermediate area	.4107407	0	.4920594	0	1	2700
thinly populated area	.2244444	0	.4172932	0	1	2700
Married	.7088889	1	.4543588	0	1	2700
Separated	.0288889	0	.1675253	0	1	2700
Widowed	.102963	0	.3039668	0	1	2700
Divorced	.052963	0	.2240011	0	1	2700
College	.1262963	0	.3322445	0	1	2700
High school	.3725926	0	.4835845	0	1	2700
Middle school	.247037	0	.4313684	0	1	2700
age	56.49704	56	6.037335	43	79	2700
ycontrib	28.32444	29	7.374937	2	46	2700
age first job	20.0437	19	6.09985	8	50	2700
years paid job	28.82963	30	8.333891	2	60	2700
employee public	.362963	0	.4809434	0	1	2700
self-employed	.2396296	0	.4269365	0	1	2700
Observations	2700					

S calculated without adding one to the age at retirement

Sample selection follows the same rules discussed in the main text, but with the different definition of S.

Table A4: First stage OLS regressions for retirement status, SILC 2007 (selected sample)

Dep var R	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	0.177**	0.231***	0.360***	0.185*	0.262***	0.462***
	(0.082)	(0.050)	(0.028)	(0.096)	(0.059)	(0.034)
$(1-D) \times S$	0.002	0.023***	0.010***	0.073***	0.023***	0.006***
	(0.022)	(0.007)	(0.001)	(0.026)	(0.008)	(0.002)
$(1-D) \times S^2$	-0.003	0.001**		0.012**	0.002**	
	(0.004)	(0.001)		(0.005)	(0.001)	
$(1-D) \times S^3$	-0.000			0.001**		
	(0.000)			(0.000)		
$D \times S$	0.166***	0.098***	0.049***	0.150**	0.133***	0.050***
	(0.053)	(0.018)	(0.004)	(0.060)	(0.019)	(0.004)
$D \times S^2$	-0.019*	-0.004***		-0.011	-0.007***	
	(0.011)	(0.001)		(0.012)	(0.002)	
$D \times S^3$	0.001			0.000		
	(0.001)			(0.001)		
Constant	0.090***	0.114***	0.086***	0.155***	0.097***	0.061***
	(0.034)	(0.020)	(0.011)	(0.042)	(0.025)	(0.012)
Observations	3864	3864	3864	2639	2639	2639
R^2	0.572	0.571	0.569	0.711	0.710	0.705
$H_0 : \gamma_D = 0$ (p-val)	0.031	0.000	0.000	0.052	0.000	0.000
- (p-val cluster)	0.011	0.003	0.000	0.001	0.000	0.000
Dong's $\widehat{\gamma_D}$	0.093	0.192	0.340	0.143	0.206	0.440
Dong's $\widehat{\gamma_D}$ (p-val)	0.374	0.001	0.000	0.233	0.002	0.000
RESET2	0.343	0.130	0.003	0.295	0.648	0.000
RESET23	0.600	0.220	0.012	0.528	0.866	0.000
G (p-value)	0.012	0.007	0.000	0.180	0.148	0.000
AIC	1762.363	1762.918	1778.439	247.773	246.998	295.813
BIC	1812.439	1800.474	1803.477	294.798	282.267	319.326

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. γ_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A5: Reduced form OLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	-77.544	6.151	20.148	43.711	186.718**	205.334**
	(67.854)	(38.901)	(37.665)	(149.655)	(89.996)	(87.732)
$(1-D) \times S$	5.259	4.084		14.344	8.270	
	(16.487)	(3.505)		(36.669)	(8.382)	
$(1-D) \times S^2$	0.103			0.534		
	(1.391)			(3.157)		
$D \times S$	58.069**	18.438***		91.387*	25.880**	
	(24.249)	(5.421)		(50.478)	(11.562)	
$D \times S^2$	-3.586			-5.928		
	(2.205)			(4.556)		
S			9.790***			15.326**
			(3.021)			(6.839)
Constant	420.104***	417.631***	450.991***	1629.994***	1617.051***	1658.994***
	(41.976)	(23.284)	(20.653)	(93.404)	(55.076)	(46.989)
Observations	3864	3864	3864	2639	2639	2639
R^2	0.018	0.017	0.016	0.036	0.036	0.035
$H_0 : \beta_D = 0$ (p-val)	0.253	0.874	0.593	0.770	0.038	0.019
- (p-val cluster)	0.002	0.804	0.614	0.548	0.014	0.024
Dong's $\widehat{\beta_D}$	-104.564	-1.025		4.113	177.913	
Dong's $\widehat{\beta_D}$ (p-val)	0.155	0.980		0.980	0.053	
RESET2	0.892	0.105	0.051	0.490	0.202	0.255
RESET23	0.954	0.265	0.050	0.593	0.436	0.300
G (p-value)	0.970	0.902	0.646	0.632	0.621	0.574
AIC	59832.011	59831.428	59834.463	43996.781	43994.811	43994.409
BIC	59869.568	59856.466	59853.241	44032.050	44018.324	44012.044

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. β_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A6: OLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	299.850***	289.181***	289.698***	700.562***	670.170***	649.006***
	(30.537)	(29.778)	(28.237)	(78.543)	(74.801)	(70.461)
$(1-D) \times S$	-24.301*	-4.442		-23.928	-2.722	
	(13.281)	(2.868)		(29.477)	(7.087)	
$(1-D) \times S^2$	-1.992*			-2.216		
	(1.177)			(2.671)		
$D \times S$	-3.215	-3.889		-30.886	-17.911*	
	(17.692)	(4.636)		(38.212)	(10.707)	
$D \times S^2$	0.267			1.584		
	(1.708)			(3.606)		
S			-4.212**			-8.664*
			(1.873)			(5.064)
Constant	323.743***	351.121***	352.479***	1500.765***	1525.696***	1490.484***
	(31.627)	(17.812)	(11.668)	(70.531)	(44.039)	(32.194)
Observations	3864	3864	3864	2639	2639	2639
R^2	0.044	0.043	0.043	0.067	0.066	0.066
RESET2	0.963	0.500	0.589	0.994	0.515	0.537
RESET23	0.906	0.239	0.257	0.791	0.237	0.098

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A7: 2SLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	-335.946	17.102	50.606	166.676	403.722**	403.119**
	(318.921)	(107.798)	(93.855)	(560.012)	(188.923)	(167.635)
$(1-D) \times S$	12.982	3.919		10.467	5.834	
	(21.909)	(4.158)		(45.114)	(8.978)	
$(1-D) \times S^2$	0.495			0.281		
	(1.642)			(3.641)		
$D \times S$	91.107*	17.598*		69.263	5.543	
	(53.498)	(9.829)		(114.490)	(19.052)	
$D \times S^2$	-5.083			-4.686		
	(3.455)			(7.791)		
S			8.506			5.732
			(5.197)			(10.385)
Constant	458.334***	416.163***	442.012***	1613.775***	1592.583***	1591.991***
	(71.766)	(29.716)	(36.100)	(133.689)	(62.196)	(72.186)
Observations	3864	3864	3864	2639	2639	2639
$H_0 : \delta = 0$ (p-val)	0.292	0.874	0.590	0.766	0.033	0.016
- (p-val cluster)	0.024	0.790	0.556	0.530	0.001	0.000
First Stage F	21.555	164.057	227.795	19.987	186.255	253.452

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The p-val cluster is calculated by clustering on S.

Table A8: 2SLS regressions for time spent on house work (in minutes per week), including covariates, SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	-440.776	49.824	55.682	329.130	485.788**	445.209**
	(390.047)	(113.734)	(109.722)	(599.743)	(200.020)	(189.224)
$(1-D)\times S$	61.986	20.655*		20.837	17.586	
	(43.465)	(11.950)		(66.827)	(20.867)	
$(1-D)\times S^2$	0.722			-0.414		
	(1.666)			(3.458)		
$D\times S$	130.612*	23.305*		45.178	-4.712	
	(72.581)	(12.650)		(129.433)	(24.993)	
$D\times S^2$	-6.528			-3.504		
	(4.057)			(7.774)		
S			21.852*			7.213
			(11.601)			(21.417)
Constant	2723.253**	1301.370***	1325.253***	2924.743*	2557.524***	2348.534***
	(1202.035)	(466.095)	(456.732)	(1707.340)	(897.942)	(910.978)
Observations	3864	3864	3864	2639	2639	2639
$H_0:\delta=0$ (p-val)	0.258	0.661	0.612	0.583	0.015	0.019
- (p-val cluster)	0.028	0.508	0.479	0.242	0.000	0.000
First Stage F	18.861	180.300	205.149	18.452	173.025	204.188

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The p-val cluster is calculated by clustering on S. Covariates include age, age at first job, years of contributions, years spent in a paid job, plus geographic area, population density, education and employment category dummies. Coefficients are available on request.

Regressions including $S = 0$.

Table A9: First stage OLS regressions for retirement status, SILC 2007 (selected sample)

Dep var R	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	0.201***	0.220***	0.328***	0.130**	0.212***	0.404***
	(0.048)	(0.034)	(0.023)	(0.062)	(0.045)	(0.030)
$(1-D) \times S$	-0.004	0.021***	0.011***	0.081***	0.034***	0.008***
	(0.024)	(0.007)	(0.001)	(0.028)	(0.009)	(0.002)
$(1-D) \times S^2$	-0.004	0.001		0.012**	0.002***	
	(0.005)	(0.001)		(0.005)	(0.001)	
$(1-D) \times S^3$	-0.000			0.001**		
	(0.000)			(0.000)		
$D \times S$	0.186***	0.111***	0.054***	0.188***	0.146***	0.057***
	(0.027)	(0.012)	(0.003)	(0.034)	(0.014)	(0.004)
$D \times S^2$	-0.025***	-0.006***		-0.019**	-0.009***	
	(0.006)	(0.001)		(0.008)	(0.001)	
$D \times S^3$	0.001***			0.001		
	(0.000)			(0.000)		
Constant	0.089**	0.117***	0.097***	0.188***	0.134***	0.078***
	(0.036)	(0.021)	(0.012)	(0.046)	(0.028)	(0.014)
Observations	4139	4139	4139	2795	2795	2795
R^2	0.549	0.547	0.541	0.680	0.679	0.667
$H_0 : \gamma_D = 0$ (p-val)	0.000	0.000	0.000	0.036	0.000	0.000
- (p-val cluster)	0.000	0.000	0.000	0.000	0.000	0.000
Dong's $\widehat{\gamma_D}$	0.102	0.174	0.306	0.072	0.154	0.379
Dong's $\widehat{\gamma_D}$ (p-val)	0.042	0.000	0.000	0.281	0.001	0.000
RESET2	0.053	0.001	0.000	0.438	0.121	0.000
RESET23	0.102	0.001	0.000	0.106	0.238	0.000
G (p-value)	0.005	0.000	0.000	0.165	0.056	0.000
AIC	2212.864	2227.963	2271.729	601.985	604.777	702.481
BIC	2263.490	2265.932	2297.042	649.470	640.391	726.224

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. γ_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A10: Reduced form OLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	20.185	27.143	23.268	68.915	187.855**	187.323**
	(51.819)	(34.058)	(34.423)	(122.670)	(80.707)	(80.818)
$(1-D) \times S$	-0.456	3.916		21.163	8.689	
	(15.671)	(3.411)		(36.312)	(8.334)	
$(1-D) \times S^2$	-0.384			1.097		
	(1.341)			(3.139)		
$D \times S$	26.847	16.217***		84.863**	24.513**	
	(17.298)	(4.797)		(38.357)	(10.088)	
$D \times S^2$	-1.064			-5.957		
	(1.713)			(3.689)		
S			9.907***			16.148**
			(2.921)			(6.485)
Constant	408.109***	417.312***	452.286***	1648.099***	1621.634***	1665.853***
	(38.963)	(22.567)	(19.968)	(91.995)	(54.467)	(44.926)
Observations	4139	4139	4139	2795	2795	2795
R^2	0.017	0.016	0.015	0.035	0.033	0.033
$H_0 : \beta_D = 0$ (p-val)	0.697	0.426	0.499	0.574	0.020	0.021
- (p-val cluster)	0.539	0.299	0.513	0.240	0.006	0.019
Dong's $\widehat{\beta_D}$	6.420	20.993		35.890	179.943	
Dong's $\widehat{\beta_D}$ (p-val)	0.903	0.544		0.773	0.027	
RESET2	0.360	0.527	0.060	0.888	0.116	0.283
RESET23	0.655	0.636	0.127	0.949	0.256	0.172
G (p-value)	0.174	0.249	0.128	0.797	0.696	0.656
AIC	64253.624	64250.188	64252.561	46647.195	46646.436	46645.947
BIC	64291.593	64275.501	64271.545	46682.808	46670.179	46663.754

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. β_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A11: OLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	298.222***	285.657***	287.267***	677.528***	657.555***	638.151***
	(29.401)	(28.573)	(27.504)	(75.147)	(71.879)	(68.582)
$(1-D)\times S$	-25.122*	-4.853*		-24.468	-2.258	
	(13.060)	(2.864)		(29.507)	(7.101)	
$(1-D)\times S^2$	-2.067*			-2.199		
	(1.167)			(2.674)		
$D\times S$	-4.132	-2.739		-18.699	-18.076*	
	(16.948)	(4.576)		(37.792)	(10.316)	
$D\times S^2$	0.478			0.346		
	(1.670)			(3.561)		
S			-3.951**			-8.582*
			(1.895)			(4.996)
Constant	322.211***	348.290***	353.577***	1495.149***	1528.237***	1490.893***
	(30.527)	(17.744)	(11.739)	(70.572)	(44.056)	(31.963)
Observations	3970	3970	3970	2700	2700	2700
R^2	0.044	0.043	0.043	0.065	0.065	0.065
RESET2	0.853	0.581	0.489	0.641	0.439	0.569
RESET23	0.859	0.134	0.205	0.582	0.371	0.171

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A12: 2SLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	91.628	82.835	74.059	325.364	465.228**	465.823**
	(232.654)	(102.845)	(108.443)	(564.692)	(194.065)	(194.940)
$(1-D)\times S$	-2.339	3.022		10.032	5.021	
	(19.175)	(4.150)		(51.094)	(9.243)	
$(1-D)\times S^2$	-0.463			0.344		
	(1.466)			(4.039)		
$D\times S$	16.698	11.773		37.346	-2.101	
	(37.779)	(9.205)		(107.193)	(18.738)	
$D\times S^2$	-0.540			-3.104		
	(2.578)			(7.309)		
S			7.562			1.645
			(6.090)			(11.984)
Constant	397.356***	409.293***	436.093***	1604.525***	1585.314***	1565.284***
	(61.876)	(29.966)	(42.226)	(155.775)	(65.026)	(83.764)
Observations	4139	4139	4139	2795	2795	2795
$H_0: \delta = 0$ (p-val)	0.694	0.421	0.495	0.564	0.017	0.017
- (p-val cluster)	0.538	0.264	0.435	0.236	0.000	0.000
First Stage F	41.282	196.389	160.522	22.223	186.016	166.154

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The p-val cluster is calculated by clustering on S.

Table A13: 2SLS regressions for time spent on house work (in minutes per week), including covariates, SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	85.259	114.716	112.074	593.785	540.507***	549.978***
	(260.648)	(106.020)	(108.330)	(566.983)	(199.277)	(201.375)
$(1-D) \times S$	10.480	11.207		3.799	20.359	
	(31.830)	(11.011)		(72.480)	(22.051)	
$(1-D) \times S^2$	-0.234			-1.207		
	(1.480)			(3.740)		
$D \times S$	25.712	13.763		-6.089	-7.747	
	(47.058)	(11.474)		(117.623)	(25.087)	
$D \times S^2$	-0.979			-0.586		
	(2.875)			(7.049)		
S			13.180			1.181
			(10.951)			(23.215)
Constant	984.105	909.302**	959.666**	2558.757	2720.757***	2250.832**
	(754.031)	(426.612)	(416.832)	(1731.577)	(953.871)	(981.538)
Observations	4139	4139	4139	2795	2795	2795
$H_0 : \delta = 0$ (p-val)	0.744	0.279	0.301	0.295	0.007	0.006
- (p-val cluster)	0.616	0.133	0.169	0.039	0.000	0.000
First Stage F	40.647	219.010	203.869	23.755	183.624	172.766

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The p-val cluster is calculated by clustering on S. Covariates include age, age at first job, years of contributions, years spent in a paid job, plus geographic area, population density, education and employment category dummies. Coefficients are available on request.

Including proxy interviews

Table A14: First stage OLS regressions for retirement status, SILC 2007 (selected sample)

Dep var R	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	0.284***	0.287***	0.363***	0.110	0.244***	0.474***
	(0.076)	(0.045)	(0.025)	(0.094)	(0.057)	(0.032)
$(1-D)\times S$	-0.000	0.018***	0.010***	0.081***	0.034***	0.008***
	(0.022)	(0.006)	(0.001)	(0.026)	(0.009)	(0.002)
$(1-D)\times S^2$	-0.003	0.001		0.012**	0.002***	
	(0.004)	(0.000)		(0.005)	(0.001)	
$(1-D)\times S^3$	-0.000			0.001**		
	(0.000)			(0.000)		
$D\times S$	0.098**	0.077***	0.047***	0.205***	0.134***	0.048***
	(0.049)	(0.016)	(0.003)	(0.057)	(0.018)	(0.004)
$D\times S^2$	-0.007	-0.003**		-0.023**	-0.008***	
	(0.010)	(0.001)		(0.011)	(0.001)	
$D\times S^3$	0.000			0.001		
	(0.001)			(0.001)		
Constant	0.085***	0.106***	0.090***	0.185***	0.130***	0.075***
	(0.033)	(0.019)	(0.010)	(0.043)	(0.026)	(0.013)
Observations	4650	4650	4650	2922	2922	2922
R^2	0.555	0.555	0.554	0.705	0.704	0.698
$H_0: \gamma_D = 0$ (p-val)	0.000	0.000	0.000	0.240	0.000	0.000
- (p-val cluster)	0.001	0.000	0.000	0.154	0.003	0.000
Dong's $\widehat{\gamma_D}$	0.234	0.256	0.344	0.042	0.192	0.454
Dong's $\widehat{\gamma_D}$ (p-val)	0.014	0.000	0.000	0.716	0.002	0.000
RESET2	0.237	0.567	0.048	0.016	0.091	0.000
RESET23	0.364	0.321	0.121	0.056	0.205	0.000
G (p-value)	0.007	0.014	0.003	0.073	0.019	0.000
AIC	2312.911	2309.807	2314.167	385.872	389.238	449.781
BIC	2364.468	2348.475	2339.945	433.713	425.118	473.701

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. γ_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A15: Reduced form OLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	16.450	30.202	42.434	14.803	212.423**	220.639***
	(58.374)	(34.302)	(33.154)	(143.613)	(85.352)	(83.025)
$(1-D) \times S$	-9.134	1.330		33.295	11.282	
	(14.570)	(3.096)		(34.677)	(7.914)	
$(1-D) \times S^2$	-0.918			1.936		
	(1.243)			(2.997)		
$D \times S$	34.132	16.286***		94.321*	19.394*	
	(21.150)	(4.969)		(48.519)	(11.010)	
$D \times S^2$	-1.625			-6.771		
	(1.947)			(4.341)		
S			7.442***			14.539**
			(2.737)			(6.479)
Constant	354.206***	376.362***	412.001***	1665.457***	1618.747***	1638.081***
	(35.635)	(20.073)	(18.342)	(87.908)	(51.791)	(44.313)
Observations	4650	4650	4650	2922	2922	2922
R^2	0.015	0.015	0.014	0.038	0.036	0.036
$H_0 : \beta_D = 0$ (p-val)	0.778	0.379	0.201	0.918	0.013	0.008
- (p-val cluster)	0.714	0.315	0.298	0.832	0.018	0.014
Dong's $\widehat{\beta_D}$	-5.301	22.724		-17.162	208.367	
Dong's $\widehat{\beta_D}$ (p-val)	0.933	0.520		0.912	0.017	
RESET2	0.893	0.401	0.020	0.422	0.209	0.593
RESET23	0.772	0.703	0.062	0.662	0.254	0.259
G (p-value)	0.499	0.554	0.216	0.646	0.546	0.589
AIC	72006.299	72003.551	72008.198	48692.412	48691.628	48690.004
BIC	72044.967	72029.329	72027.532	48728.292	48715.548	48707.944

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. β_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A16: OLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	303.760***	292.926***	294.944***	666.409***	648.092***	623.117***
	(27.456)	(26.819)	(25.737)	(72.063)	(68.932)	(65.728)
$(1-D)\times S$	-25.973**	-6.166**		-12.157	0.907	
	(12.046)	(2.631)		(28.317)	(6.772)	
$(1-D)\times S^2$	-2.011*			-1.360		
	(1.071)			(2.561)		
$D\times S$	-4.269	-3.797		-26.544	-19.250*	
	(15.196)	(4.167)		(36.005)	(9.909)	
$D\times S^2$	0.390			0.907		
	(1.506)			(3.393)		
S			-5.168***			-7.112
			(1.750)			(4.773)
Constant	290.534***	316.652***	322.506***	1514.987***	1530.632***	1483.152***
	(27.814)	(15.835)	(10.418)	(67.954)	(42.143)	(30.575)
Observations	4650	4650	4650	2922	2922	2922
R^2	0.044	0.043	0.043	0.066	0.066	0.065
RESET2	0.534	0.923	0.793	0.539	0.274	0.419
RESET23	0.782	0.067	0.116	0.463	0.324	0.117

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table A17: 2SLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	57.410	83.142	107.703	60.603	448.289**	428.896***
	(201.716)	(93.251)	(82.896)	(583.605)	(174.863)	(157.232)
$(1-D)\times S$	-10.160	0.511		31.254	7.853	
	(16.797)	(3.644)		(48.343)	(8.625)	
$(1-D)\times S^2$	-0.958			1.797		
	(1.317)			(3.830)		
$D\times S$	29.704	12.344		86.183	-1.997	
	(33.859)	(8.591)		(117.636)	(17.516)	
$D\times S^2$	-1.470			-6.297		
	(2.319)			(8.030)		
S			4.728			4.359
			(4.622)			(9.793)
Constant	348.092***	368.918***	392.717***	1657.571***	1585.172***	1565.001***
	(51.488)	(25.755)	(31.976)	(146.912)	(60.040)	(68.552)
Observations	4650	4650	4650	2922	2922	2922
$H_0:\delta=0$ (p-val)	0.776	0.373	0.194	0.917	0.010	0.006
- (p-val cluster)	0.687	0.246	0.177	0.822	0.001	0.000
First Stage F	41.076	206.426	269.863	18.434	217.315	283.785

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The p-val cluster is calculated by clustering on S.

Table A18: 2SLS regressions for time spent on house work (in minutes per week), including covariates, SILC 2007 (selected sample)

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
R	27.954	109.649	121.905	373.599	542.957***	488.221***
	(213.835)	(93.954)	(90.498)	(603.482)	(178.946)	(170.072)
$(1-D)\times S$	6.671	8.869		31.916	21.604	
	(26.425)	(9.791)		(69.263)	(20.076)	
$(1-D)\times S^2$	-0.647			0.152		
	(1.321)			(3.577)		
$D\times S$	43.197	15.469		37.547	-9.968	
	(39.290)	(10.412)		(132.064)	(23.049)	
$D\times S^2$	-2.034			-3.257		
	(2.428)			(8.072)		
S			12.195			6.673
			(9.450)			(20.207)
Constant	1065.282*	863.466**	936.615**	3119.749*	2717.378***	2414.561***
	(641.234)	(390.944)	(380.613)	(1707.816)	(887.207)	(892.189)
Observations	4650	4650	4650	2922	2922	2922
$H_0:\delta=0$ (p-val)	0.896	0.243	0.178	0.536	0.002	0.004
- (p-val cluster)	0.838	0.104	0.079	0.226	0.000	0.001
First Stage F	42.991	240.170	269.838	17.862	213.232	247.816

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The p-val cluster is calculated by clustering on S. Covariates include age, age at first job, years of contributions, years spent in a paid job, plus geographic area, population density, education and employment category dummies. Coefficients are available on request.

Main results including covariates

Table A19: First stage OLS regressions for retirement status, SILC 2007 (selected sample), including covariates

Dep var R	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	0.261***	0.293***	0.392***	0.071	0.216***	0.444***
	(0.074)	(0.044)	(0.025)	(0.091)	(0.055)	(0.032)
$(1-D) \times S$	0.038*	0.069***	0.058***	0.119***	0.072***	0.052***
	(0.022)	(0.007)	(0.003)	(0.025)	(0.009)	(0.005)
$(1-D) \times S^2$	-0.006	0.001*		0.012**	0.002***	
	(0.004)	(0.001)		(0.005)	(0.001)	
$(1-D) \times S^3$	-0.000*			0.001**		
	(0.000)			(0.000)		
$D \times S$	0.157***	0.098***	0.060***	0.241***	0.161***	0.069***
	(0.047)	(0.015)	(0.003)	(0.055)	(0.018)	(0.005)
$D \times S^2$	-0.016*	-0.003***		-0.026**	-0.008***	
	(0.009)	(0.001)		(0.011)	(0.001)	
$D \times S^3$	0.001			0.001*		
	(0.001)			(0.001)		
Constant	2.379***	2.412***	2.370***	2.488***	2.436***	2.402***
	(0.143)	(0.142)	(0.141)	(0.228)	(0.227)	(0.227)
Observations	3970	3970	3970	2700	2700	2700
R^2	0.646	0.645	0.644	0.740	0.739	0.732
$H_0 : \gamma_D = 0$ (p-val)	0.000	0.000	0.000	0.432	0.000	0.000
- (p-val cluster)	0.001	0.000	0.000	0.303	0.004	0.000
Dong's $\widehat{\gamma_D}$	0.200	0.278	0.392	0.004	0.170	0.435
Dong's $\widehat{\gamma_D}$ (p-val)	0.033	0.000	0.000	0.974	0.006	0.000
RESET2	0.000	0.000	0.000	0.971	0.823	0.413
RESET23	0.000	0.000	0.000	0.000	0.000	0.000
G (p-value)	0.025	0.012	0.000	0.112	0.019	0.000
AIC	1203.924	1205.079	1215.765	51.610	56.576	124.663
BIC	1335.941	1324.523	1322.636	175.531	168.695	224.980

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. γ_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes. Covariates include a constant, age, age at first job, years of contributions, years spent in a paid job, plus geographic area, population density, education and employment category dummies. Coefficients are available on request.

Table A20: Reduced form OLS regressions for time spent on house work (in minutes per week), SILC 2007 (selected sample), including covariates

Dep var Y	(1) Men	(2) Men	(3) Men	(4) Women	(5) Women	(6) Women
D	-28.388	31.900	35.064	83.966	234.174***	220.917***
	(63.825)	(38.019)	(37.068)	(145.732)	(86.691)	(84.711)
S			19.931***			38.004***
			(5.612)			(13.291)
$(1-D)\times S$	15.521	16.568**		47.868	47.823***	
	(16.981)	(7.677)		(37.428)	(15.894)	
$(1-D)\times S^2$	-0.122			0.003		
	(1.328)			(3.050)		
$D\times S$	53.367**	21.904***		105.685**	30.847**	
	(23.575)	(6.179)		(50.302)	(14.697)	
$D\times S^2$	-2.842			-6.777		
	(2.142)			(4.347)		
Constant	1120.907***	1106.114***	1191.735***	4037.526***	4041.642***	3805.963***
	(344.016)	(341.921)	(292.806)	(760.950)	(760.473)	(719.864)
Observations	3970	3970	3970	2700	2700	2700
R^2	0.036	0.036	0.036	0.104	0.103	0.103
$H_0:\beta_D=0$ (p-val)	0.656	0.401	0.344	0.565	0.007	0.009
- (p-val cluster)	0.505	0.351	0.301	0.209	0.004	0.002
Dong's $\widehat{\beta_D}$	-47.765	29.232		53.927	242.663	
Dong's $\widehat{\beta_D}$ (p-val)	0.493	0.456		0.731	0.006	
RESET2	0.618	0.915	0.752	0.834	0.866	0.901
RESET23	0.015	0.003	0.002	0.063	0.032	0.034
G (p-value)	0.157	0.163	0.184	0.838	0.766	0.733
AIC	61566.021	61564.252	61562.835	44875.678	44874.482	44873.875
BIC	61685.465	61671.123	61663.419	44987.797	44974.799	44968.292

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. β_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction, while the p-val cluster is calculated by clustering on S. AIC is the Akaike criterion; BIC is the Bayesian criterion; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes. Covariates include a constant, age, age at first job, years of contributions, years spent in a paid job, plus geographic area, population density, education and employment category dummies. Coefficients are available on request.

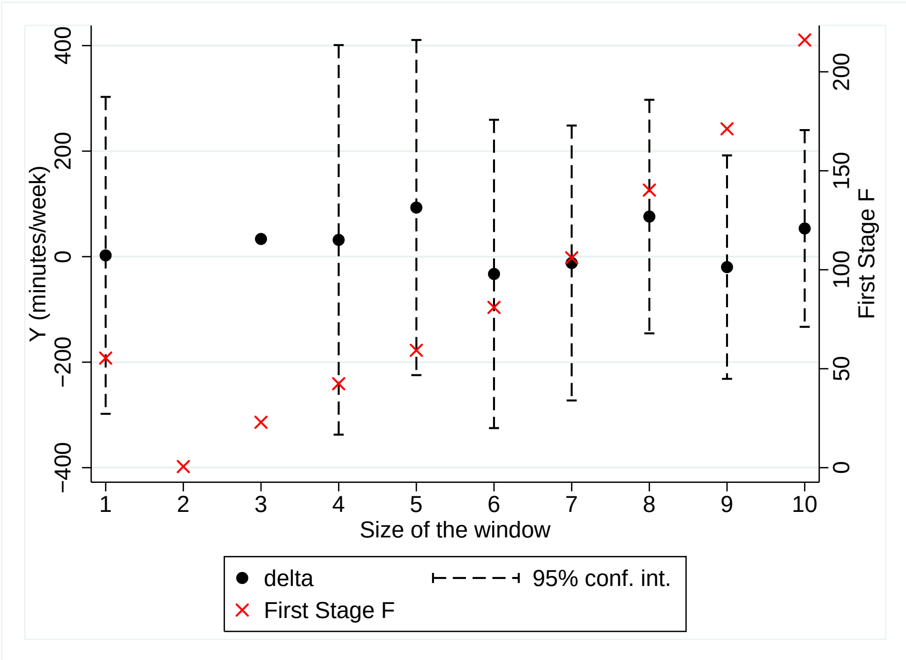
Table A21: 2SLS estimates splitting the sample by education, area, employment category and marital status, with other covariates, SILC 2007 (selected sample)

		Men				Women			
		$\widehat{\delta}$	p-value	First stage F	obs	$\widehat{\delta}$	p-value	First stage F	obs
By education:	Middle school or less	127.483	0.331	162	2020	676.387	0.008	133	1353
	High school	-4.075	0.978	89	1492	513.965	0.066	88	1006
	College	116.972	0.698	18	458	-865.023	0.260	10	341
By area:	North	181.041	0.103	185	1953	633.643	0.005	133	1369
	Centre	84.752	0.640	60	966	395.953	0.266	59	717
	South	-138.846	0.630	39	1051	198.879	0.654	41	614
By degree of urbanization:	Densely populated	234.433	0.086	100	1360	680.701	0.004	89	985
	Intermediate area	38.323	0.804	110	1689	599.091	0.048	99	1109
	Thinly populated	-73.597	0.724	57	921	-244.665	0.598	38	606
By category:	Private employee	120.494	0.221	261	2080	627.305	0.005	182	1073
	Public employee	83.872	0.796	25	761	273.831	0.403	64	980
	Self-employed	26.534	0.927	24	1129	754.643	0.141	26	647
By marital status:	Married, living with partner	10.753	0.910	239	3132	410.661	0.032	213	1827
	Not married, living with partner	634.506	0.299	5	105	-1550.56	0.449	4	55
	Not living with partner	552.503	0.066	30	733	968.210	0.036	29	818

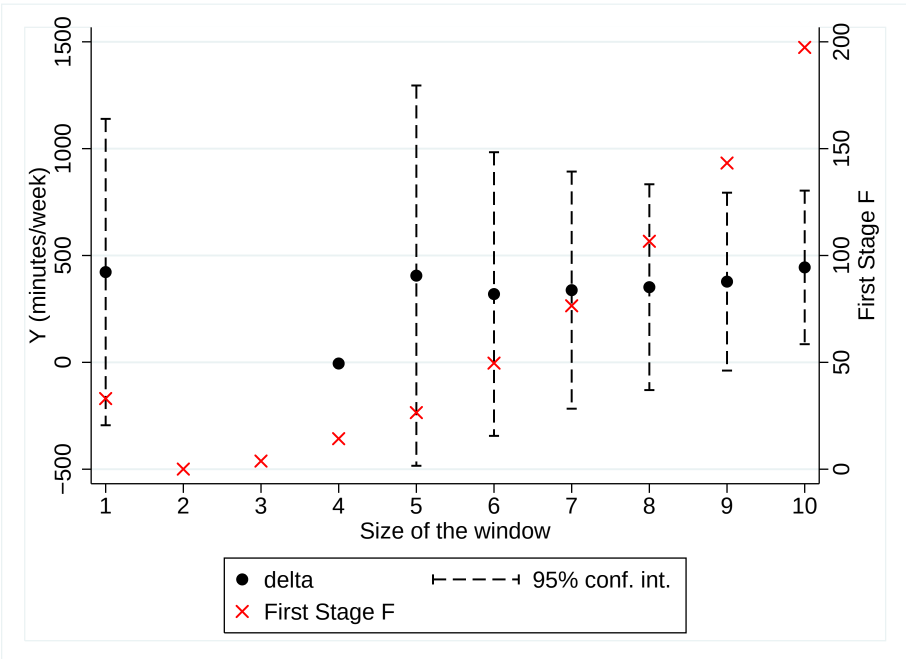
Note: all estimates include S plus other covariates, while R is instrumented by D. The p-value and the first stage F are calculated using robust standard errors.

Choice of window size

Figure A1: 2SLS estimates for different windows (for windows $|S| \in [2, 10]$, regressions include $D \times S$ and $(1 - D) \times S$; for $|S| = 1$ they are a Wald estimator with no covariates; when confidence interval or estimates are not shown they are larger than the graph interval).



(a) Men

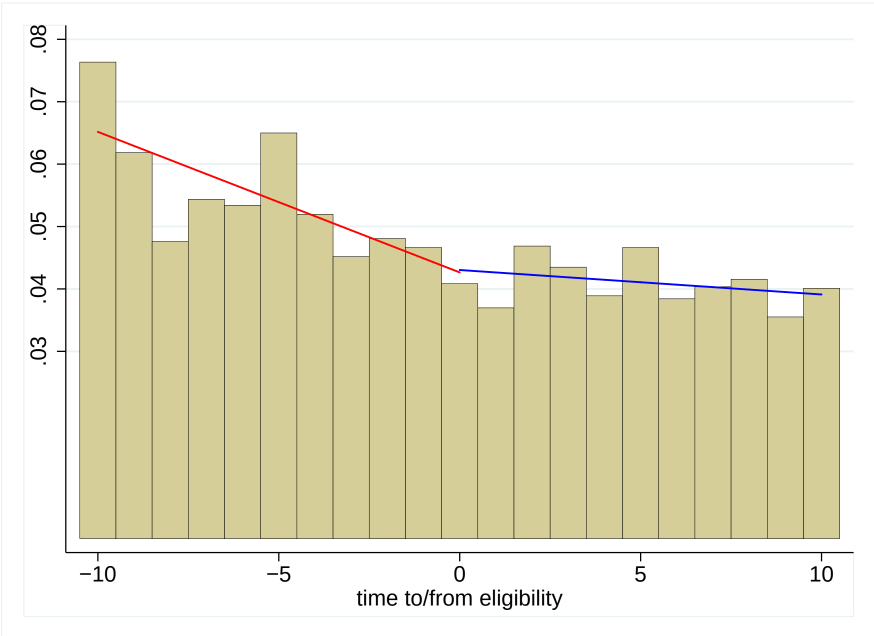


(b) Women

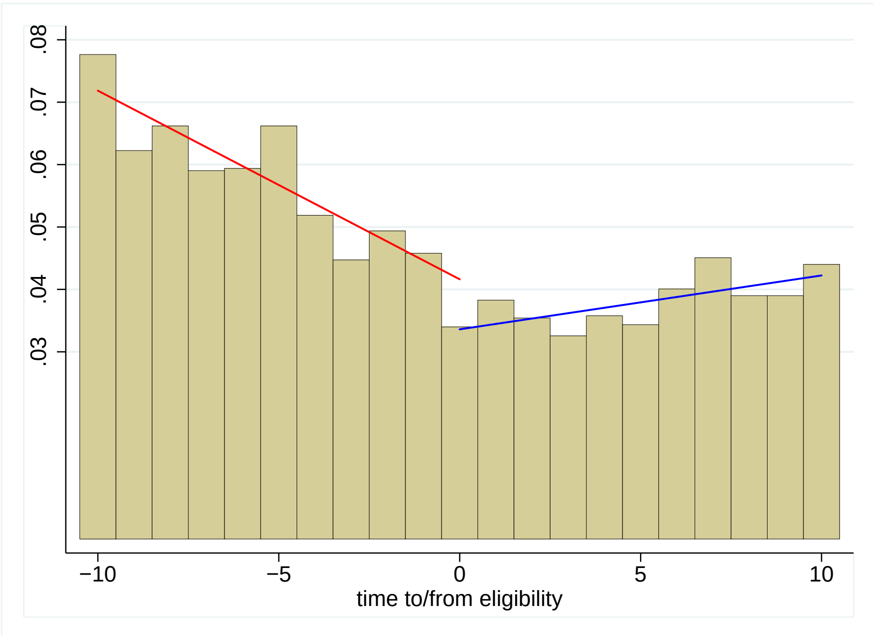
Appendix B: density plots and discontinuities in baseline covariates

Density plots

Figure B1: Density, SILC 2007 (selected sample)



(a) Men



(b) Women

Discontinuities for men

Table B1: Regressions for different socio-economic variables, men, SILC 2007 (selected sample)

	(1) North	(2) Centre	(3) South	(4) Densely pop area	(5) Interm. area	(6) Thinly pop area
D	-0.046	0.059*	-0.014	-0.030	0.026	0.004
	(0.035)	(0.030)	(0.030)	(0.034)	(0.035)	(0.029)
Observations	3970	3970	3970	3970	3970	3970
R^2	0.003	0.001	0.002	0.001	0.001	0.004
$H_0 : \gamma_D = 0$ (p-val)	0.196	0.052	0.656	0.368	0.458	0.881
G (p-value)	0.249	0.508	0.042	0.680	0.048	0.229
RESET2	0.591	0.291	0.699	0.258	0.524	0.786
RESET23	0.157	0.368	0.909	0.496	0.519	0.964

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. The regressions include $(1-D)\times S$, $D \times S$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table B2: Regressions for different socio-economic variables, men, SILC 2007 (selected sample)

	(1) College	(2) High school	(3) Middle sch.	(4) Primary sch.	(5) Y. of schooling	(6) Age highest edu - 6
D	-0.015	0.084**	-0.065**	-0.004	0.282	-1.180
	(0.023)	(0.034)	(0.032)	(0.029)	(0.282)	(0.721)
Observations	3970	3970	3970	3970	3970	3970
R^2	0.002	0.033	0.005	0.094	0.052	0.002
$H_0 : \gamma_D = 0$ (p-val)	0.522	0.013	0.041	0.889	0.318	0.102
G (p-value)	0.627	0.051	0.048	0.460	0.769	0.318
RESET2	0.273	0.369	0.593	0.595	0.528	0.463
RESET23	0.188	0.073	0.813	0.413	0.606	0.762

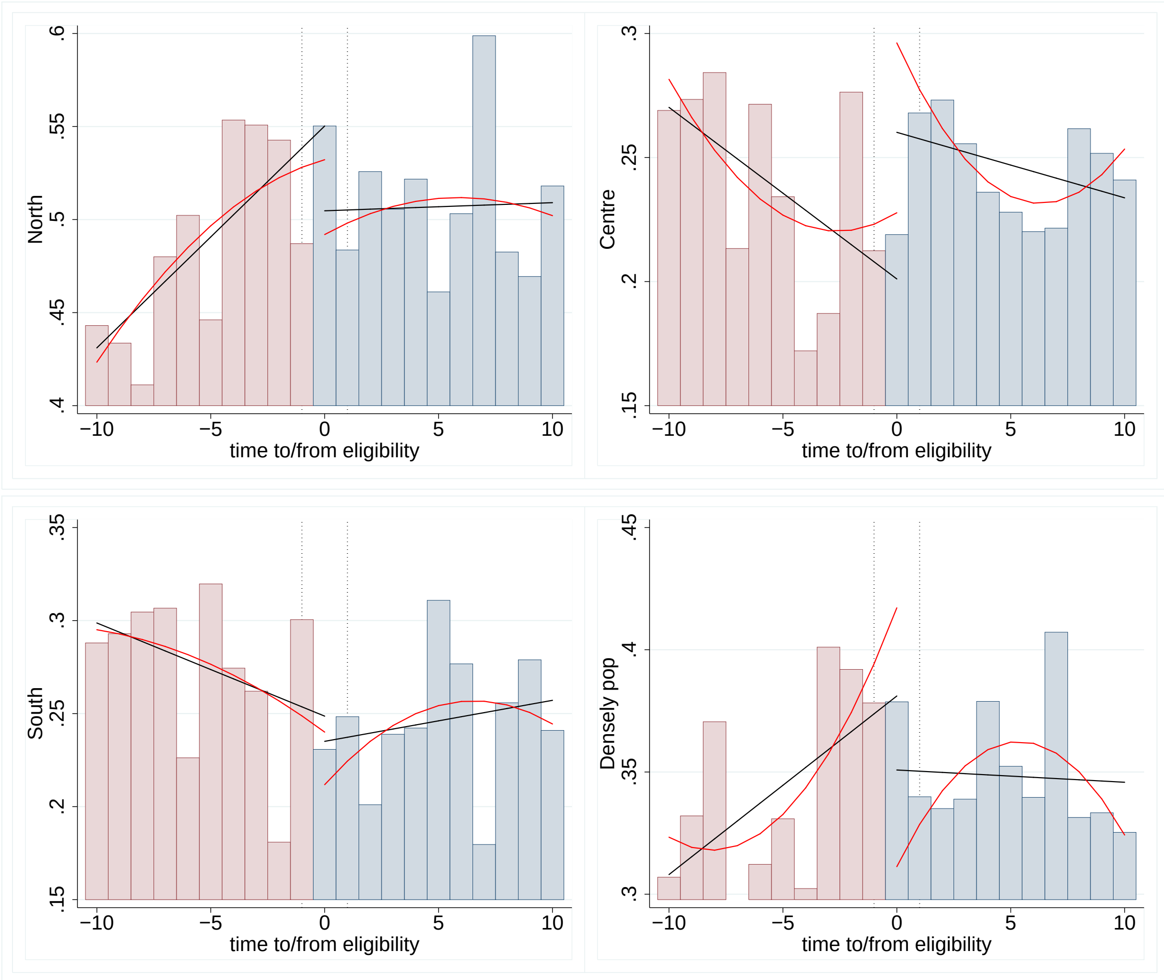
* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. The regressions include $(1-D)\times S$, $D \times S$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table B3: Regressions for different socio-economic variables, men, SILC 2007 (selected sample)

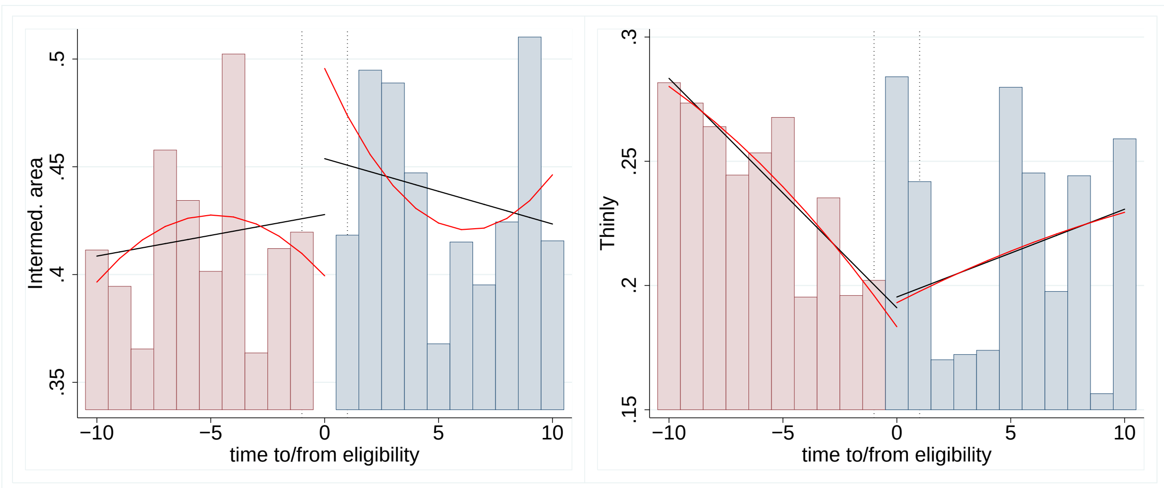
	(1) Age	(2) Y. contrib	(3) Age at retirement	(4) Age at first job	(5) Years in paid job	(6) Private	(7) Public	(8) Self-empl.
D	-0.123	0.175	0.064	0.060	0.976**	-0.013	-0.055**	0.068**
	(0.227)	(0.326)	(0.295)	(0.342)	(0.412)	(0.035)	(0.027)	(0.032)
Observations	3970	3970	3970	3970	3970	3970	3970	3970
R^2	0.708	0.398	0.339	0.004	0.385	0.008	0.009	0.002
$H_0 : \gamma_D = 0$	0.589	0.592	0.828	0.860	0.018	0.704	0.044	0.031
G (p-value)	0.036	0.284	0.064	0.798	0.304	0.020	0.469	0.005
RESET2	0.210	0.932	0.646	0.445	0.068	0.567	0.468	0.422
RESET23	0.087	0.803	0.899	0.353	0.098	0.772	0.403	0.242

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. The regressions include $(1-D)\times S$, $D \times S$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Figure B2: Discontinuities in other variables with respect to distance to/from eligibility (in years), men, SILC 2007 (selected sample)

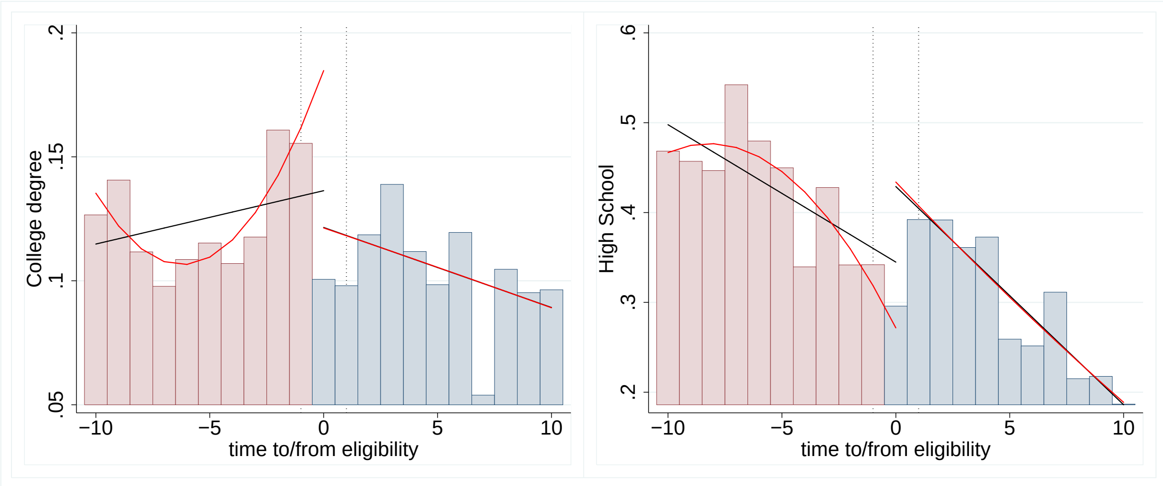


(a)

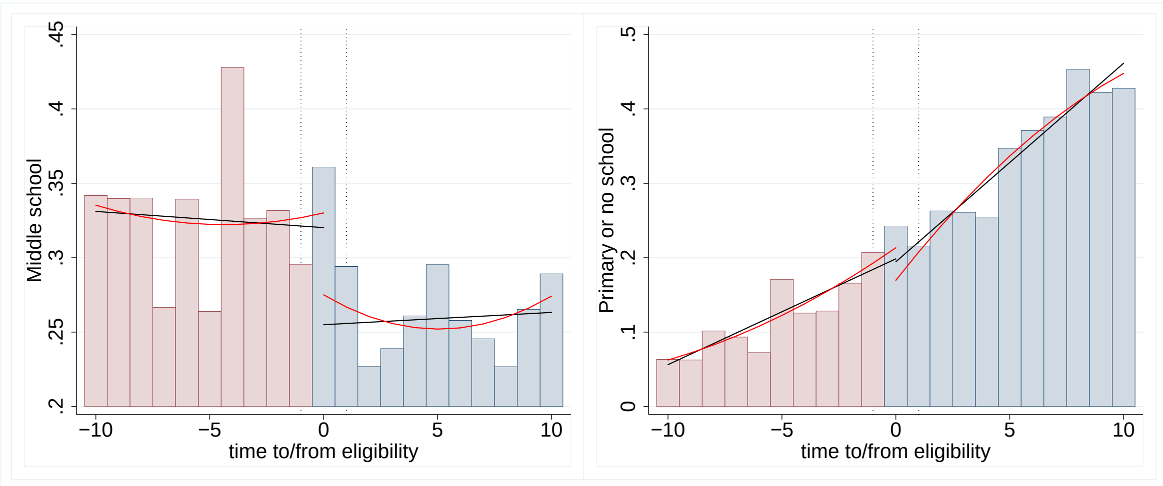


(b)

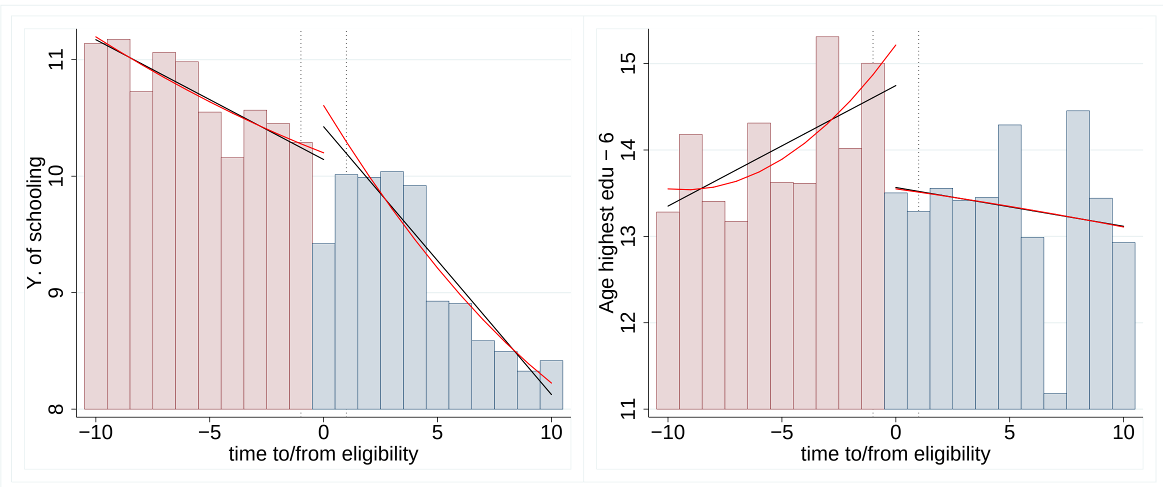
Figure B3: Discontinuities in other variables with respect to distance to/from eligibility (in years), men, SILC 2007 (selected sample)



(a)

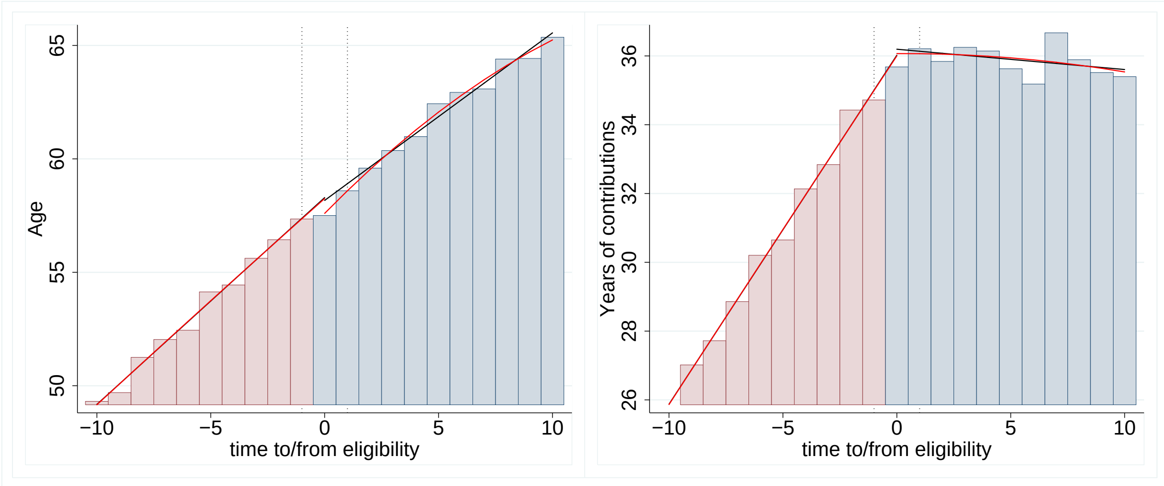


(b)

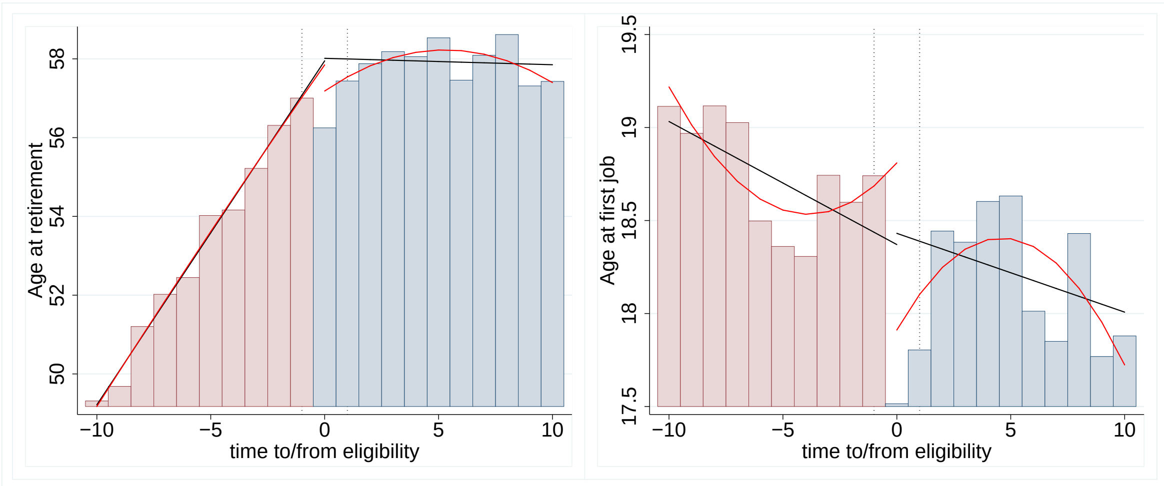


(c)

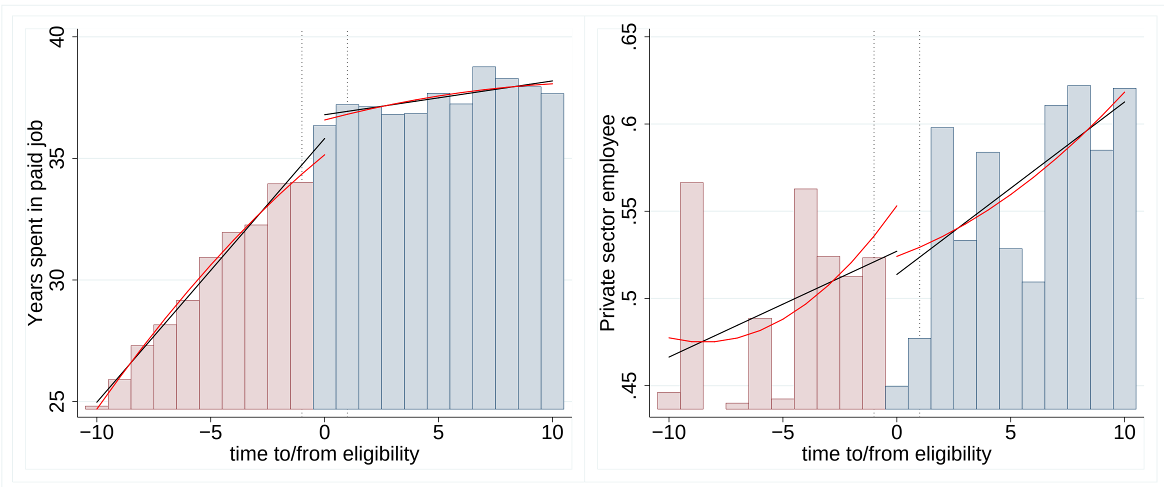
Figure B4: Discontinuities in other variables with respect to distance to/from eligibility (in years), men, SILC 2007 (selected sample)



(a)

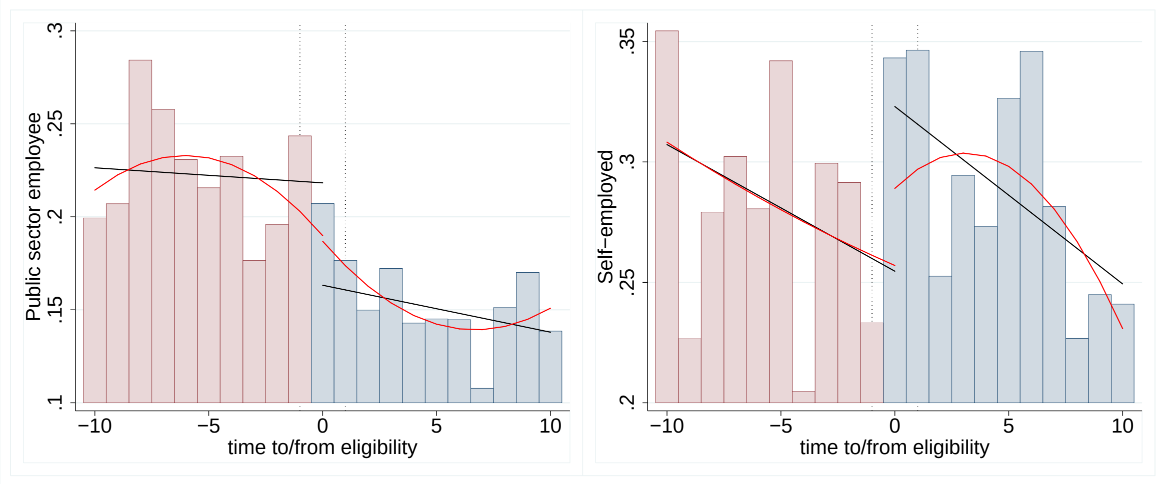


(b)



(c)

Figure B5: Discontinuities in other variables with respect to distance to/from eligibility (in years), men, SILC 2007 (selected sample)



(a)

Discontinuities for women

Table B4: Regressions for different socio-economic variables, women, SILC 2007 (selected sample)

	(1) North	(2) Centre	(3) South	(4) Densely pop area	(5) Interm. area	(6) Thinly pop area
D	-0.035	0.013	0.022	0.087**	-0.068	-0.019
	(0.044)	(0.040)	(0.037)	(0.043)	(0.044)	(0.036)
Observations	2700	2700	2700	2700	2700	2700
R^2	0.002	0.000	0.001	0.002	0.001	0.001
$H_0 : \gamma_D = 0$ (p-val)	0.430	0.743	0.561	0.042	0.120	0.590
G (p-value)	0.975	0.060	0.373	0.284	0.790	0.279
RESET2	0.748	0.735	0.591	0.349	0.089	0.034
RESET23	0.329	0.897	0.845	0.305	0.235	0.099

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. The regressions include $(1 - D) \times S$, $D \times S$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table B5: Regressions for different socio-economic variables, women, SILC 2007 (selected sample)

	(1) College	(2) High school	(3) Middle sch.	(4) Primary sch.	(5) Y. of schooling	(6) Age highest edu - 6
D	0.016	0.005	0.004	-0.025	0.508	0.736
	(0.031)	(0.042)	(0.037)	(0.038)	(0.371)	(0.814)
Observations	2700	2700	2700	2700	2700	2700
R^2	0.013	0.059	0.006	0.145	0.115	0.012
$H_0 : \gamma_D = 0$ (p-val)	0.608	0.904	0.921	0.523	0.170	0.366
G (p-value)	0.893	0.318	0.970	0.370	0.828	0.007
RESET2	0.507	0.579	0.882	0.842	0.976	0.183
RESET23	0.453	0.171	0.830	0.185	0.999	0.037

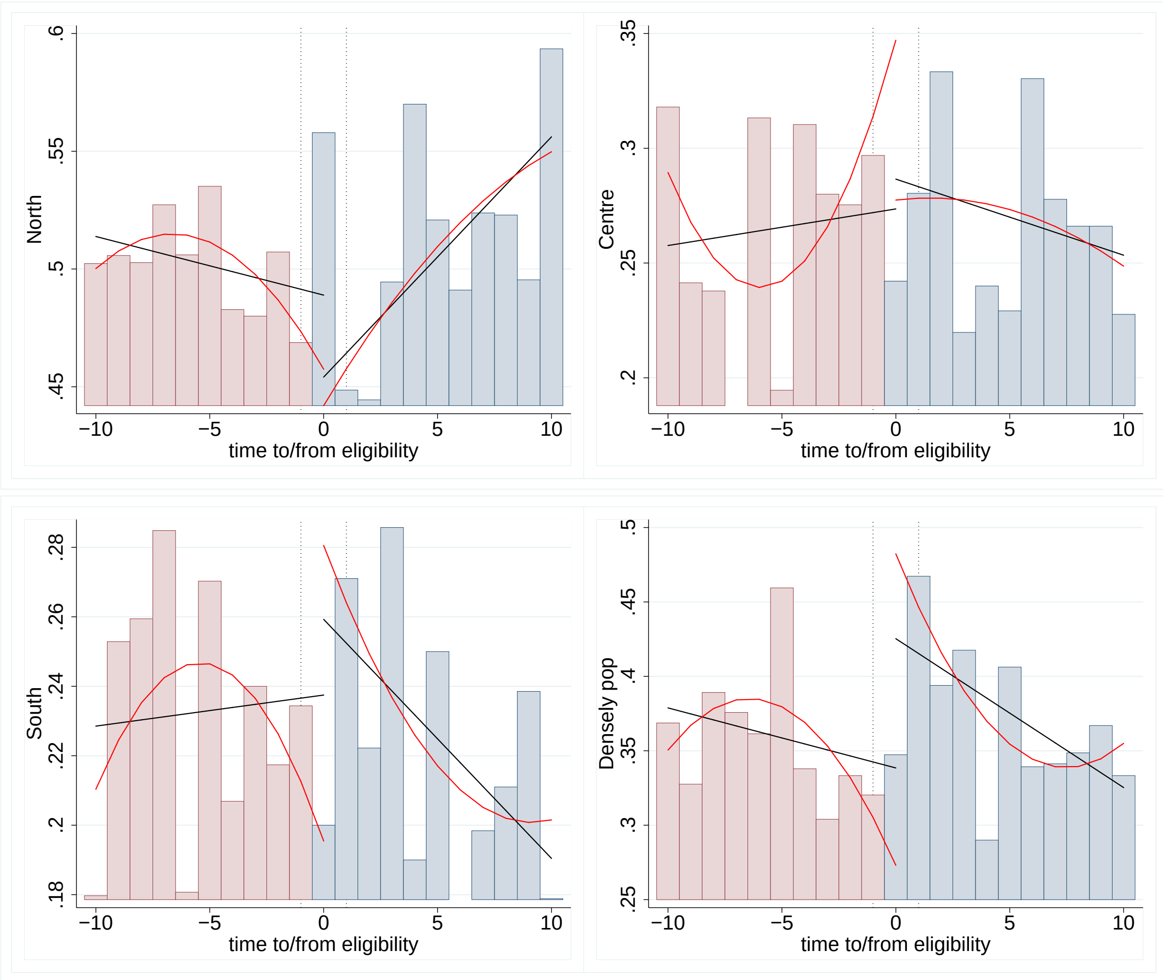
* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. The regressions include $(1-D)\times S$, $D \times S$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table B6: Regressions for different socio-economic variables, women, SILC 2007 (selected sample)

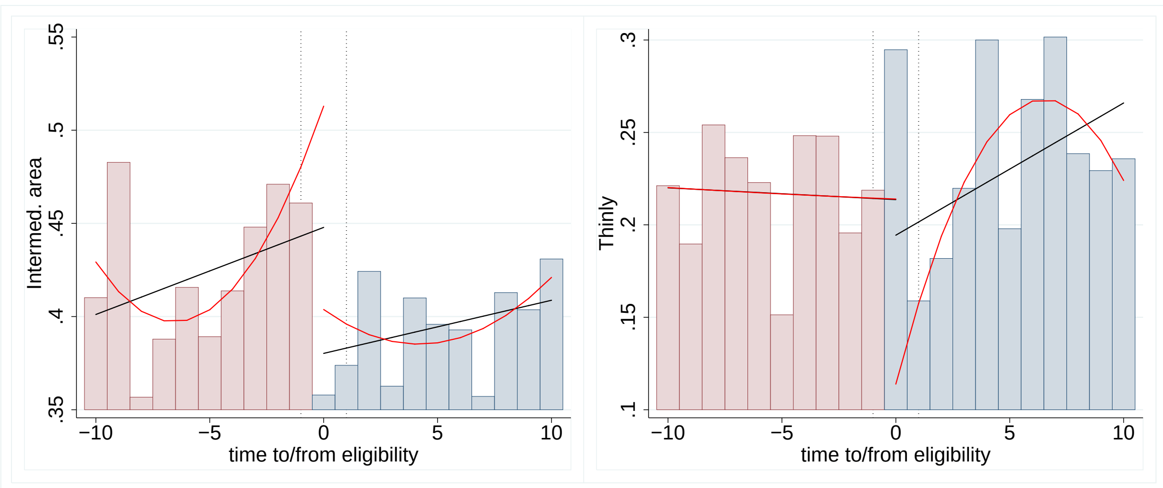
	(1) Age	(2) Y. contrib	(3) Age at retire-woment	(4) Age at first job	(5) Years in paid job	(6) Private	(7) Public	(8) Self-empl.
D	-0.131	-0.220	-0.087	1.175**	1.186*	-0.071*	0.024	0.048
	(0.201)	(0.643)	(0.480)	(0.547)	(0.703)	(0.042)	(0.043)	(0.038)
Observations	2700	2700	2700	2700	2700	2700	2700	2700
R^2	0.846	0.128	0.192	0.005	0.190	0.014	0.042	0.016
$H_0 : \gamma_D = 0$	0.516	0.732	0.855	0.032	0.092	0.091	0.579	0.212
G (p-value)	0.670	0.125	0.285	0.725	0.325	0.485	0.194	0.210
RESET2	0.390	0.009	0.382	0.959	0.423	0.122	0.860	0.119
RESET23	0.616	0.034	0.624	0.040	0.619	0.299	0.174	0.016

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. The regressions include $(1-D)\times S$, $D \times S$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Figure B6: Discontinuities in other variables with respect to distance to/from eligibility (in years), women, SILC 2007 (selected sample)

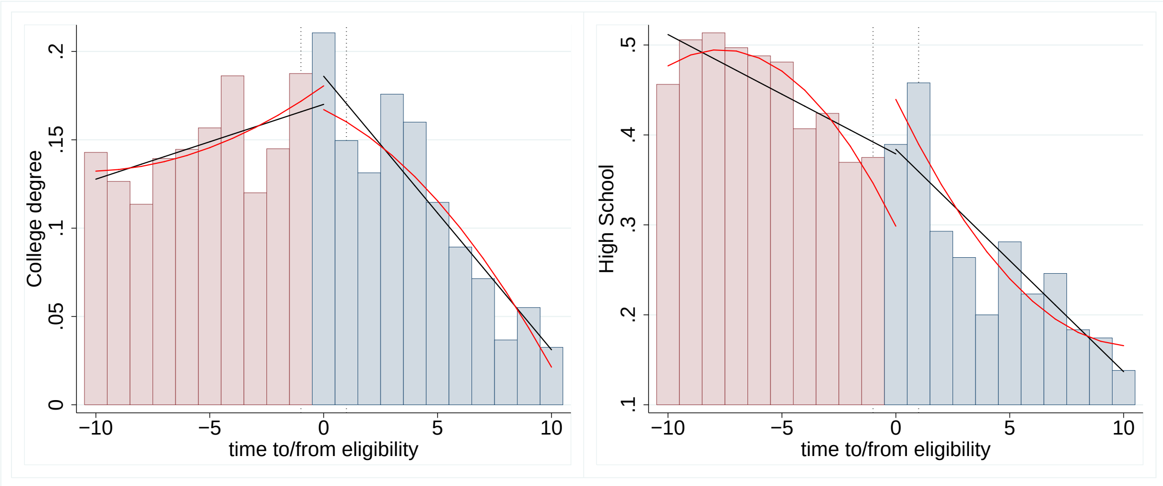


(a)

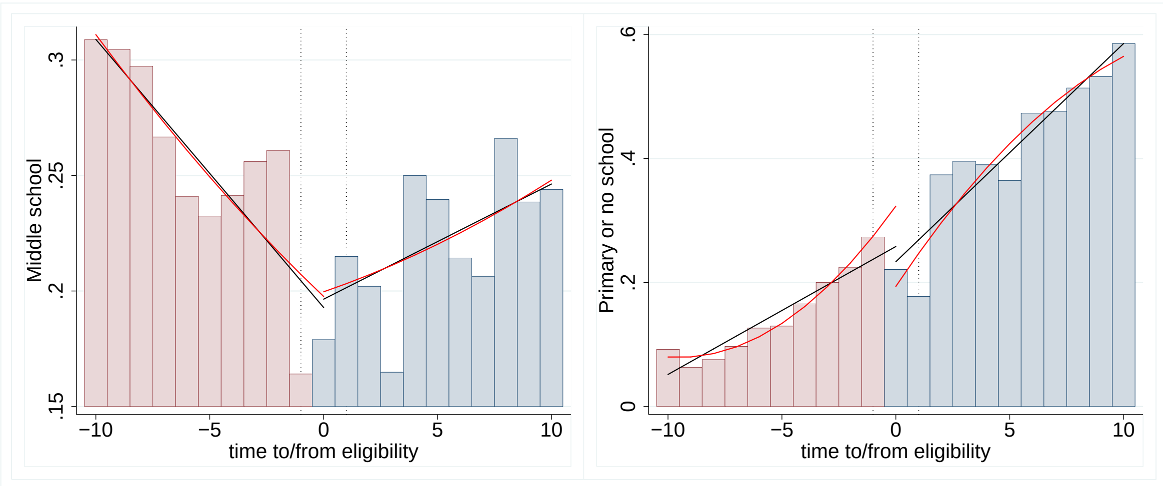


(b)

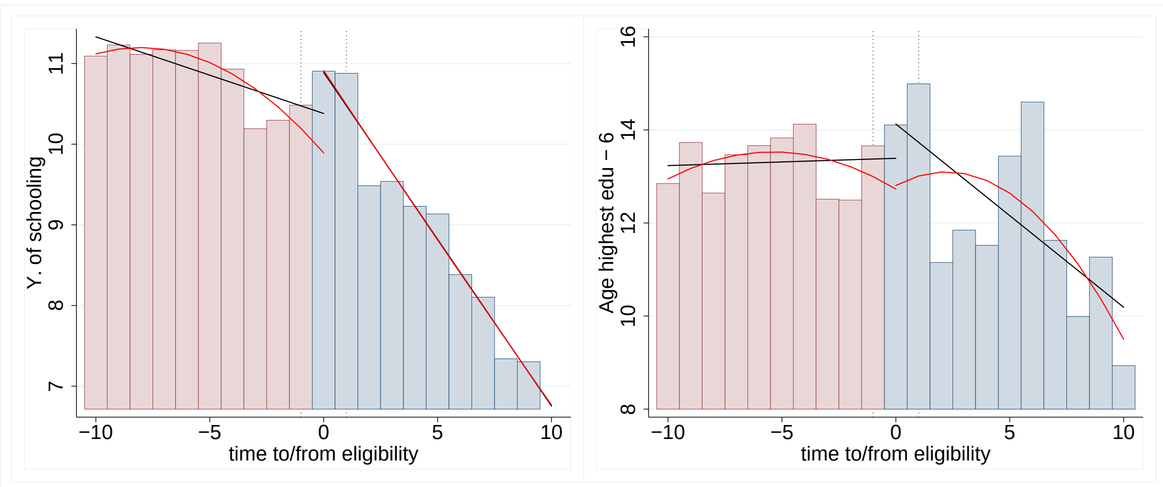
Figure B7: Discontinuities in other variables with respect to distance to/from eligibility (in years), women, SILC 2007 (selected sample)



(a)

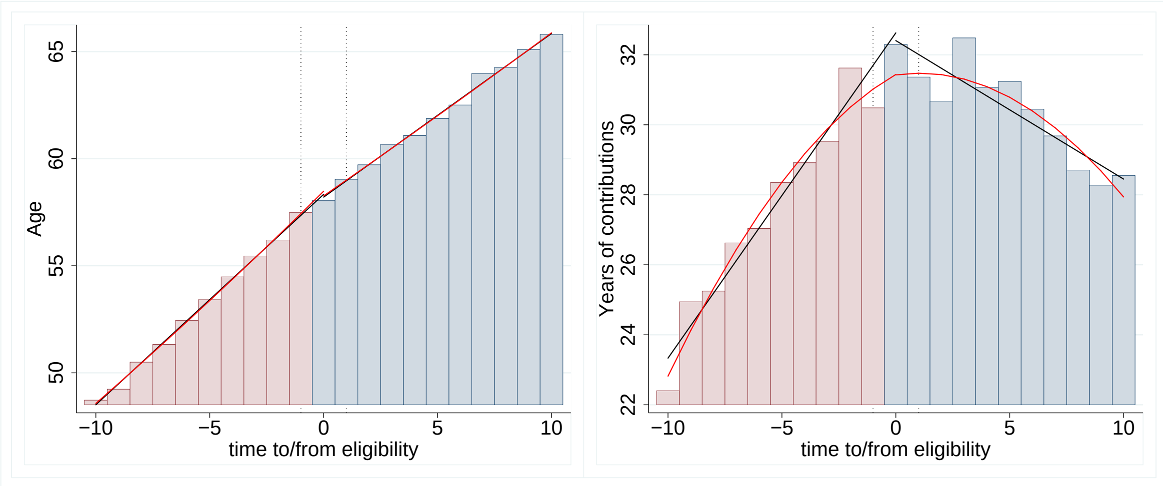


(b)

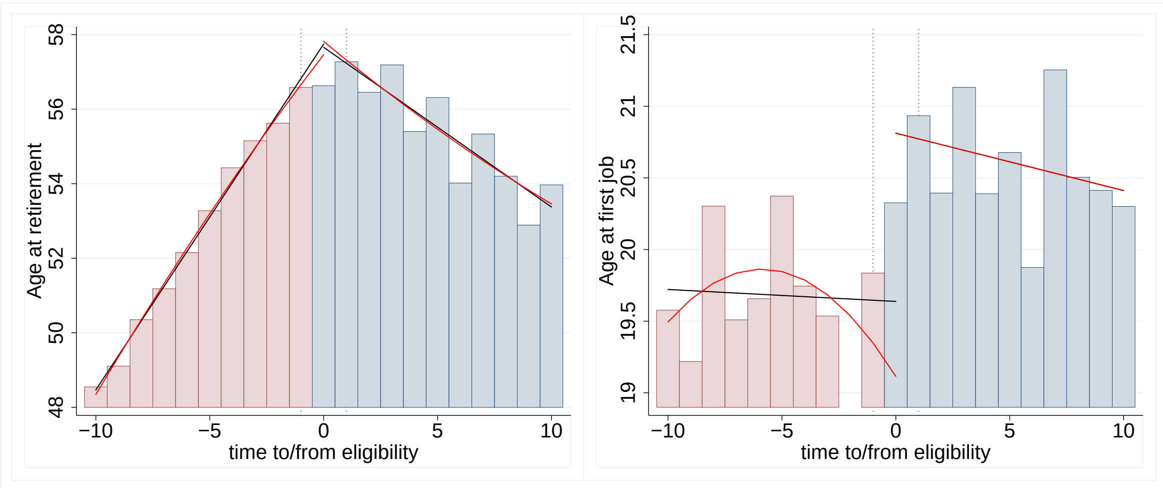


(c)

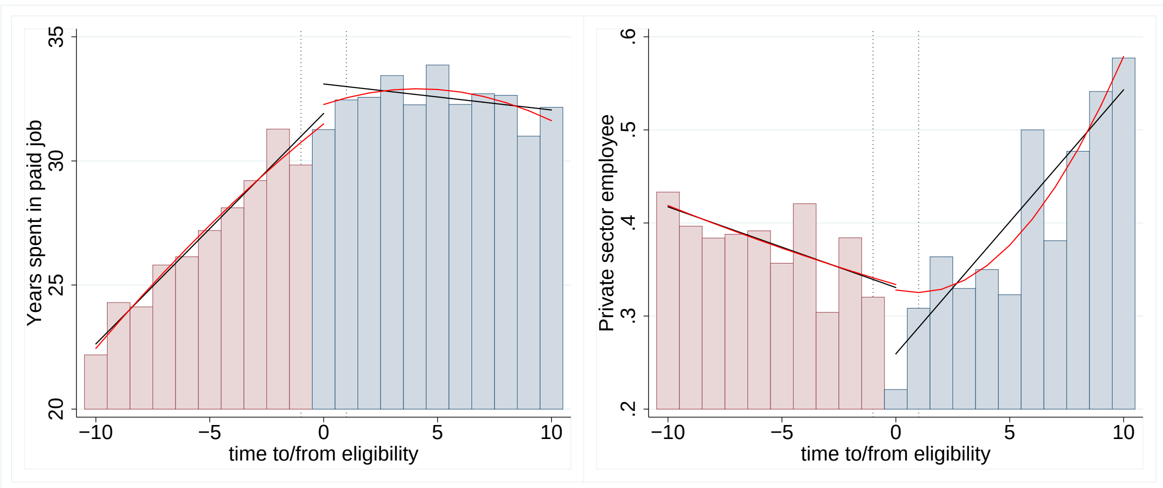
Figure B8: Discontinuities in other variables with respect to distance to/from eligibility (in years), women, SILC 2007 (selected sample)



(a)

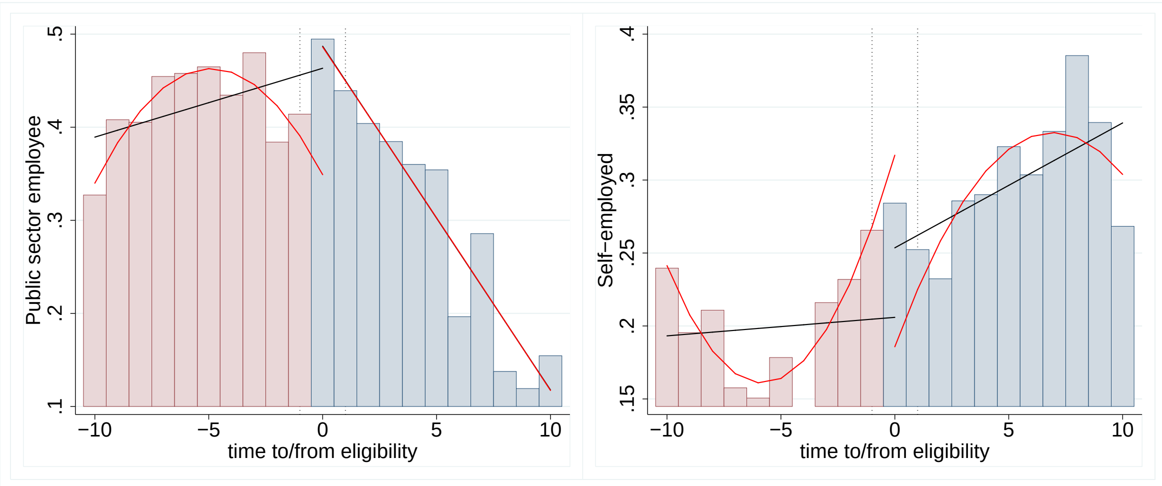


(b)



(c)

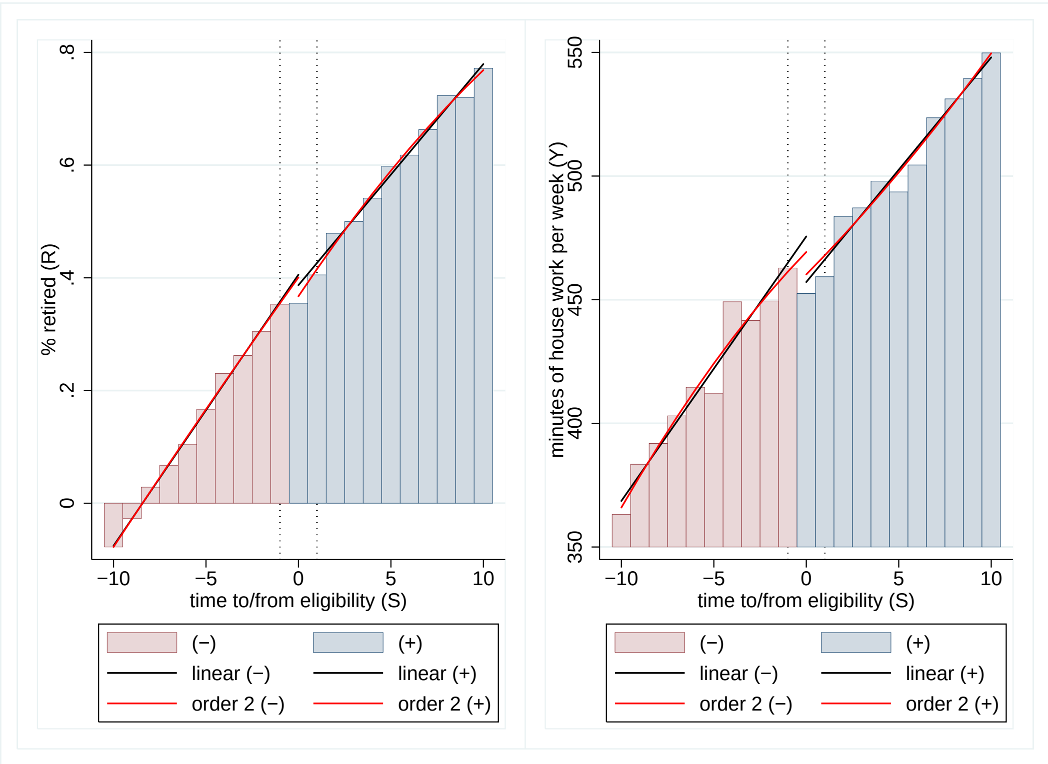
Figure B9: Discontinuities in other variables with respect to distance to/from eligibility (in years), women, SILC 2007 (selected sample)



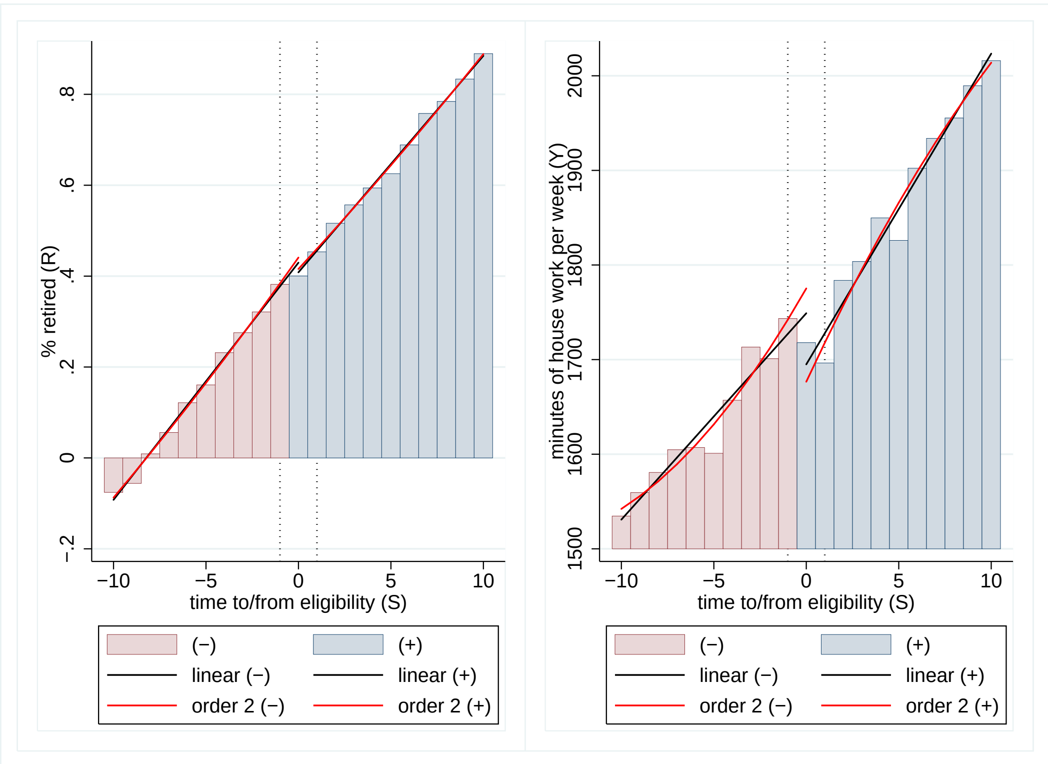
(a)

Plots of fitted values from regressions on X

Figure B10: Fitted values for retirement and domestic work from a regression on X, SILC 2007 (selected sample).



(a) Men



(b) Women

Appendix C: regressions using age as running variable

In 2007, 80% of women who went into retirement exploited the rules for the NRA path (*pensione di vecchiaia*). In most of the cases, they required at least 60 years of age and 15 of social contributions, although for younger workers the latter requirement was increased up to 20 years.[7] One possibility is to ignore all contributory requirements and focus only on the discontinuity at age 60, although this threshold does not correspond to a precise eligibility condition. In this way, I can avoid measurement error in S and I also reduce the influence of rounding, because age is available in quarters.[8]

For retirement, graphical analysis (Figure C1) clearly indicates a jump at age 60. The Akaike and Bayesian criteria indicate a preference for the quadratic regression, which passes the RESET and the G test (Table C1). The estimated discontinuity is 0.170 (s.e. 0.046), not far from $\widehat{\gamma_D}$ obtained using S as running variable.

The evidence for house work is less clear (Figure C1). The figure with age in quarters has a large dispersion, while if I aggregate age at intervals of one year I observe a jump at age 60 by around 200 minutes/week. In both cases, fitted linear polynomials predict a similar discontinuity (Table C2). Using quarters, the point estimate is 179.5 (p-value 0.018). The resulting 2SLS estimate for δ is 503.8 minutes/week (p-value 0.016, see Table C3).

Differently, a quadratic polynomial suggests no jump, and it is preferable according to the Akaike criterion (though not according to the Bayesian). Nevertheless, there is evidence of a kink at eligibility. Indeed, one possible reason for the different result is that the proportion of retired women shows already a large increase from age 57, because they can start going into retirement following the seniority path. This change is associated with a steeper slope in the average Y in the interval $[57, 60]$, while the curve becomes flatter after age 60. One alternative would be to exploit this kink as

[7]To be precise, the requirement was increased for younger workers who started paying social contributions before 1995, while it was decreased to 5 years to those who started later.

[8]I still do not consider observations at exactly age 60. I cannot exclude the presence of rounding at quarter level and, actually, the exact NRA in 2007 was 60 years and 2 months.

an instrument, assuming that without retirement the average house work would have had a continuous slope at eligibility (see Card et al., 2009; Dong, 2010). In Table C4 I also show regressions using $(age - 50)$ as running variable, and using the kinks at 57 and 60 together with the jump at 60 as instruments for R. The point estimate is 660.0 (s.e. 179.0), quite large and more similar to OLS results.

There are two main reasons to prefer the estimates using S as running variable. First of all, we can interpret them as the local average treatment effect for those individuals who go into retirement as soon as eligible. Differently, the discontinuity at age 60 does not have such a clear interpretation, because a relevant group of women could go into retirement earlier than that. Secondly, there is evidence of discontinuities in baseline covariates at age 60, which are stronger than the ones found at the time of eligibility (see Tables C5 and C6). If I introduce covariates in the 2SLS regression exploiting the jump at age 60 by mean of a linear polynomial, I obtain an estimate for δ of 418.6 minutes/week (s.e. 215.6), very similar to my main result using S, though significant only at the 10% level.

Table C1: First stage OLS for retirement status, SILC 2007, women with age between 50 and 70

Dep var R	(1)	(2)	(3)	(4)
$\mathbf{1}[age \geq 60]$	0.220**	0.155**	0.170***	0.356***
	(0.087)	(0.065)	(0.046)	(0.029)
$(1-\mathbf{1}[age \geq 60]) \times (age-60)$	0.007	0.033***	0.033***	0.013***
	(0.020)	(0.009)	(0.003)	(0.001)
$(1-\mathbf{1}[age \geq 60]) \times (age-60)^2$	-0.002	0.001	0.000***	
	(0.002)	(0.000)	(0.000)	
$(1-\mathbf{1}[age \geq 60]) \times (age-60)^3$	-0.000*	0.000		
	(0.000)	(0.000)		
$(1-\mathbf{1}[age \geq 60]) \times (age-60)^4$	-0.000*			
	(0.000)			
$\mathbf{1}[age \geq 60] \times (age-60)$	0.013	0.017**	0.013***	0.006***
	(0.017)	(0.008)	(0.003)	(0.001)
$\mathbf{1}[age \geq 60] \times (age-60)^2$	0.000	-0.000	-0.000***	
	(0.002)	(0.000)	(0.000)	
$\mathbf{1}[age \geq 60] \times (age-60)^3$	-0.000	0.000		
	(0.000)	(0.000)		
$\mathbf{1}[age \geq 60] \times (age-60)^4$	0.000			
	(0.000)			
Constant	0.515***	0.571***	0.570***	0.432***
	(0.065)	(0.049)	(0.036)	(0.023)
Observations	3379	3379	3379	3379
R^2	0.624	0.624	0.623	0.616
$H_0 : \gamma_D = 0$ (p-value)	0.012	0.018	0.000	0.000
H_0 : " (p-val clust)	0.001	0.001	0.000	0.000
Dong's $\widehat{\gamma_D}$	0.218	0.163	0.180	0.360
Dong's $\widehat{\gamma_D}$ (p-value)	0.013	0.012	0.000	0.000
RESET2	0.551	0.085	0.751	0.000
RESET23	0.022	0.194	0.276	0.000
G (p-value)	0.221	0.168	0.204	0.000
AIC	1599.369	1600.065	1596.416	1659.753
BIC	1660.623	1649.068	1633.168	1684.255

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. Age is measured in quarters. The selected sample includes only workers or retirees and excludes proxy interviews, missing house work and observations with age exactly equal to 60. γ_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction. AIC is the Akaike criterium; BIC is the Bayesian criterium; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table C2: Reduced form OLS for time spent in domestic work (minutes/week), SILC 2007, women with age between 50 and 70

Dep var Y	(1)	(2)	(3)	(4)
$\mathbf{1}[age \geq 60]$	146.362	188.426	38.621	179.504**
	(219.935)	(164.433)	(117.553)	(75.775)
$(1 - \mathbf{1}[age \geq 60]) \times (age - 60)$	1.636	7.856	28.030***	6.672***
	(46.496)	(22.107)	(8.516)	(2.083)
$(1 - \mathbf{1}[age \geq 60]) \times (age - 60)^2$	-1.369	-0.700	0.516**	
	(4.567)	(1.267)	(0.200)	
$(1 - \mathbf{1}[age \geq 60]) \times (age - 60)^3$	-0.045	-0.020		
	(0.166)	(0.020)		
$(1 - \mathbf{1}[age \geq 60]) \times (age - 60)^4$	-0.000			
	(0.002)			
$\mathbf{1}[age \geq 60] \times (age - 60)$	3.610	-20.034	1.534	2.648
	(55.214)	(26.333)	(10.163)	(2.365)
$\mathbf{1}[age \geq 60] \times (age - 60)^2$	-1.182	1.318	0.027	
	(5.307)	(1.466)	(0.241)	
$\mathbf{1}[age \geq 60] \times (age - 60)^3$	0.073	-0.021		
	(0.193)	(0.023)		
$\mathbf{1}[age \geq 60] \times (age - 60)^4$	-0.001			
	(0.002)			
Constant	1844.494***	1857.992***	1928.870***	1780.308***
	(133.376)	(102.780)	(77.637)	(52.165)
Observations	3379	3379	3379	3379
R^2	0.035	0.035	0.035	0.033
$H_0 : \beta_D = 0$ (p-value)	0.506	0.252	0.743	0.018
H_0 : " (p-val clust)	0.443	0.247	0.737	0.033
Dong's $\widehat{\beta_D}$	145.406	202.707	51.787	181.516
Dong's $\widehat{\beta_D}$ (p-value)	0.530	0.230	0.663	0.017
RESET2	0.412	0.351	0.330	0.011
RESET23	0.662	0.624	0.525	0.032
G (p-value)	0.060	0.079	0.083	0.051
AIC	56705.229	56701.525	56699.321	56701.187
BIC	56766.482	56750.527	56736.073	56725.688

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. Age is measured in quarters. The selected sample includes only workers or retirees and excludes proxy interviews, missing house work and observations with age exactly equal to 60. β_D is the coefficient for the discontinuity at eligibility. G (p-value) is Lee and Card (2008) statistic. Dong's refer to Dong (2014) correction. AIC is the Akaike criterium; BIC is the Bayesian criterium; RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table C3: 2SLS regressions for time spent on house work (in minutes per week), SILC 2007, women with age between 50 and 70

Dep var Y	(1) No X	(2) No X	(3) No X	(4) With X	(5) With X	(6) With X
R	1219.5	227.1	503.8**	1422.1	178.4	418.6*
	(1095.0)	(681.4)	(208.3)	(1100.7)	(691.2)	(215.6)
$(1-\mathbf{1}[age \geq 60]) \times (age-60)$	-32.4	20.6	0.2	-40.9	20.6	1.8
	(52.6)	(27.8)	(4.2)	(53.6)	(27.6)	(4.4)
$(1-\mathbf{1}[age \geq 60]) \times (age-60)^2$	-1.3	0.4		-1.5	0.4	
	(1.6)	(0.5)		(1.7)	(0.4)	
$(1-\mathbf{1}[age \geq 60]) \times (age-60)^3$	-0.0			-0.0		
	(0.0)			(0.0)		
$\mathbf{1}[age \geq 60] \times (age-60)$	-40.4	-1.4	-0.3	-53.1	-3.8	-1.2
	(41.2)	(17.2)	(3.2)	(37.3)	(16.5)	(3.2)
$\mathbf{1}[age \geq 60] \times (age-60)^2$	1.8	0.1		2.4	0.1	
	(1.8)	(0.3)		(1.6)	(0.3)	
$\mathbf{1}[age \geq 60] \times (age-60)^3$	-0.0			-0.0		
	(0.0)			(0.0)		
Constant	1161.2*	1799.5***	1562.5***	1018.4	2132.8***	1857.7***
	(701.6)	(442.4)	(130.4)	(1011.6)	(641.0)	(232.5)
Observations	3379	3379	3379	3379	3379	3379
$H_0 : \delta = 0$ (p-val)	0.265	0.739	0.016	0.196	0.796	0.052
First Stage F	5.646	13.485	151.490	5.999	12.989	137.276

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by $\mathbf{1}[age \geq 60]$. δ is the coefficient on R. Age is measured in quarters. The selected sample includes only workers or retirees and excludes proxy interviews, missing house work and observations with age exactly equal to 60. Covariates X include a constant, age at first job, years of contributions, years spent in a paid job, plus geographic area, population density, education and employment category dummies. Coefficients are available on request.

Table C4: 2SLS regressions for time spent on house work (in minutes per week), SILC 2007, women age between 50 and 70

	No X	With X	No X	With X	No X	With X
	FIRST STAGE: dependent variable R					
$\mathbf{1}[age \geq 57]$	0.20***	0.21***	-0.30	-0.27		
	(0.03)	(0.03)	(0.21)	(0.20)		
$\mathbf{1}[age \geq 60]$	0.32***	0.30***	0.81***	0.79***	0.52***	0.52***
	(0.03)	(0.03)	(0.21)	(0.21)	(0.04)	(0.04)
$(age - 50)$	0.03***	0.03***	0.02***	0.03***	0.02***	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\mathbf{1}[age \geq 57]\,(age - 50)$			0.06**	0.06**	0.03***	0.03***
			(0.03)	(0.02)	(0.00)	(0.00)
$\mathbf{1}[age \geq 60]\,(age - 50)$			-0.06**	-0.06**	-0.02***	-0.03***
			(0.03)	(0.02)	(0.00)	(0.00)
Constant	-0.02**	0.25***	-0.01	0.25***	-0.01	0.25***
	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.03)
R^2	0.612	0.633	0.614	0.635	0.613	0.634
First Stage F	129.629	127.987	69.328	67.213	91.046	88.790
	SECOND STAGE: dependent variable Y					
R	659.54***	636.96***	653.08***	631.53***	659.95***	637.42***
	(181.32)	(183.85)	(178.03)	(180.65)	(179.03)	(181.83)
$(age - 50)$	-9.88	-12.33	-9.47	-11.99	-9.91	-12.36
	(12.02)	(12.07)	(11.81)	(11.89)	(11.88)	(11.96)
Constant	1574.72***	1752.29***	1574.11***	1753.33***	1574.75***	1752.21***
	(37.79)	(126.50)	(37.55)	(126.67)	(37.66)	(126.55)
Observations	3412	3412	3412	3412	3412	3412
Hansen's test (p-val)	0.377	0.189	0.756	0.630	0.575	0.430

* p<.10 ** p<.05 *** p<.01. Robust standard error in brackets. Age is measured in quarters. The selected sample includes only workers or retirees and excludes proxy interviews, and missing house work. Covariates X include a constant, age at first job, years of contributions, years spent in a paid job, plus geographic area, population density, education and employment category dummies. Coefficients are available on request.

Table C5: Regressions for different socio-economic variables, SILC 2007, women with age between 50 and 70

	(1) North	(2) Centre	(3) South	(4) College	(5) High school
$\mathbf{1}[age \geq 60]$	0.074**	-0.018	-0.056**	-0.065***	-0.067**
	(0.035)	(0.030)	(0.028)	(0.024)	(0.032)
Observations	3379	3379	3379	3379	3379
R^2	0.004	0.000	0.005	0.026	0.067
$H_0 : \gamma_D = 0$ (p-val)	0.034	0.544	0.049	0.006	0.039
G (p-value)	0.199	0.087	0.443	0.254	0.307
RESET2	0.823	0.374	0.830	0.066	0.560
RESET23	0.823	0.515	0.820	0.025	0.826

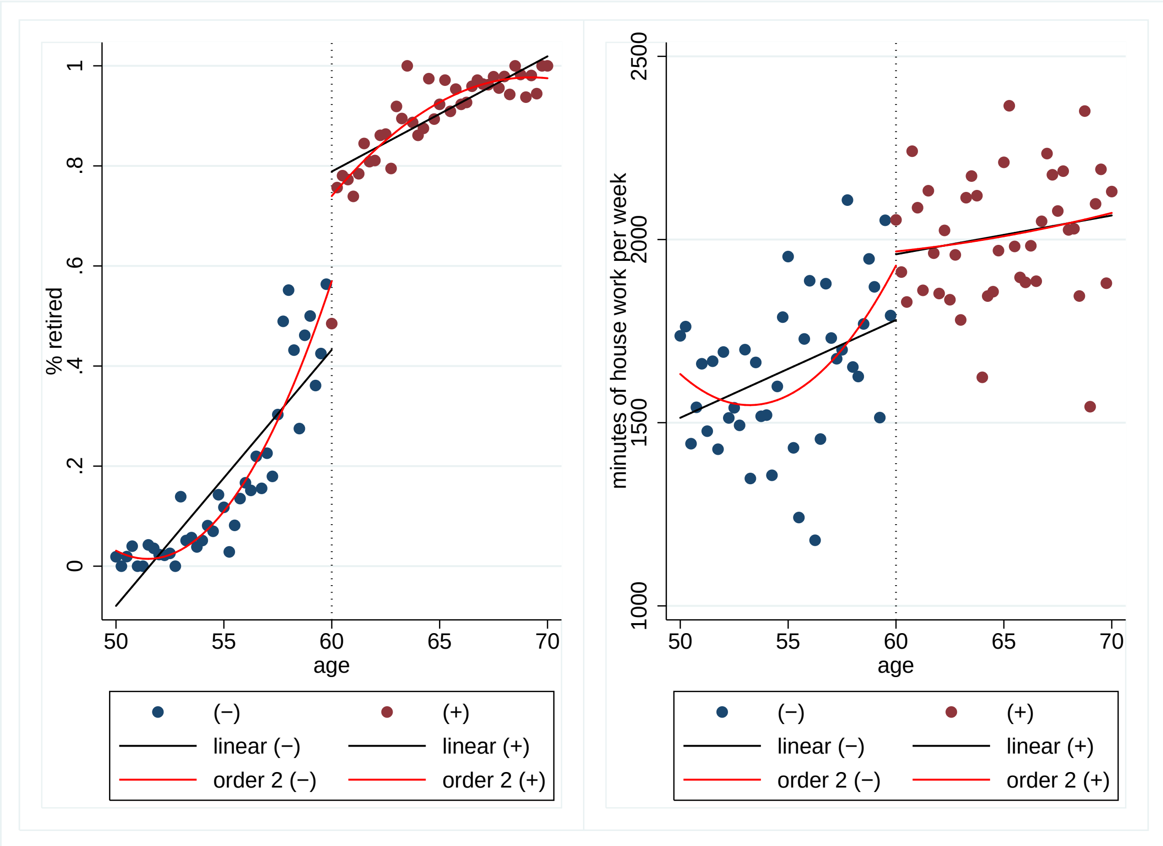
* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. The regressions include $(1 - D) \times (age - 60)$, $D \times (age - 60)$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Table C6: Regressions for different socio-economic variables, SILC 2007, women with age between 50 and 70

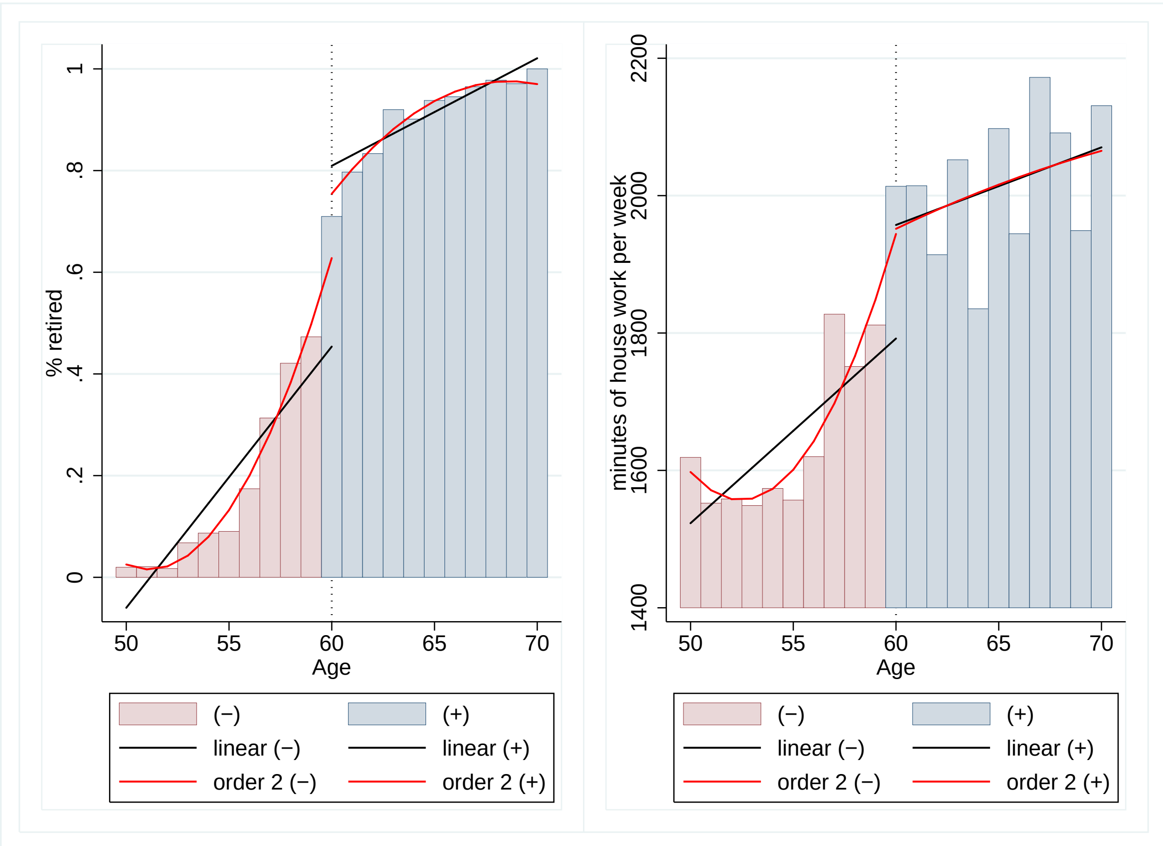
	(1) Middle sch.	(2) Primary sch.	(3) Private	(4) Public	(5) Self-empl.
$\mathbf{1}[age \geq 60]$	0.059**	0.073**	0.119***	-0.098***	-0.021
	(0.029)	(0.032)	(0.034)	(0.034)	(0.029)
Observations	3379	3379	3379	3379	3379
R^2	0.004	0.153	0.026	0.048	0.004
$H_0 : \gamma_D = 0$ (p-val)	0.044	0.024	0.001	0.004	0.470
G (p-value)	0.443	0.034	0.760	0.071	0.529
RESET2	0.758	0.643	0.497	0.819	0.212
RESET23	0.890	0.084	0.154	0.958	0.356

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. The regressions include $(1 - D) \times (age - 60)$, $D \times (age - 60)$ and a constant. γ_D is the coefficient for the discontinuity at eligibility. RESET2 is the p-value from the RESET test adding square of fitted values, while RESET23 adding squares and cubes.

Figure C1: Retirement and house work with respect to age, SILC 2007, women with age $\in [50, 70]$



(a) Age in years and quarters



(b) Age in years

Appendix D: poor health, household size and marital status

Other changes caused by retirement may have an off-setting effect on the increase in home production. Coe and Zamarro (2011), using data from the Survey of Health, Ageing and Retirement in Europe (SHARE), found evidence of a health preserving effect of retirement for men, with a 35 percentage points decrease in the probability of reporting fair, poor or very poor health. In my dataset, I used 2SLS with, as a dependent variable, a dummy for these answers to the general health question. In the linear specification including $(1 - D) \times S$ and $D \times S$, the estimated effect is indeed negative, but much smaller (8.4 percentage points) and not statistically significant at conventional significance levels (Table D1).[9] For women the estimate is also negative, but still smaller (6.7 percentage points), and not statistically significant (Table D2). The reason for such a difference is not clear, and further research might try to extend the analysis using other waves.

Battistin et al. (2009) found a reduction in household size by 0.3, explained by adult children leaving the parental home. The estimated effect of retirement in SILC is actually positive for men, but very small (Table D1). For women it is negative, but still far from Battistin et al. (2009) and not statistically significant (Table D2). There is also no evidence of a change in the probability of being married. Battistin et al. (2009) estimates refer to years 1993-2004. On the one hand, between 1993 and 2007 there was a rise in retirement age. This may imply that, at the time that parents retired, children were older in 2007 than in 1993, so that they were more likely to leave the household. On the other hand, the deterioration of expectations about economic growth after 2007 could have reduced the incentives for adult children to form independent households. The latter trend may have offset the former.

[9] Estimates with polynomials up to the 3rd order are never statistically significant.

Table D1: 2SLS regressions for health, household size and marital status, men, SILC 2007 (selected sample)

	(1) Health fair or poor	(2) Hh size	(3) Married
R	-0.0842	0.0600	0.0867
	(0.0858)	(0.2046)	(0.0678)
$(1-D)\times S$	0.0130***	-0.0360***	0.0011
	(0.0039)	(0.0104)	(0.0034)
$D\times S$	0.0263***	-0.0612***	-0.0042
	(0.0073)	(0.0163)	(0.0056)
Constant	0.3714***	2.9511***	0.8030***
	(0.0285)	(0.0721)	(0.0236)
Observations	3930	3970	3970
R^2	0.026	0.053	-0.007
Average dep var for $R=0$	0.2972	3.1288	0.8173
$H_0:\delta=0$ (p-val)	0.3267	0.7693	0.2010
$H_0:$ " (p-val clust)	0.1239	0.7207	0.1207
First Stage F	210.9310	216.0805	216.0805

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The selected sample for column (1) excludes missing values in general health. The dummy for health equals one for fair, poor or very poor health.

Table D2: 2SLS regressions for health, household size and marital status, women, SILC 2007 (selected sample)

	(1) Health fair or poor	(2) Hh size	(3) Married
R	-0.0674	-0.0659	-0.0381
	(0.0936)	(0.1922)	(0.0854)
$(1-D)\times S$	0.0138***	-0.0646***	0.0029
	(0.0046)	(0.0108)	(0.0043)
$D\times S$	0.0236***	-0.0206	-0.0004
	(0.0090)	(0.0181)	(0.0083)
Constant	0.4127***	2.5713***	0.7333***
	(0.0329)	(0.0733)	(0.0302)
Observations	2671	2700	2700
R^2	0.030	0.073	-0.002
Average dep var for $R=0$	0.3280	2.8833	0.7032
$H_0:\delta=0$ (p-val)	0.4715	0.7316	0.6556
$H_0:$ " (p-val clust)	0.1802	0.6279	0.6640
First Stage F	189.6454	197.3799	197.3799

* p<.10 ** p<.05 *** p<.01. Robust standard errors in brackets. R is instrumented by D. δ is the coefficient on R. The selected sample for column (1) excludes missing values in general health. The dummy for health equals one for fair, poor or very poor health.