Mathematical appendix

This Appendix is organized as follows. First, we derive the identification result discussed in Section I. Second, we present the distribution of S resulting from the imputation strategy discussed in Section C. Finally, we derive the conditions on the distribution of the measurement error in S that allow to recover the drop in consumption around retirement within a standard instrumental variables approach.

Identification through discontinuities in the eligibility rule

As the decision to retire is entirely up to eligible individuals, it is crucial to discuss how the resulting endogeneity problem can be accounted for in the context of discontinuity designs. We will consider the case of a sharp discontinuity first, and discuss the fuzzy case further below in this section. Let:

$$Y = Y_0 + R(s)\beta \tag{10}$$

be the observed outcome as it results from the retirement decision: it follows from the last expression that we have either $Y \equiv Y_1$ or $Y \equiv Y_0$ depending on the retirement status of individuals ($R = 1$ and $R = 0$, respectively). The dependence of the retirement status R on S is stressed by writing $R(s)$. The difference of mean outcomes for individuals marginally above and below the threshold $\bar{s}$:

$$E\{Y|\bar{s}^+\} - E\{Y|\bar{s}^-\} \tag{11}$$

can be written as:

$$E\{Y_0|\bar{s}^+\} - E\{Y_0|\bar{s}^-\} + E\{R(s)\beta|\bar{s}^+\} - E\{R(s)\beta|\bar{s}^-\}, \tag{12}$$

which simplifies to:

$$E\{Y_0|\bar{s}^+\} - E\{Y_0|\bar{s}^-\} + E\{\beta|\bar{s}^+\} \tag{13}$$

since $R = \mathbb{1}(S \geq \bar{s})$. Condition 1 in Section I is then sufficient for the mean impact of the treatment at $\bar{s}^+$ to be identified with a sharp discontinuity. This condition requires that in the counterfactual world of no retirement, no discontinuity would take place at the threshold for selection. Intuitively, in order to give a causal interpretation to discontinuities of Y around the threshold for eligibility $\bar{s} = 0$ it has to be the case that in the absence of retirement no discontinuity would be observed in the outcome Y around $\bar{s} = 0$. This amounts to assuming a smooth consumption profile at $\bar{s} = 0$ in a counterfactual world of no retirement. If this condition holds, we can write:

$$E\{\beta|\bar{s}^+\} \equiv E\{Y|\bar{s}^+\} - E\{Y|\bar{s}^-\}, \tag{14}$$

so that the difference in expected consumption expenditures above and below the threshold for eligibility identifies the average causal effect of retirement on consumption for marginally eligible subjects.

When the treatment status is not the result of a sharp design, the discontinuity in the probability to retire around the threshold is less than one thus defining a fuzzy design. It follows that the mean impact at $\bar{s}$ cannot be identified by simply comparing the mean outcome for marginal retired to the mean outcome for marginal non-retired households. In general, additional conditions are required to recover meaningful causal

parameters from (11), thus loosing much of the simplicity of the design. Hahn, Todd and van der Klaauw (2001) as well as many other authors in the literature point out that assumptions can be made to recover causal effects for a particular group of individuals around the threshold $\bar{s}$. Such assumptions qualify S as an instrumental variable for R around $\bar{s}$, so that a LATE (Imbens and Angrist, 1994) parameter can be estimated for the group of compliers. James J. Heckman, Robert Lalonde and Jeffrey Smith (1999) emphasize this point by saying that much of the simplicity of the design is lost moving from a sharp design to a fuzzy design.

In the context of this paper self-selection of households into retirement fits the partially fuzzy design described by Battistin and Rettore (2008). As a result of the eligibility rule and of self-selection, the probability of retirement for those heads scoring a value of S below the threshold $\bar{s}$ is zero by definition, since they are not eligible for retirement. The probability of retirement for those scoring above $\bar{s}$ is smaller than one because retirement is not mandatory. This implies that the probability of retirement is discontinuous at the threshold for eligibility and the size of the discontinuity is less than one (i.e. according to the terminology introduced in the previous section, a fuzzy design is defined). As pointed out by Battistin and Rettore (2008), despite the fuzziness of this design the existence of a sharp eligibility rule can help recover much of the simplicity of the design.

To recover the regularity conditions required for identification consider again the difference in (11). Since participation is precluded to marginally ineligibles ($R(\bar{s}^-) = 0$), the expression in (12) can be written as:

$$E\{Y_0|\bar{s}^+\} - E\{Y_0|\bar{s}^-\} + E\{R(s)\beta|\bar{s}^+\}. \tag{15}$$

If Condition 1 holds, by using the law of iterated expectations and by noting that $E\{R(s)\beta|R = 0, \bar{s}^+\} = 0$ the previous expression yields:

$$E\{R(s)\beta|\bar{s}^+\} = E\{\beta|R = 1, \bar{s}^+\}Pr\{R = 1|\bar{s}^+\}, \tag{16}$$

so that the mean impact on retirees in a right-neighborhood of $\bar{s}$ is identified by:

$$E\{\beta|R = 1, \bar{s}^+\} = \frac{E\{Y|\bar{s}^+\} - E\{Y|\bar{s}^-\}}{E\{R|\bar{s}^+\}}. \tag{17}$$

In other words, Condition 1 is sufficient for the effect of retirement on the retirees to be identified locally at the threshold for eligibility $\bar{s}$.

It turns out that, despite the *prima facie* fuzzy nature of this set-up, the LATE (Imbens and Angrist, 1994) at the discontinuity point is identified under the same condition used to estimate the average treatment effect in the sharp design. The result rests upon the fact that the probability of retirement on the left-hand side of $\bar{s}$ is zero by design, and this simplifies the expression in (12) without further assumptions on individuals' behavior. It also follows that (17) can be estimated following an instrumental variable procedure, where eligibility is used as an instrument for the actual status R conditional on S.

The measurement of years of contributions

Figure V shows the distribution of years of contributions, separately for men and women, resulting from the imputation procedure described in Section C. For men there is a high percentage of cases with contribution spells above 30 years - with a relevant spike at 35 years. For women the distributions is much more dispersed and in many cases there are only 5 or 10 years of contributions completed.

It follows that the variable which measures the distance from the first eligibility year (S) is computed by first establishing on the basis of recorded age and years of contributions the eligibility year and then taking the difference with the survey year. The distribution of this variable in the range (-10, 10) for males and females is reported in Figure V. Negative values of this variable imply that eligibility for retirement has not yet been attained. Positive values measure the time from the first year of eligibility. While for men the distribution of S is rather even over the range of negative and positive values, for women there is a prevalence of positive values of S, indicating that a large percentage of women have past their first eligibility year (and they are presumably retired).

Allowing for measurement error in S

Consider the regression of R on S_{obs}, where:

$$S_{obs} = S^* Z + S(1 - Z). \tag{18}$$

To begin with consider the case in which S_{obs} is equal to $S^* + u$, with u a classical measurement error. Since $E\{R|S_{obs}\}$ is equal to:

$$\int_{\bar{s}}^{+\infty} E\{R|S^*\} f_{S^*|S_{obs}}(S^*) dS^* \tag{19}$$

with $f_{S^*|S_{obs}}$ a density function continuous on the real line. This measurement error model implies that even if $E\{R|S^*\}$ were discontinuous at $\bar{s}$ such discontinuity would be smoothed out by the measurement error and as a result $E\{R|S_{obs}\}$ would be smoooth at $\bar{s}$. This is in sharp contrast with the apparent discontinuity we do observe at the threshold for eligibility (see Table 4 and Figure A).

From the definition of S_{obs} and by using the law of iterated expectations we have:

$$\begin{aligned}
(20)\quad E\{R|S_{obs} = s_{obs}\} &= E\{R|S^* = s_{obs}, Z = 1\} E\{Z|S_{obs} = s_{obs}\} \\
(21)\quad &+ E\{R|S = s_{obs}, Z = 0\}(1 - E\{Z|S_{obs} = s_{obs}\}), \\
(22)\quad &= E\{R|S^* = s_{obs}, Z = 1\} E\{Z|S_{obs} = s_{obs}\} \\
(23)\quad &+ (1 - E\{Z|S_{obs} = s_{obs}\}) \\
(24)\quad & \int E\{R|S = s_{obs}, S^* = \tau, Z = 0\} f_{S^*|S,Z}(\tau|s_{obs}, 0) d\tau.
\end{aligned}$$

Under the assumption $(Z, S) \perp R|S^*$ which states the ignorability of measurement error for the generating process of R conditional on the true value of the eligibility variable S^* the last expression becomes:

$$\begin{aligned}
(25)\quad E\{R|S_{obs} = s_{obs}\} &= E\{R|S^* = s_{obs}\} E\{Z|S_{obs} = s_{obs}\} \\
(26)\quad &+ (1 - E\{Z|S_{obs} = s_{obs}\}) \\
(27)\quad & \int_{\bar{s}}^{+\infty} E\{R|S^* = \tau\} f_{S^*|S,Z}(\tau|s_{obs}, 0) d\tau,
\end{aligned}$$

as $E\{R|S^* = \tau\} = 0$ when $\tau < \bar{s}$. Note that, in general, the measurement error in S does not need to be classical, though it has to be non-differential, i.e. it must contain no information on R (and on Y, see below) once the true value S^* has been controlled for (see John Bound, Charlie Brown and Nancy Mathiowetz, 2001).

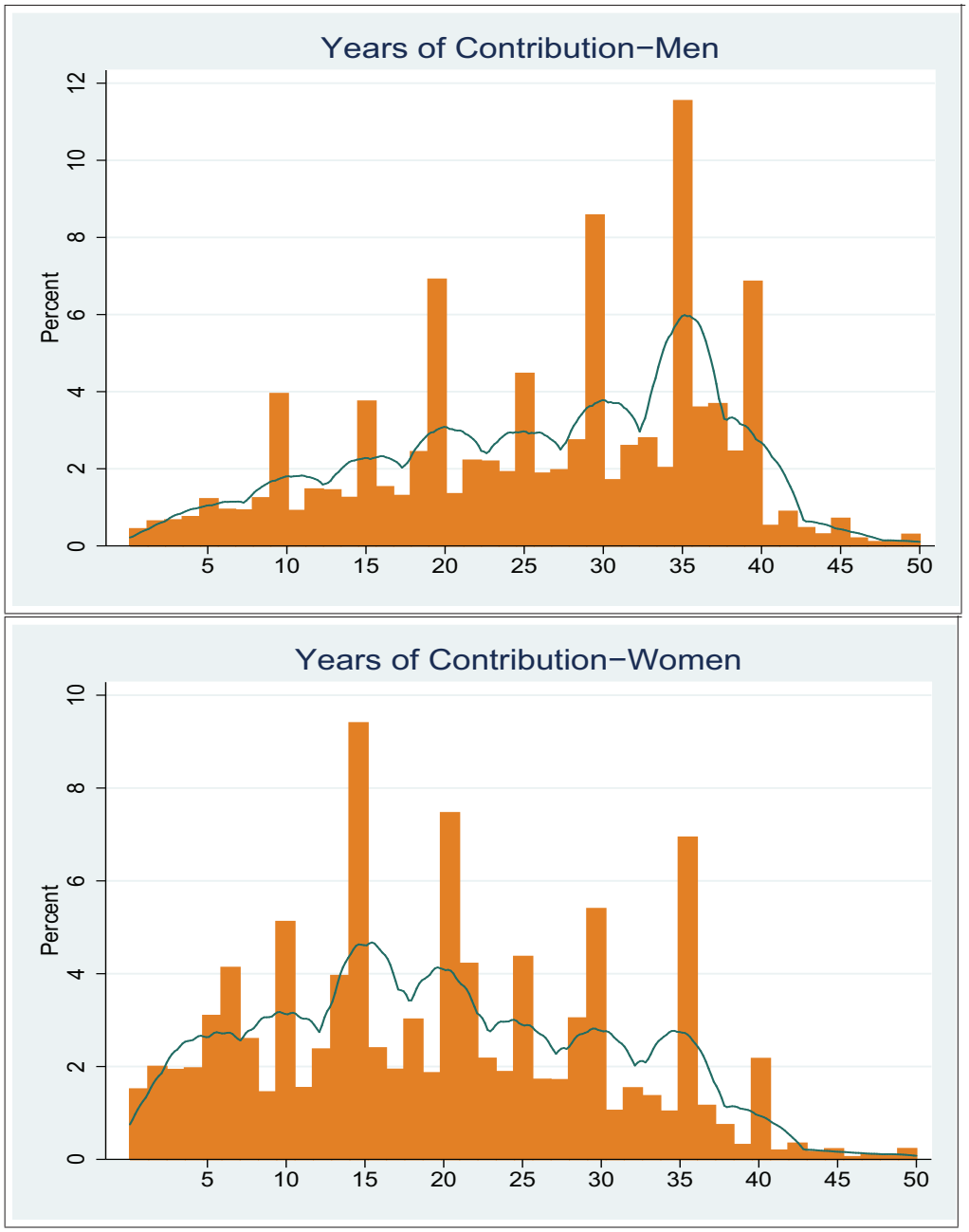

Figure 2: Contributive years - waves 1993 to 2004.

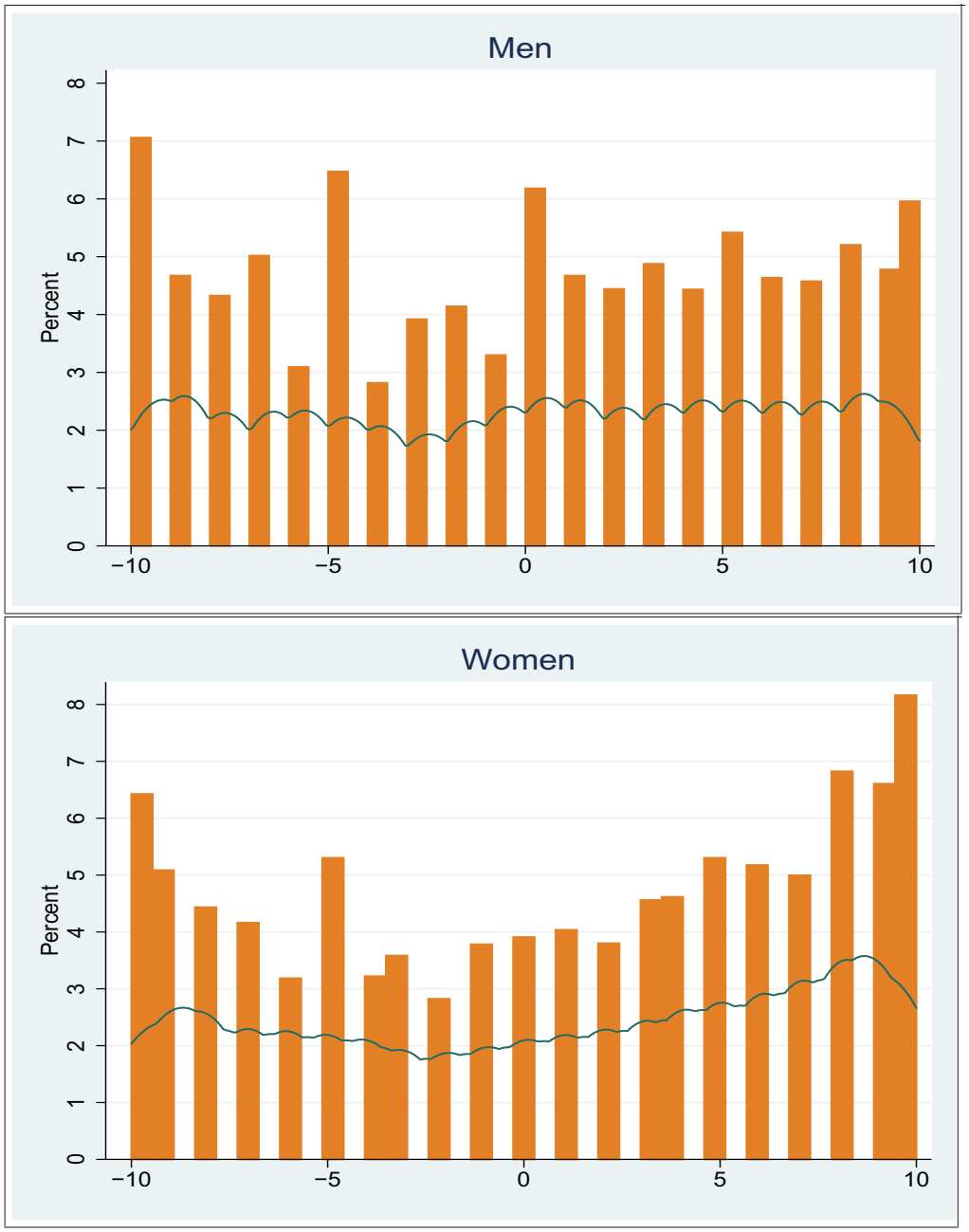

Figure 3: Distribution of S.

Under smoothness conditions of the distribution of (S, S^*) around $(\bar{s}, \bar{s})$, it follows from the last expression that:

$$E\{R|S_{obs} = \bar{s}^+\} - E\{R|S_{obs} = \bar{s}^-\} = E\{R|S^* = \bar{s}^+\}E\{Z|S_{obs} = \bar{s}^-\}, \tag{28}$$

implying that the discontinuity in the retirement probability observed around the threshold for eligibility understates the true discontinuity by the factor $E\{Z|S_{obs} = \bar{s}^-\}$. It therefore follows that the estimated discontinuity is downward biased for the true discontinuity. The bias term can be estimated from the proportion of heads who self-report being retired though marginally ineligible according to S_{obs}, $P\{R = 1|S_{obs} = \bar{s}^-\}$.

By applying a similar argument to the regression function of Y on S_{obs} under the condition $(Z, S)\perp Y|S^*$ we obtain:

$$E\{Y|S_{obs} = \bar{s}^+\} - E\{Y|S_{obs} = \bar{s}^-\} = (E\{Y|S^* = \bar{s}^+\} - E\{Y|S^* = \bar{s}^-\})E\{Z|S_{obs} = \bar{s}^-\}, \tag{29}$$

which implies that the discontinuity in consumption expenditures estimated around $S_{obs} = \bar{s}$ is still downward biased for the true discontinuity by the same factor $E\{Z|S_{obs} = \bar{s}^-\}$ responsible for the bias in the estimation of discontinuity in the retirement probability.

The parameter in (17) can then be written as:

$$E\{\beta|R = 1, S^* = \bar{s}^+\} = \frac{E\{Y|S^* = \bar{s}^+\} - E\{Y|S^* = \bar{s}^-\}}{E\{R|S^* = \bar{s}^+\}}, \tag{30}$$

which depends on the joint distribution of (Y, R, S^*). Because of measurement error in S_{obs}, (28) and (29) imply that the estimator constructed by using the empirical analogues of the quantities in the last expression from raw data (Y, R, S) is *not* consistent for the parameter of interest.

However, under the condition $(Z, S)\perp(R, Y)|S^*$ on the measurement error it is immediate to see how the following ratio:

$$\frac{E\{Y|S_{obs} = \bar{s}^+\} - E\{Y|S_{obs} = \bar{s}^-\}}{E\{R|S_{obs} = \bar{s}^+\} - E\{R|S_{obs} = \bar{s}^-\}}, \tag{31}$$

identifies the causal effect of retirement on consumption since the bias factor cancels out. As pointed out by Imbens and Angrist (1994), the latter expression can be interpreted as an instrumental variable estimand, where the eligibility status is used to correct for the endogeneity of R. The sample analogue of (31) can be obtained by taking the ratio of the discontinuity pictured in the top panel of Figure V to the discontinuity in the bottom panel of the same figure (see Section B in the paper for a description of how data points are derived).

Additional references

- Bound, John, Brown, Charlie and Nancy A. Mathiowetz, N. (2001), "Measurement error in survey data", in J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics. Vol. 5*, Amsterdam: North-Holland, 3705-3843
- Heckman, James J., Robert Lalonde and Jeffrey Smith (1999), "The Economics and Econometrics of Active Labor Market Programs", *Handbook of Labor Economics*, Volume 3, Ashenfelter, A. and Card, D. (eds.), Amsterdam: Elsevier Science

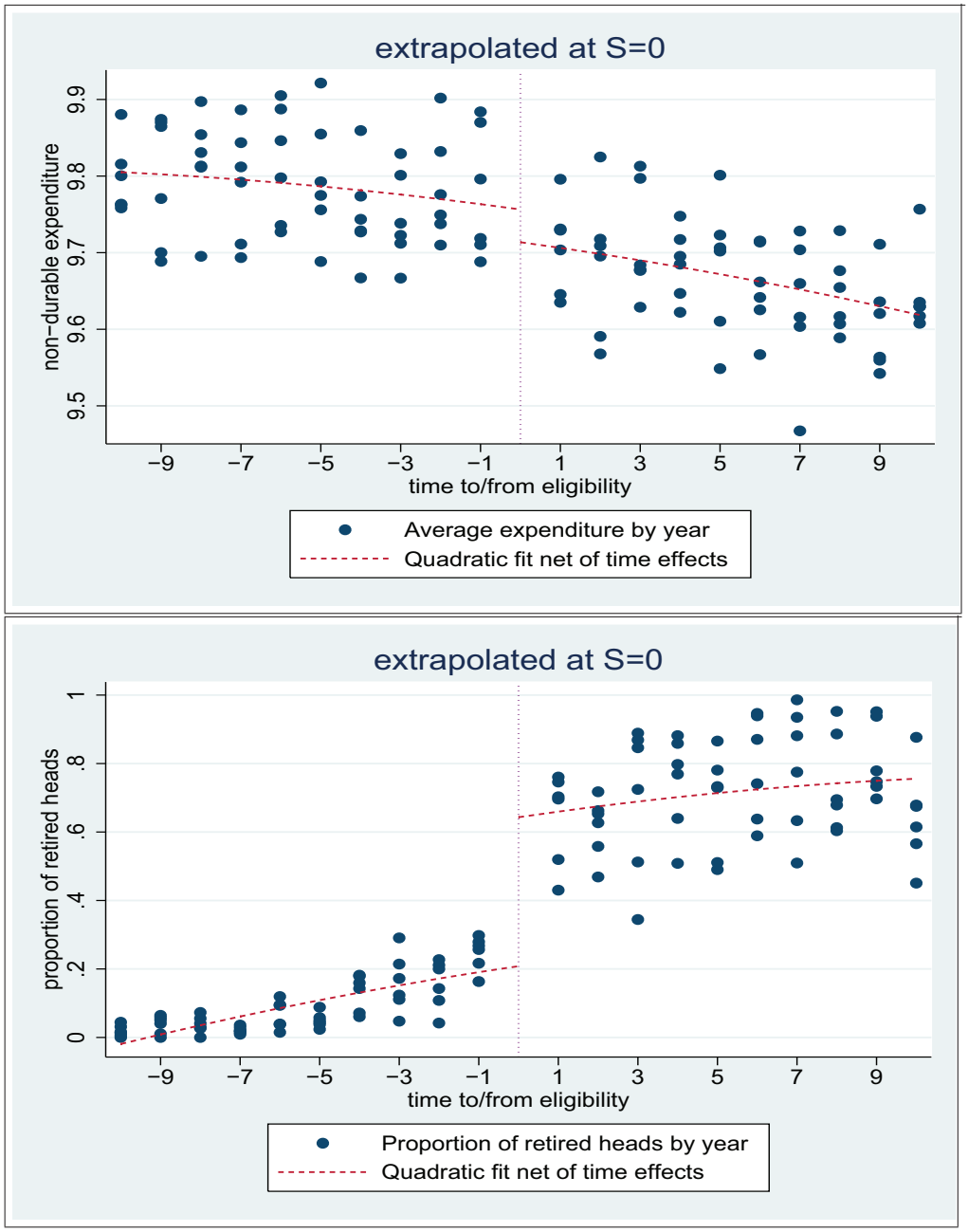

FIGURE 4: ESTIMATION OF THE CAUSAL EFFECT OF ELIGIBILITY ON RETIREMENT AND ON NON-DURABLE CONSUMPTION EXPENDITURES.