Caregiving to Elderly Parents and Employment Status of European Mature Women*

Laura Crespo & Pedro Mira
CEMFI, Madrid

This version: April 16, 2010

Abstract

We study the prevalence of informal caregiving to elderly parents by their mature daughters in Europe and the effect of intense (daily) caregiving on the employment status of the daughters. We use data from the first two waves of SHARE. We group the data into three country pools (North, Continental and South) which strongly differ in the availability of public formal care services and female labour market attachment. We use a time allocation model to provide a link to an empirical IV-treatment effects framework, to discuss the parameters of interest which can be identified in that framework and to interpret the differences in results across country pools. We argue that the Local Average Treatment effect of daily care on labor supply, as identified by variation in parental health, is a parameter of interest. We find that there is a clear and robust North-South gradient in the (positive) effect of parental ill-health on the probability of daily care-giving. There is a strong negative correlation between poor health of parents and employment of daughters, but this is not so robust to controls for the human capital of daughters. The effects linked to longitudinal variation in the health of parents are stronger than those linked to cross-sectional variation. The aggregate loss of employment that can be attributed to daily informal caregiving seems negligible in northern and central European countries but not in southern countries, and larger and significant impacts are found for particular combinations of daughter characteristics and parental disability conditions.

*We would like to thank Olympia Bover and Agar Brugiavini for insightful discussions and Manuel Arellano and Stephane Bonhomme for useful conversations. This paper is based on and supersedes earlier work entitled "Caring for Parents and Employment Status of European Middle-Aged Women" by Crespo (2008), which was part of her Ph.D dissertation at the University of Alicante. This author would also like to thank Lola Collado, Raquel Carrasco, Martin Browning, Edward Norton, Frederic Vermeulen, Ernesto Villanueva, Francois-Charles Wolff and Julie Zissimopoulos for helpful comments and suggestions. Thanks are also due to Dimitrios Christelis, Marcel Das, Maite Martínez-Granado, Edith Madsen, Connie Nielsen, Omar Paccagnella, and Roberta Rainato for their help with the data. The usual disclaimer applies. Financial support provided by IMSERSO for this research project is also gratefully acknowledged.

Keywords: informal care, employment, instrumental variables, treatment effects.
JEL: J2, C3, D1

1 Introduction

Population ageing is one of the most important demographic changes and challenges in all European countries. As a result of ageing the demand for care by the elderly is already very high and is expected to increase in the future. Regarding how the disabled elderly get their care, it is also well know that the family represents one of the most important sources of help, specially daughters in their mature age (Attias-Donfut et al. (2005)). In this paper we use recently released data from the first two waves of the Survey of Health, Ageing and Retirement in Europe (SHARE) to study the prevalence of informal caregiving to disabled parents by their mature daughters across European countries, as well as the effect of intense caregiving on the employment status of the daughters.

Evaluating the prevalence of women who take up the caregiving of their elderly and the opportunity costs that this may represent for them in terms of reduced employment is of policy interest in the design of optimal public long-term care systems and in the implementation of programs to support informal caregivers. Furthermore, the analysis of this question across European countries is of particular interest. On the one hand, the results provided by the European Commission and the Council (2003) show a substantial degree of heterogeneity among European countries with respect to the availability and generosity of public formal care services and long-term care benefits, with the northern countries having extremely generous and universal long-term care systems and the southern countries covering only basic needs of the poorest elderly. On the other hand, there is an important difference in the degree of labour force attachment and the level of education that runs from northern to southern countries with northern mature-aged women having much higher employment rates. These two factors are important source of variation for the question under study. For example, one may hypothesize that variation in the availability of alternative sources of caregiving may lead to variation in the prevalence of women willing to undertake informal care. Furthermore, a stronger labour force attachment may be reflected in a lower prevalence of informal caregivers but also in higher opportunity costs in terms of reduced employment for caregiver women. The aim of this paper is to exploit this cross-country variation represented in the SHARE data to gain new insights into the relationship between parental ill health, informal caregiving and labor supply of mature European women.[1]

[1] This paper uses data from SHARE Waves 1 & 2 as of December 2008. SHARE data collection in 2004-2007 was primarily funded by the European Comission through the 5^{th} and 6^{th} framework programme (project numbers QLK6-CT-2001-00360;RII-CT-2006-062193; CIT5-CT-2005-028857). Additional funding came from the US National Institute on Aging (grant numbers U01 AG09740-13S2; P01 AG005842; P01 AG08291, P30 AG12815; Y1-AG-4553-01; OGHA 04-064; R21 AG025169) as well as by various national sources is gratefully acknowledged (see http://www.share-project.org for a full listing of funding institutions).

Literature review: Most of the studies in the literature analysing the effect of informal caregiving on labor supply refer to the US (i.e. Ettner (1995, 1996), Johnson and Lo Sasso (2000), Pezzin and Schone (1999), Wolf and Soldo (1994)). Furthermore, the evidence provided by these studies is mixed. On the one hand, some of them conclude that there exists a negative effect of caregiving on labour supply. For example, Ettner (1995) uses data from the Survey of Income and Program Participation (SIPP) for the period 1986-1988. Applying an instrumental variable (IV) technique to control for the potential endogeneity of caregiving, she finds that living with a dependent parent has a significantly negative effect on female labour supply. Ettner (1996) distinguises between care provided to coresidential and non-coresidential parents. Using data from the 1987 National Survey of Families and Households (NSFH) she concludes that caregiving activities do not have a significant negative effect on male labour supply whereas female labour supply is only significantly negatively affected by the caregiving activities to parents not living at home. Johnson and Lo Sasso (2000) estimate a simultaneous panel data model of annual hours of paid work and the provision of time assistance to parents. They use a sample of men and women aged 53 to 65 drawn from the second and third waves of the Health and Retirement Study (HRS). Their results suggest that time devoted to parent caregiving significant and substantially reduces labour supply for both women and men.

On the other hand, other researchers do not find any statistically significant effect of caregiving on labour supply. For example, Wolf and Soldo (1994) estimate a simultaneous equations model of employment, hours of work, and the provision of care to an elderly parent. They also use data drawn from the 1987-88 National Survey of Families and Households (NSFH) but they focus on a sample of married women. Even though labour supply behaviour of married women is usually more elastic, they find no evidence of reduced propensity to be employed or reduced conditional hours of work due to the provision of care to frail parents. Finally, Pezzin and Schone (1999) estimate a simultaneous, multi-equation, endogenous switching model of informal care to elderly parents, coresidence, and female labour supply using data from the 1986-1987 matched Hebrew Rehabilitation Center for the Aged (HRCA) Survey of the Elderly in Massachusetts and HRCA-NBER Child Survey. They find that the correlation between informal care and labour force participation was negative but small, which reflects a modest trade-off between both variables for adult daughters. However, the possibility of extending their results is limited since their data consist of a small sample of parent-daughter pairs from a single state.

For Europe, there is less work on this topic. Heitmueller and Michaud (2006) develop a multivariate dynamic panel data model to identify the causal link from informal care to employment for men who are aged 16 to 64 and women who are aged 16 to 59 in England. Using data from the British Household Panel Study (BHPS) from 1991 to 2003 they find that caring only reduces employment probabilities by up to 6 percentage points for individuals caring within their own homes and no significant effect is found for the extra-residential carers. These small effects could be driven by the fact that no information about the intensity of the care is considered in the analysis. Therefore, they may be including caregiving activities that are not very time consuming and do not represent a significant burden for caregivers. From a cross-country perspective, Spiess and Schneider (2003) use data drawn from the European Community Household Panel for 12 EU-countries to decribe the relationship between the changes in weekly work hours and changes in caregiving for women aged between 45 and 59 years old. They show that a change in work hours is significantly and negatively associated with the start or the increase of hours of informal caregiving only in northern countries. However, their analysis does not take into account the potential simultaneity of these two decision variables. Finally, Bolin et al. (2008) use the first wave of the SHARE data to estimate the effect of hours of informal care provided to elderly parents on employment, hours of work and wages for men and women aged between 50 and 64 years old. Their results imply that one extra (weekly) hour of informal care has a negative effect on the probability of employment of -0.032 percent and -0.028 percent for men and women, respectively, and significantly different from zero at 10 percent level. In their main specification informal care is found to be exogenous in the employment equation and it is assumed that its is homogenous for all countries. When including group dummies to account for differential effects relating to the North-South gradient in the availability of publicly financed long-term care services their estimates do not reveal any patterns that can be linked to institutional differences.

Contribution: In this paper we revisit the estimation of the effect of the provision of informal care to elderly parents on employment of their daughters.[2] We make the following contributions: 1) Our empirical work is based on an instrumental variable-treatment effects framework (IV-TE), as in Imbens and Angrist (1994). [and Heckman and Vytlacil (2005).] The IV-TE framework emphasizes heterogeneity of treatment effects and shows what causal parameters can be (non-parametrically) identified by IV estimates when selection into treatment is not random. This is relevant in our context because, given the extent of variation in labor market behavior of mature daughters within and across European countries, it is highly implausible that the effect of providing informal care on employment is homogenous. 2) We provide a simple model of time allocation decisions of the daughters between labour supply and informal care which includes the utility derived from the well being of the care recipient. We use the model to make a link to the empirical IV- treatment effects framework and to improve the discussion of what are causal parameters of interest and the different sets of assumptions needed to estimate each of them. The model predicts that the reservation wage when caring is higher than when not caring. Thus the 'treatment effect' of daily caring on employment is likely to be non-monotonic in potential wages, i.e., zero for low and high wages and -1 between the two reservation wages. We argue that the Local Average Treatment effect of daily care on labor supply, as identified by shocks to parental health, is a parameter of interest. We are interested in LATE's two components as much as on the treatment effect itself, i.e. we focus on the impact of parental ill health on the employment rates of daughters, and on its impact on the probability of daily caregiving (the "first stage"). We decompose the population of daughters into 'always-taker', 'complier' and 'never taker'

[2] A short progress report of the first stages of this research can be found in Crespo and Mira (2008) which was prepared for the First Results Book that was released with the second wave of SHARE.

subpopulations (Imbens and Angrist (1994)) based on the relationship between informal care and parental health. We argue that these decompositions are of substantial interest in this context and we note that LATE's two components can be consistently estimated even if the parental health instrument does not satisfy exclusion restrictions. 3) The comparison across country groups defined by the afore-mentioned North-South gradient in the availability and generosity of public long-term care benefits has center stage in our paper. In particular, we perform all our estimations separately for each group of countries rather than including a few interaction effects in a pooled estimation. We use the behavioral model as a guide to interpret and rationalize the differences found across countries. 4) We exploit the richness of the SHARE data, including its longitudinal dimension and the availability of multidimensional measures of the health of parents and of the care they receive from sources other than their daughter.

Our analysis is limited to binary indicators of labor supply and informal care. Our measure of labor supply is an employment indicator, and we focus on informal care provided on a *daily* basis because this help is much more likely to represent a significant burden competing with labor supply in the time allocation of these women.[3] We show that these extensive margins are very important in the data, so the alternative would be to consider mixed discrete - continuous models for both outcomes. This alternative seems infeasible or much more difficult because we aim to use an empirical IV-TE framework and to provide careful interpretation of the results within an explicit behavioral model.

Overview of main findings: 1) There is a clear North-South gradient in the (positive) effect of parental ill-health on the probability of daily informal caregiving by daughters. This gradient is robust to different specifications and samples and mirrors the North-South gradient in the availability of public long-term formal care. Most women in all countries will never take up daily caregiving, but in Southern countries there is a sizeable group who do provide daily care. 2) We also observe a strong negative correlation between poor health of parents and labor supply of daughters with a North-South gradient, but the employment effect is less robust to controls for the human capital of daughters. 3) The employment and daily caregiving effects linked to longitudinal variation in the health of parents are stronger than those linked to cross-sectional variation. 4) The aggregate loss of employment for that can be attributed to daily informal caregiving for women between ages 50 and 60 seems negligible in northern and central European countries but not in southern countries. However, estimates of employment effects and LATE are not very precise. 5) Larger and significant impacts are found for particular combinations of daughter characteristics and parental disability conditions, e.g. low-skilled daughters who work but are close to the margin of non-participation, or daughters whose parents suffer from dementia. Our model offers plausible interpretations of most of our findings.

The structure of the paper is as follows: Section 2 describes the data: samples, variables, descriptive statistics and correlations. Section 3 contains the conceptual framework: we present a simple time allocation model and we discuss the parameters of interest, the

[3]This is an advantage of our paper relative to other studies which had no information on the intensity of informal care (Ettner (1996), Heitmueller and Michaud (2006), Wolf and Soldo (1994)).

assumptions needed to estimate them and the predictions of the model about differences across country pools. Section 4 reports the empirical results: first, evidence based in cross-sectional variation in the health of parents; second, evidence based on longitudinal variation in the health of parents; and third, evidence based on multiple measures of parental disability.

2 The Data

The data used in this analysis comes from the first two waves of SHARE. Specifically, we use data from Wave 1 and Wave 2, that were collected by personal interviews in 2004 and 2006/07 respectively. The main purpose of this survey is to provide detailed and specific information about the living conditions of people aged 50 and older for several countries in Europe. SHARE collects information on demographics, employment and retirement, physical and mental health, social support and networks, housing, income and consumption, both at household and individual level.

The target population of this study is women at risk of having to combine the provision of care to elderly parents and paid employment. We are interested in women because daughters are often named as the most important source of help by elders. This is supported by Figure 1 which shows how daughters in their mature age become the main caregivers of the elderly in the family in northern, continental and southern European countries (SHARE, 2004).[4] Specifically, we focus on women aged between 50 and 60 with at least one living parent at the moment of the interview. Women in this range of age are the most likely to be involved in personal care mainly with their elderly parents (Attias-Donfut et al. (2005)) and, at the same time, they can be still part of the labour force. We exclude women older than 60 to minimize issues related to retirement decisions.[5]

Samples: Given the information provided by SHARE one may think of drawing two different samples of women with elderly living pa rents. The first possibility is to consider a sample of women between ages 50 and 60 who are age-elegible respondents of the survey (the "daughers-sample"), who provide some information on their living natural parents, such as their age, health status, and closeness of residence. The second possibility is to construct a sample of women in the same age interval who are daughters of (older) age-eligible respondents (the "parents-sample"). In this case, the respondents are the elderly parents. This sample can be identified since each respondent at the couple level provides some information about their living children (gender, age and residence closeness, type of children, marital status, frequency of contact, occupation status, education and number of children).[6] Both samples are potentially useful for analysing the question at hand

[4] In SHARE, both members of the couple provide information about their living parents. However, in this analysis we do not consider caregiving to parents-in-law given that a substantial percentage of spouses/partners did not complete the interview in countries like Italy and Spain.

[5] We exclude from the sample those women who report to be permanently sick or disabled or retired as their current job status.

[6] The information about type of children, marital status, frequency of contact, occupation status, education and number of children is only asked about up to four children. When there are more than

since they are composed by women from the same cohorts and population. However, the variables available in each case are not exactly the same. Each of these samples presents some advantages and disadvantages. On the one hand, in the "daughters-sample" there is better information on the daughter including age, education, current marital status, health status, income, living children and siblings, employment status and hours worked, and informal care given. With respect to their parents we observe age, proximity, and a categorical variable on their general health status as perceived by the daughters. On the other hand, the main advantage of the "parents-sample" is that it provides comprehensive information reported by the elderly parents themselves on their health status and their access to different sources of care, in addition to informal care provided by their daughter. In addition to the self-reported general health, more objective health measures based on self reported diagnosed chronic conditions, functional limitations, ADL and IADL limitations, symptoms and mental health are available. This allows us to contruct more detailed parents' health indicators. Besides, we observe each parent's age and income, and the selected daughters' employment status and age, education, current marital status, children, siblings and proximity. However,in this sample we do not observe the daughters' own health status or financial situation. We decided to use the "daughters-sample" for the main part of our analysis because the most relevant information relating to employment and caregiving decisions is reported by the daughthers, who are the decision makers in our analysis. Nevertheless, the main results are replicated using the "parents-sample" and exploting additional information included therein.[7] The results for the parents sample are shown in section 4.3.[8]

Country pools: Since samples sizes are too small at the country level we group countries according to the availability and generosity of public formal care services and long-term care benefits. The results provided by the European Commission and the Council (2003,a) show that there exists a substantial degree of heterogeneity among European countries with respect to the availability and generosity of public formal care services and long-term care benefits. On the one hand, northern countries like Denmark, Sweden, and The Netherlands are characterized by extremely generous and universal long-term care systems. In fact, these countries exhibit the highest levels of public expenditure on long-term care as a percentage of GDP (from 3 percent in Denmark to 2.5 percent in The Netherlands). On the other hand, southern countries like Greece, Italy and Spain have

four children, the selection is not random but follows a set of criteria. First, children are sorted in ascending order by minor, proximity, and birth year, where minor is defined as 0 for all children aged 18 and over and 1 for all others. Second, the first four are picked. When all sorting variables are equal, a child is selected randomly.

[7] Another important advantage of the "daughters-sample" is that it is much easier to build longitudinal linkages between waves since in this sample the daughters are the respondents of the survey. However, for the "parents-sample", this linkage is very complicated since children do not have to be reported in the same order and do not have identification numbers to be uniquely identified between waves. Therefore, the longitudinal analysis of the data is just based on the "daughters-sample".

[8] For the "daughters-sample", we use data from Wave 1 release 2.0.1 and Wave 2 release 0. For the "parents-sample", we use data from Wave 2 release 1.0.1.

been characterized until very recently by social assistance systems providing public care to meet very basic needs of poor elderly. Therefore, in these countries the public provision of formal care has been very limited in quality and quantity. In fact, according to European Commission and the Council (2003,a)'s results, these countries exhibit the lowest levels of public expenditures on long-term (0.6 percent for Italy and even lower for Greece and Spain). Moreover, the informal help provided by the family, especially by women, has been the most important pattern of social support to the elderly in these societies. Finally, continental European countries like Austria, Belgium, France, Germany and Switzerland fall in an intermediate situation. Regarding the level of public expenditure on long-term care as a percentage of GDP, this indicator ranges from 1.2 percent in Germany to 0.7 in Austria and France. This North-Continental-South gradient in the patterns of social support to dependent elderly is also reflected in Figure 2, based on data from the first wave of SHARE. In particular, this figure shows striking differences in the use of formal care services (i.e, being in a nursing home or receiving formal care at home) in these three groups of countries. In the northern countries, more than 80 percent of respondents aged 80+ who report receiving help in a regular basis had formal care. In continental countries, these were 70 percent, and in southern countries, this percentage does not reach 30 percent. An inverse picture is obtained for the use of regular informal care by these elders. Based on this we group the SHARE longitudinal countries into the following pools: the northern countries (NC) including Denmark, Sweden and The Netherlands; the central countries (CC), including Austria, Belgium, France, Germany and Switzerland; and the southern countries (SC) including Greece, Italy and Spain.

Main variables: The main variables of interest are those that measure the daughters' decisions about labour supply and caregiving activities. Regarding the participation decision, SHARE respondents are asked about their current job situation. Based on this information, the employment decision is defined by an indicator variable, *LP,* that equals 1 if the woman reports to be employed or self-employed (including working for family business) and 0 otherwise.[9] Even though those who are working are also asked about the number of contracted and usual weekly hours of work in all jobs, we will only focus on the employment decision. The main reason for this is that the extensive margin is the most important source of variation in labor supply. This is specially the case for the Mediterranean countries given lower labor market attachment and the especially high prevalence of full-time jobs with fixed working-schedules. To assess whether the intensive margin of labour supply may play an important or different role in these three groups of countries, Table A1.1 and Figure A1.1 in Appendix A1 show some summary statistics and kernel density estimates of the distribution of weekly hours worked conditional on participation across country pools. From this comparison we can highlight several facts. First, differences in weekly hours worked are negligible between northern and continental countries. Second, differences between the former and southern countries are small and

[9] Our *LP* binary indicator is equal to 0 for unemployed women since our focus is on the employment decision and unemployment is not modeled in our theoretical framework given its low prevalence in our sample (5 percent for NC, 8.7 percent in CC and 3.8 percent in SC). Therefore, the variable *LP* should be interpreted as women's employment status taking into account these considerations.

attributable to a smaller prevalence of part-time in Mediterranean countries.[10] However, variation in the intensive margin does not seem to be crucial given these figures.

Parental caregiving activities are identified from the information reported by each respondent about the provision of help to elderly parents living inside or outside the household in the last twelve months. This help refers to personal care, practical household help, and help with paperwork. Respondents that reported to have provided care to someone living outside the household also report information about the frequency of this care (i.e., almost daily, almost every week, almost every month, less often) and its intensity (hours). For those that reported to have provided care to an elderly parent living in the same household, it has to be daily because a 'daily' filter is included in the opening question but no information on hours is reported in this case. Table 1 shows the prevalence of caregiving activities in our sample for the three groups of countries. The variable *Caregiver* indicates whether the woman has provided any help to at least one elderly parent in the last 12 months regardless of the frequency of this activity.[11] We observe that the prevalence of being a caregiver is high. Furthermore, according to this measure northern women are more likely to be caregivers whereas southern countries show the lowest percentage. However, information on the intensity or the frequency of the provision of informal care may be crucial in this context to focus on those caregiving activities that are more likely to represent a significant burden for these women. In line with this, the top panel of Table 1 provides the percentages of women who report providing care to elderly parents on a daily or weekly basis and of those that do it daily within the sample of caregivers. These are the so-called *intensive caregivers* (IC). Once we condition on being a caregiver a different gradient emerges among these three groups of countries. Specifically, the gradient runs clearly from the southern countries where more than 80 percent of women who report taking care of elderly parents have done it on a daily or weekly basis to the northern countries where only 41 percent do so regularly. This suggests that women in the southern countries are much more likely to be involved in intensive caregiving activities. However, the bottom panel of Table 1 shows that this measure of intensive caregiving may still not be homogeneous since within the sample of daily/weekly caregivers only 12 percent of women in northern countries are daily caregivers whereas this percentage is higher than 50 percent in southern countries. Therefore, hereafter in our analysis we define intensive caregivers as those who have provided care on a daily basis in order to obtain a homogeneous measure of the burden of caregiving. To further check whether daily caregiving implies similar burdens in terms of

[10] Regarding part-time, the percentage of women who work between 10 and 20 hours per week in the sample of workers is the following: 15.36 for northern countries, 18.89 for continental and 8.72 for the southern.

[11] One may argue that co-residential and extra-residential care should not be pooled in the same caregiving measure. However, in our case this does not constitute a major limitation since in our sample of mature women the number of respondents that report to provide care to a coresident elderly parent is very low. In northern countries the fraction of respondents that gave informal care to a parent in the household was zero whereas in continental countries and southern countries is 1.04 and 2.48, respectively. By country, the proportion ranges from zero in Denmark, Sweden and The Netherlands to near 6 percent in Spain, which presents the highest rate. This is consistent with Bolin et al. (2008).

daily hours in these three pools of countries, Table A1.1 and Figure A1.2 in Appendix A1 show some summary statistics and kernel density estimates of the distribution of weekly hours of care conditional on proving care daily to at least one parent living outside the household. In particular, these figures show that weekly hours of care for these caregivers are somewhat larger in the South, but distributions are not very different among the three pools of countries.

Table 2 shows the joint distribution of the employment and the intensive caregiving decisions. This gives a first insight about the relationship between both variables. In particular, these simple cross-tabulations show that in all countries women who take up intensive caregiving to an elderly parent are less likely to be employed on average than women who do not. This difference is specially remarkable for continental countries where 57 percent of daily caregivers are employed, compared to 71 percent among non-daily caregivers.

Of central importance for this study is the use of some measure of the health status of elderly parents as an instrumental variable for the caregiving decision. Specifically, SHARE asks respondents to rate their living parents' health status according to a categorical variable. However, different versions of this item are applied in Wave 1 and Wave 2. Whereas in Wave 1 the EU (European) version (Very Good, Good, Fair, Poor, and Very Poor) is used, in Wave 2 the US (United States) version (Excellent, Very Good, Good, Fair, and Poor) is applied. Based on results shown in Jürges et al. (2007), a simple and quite accurate way of mapping one scale into the other is to collapse the two top categories of the US version as category "Very Good", and the two bottom categories of the EU version as category "Poor". This results in a four-point comparable scale (Very Good, Good, Fair, Poor). In particular, for the "daughters- sample" our instrument is defined by a binary variable, *Parental Health* $(\overline{PH})$ that equals to 1 if at least one parent is in a poor health status. In section 4.1 we show how we use the richer information on the health of parents which is available in the "parents-sample".

Other covariates: Apart from the potential simultaneous relationship between employment and caregiving activities, both decisions are functions of other variables that account for preferences and other daughters' characteristics like education, marital status, children, health status, age, non-labour income, and siblings. Definitions and more specific details about these control variables are provided in Appendix A1.

Table 3 reports the means of the variables used in the analysis for the resulting sample of 2429 women drawn from Wave 2. These results show a remarkable North-Central-South gradient in some characteristics of these women in their mature age. For example, regarding employment this difference runs from the highest employment rates in northern countries (83 percent) to the lowest rates in the southern countries (45 percent). A similar gradient is observed for education where northern women are more educated (the percentage of women with the lowest level of education is 3.6 in the northern area and 32.2 in the southern area whereas the percentage of the highest educated women is 47.4 in the northern area and 20.7 in the southern area), and for health where the percentage of women reporting an excellent or very good health status is also substantially higher in northern countries. With respect to income variables, northern and continental women have on average higher non-wage income. However, there is not a remarkable difference in the prevalence of parents in bad health. Overall, around a 20 percent of women have at least one parent in this status.

Next, we compare the employment status and other individual characteristics between the sub-samples of daily caregivers and non-daily caregivers. The results from this comparison are shown in Table 4.[12] As we noted above, daily caregivers are less likely to be employed than women who do not provide daily care. They are also more likely to have parents in poorer health status. With respect to characteristics related to labour market attachment, we can see that northern and southern daily caregivers are less educated on average than non-intensive caregivers whereas no difference is found for continental women. Moreover, daily caregivers are more likely to be married than non-daily caregivers in the three pools of countries. The availability of alternative sources of care are measured by the variables *Sisters* and *Brothers*, which indicate the number of living sisters and brothers, respectively. Regarding this, our cross-tabulates suggest that daily caregivers have less sisters on average whereas the same result holds for brothers only in continental and southern countries.

Finally, given that we will exploit the health status of elderly parents as a source of variation in the care-giving and labour supply choices, we compare the prevalence of these two decisions and other individual characteristics between women with parents in poor health and those without parents in such situation. Results are shown in Table 5. From these simple cross-tabulations we can see that in all pools women with parents in bad health are less likely to be employed. This difference is particularly remarkable for southern countries where 38 percent of women with parents in poor health are employed, compared to 47 percent for women with no parents in such health status. With respect to the provision of intensive informal care, the table clearly shows that there exists a positive relationship between having parents in bad health and providing daily care for all groups of countries, especially for the South.

3 Conceptual Framework

3.1 A simple behavioural model

The relationship between employment and caregiving can be studied using a standard model of the daughter's time allocation decisions. The daughter is altruistic towards her parent, deriving utility from own consumption and leisure and from the well being of the parent as follows:

$$U = C - \alpha_{12}C^2 + \alpha_2 W_p + \alpha_{31}\widetilde{h} - \alpha_{32}\widetilde{h}^2 + \alpha_4 C W_p + \alpha_5 C\widetilde{h} + \alpha_6 \widetilde{h} W_p \tag{1}$$

where C is consumption, W_p is parental welfare, $\widetilde{h}$ is leisure. Parental welfare is

[12] We should note that some of these descriptive results could be affected by the extremely small size of some samples, especially for daily caregivers in northern countries.

$$W_p = f(PH, IC, FC, OC) \tag{2}$$

where PH is a binary indicator of parental ill health, IC is informal care provided by the daughter (time) and FC is formal care purchased by the daughter. The variable OC represents other inputs into parental welfare which are not directly controlled by the daughter, e.g., any formal care not paid by the daughter, or informal care provided by siblings, etc. The derivatives of f are $f_1 < 0$, $f_2 > 0$, $f_3 > 0$, and we assume that the second cross derivative $f_{21} > 0$. The ill-health indicator PH should be interpreted as a summary measure of disability or "need" of care, which is unaffected by IC itself.[13]

The daughter's time endowment T is allocated to $\widetilde{h}$, IC and market work h. An implicit assumption is that the disutility of work and informal care are the same. The budget constraint is

$$C = y + wh + \beta_1 IC - \beta_2 FC \tag{3}$$

where y is non-labour income, w is the daughter's wage, β_1 represents any transfers received by the daughter from the state or from her parent in exchange for providing informal care, and β_2 is the price of formal care paid for by the daughter for her parent.

In this paper, we focus on the daughter's binary choices $IC \in \left\{0, \overline{IC}\right\}$ and $LP \in \left\{0, \overline{h}\right\}$, where $\overline{h}$ is the fixed proportion of time engaged in market activity and $\overline{IC}$ is the proportion of time devoted to provide informal care. Therefore, in the discrete choice version of this model, the daughter makes the binary choices IC and LP, as well as FC which we need not treat as binary, taking OC and PH and all prices as given.[14] Because our focus is on the binary choices and not on formal care it will be useful to define an 'indirect' formal care function which gives the optimal choice of formal care conditional on any pair $(LP; IC)$. Let this function be $FC(LP, IC; PH, y, w, OC; \alpha, \beta)$.

3.2 Discussion of parameters and empirical models

Heterogeneity: The optimal decision rules for employment and care are a pair of binary-valued functions with parameters and arguments $(\alpha, \beta, \overline{h}, \overline{IC}; PH, y, w, OC)$. In our empirical work we do not perfectly observe (w, OC). Our econometric models approximate decision rules as functions of parental health PH and a vector of controls X which includes non labor income y, preference and price shifters, observable determinants of wages (e.g.,

[13] Absence of reverse causality is a maintained assumption. As explained in section 2, PH can be obtained from the answer to a question on overall health, i.e., a 'subjective' measure of health. Or alternatively, we may observe a vector of measures of parental disability and use all of them as instruments for caregiving. In that case the conceptual framework described in this section is still useful if we reinterpret PH as a binary variable which takes values 0 or 1 for particular subsets of values of the vector of instruments reflecting different 'need' levels. This case is considered in section 4.3.

[14] The daughter's choices and OC may be jointly determined as the outcome of a game played by different units of an extended family, e.g. see Stern et al (2009). If outcomes are determined in Nash equilibrium, then OC would be an argument of the daughter's decision rule but other controls for the characteristics of the extended decision unit might be also relevant.

education) and observables relating to other sources of care (e.g., number of siblings). Conditional on (PH, X), the data give joint probability distributions for the discrete pair $(LP; IC)$. We interpret these distributions as the integrals of the model's decision rules over the distribution of unobserved components of $(a, \beta, \overline{h}, \overline{IC}, w, OC)$. All the empirical work we report in Section 4 consists of estimates of the impact of PH in these decision rules, based on non-parametric and parametric approximations, and ratios of these estimates which are local average treatment effects. The rest of this section uses the behavioral model to guide a detailed discussion of the assumptions needed to give a causal interpretation to these estimates and to make predictions about their sign and size. We argue that these estimates can answer the following questions of interest.

Questions & parameters of interest: 1) What is the effect of a change in parents' health status on daily caregiving and employment decisions of their mature daughters? 2) Does daily caregiving reduce employment? Can all the employment loss be attributed to ill-health of parents, or are some daughters providing daily care to parents in reasonably good health? 3) Are the answers to these questions different across our pools of countries - and why?

IV-treatment effects and the behavioral model: In order to further clarify the questions we have posed and the interpretation of our estimators it is useful to link our behavioral model to the framework described in Imbens and Angrist (1994) on the identification and estimation of treatment effects using a binary instrument.[15] In our case, the 'daily care' variable IC is the indicator of treatment and the parental ill-health indicator PH is the instrument. Researchers have used this and other instruments such as the number of siblings to identify the effect of caregiving on employment. We argue that, even if other instruments are thought to be 'relevant' and 'valid', the causal effects identified by different instruments are not on an equal footing in terms of their ability to inform policy discussions. The PH instrument is more important because the opportunity costs of caregiving are more relevant when take-up of care-giving is a direct consequence of parental disability.

The treatment effects framework defines causal effects in terms of potential outcomes or counterfactuals without relying on any functional form or distributional assumption. Define $LP(1)$ as the employment decision of a woman if she were to provide care. Similarly, $LP(0)$ represents the woman's employment decision if she does not provide care. Specifically, $LP(1)$ and $LP(0)$ are called *potential outcomes* or *counterfactuals* because they are not observed together for the same individual. For instance, if $IC = 1$ turns out to be chosen we observe $LP(1)$ but not $LP(0)$. Our behavioral model can be mapped into this framework as follows. Let $U(i, j)$ be the utility derived from choosing $LP = i$ and $IC = j$. In order to evaluate $U(i, j)$ we need to know the values of all structural parameters in (1)-(3) and the indirect formal care function $FC(.)$. The optimal (LP, IC) pair is obtained by comparing the four utilities $U(1, 0), U(1, 1), U(0, 0), U(0, 1)$. Instead, the potential outcome $LP(1)$ is obtained from the comparison of $U(1, 1)$ and $U(0, 1)$, and

[15] Our analysis in this section owes to the discussion of IV estimation of the effect of fertility on labor supply contained in Rosenzweig and Wolpin (2000).

the potential outcome $LP(0)$ from the comparison of $U(1,0)$ and $U(0,0)$. Given X, the distribution over preference parameters etc and the behavioral model determine a distribution of potential outcomes. To complete the framework in I&A, define $IC(1)$ and $IC(0)$ as potential outcomes for the treatment status given the instrument. Again, in our model $IC(1) = 1$ if $\max[U(1,1), U(0,1)] - \max[U(1,0), U(0,0)] > 0$, and $IC(1) = 0$ otherwise, where all for utilities are evaluated for $PH = 1$. The instrument PH is valid if, conditional on X, the two pairs of potential outcome ($LP(1)$, $LP(0)$) and ($IC(1)$, $IC(0)$) are independent of PH. Every woman in the population belongs to one of four 'compliance types': *always takers* ($IC(1) = IC(0) = 1$), *never takers* ($IC(1) = IC(0) = 0$), *compliers* ($IC(1) = 1, IC(0) = 0$) and *defiers* ($IC(1) = 0, IC(0) = 1$). The instrument is called monotone if $IC(1) \geq IC(0)$. This means that any woman who provides care when her parents are not in bad health will also provide care if at least one parent experiences this contingency. Notice that this implies the non-existence of *defiers*.

Imbens and Angrist (1994) show that if the 'treatment' regressor is binary and if there exists an instrument which is binary and monotone, an IV estimate can be interpreted as a *local average treatment effect* (LATE) specific to the instrument. More formally, the LATE parameter is given by

$$\widehat{\beta}_{IV} = \widehat{E}[LP(1) - LP(0) \mid IC(1) - IC(0) = 1] \tag{4}$$

which is the average effect of daily care on the probability of employment for the subpopulation of *compliers*.[16] These are the women whose caregiving decision is changed by the value of the health instrument. In particular, they would not provide daily care in the absence of parents in bad health, but they choose to provide care when there is such a situation.

Is PH a monotone and valid instrument if the data are generated by our behavioral model? We discuss: a) the assumptions that need to be imposed on the behavioral model and our empirical approximations to it; b) the plausibility of these assumptions.

Assumption - Exogeneity of PH: The distribution of $(a, \beta, \overline{h}, \overline{IC}, w, OC)$ conditional on X is independent of PH. This assumption requires that parental health status not be correlated with unobservable determinants of daughters' employment or caregiving decisions relating to preferences or human capital or labour market attachment. It seems likely that health capital of parents is in fact correlated with the human capital of their daughters. If so, it is more difficult to draw causal inferences from the correlations between

[16] In the absence of conditioning variables X, very simple IV regression techniques can be used to compute the LATE parameter. In particular, in the linear regression of LP on IC plus a constant term the IV or Wald estimate of β, the regression coefficient of IC is

$$\widehat{\beta}_{IV} = \frac{\widehat{E}(LP|PH=1) - \widehat{E}(LP|PH=0)}{\widehat{E}(IC|PH=1) - \widehat{E}(IC|PH=0)} = \frac{\overline{LP}_1 - \overline{LP}_0}{\overline{IC}_1 - \overline{IC}_0}$$

where in the numerator $\overline{LP}_1$ is the average of LP for those women with at least one parent in bad health and $\overline{LP}_0$ is the average of LP for those women with no parents in this situation. Likewise, the denominator is the difference in the proportions providing care with and without parents in bad health.

labor supply, daily care and parental health. Exogeneity is plausible only if X includes apropriate controls for the daughter's own human capital.

Assumptions - Exclusion restrictions on PH: Let $U(1,0;IC;.) = U(1,IC;.) - U(0,IC;.)$ be the utility difference between working and not working, conditional on the choice of IC. In order for PH to be a valid instrument, in addition to exogeneity we need an exclusion restriction to be satisfied. In particular, the utility difference $U(1,0;IC)$ should not depend on PH. The decision to work trades off the marginal utility of increased consumption against the marginal disutility of reduced leisure. If utility is concave in leisure, the disutility of reduced leisure from work is even greater if the woman is allocating time to caregiving, an this is the main mechanism through which caregiving reduces the propensity to work. An analysis of the utility difference shows that the exclusion of PH requires the following:

Separability: $\alpha_4 = \alpha_6 = 0$, i.e., the marginal utilities of consumption and leisure do not depend on parental welfare.[17]

Exclusions in formal care: $FC(LP, IC, PH;.) = FC(IC)$. Conditional on her choice of informal care, spending by the daughter on formal care does not vary with employment or with parental health.[18] An example of behavior that would violate this assumption is as follows: suppose a daughter decides not to provide care; having decided this, if her parent is in poor health and does not have another source of care she will pay for formal care but she can only afford it if she is working. Then, this would introduce a positive bias in the impact of PH on employment. This type of behavior seems more likely to occur in southern countries. Even if failures of exclusions are plausible the bias in IV estimates is likely to be small in practice to the extent that it is unusual for daughters to pay for formal care out of their own pocket.[19] Finally, note that any alternative instrument operating through the production function of parental welfare, such as the number of siblings, will require similar exclusion restrictions in order to be valid.

Assumptions - Monotonicity: *The treatment IC is monotone in the instrument PH.* Monotonicity is highly plausible as long as $f_{21} > 0$, that is as if the marginal productivity of the daughter's daily care increases when her parent's health deteriorates. Consider the following example: A parent in "not too good but not too bad" health which we classify as $PH = 0$ receives daily care from his/her daughter. The parent's condition

[17] If they do, the marginal utility and the marginal disutility from working will depend on parental welfare W_p, and the instrument PH has a direct effect on this.

[18] Consider first the exclusion of LP. If spending by the daughter on formal care depends on whether she works or not, even after conditioning on IC and PH, then the gain from working will include an increase in the welfare of parents the size of which depends on their health. To see why the exclusion of PH in the $FC()$ function may also be needed, note that the marginal utility of consumption from working will in general depend on the 'baseline' level of consumption when not working, and this in turn depends on formal care. Therefore, we need either $FC()$ to be independent of PH given IC, or else additional restrictions on utility such as $\alpha_5 = \alpha_{12} = 0$.

[19] The exclusion restriction on $FC()$ is not easy to test empirically. First, beacuse it would require data on expenditures on formal care by the daughter. Second, even if such data were available the $FC()$ function describes potential outcomes so estimating the coefficients on LP and PH poses the same kind of challenges we are trying to deal with in first place.

deteriorates to the point that the parent is institutionalized and ceases to receive daily care from the daughter. This 'defier' behavior is not implausible and it would seem to violate monotonicity. However, the following considerations should also be taken into account: First, a suitably "conservative" definition of "not in bad health" ($PH = 0$) would essentially rule out this defier behavior. Second, the behavioral model conditions on any sources of care OC which are taken as a given by the daughter. Therefore, behavior in the preceding example violates monotonicity only if we do not control for those other sources or if it is the daughter who pays for care at an institution.[20]

Estimation of parameters of interest:

Most papers cited in our review of the literature have specified a reduced form parametric approximation to the model of potential outcomes $(LP(1), LP(0)|X)$. Usually the focus of the empirical investigation was the sign and significance of the coefficient on IC and the calculation of an 'average treatment effect' of care on employment. In order to deal with the endogeneity of care a number of instruments were proposed and their relevance and validity were informally discussed. A limitation of this approach is that it does not pay much attention to selection into 'treatment'. In this context, selection into treatment is important for two reasons. First, the prevalence of daily care can be linked more naturally to the decision rules of a behavioral model. Its empirical investigation does not require a model of potential outcomes, and it is arguably as interesting as the effect of care on labor supply. Second, if the effect of caregiving on labor supply varies across daughters it is doubtful that an average treatment effect is a parameter of much interest. To see why consider the following examples. Suppose that, if provided, daily caregiving would lead most women to drop out of the labor force but that very few women are actually willing to provide daily care. The average treatment effect would be large but this does not seem very relevant in the sense that very few women will actually change their employment status. Alternatively, suppose a negative shock to the health of their parents leads many women to take up daily care, and about a third drop out of the labor force as a result. In this case the average treatment effect would be smaller but the loss of employment linked to caregiving could be important.

We now discuss what we believe are the parameters of interest and their estimation:

1. Under monotonicity and exogeneity, one can estimate the population proportions of compliers, never-takers and always takers from the population distribution of treatment

[20] More generally, violation of monotonicity will occur if the change in parental health PH leads the daughter to increase spending on formal care FC so that the marginal productivity of her own informal care is lower now with $PH = 1$ than it was before with $PH = 0$. Testing for monotonicity is not simple. One possibility would be to exploit longitudinal variation in PH but a formal test would have to allow for changes in unobservable determinants of the behavior of daughters between waves 1 and 2.

and instrument status:

$$\begin{aligned}\Pr(\text{compliers}) &= \int [E(IC \mid PH=1,X) - [E(IC \mid PH=0,X)]dF(X) \\ \Pr(\text{always takers}) &= \int [E(IC \mid PH=0,X)]dF(X) \\ \Pr(\text{never takers}) &= 1 - \int [E(IC \mid PH=1,X)]dF(X)\end{aligned}$$

This decomposition is interesting for two reasons. First, the sum of always-takers and compliers measures the quantity of daily care services supplied to disabled parents by the population of daughters, given current characteristics of the population and the environment. Second, estimating the mass of always-takers allows us to separate the fraction of daily care services which is induced by true parental disability from that which is not, given the definition of disability implicit in the instrument.

2. The LATE parameter: This is the average treatment effect for compliers, as defined above. The complier subpopulation is of special interest because women who are driven to provide daily care *because* their parents suffer from bad health are the obvious target of any policy aimed at reducing the opportunity costs of informal care. If we consider the controls X suggested by the behavioural model, and if PH is valid as an instrument and monotone, then the average treatment effect for the overall subpopulation of compliers is given by

$$\beta = \int \beta(X)dF(X \mid compliers) = \frac{\int [E(LP \mid PH=1,X) - E(LP \mid PH=0,X)]dF(X)}{\int [E(IC \mid PH=1,X) - E(IC \mid X=0,X)]dF(X)} \tag{5}$$

where the denominator is the proportion of compliers, as shown by Froelich and Melly (2007). This parameter can be estimated as a ratio of two non-parametric matching estimators as follows

$$\widehat{\beta} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{LP_i PH_i}{\widehat{\Pi}(X_i)} - \frac{LP_i(1-PH_i)}{1-\widehat{\Pi}(X_i)}\right]}{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{IC_i PH_i}{\widehat{\Pi}(X_i)} - \frac{IC_i(1-PH_i)}{1-\widehat{\Pi}(X_i)}\right]} \tag{6}$$

where if we consider the numerator and the denominator separately we can think of PH as another treatment indicator and $\widehat{\Pi}(X_i)$, the propensity score, as the conditional probability of receiving this treatment.[21] Alternatively, it would be also possible to consider

[21] The idea of the weighting using the propensity score is to create balance between treated and control units given that the distribution of X may be different in these two groups (Hirano, Imbens and Ridder (2003)).

parametric versions of this estimation based on a biprobit (see more details in Appendix A3).

3. If the exclusion restrictions are not valid but exogeneity is satisfied then the numerator and the denominator of the last expression above still measure parameters of interest, i.e., the causal effect of parental ill-health PH on employment and intensive care. Estimates of these parameters can be obtained from the non-parametric matching estimators. Alternatively, as a parametric approximation, one can also compute the corresponding marginal effects of PH on LP and IC from the estimation of a bivariate probit model for labour market participation and daily care-giving with PH as a regressor.[22]

Some issues in specification and causal interpretations:

a) Using the longitudinal dimension: Our first set of estimates reported in section 4.1 is obtained using the second wave of the SHARE data. One concern is that, in the cross section, parental health could be correlated with unobservable determinants of LP and IC (preference shifters, human capital), even after controlling for the daughter's own health and education, etc. One may argue that longitudinal variation in the parental health instrument is less likely to be subject to this problem. In particular, reestimation of the parameters of interest using the subsample of women in 2006 for which $PH_{2004} = 0$ would allow us to mitigate any systematic correlation that may exist between PH and unobservables factors of daughter's preferences or human capital and to move closer to the 'ideal' experiment in which we observe the (caeteris paribus) effect of an exogenous shock to the health of parents. Furthermore, results obtained from this sample remain interpretable in terms of our static model. Finally, using the first two waves we can control for the daughter's lagged participation for which there is a solid basis in labor economics. Estimates based on the longitudinal dimension using the first and second waves are reported in section 4.2.

b) Co-residence and distance between parents and daughters: The value of the parameter $\overline{IC}$, the time cost of daily care, clearly depends on the distance separating daughters and parents. This distance varies considerably in the population, and some of this variation may be the outcome of choices made simultaneously with labor supply and daily caring. For instance, daughters may decide to co-reside because they plan to provide daily care. Joint modelling of employment, informal and formal care and location of residence is beyond the scope of this paper. But given that distance between daughter and parents is observable, should we include it as a control? If distance is chosen jointly with informal care, then conditioning on distance is likely to produce systematic correlation between PH and other components of $(a, \beta, \overline{h}, \overline{IC}, w, OC)$. On the other hand, heterogeneity in distance and $\overline{IC}$ caused by caregiving does not invalidate causal interpretations of our estimators as long as the joint distribution of $(a, \beta, \overline{h}, \overline{IC}, w, OC)$ conditional on X is still independent of PH. Therefore, we argue that it is better not to include distance as a control while keeping in mind that part of the measured impact of PH on LP and IC may operate through the choice of $\overline{IC}$.

[22] The index restrictions implicit in the bivariate probit are equivalent to the monotonicity assumption. [REF?]

c) Other sources of care: The behavioral model suggests that the effect of PH on employment and daily care should be measured net of other sources of care, which the daughter takes as given. In the daughters sample this information is not available but we include the number of sisters as a proxy. In the parents sample information on other formal and informal care received is available. If parental disability PH correlates positively with the receipt of other care and measures of OC are omitted, the estimates of the effect of PH on the daughter's employment and care could be smaller (in absolute value) than the behavioral model's parameters. On the other hand if OC is determined jointly with the daughter's choices within an extended decision unit failure to include common determinants can lead to biases. However, it is not clear that the bias on coefficients of interest would be important. For these reasons we compare specifications with and without controls for OC.

Using the model to make predictions about the parameters of interest: Let us assume hereafter that the exclusion, monotonicity and exogeneity assumptions hold. In Appendix A2 we derive expressions for the utility differences which measure: (A) The propensity to work conditional on caregiving status, $U(1, IC) - U(0, IC)$, which determines 'potential outcomes'. (B) The propensity to provide informal care or 'propensity–to-care' index, $\max[U(1,1), U(0,1)] - \max[U(1,0), U(0,0)]$. Based on those expressions we characterize the solution to the discrete choice time allocation model. In order to make predictions about populations we need to be more explicit about heterogeneity in the parameters $(a, \beta, \overline{h}, \overline{IC}, w, OC)$. The simplest way of doing this is to consider a population with fixed values of $(a, \beta, \overline{h}, \overline{IC}, OC)$ and heterogeneity in wages. We then obtain the following results.

A. Potential outcomes and treatment effects: First, let us assume that the propensity-to-work index is monotonic in w.This is a very weak assumption since it only imposes that the value of work increases with the wage. Second, we show that the effect of informal care on the propensity-to-work index is negative as long as utility is concave in consumption and leisure. Therefore,

Result 1: There exist two reservation wages $w_{r1} < w_{r2}$ partitioning the support of wages into 3 intervals within which the treatment effect of daily caregiving on employment is 0 (for low wages), -1 (for intermediate wages) and 0 (for high wages).

Discussion: Essentially this result says that, because daily caregiving increases the marginal utility of leisure, the reservation wage which induces the daughter to work is higher if she is providing care than if she is not. Therefore, daughters with very low potential wages are 'never workers' who do not work regardless of their caregiving choice. At the other end, daughters with sufficiently high wages are 'always workers' who work even if they have to provide dayly care. In between, daughters work as long as they are not providing care and quit if they have to take up daily care.

Empirical predictions: The estimated treatment effects of daily caregiving on employment have to be non-positive. Together with monotonicity of daily care in the instrument PH, this implies that the effect of parental disability on the probability that the daughter works should also be non-positive. The LATE parameter should be highest (in absolute value) for subpopulations of women who are observed to work when their parents are in 'good' health but have 'marginal' attachment to the labor market, e.g, low-skilled working women.

B. Daily care and compliance types: The propensity-to-care index is (negatively) monotonic in the daughter's potential wage. Combined with monotonicity of the instrument PH, this implies the second main result:

Result 2: There exist up to two thresholds w_{c1} and w_{c2} partitioning the support of wages into 3 intervals within which all women are always-takers (for low wages), compliers (intermediate wages) and never-takers (high wages). Special cases arise if there is only one threshold separating compliers and never takers (with no always takers), or if there is no threshold because all individuals are never takers.

Discussion: Figure 3 illustrates our two main results. The figure shows the daughter's propensity-to-care index as a function of her wage, conditional on the health of parents. Consider the graph of the index which conditions on parental bad health ($PH = 1$). If the daughter faces a very low wage the utility from caring is assumed greater than that of not caring and she provides daily care. As long as her potential wage is below the reservation wage w_{r1} of Result 1, the daughter is a 'never worker' so the opportunity cost of providing care does not depend on the wage. Therefore, the propensity-to-care is flat as a function of the wage in this range. If the wage is between the two reservation wages, then caring induces the daughter to quit work. Therefore, the opportunity cost of caring includes the value of work and in this range the propensity-to-care is positive (in this example), but decreasing in the wage. For 'always worker' women with wages above w_{r2}, daily caring does not change their employment status but the propensity-to-care is still decreasing in the wage.[23] If the daughter earns a sufficiently high wage ($w > w_{c2}$), the propensity-to-care becomes negative and she will not provide daily care even if her parent is in bad health.

The second function drawn in Figure 3 is the propensity-to-care when parents are in good health ($PH = 0$). Note that it has kinks at the same reservation wages but it it is below the first one for every wage because the marginal utility of caring is lower. Therefore, when parents are in good health the propensity-to-care becomes negative at lower wages than when parents are in poor health. In this example, the women with wages between w_{c1} and w_{c2} are the compliers who take up caring only when their parents are in poor health. Their number (the causal impact of parental disability on the probability of daily care) is given by the proportion of wage offers in that range, $F(w_{c2}) - F(w_{c1})$ where $F()$ is the cdf of potential wage offers. The number of never takers is $1 - F(w_{c2})$ and the number of always takers is $F(w_{c1})$. In the example, complier women with wages between w_{c1} and w_{r2} will quit work when they comply (treatment effect -1), and those with higher wages will not (treatment effect 0). Thus the impact of parental disability on the daughter's probability of employment is $-[F(w_{r2}) - F(w_{c1})]$. The LATE parameter is the ratio of the employment and care probability impacts, $\frac{-[F(w_{r2})-F(w_{c1})]}{F(w_{c2})-F(w_{c1})}$. This is also the proportion of complier women who quit work. More generally, the impact

[23] We show that this is the most plausible scenario in Appendix A2. A sufficient condition is that consumption and leisure be complements in utility ($\alpha_5 > 0$).

of parental disability on the probability that the daughter works can be shown to be $-[F(\min[w_{r2}, w_{c2}]) - F(\max[w_{r1}, w_{c1}])]$, its impact on the probability of providing daily care is $F(w_{c2}) - F(w_{c1})$ and LATE is the ratio of the two.

Finally, in Figure 3 suppose the propensity-to-care- index when parents are disabled is still the same as before but the index conditional on good health is instead the function labeled "" which is negative for all wages. In this case daughters never provide daily care when their parents are in good health. There are no always takers and we say $w_{c1} = 0$. The impact on the employment probability impact simplifies to $-[F(\min[w_{r2}, w_{c2}]) - F(w_{r1})]$, the impact on the caregiving probability is $F(w_{c2})$ and the LATE parameter is again the ratio of the two. As shown in section 5.3 below, the proportion of always takers ranges from negligible to small depending on the samples and/or the instruments considered. Therefore this scenario is empirically relevant and we will take it into account to derive some of the comparative statics results which follow.

C. Some comparative statics:

(C1) Caeteris paribus, an increase in other sources of care OC received by parents (e.g. public formal care) has no effect on the daughter's reservation wages.[24] However, the increased availability of other care reduces the marginal utility of her own informal care and shifts the two propensity-to-care functions of Figure 3 downwards. Therefore, the compliance type thresholds w_{c1} and w_{c2} move left. The mass of compliers plus always-takers decreases. Furthermore, if there are no always takers we get a sharper result: increased availability of other care (weakly) reduces the impact of parental health on the employment probability of daughters and it also reduces the impact on the proability of caregiving, i.e., the mass of compliers. However, the effect on LATE is ambiguous.

(C2) A reduction of the time-cost of informal care $\overline{IC}$ (e.g., because the daughter lives closer to her parents) shifts the propensity-to-care functions upwards and reduces w_{r2}, narrowing the range of wages for which the treatment effect is -1. Therefore, the mass of compliers plus always-takers increases but the effect on the employment impact and on LATE are ambiguous.

(C3) Suppose monetary payments are offered to daughters providing daily caregiving. That is, $\beta_1 > 0$. This shifts the propensity-to-care functions upwards and increases w_{r2}, widening the range of wages for which the treatment effect is -1. This type of support has been put in place, for instance, as part of the new public long-term care system in Spain. Our simple model predicts that it should increase the supply of daily caregiving and reduce labor supply.

D. Comparisons across country pools: Suppose the main differences across the three country pools (North-Central-South) are: i) The availability of formal care, interpretable as variation in the distribution of OC; to focus, consider the simplest case where there is a single value of OC within pools, which grows from South to North. ii) Differences in labour market attachment of daughters - interpretable as differences in the distribution of wages w. Let the distributions of wages be ordered from North to South, in the sense of stochastic dominance.We obtain the following **empirical predictions**:

[24] An implication of this is that increased availability of care has no effect on the average treatment effect of care on employment.

(1) The mass of compliers plus always takers should increase from North to South. This follows both from comparative statics result (C1) and from the ordering of the distribution of wages. This prediction is reinforced by comparative statics result (C2) to the extent that the 'average' time cost of daily care ($\overline{IC}$) is smaller in the South because daughters tend to live closer to their parents.

(2) The estimated impact of parental disability on the employment probability of daughters should grow from North to South. This sharp prediction obtains as long as the proportion of daughters who are always takers is zero or very small. It follows partly from comparative statics result (C1). Furthermore, recall that the impact on employment probability predicted by the model is $-[F(\min[w_{r2}, w_{c2}]) - F(w_{r1})]$ which measures the proportion of daughters just above the margin of participation but close to it. We also expect this to grow from North to South.

(3) There is no clear prediction ordering the LATE parameters across country pools.[25]

4 Empirical Results

We report estimates of the impact of a change in the parental disability instrument ($PH = 0$ to $PH = 1$) on the daughter's employment ('numerator') and daily caregiving choices ('denominator'), as well as the ratio of the two impacts which is the LATE parameter. We compute non-parametric and bivariate probit estimates with different sets of controls for different samples and different definitions of the instrument. In every case we first report impact estimates with no controls which show the unconditional correlations in the data. Ideally we would next introduce as many controls as suggested by theory in non-parametric matching estimators, which impose minimal assumptions on the distribution of impacts. Because sample sizes limit the precision of non-parametric estimates it is clear that there is a tradeoff between increasing the number of controls and allowing for more flexibility in the specification of causal impacts. Our strategy is to explore increasing sets of controls, to compare non-parametric estimates to those obtained from bivariate probit models and to switch to bivariate probits when sample sizes are too small, e.g., to obtain estimates for specific subpopulations of daughters.

4.1 Evidence from cross-sectional variation in parental health

The top panel of Table 6 shows the components of the Wald estimate and the non-parametric estimates conditional on a reduced vector of controls X for the sample of daughters interviewed in wave 2 (2006), who were between 50 and 60 at the time of the interview and had at least one living parent. The controls are the daughter's age, education and the number of living sisters. We do not report other estimates which we computed using the bivariate probit model and/or including a more extensive set of controls. The results were very similar, from which we concluded that the normality and

[25] One can show that the predictions on the ordering of ATE (the average treatment effect) across country pools are not any sharper.

functional form assumptions in the biprobit model and the use of a reduced set of controls provide good approximations.[26] The lower panel shows estimates for the longitudinal subsample of women who were interviewed in both waves. In this case we report the components of the Wald estimate and biprobit estimates conditional on two different vectors of controls. First we condition on the same controls as in the cross-sectional sample. Next we add the first lag of LP. On both theoretical and empirical grounds this is a potentially relevant variable which is correlated with PH and omitted from the cross-sectional specification.[27]

Columns 4-6 of the table report estimators of the denominators. For the Wald estimate these are just the difference in the proportion of daily caregivers between women with and without a parent in bad health. For the matching estimators and the biprobits the denominator is obtained by averaging the differences in conditional means across the distribution of the covariate vector X. If PH is exogenous this estimate gives the causal effect of having a parent in bad health ($PH = 1$) on the daughters' decision to provide daily care. It is also the mass of compliers. For Wald estimators the mass of compliers is positive and significantly different from zero in all three groups of countries, ranging between 4.4% in the North to 17.2% in the South in the first row of Table 6. The other rows show that the results are qualitatively the same when we introduce the controls: there is a large and significant effect of PH on IC in southern countries, and smaller but still significant effect in central and northern countries. We thus find a North-South gradient in the proportion of compliers which mirrors the negative North-South gradient in the development of public long-term care systems. As predicted by the time allocation model, the greater the availability of public formal care the smaller the proportion of women who are induced to take up informal care.[28]

The first three columns of Table 6 report the estimators of the numerators. This parameter is the causal effect of having parents in poor health on the employment rates of women under the assumption of exogeneity of PH. When no controls are included all estimates are negative as suggested by theory and increasing (in absolute value) from North to South. For southern countries, we obtain that women with at least one parent in bad health are 9 percent less likely to be at work than women with no parents in that situation. However, the size and significance and (to a lesser extent) the gradient of the PH effect on LP do not seem robust to the inclusion of covariates which account

[26] The complete set of controls includes the woman's age, number of children, dummies for different education level, number of brothers and sisters, annual non-wage income, and dummies for health status. For a more detailed description on these covariates, see Appendix A1. Results on the estimations including the complete set of controls are available upon request.

[27] Lagged participation is a good proxy for labor market attachment and potential wages. Furthermore it may reduce search costs and is well known to be a strong predictor of current participation (see Eckstein and Wolpin (1989), Hyslop (1999)). On the other hand including lags of the endogenous variables LP and IC could bias the estimate of the causal impact of PH when PH is serially correlated. On balance we choose to include lagged LP but not lagged IC.

[28] The prediction we derived from the model applied to the sum of always takers and compliers and we are only looking at compliers here. However, in section 4.3 we show that the North-South gradient is also observed for always takers and that the sum of always takers and compliers is dominated by the latter.

for the human capital and labor market attachment of daughters. In particular, the estimated impacts decrease substantially and are no longer significant when we control for the daughter's education, and they almost fade away when we also control for lagged participation in the longitudinal sample. In summary our point estimates show negative impacts of PH on employment of daughters across specifications and country pools, but based on this table the effect of ill-health of parents on the aggregate employment rates of their daughters would seem very small.

The last columns of Table 6 shows the estimates of LATE for southern countries. This parameter attributes any effect of parents' bad health on the employment rate of women to its effect on the provision of daily informal care, under exogeneity and exclusion restrictions. We first note that our estimates of LATE are very imprecise and for this reason we do not report estimates for northern and central country pools.[29] Second, point estimates are negative in most cases but they are small and not significantly different from zero once we introduce controls. Rather than standard errors we report the percentage of bootstrapped replicas for which point estimates fall in the intervals $(-\infty,-1),[-1,-0.5),[-0.5,0),[0,\infty)$. For instance, the point estimate of LATE in the longitudinal sample is -0.230 and there is 59 % probability that the true value of the parameter is between 0 and -0.5. That is, the largest mass of the distribution of the estimator of LATE is in negative values close to zero. The most plausible interpretation in our view is that LATE for the whole population is not very large because most women in their 50's who take up daily informal care are never workers or always workers.[30]

4.2 Evidence from longitudinal variation in parental health

Table 7 reports new estimates of the impact of the health of parents in 2006 on the employment and daily care choices of their daughters in 2006. The estimates are now obtained for the subsample of women who had parents in good health in 2004. Estimation on this "conditional" longitudinal sample exploits longitudinal variation in the instrument. As argued above this should move us closer to the 'ideal' experiment in which we observe the (caeteris paribus) effect of an exogenous shock to the health of parents, while the estimates have exactly the same (causal) interpretation in terms of the static behavioral model.[31] On the down side the sample size is smaller and the parental health instrument

[29] Lack of precision is partly explained by the relatively small mass of compliers - see Appendix A4 for further comment.

[30] However, the exclusion restriction is a maintained assumption so we cannot rule an alternative interpretation which is that the causal effect of daily care on employment is considerably larger but the exclusion restriction is not valid and parental ill-health increases employment of daughters through channels other than informal care.

[31] If there is no systematic correlation between PH and unobservable determinants of LP and IC in the cross-section, then the estimates obtained from the "conditional" longitudinal sample and from the full cross-section sample (top panel of Table 6) with the same set of controls should be similar. However, there is an obstacle to making this comparison because we have evidence of important differences in the distribution of observables between the cross-section and the longitudinal samples for 2006 due to differential panel attrition by observables. This problem is of special relevance in southern countries.

has less variation in the longitudinal dimension than in the cross section as it is shown in the last panel of Table 7, both of which reduce the precision of the estimates.

In the first panel we show the components of the Wald estimate and the bivariate probit models controling for age, education, number of sisters and lagged participation. With and without controls, our estimate of the impact of an adverse shock to parental health on the probability of daily caregiving by the daughter in southern countries is large (+24 %) and significant, and substantially smaller (+9 %) and marginally significant in central and northern countries. As to the effect of the health shock on the probability of employment, the estimate without controls in southern countries is very large (-26 %). In this case a substantial effect of remains when we introduce controls: the rate of employment is reduced by 12 %. The estimate of LATE is -0.49. The point estimates of the employment effect and of LATE are significantly different from zero at the 10% and 5% levels respectively. In spite of the limited precision of point estimates we conclude that there is a significantly negative effect of the health of parents on the aggregate employment rate of their daughters, mediated through the provision of daily informal care. We obtain small (and not significant) employment effects in the other country pools.

The empirical North-South gradients which we find in caregiving and employment impacts are in line with the predictions of the behavioral model. Both employment and caregiving effects in the South are stronger than those we obtained from cross-sectional variation in the health of parents. This is remarkable because one might expect any failure of exogeneity in cross-sectional variation to bias the estimates of employment impacts upwards in absolute value. One plausible interpretation, beyond the scope of our model, is that it takes time for families and daughters in the South to adjust to shocks to the health of parents and that longitudinal variation identifies the short-run effect while cross-sectional variation identifies a smaller long-term effect. Another consideration is that our estimates of aggregate effects integrate group-specific impacts over the marginal distribution of covariates and that this distribution varies from the cross-sectional sample to the "conditional" longitudinal sample. However, the exercise we report in footnote 31 suggests that this composition effect is not the main driver of the difference between cross-sectional and longitudinal impacts.

In the second panel of Table 7 we shift the focus to impact estimates for particular subgroups. First, we empirically confirm a conjecture we made in the analysis of the behavioral model that the largest impact of poor health of parents on the employment rates of daughters should be found for those who are above (but close to) the margin of

In order to address this issue we also did the following. We estimated the biprobit model for 2006 including the vector of 3 covariates and using the longitudinal sample of women with parents in good health in 2004. The estimated coefficients from this conditional subsample were used to compute our parameters of interest using both the cross-section and the full longitudinal sample. Then the two vectors of estimated parameters would be compared to our baseline estimates. This strategy allows us to compare the estimates obtained exploiting the cross-sectional or longitudinal variation of the instrument but keeping the distribution of observables of $X's$ constant. Significant differences between these two sets of estimated parameters would point to a problem of the validity of our cross-sectional instrument as it is defined in our paper. The results of this exercise suggest that changes in the distribution of $X's$ do not play a major role.

participation. To approximate this condition we select southern women who were working in the previous period but have low education. For this group, the estimated impact of $PH = 1$ on the employment rate of daughters is -23 %, which is twice as much as the aggregate impact and stands in even sharper contrast with the impact for women who were not working in the previous wave (- 4 %). [The employment impact is also smaller for lagged participants with higher education, who are presumably not as close to the margin of participation.]

There is some concern that, in the absence of other changes, the supply of informal caregivers will decline in southern european countries with the gradual increase of education and labor market attachment of daughters who are the primary providers. Motivated by this we compute the proportion of daughters who are always takers or compliers, a measure of the potential supply of daily informal caregivers, and we compare this quantity across 'high' and 'low' education classes. For each education class we integrate over the marginal distribution of the other covariates. This includes lagged participation, which reduces the propensity to care and is systematically higher for more educated women. We find that the 'supply' of daily caregivers in southern countries decreases from 35 % to 28 % when comparing "low" and "high" education categories.[32]

4.3 Using multiple measures of parental disability

In this section we obtain additional evidence using the "parents-sample" described in Section 2. This is a sample of women aged between 50 and 60 who are the daughters of sample respondents. It provides more comprehensive information, reported by the elderly parents themselves, on their health status and their access to different sources of care. In addition to self-reported general health, other health measures are available such as (self-reported) diagnosed chronic conditions, functional limitations, ADL and IADL limitations, problems with mobility, depressive symptoms and mental health. Even though subjective self-reports of general health have proved to be informative about an individuals' health, using multiple indicators is preferable. Information on limitations with daily living activities and chronic diseases like mental health problems should capture with more accuracy symptoms and problems related to dependency or need of care. Using these instruments can give us a finer picture of the impact of parental health on the time allocation of daughters, e.g. the proportion of "complier" women who are induced to take up daily care and to quit work for different types and intensities of parental disability.[33]

Specifically, we consider the following four indicators. First, we define a binary subjective indicator based on the categorical variable on self-reported general health provided by the parents. In particular, we follow the same definition applied for the "daughters-

[32] In particular, the estimates (standard errors) for the mass of compliers and always takers for low educated women are 0.259 (0.101) and 0.09 (0.026), respectively. For high educated women these estimates are 0.218 (0.099) and 0.061 (0.017).

[33] Furthermore it is likely that objective measures are less subject to potential response scale biases arising if respondents from different countries, cultures or socioeconomic groups have different reference levels of health or different response scales when answering subjective questions on their health status.

sample" and we define the binary variable $POOR$ that equals to 1 if at least one parent is in a poor health status. Second, we construct a binary variable ADL that indicates whether at least one parent has difficulties with at least one of the following six activities of daily living: dressing, including putting on shoes and socks; walking across a room; bathing or showering; eating, such as cutting up your food; getting in and out of bed; and using the toilet, including getting up or down. Third, we also include a binary variable $DEMENTIA$ that equals to 1 if at least one parent suffers from Alzheimer, dementia or other problems with memory. Fourth, the indicator $MOBILITY$ is set to 1 if there is at least one parent with 3 or more functional limitations, i.e., walking 100 meters, sitting for two hours, getting up from a chair, climbing several flights of stairs, climbing one flight of stairs.

The main outcome variables (employment and daily caregiving) are analogous to those defined from the information reported by daughters. Definitions and more specific details of these variables and other daughters' characteristics (education, age, marital status, number of children and siblings) are provided in Appendix A1. The descriptive statistics presented in Table A1.2 of Appendix A1 for this sample are in almost every case very similar to those of the "daughters-sample", i.e., the samples seem quite comparable. In addition, the table also shows the important gradient in the use of formal care services among the three groups of countries, which increases from South to North. Regarding the prevalence of each of the health conditions considered, there is also a clear gradient increasing from South to North. It is noteworthy that no such clear health gradient was present in the data reported by the daughters.

With respect to parameters of interest we do not have a binary instrument anymore but a vector of binary instruments $Z = (POOR, ADL, DEMENTIA, MOBILITY)$. Different values of Z describe different forms and intensities of disability and parents' need of care. The conceptual framework described in section 3 is still useful if we reinterpret the variable PH in the behavioral model as an indicator which takes values 0 or 1 for particular values or subsets of values of the vector of instruments Z reflecting different (and increasing) levels of disability.

In Panel A of Table 8 we show estimates of the proportion of daughters who are 'always takers', 'compliers' and 'never takers' when the health of their parents is summarized in a single binary instrument so the events $PH = 0$ and $PH = 1$ are complements. We focus on the sensitivity of the mass of always takers and compliers to changes in the definition of the 'good health' state. In the first row we only use one of the indicators in Z, the subjective report of health. As we did for the daughters sample, $PH = 1$ correponds to "at least one parent is in poor or very poor health" and $PH = 0$ is the complement. We confirm the increasing North-South gradient in the proportion of compliers (from 0 to 8 % to 19 %) and we observe the same gradient in the proportion of always takers (from 1% to 3% to 10%). However, the within-country pool comparisons are more interesting. In the second row $PH = 0$ if 3 of the indicators in Z - $POOR, ADL, DEMENTIA$ - are all zero, and in the third row $PH = 0$ if all four indicators are zero. The behavioral model predicts that, as we consider increasingly strict definitions of 'good health', the marginal utility of caregiving when $PH = 0$ should fall and the proportion of 'always taker' daughters should decrease. This is confirmed in the data. There are two additional implications of the model: first, daughthers who are not always takers any more become compliers; and second, it is quite likely that the treatment effect is -1 for women whose compliance status switches from always taker to complier. Therefore a more subtle prediction of the model is that a broader definition of the benchmark 'good health' can reduce the estimated impact of parental disability on the employment rates of daughters. We also see that as we narrow the definition of good health the prevalence of 'poor health' increases and the estimated proportion of compliers tends to fall. The former is a mechanical effect, while the latter is not: on the one hand, some always takers are now classified as compliers as explained above; on the other hand, 'poor health' has been broadened to include less serious conditions which induce fewer compliers. Thus when the summary instrument uses all four disability conditions, the proportion of never takers is as high 98 % in the North, 93 % in the Center and 82 % in the South.

We now consider the impact of parental health changes from the benchmark value $Z^0 = (0,0,0,0)$, the best level of health that is observable, to other values of the vector, say Z^j.[34] There are two good reasons to use Z^0 as benchmark: first, changes from that particular benchmark make the assumption of monotonicity most plausible; and second, as explained above using a broader definition of good health may re-classify some compliers as always takers and hide part of the impact of parental disability on the employment of daughters. The LATE parameter for the particular subpopulation of compliers defined as those women whose caregiving decision is changed when going from the state $Z^0$to Z^j is given by

$$\beta_j = \frac{\int [E(LP|Z = Z^j, X) - E(LP|Z = Z^0, X)]dF(X)}{\int [E(IC|Z = Z^j, X) - E(IC|Z = Z^0, X)]dF(X)}$$

where the denominator measures the mass of compliers among daughters whose parents transit from good health to condition Z^j. With 4 disability indicators there are $2^4 - 1 = 31$ possible values of Z^j. Alternatively, with $PH = 0$ still meaning $Z = Z^0$ we can attach the label $PH = 1$ to any arbitrary set of values of Z which does not include Z^0. Let $\widetilde{Z}$ denote one such set. For instance, if we define $PH = 1$ as "having at least one parent with poor mental health" the set $\widetilde{Z}$ includes all vectors of the form $(.,.,1,.,.)$. The LATE parameter is

$$\widetilde{\beta} = \frac{\int [E(LP|Z \in \widetilde{Z}, X) - E(LP|Z = Z^0, X)]dF(X)}{\int [E(IC|Z \in \widetilde{Z}, X) - E(IC|Z = Z^0, X)]dF(X)}$$

Panel B of Table 8 shows estimates for four different $\widetilde{Z}$ sets which partition the whole

[34] Note than in this case the events $PH = 0$ and $PH = 1$ are not necessarily complements.

space of conditions.[35] The first row ("ADL") corresponds to Z values of the form (0,1,0,.). The ADL condition by itself has significant impacts on the probability of daily care in the Center and South, but negligible effects on employment. The second row adds the condition "POOR" and corresponds to (1,0,0,.) or (1,1,0,.). Again, we find significant impacts on the probability of daily care in the Center and South, and quite large in the latter at 24 %. Furthermore, in this case there is a significant and far from negligible effect (-14%) on employment in the South. The third row corresponds to (.,.,1,.), i.e., DEMENTIA combined with any other condition. This condition has a prevalence of 9% in the South, and its estimated effects stand out among all others. The proportion of daughters taking up daily care is 42 % in the South and 25 % in the Center., and there is a strong employment impact of -17 % in the South. The fourth row reports the effect of MOBILITY alone, that is (0,0,0,1). This conditions only induces significant (and smaller) effects in the South. Finally, the fifth row shows the impacts linked to the 'summary instrument', i.e., at least one parent has at least one of the four disability conditions considered. We find significant employment effects of about -10 % in the South. Therefore, the highlights of this Panel are: a) the remarkable impact of parental dementia on daughters in the South and Center; and b) We find significant effects of several forms of parental disability on the employment of daughters in the South. This is in contrast with cross-sectional estimates obtained from the sample of daughters, and one possible explanation is our use of a stricter 'benchmark'of good health as conjectured above. As a result, we obtain significant negative estimates of LATE associated with subjectively reported poor health, dementia and the summary instrument. The estimated LATE associated with dementia is smaller.

An additional advantage of the use of the "parents-sample" is that it also allows us to include in the analysis some information on parents' access to other sources of care, both formal and informal. In particular, the binary variable *Fcare* indicates whether at least one parent has been in a nursing home overnight or has received home care in the last twelve months prior to the interview.[36] We also define an indicator *Icare_ other* which is one if at least one parent is receiving informal care from sources other than the daughter.

[35] One can show that

$$\widetilde{\beta} = \frac{\int \sum_{Z^j \in \widetilde{Z}} (([E(LP|Z = Z_j, X) - E(LP|Z = Z_0, X)] \Pr(Z = Z_j, X))/ \sum_{Z^j \in \widetilde{Z}} \Pr(Z = Z_j, X))dF(X)}{\int \sum_{Z^j \in \widetilde{Z}} (([E(LP|Z = Z_j, X) - E(LP|Z = Z_0, X)] \Pr(Z = Z_j, X))/ \sum_{Z^j \in \widetilde{Z}} \Pr(Z = Z_j, X))dF(X)}$$

where the numerator and the denominator are weighted sums of the effects on employment and caregiving decisions of changes in health from Z^0 to each Z^j in $\widetilde{Z}$ with weights given by the conditional probabilities of this parental ill-health state.

[36] A "nursing home" is defined in SHARE as an institution sheltering older persons who need assistance in activities of daily living, in an enviroment where they can receive nursing care, for short or long stays. Home care is professional or paid nursing or personal care, professional or paid home help for domestic tasks that the individual could not perform himself due to health problems, and meals-on-wheels. We do not observe who pays for these services, or if they are publicly provided.

These variables can replace the number of sisters which we have been using as a proxy for 'Other Care'. The estimates reported in Panel C of Table 8 suggest that our results were not driven by the ommission of better measures of other care.

References

[1] Attias-Donfut, C., Ogg, J., and Wolff, F.C. (2005): "Family Support" iin Börsch-Supan, A., Brugiviani, A., Jürges, H., Mackenbach, J., Siegrist, J., and Weber, G. (Eds.), *Health, Ageing and Retirement in Europe. First Results from the Survey of Health, Ageing and Retirement in Europe*. Mannheim Research Institute for the Economics of Ageing.

[2] Byrne, D., Goeree, M., Hiedemann, B., and Stern, S. (2008): "Formal Home Health Care, Informal Care, and Family Decision-Making" mimeo.

[3] Bolin, K., Lindgren, B., and Lundborg, P. (2008): "Your next of kin or your own career? Caring and working among the 50+ of Europe", *Journal of Health Economics*, Vol. 27, No. 3, pp. 718-738.

[4] Eckstein, Z. and Wolpin, K.I. (1989): "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience ", *Review of Economic Studies*, Vol.56, No. 3, pp. 375-390.

[5] Ettner, S.L. (1995): "The Impact of "Parent Care" on Female Labor Supply Decisions", *Demography*, February, Vol. 32, No. 1, pp. 63-80.

[6] Ettner, S.L. (1996): "The Opportunity Costs of Elder Care", *The Journal of Human Resources*, Winter, Vol. 31, No. 1, pp. 189-205.

[7] European Comission and the Council. (2003a): "Supporting National Strategies for the Future of Health Care and Care for the Elderly", Brussels.

[8] Frölich, M., and Melly, B. (2007): "Unconditional Quantile Treatment Effects under Endogeneity", mimeo.

[9] Heckman, J. and Vytlacil, E. (2005): "Structural Equations, Treatment Effects and Econometric Policy Evaluation", *Econometrica*, Vol. 73, No. 3, pp.669-738.

[10] Heitmueller, A. and Michaud, P-C. (2006): "Informal Care and Employment in England: Evidence from the British Household Panel Survey", IZA Discussion Paper No. 2010.

[11] Hirano, K., Imbens, G. and Ridder, G. (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", *Econometrica*, Vol. 71, No. 4, pp.1161-1189.

[12] Hyslop, D.R. (1999): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women", *Econometrica*, Vol. 67, No. 6, pp. 1255-1294.

[13] Imbens, G.W. and J. Angrist. (1994): "Identification and Estimation of Local Average Treatment Effects", *Econometrica* Vol. 62, No. 2, pp. 467-475.

[14] Johnson, R.W. and Lo Sasso, A.T. (2000): "The Trade-Off between Hours of Paid Employment and Time Assistance to Elderly Parents at Mid-Life", Report from The Urban Institute, Washington D.C.

[15] Jürges, H., Avendano, M., and Mackenbach, J. (2008): "Are different measures of self-rated health comparable? An assessment in five European countries", European Journal of Epidemiology, Vol. 23, No. 12, pp. 773-781.

[16] Pezzin, L.E. and Schone, B.S. (1999): "Intergenerational Household Formation, Female Labor Supply and Informal Caregiving: A Bargaining Approach", *The Journal of Human Resources*, Vol. 34, No. 3, pp. 475-503.

[17] Rosenzweig, M.R. and Wolpin, K.I. (2000): "Natural "Natural Experiments" in Economics", *Journal of Economic Literature*, Vol. 38, No. 4, pp. 827-874.

[18] Spiess, C.K. and Schneider, A.U. (2003): "Interactions between care-giving and paid work hours among European midlife women" *Ageing & Society*, Vol. 23, No. 1, pp. 41-68.

[19] Wolf, D.A. and Soldo, B.J. (1994): "Married Women's Allocation of Time to Employment and Care of Elderly Parents", *The Journal of Human Resources*, Vol. 29, No.4, pp. 1259-1276.

FIGURES

Figure 1: The importance of different relatives as informal caregivers of people aged 80 and over who receive informal care in a daily or weekly basis (%, SHARE 2004)

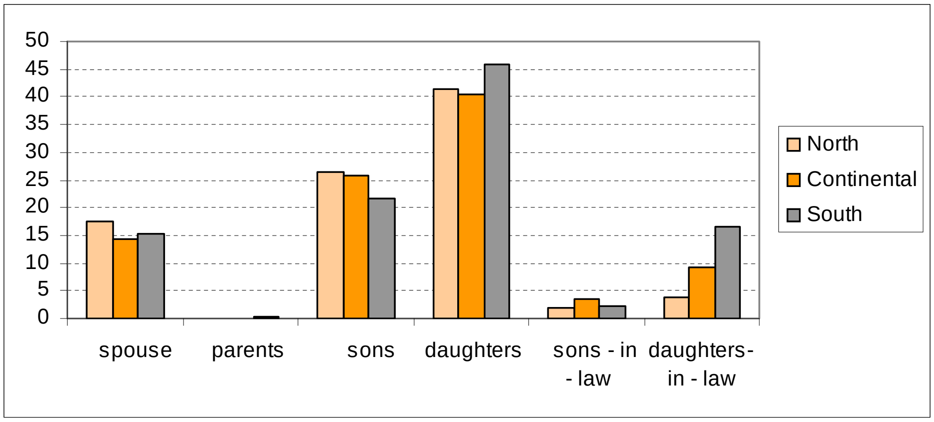
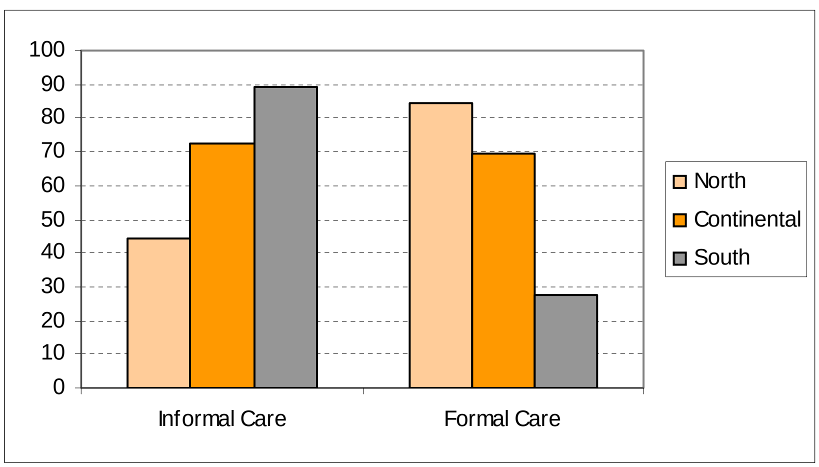


Figure 2: Prevalence of informal and formal care among respondents aged 80 and over who receive care in a daily or weekly basis (%, SHARE 2004)



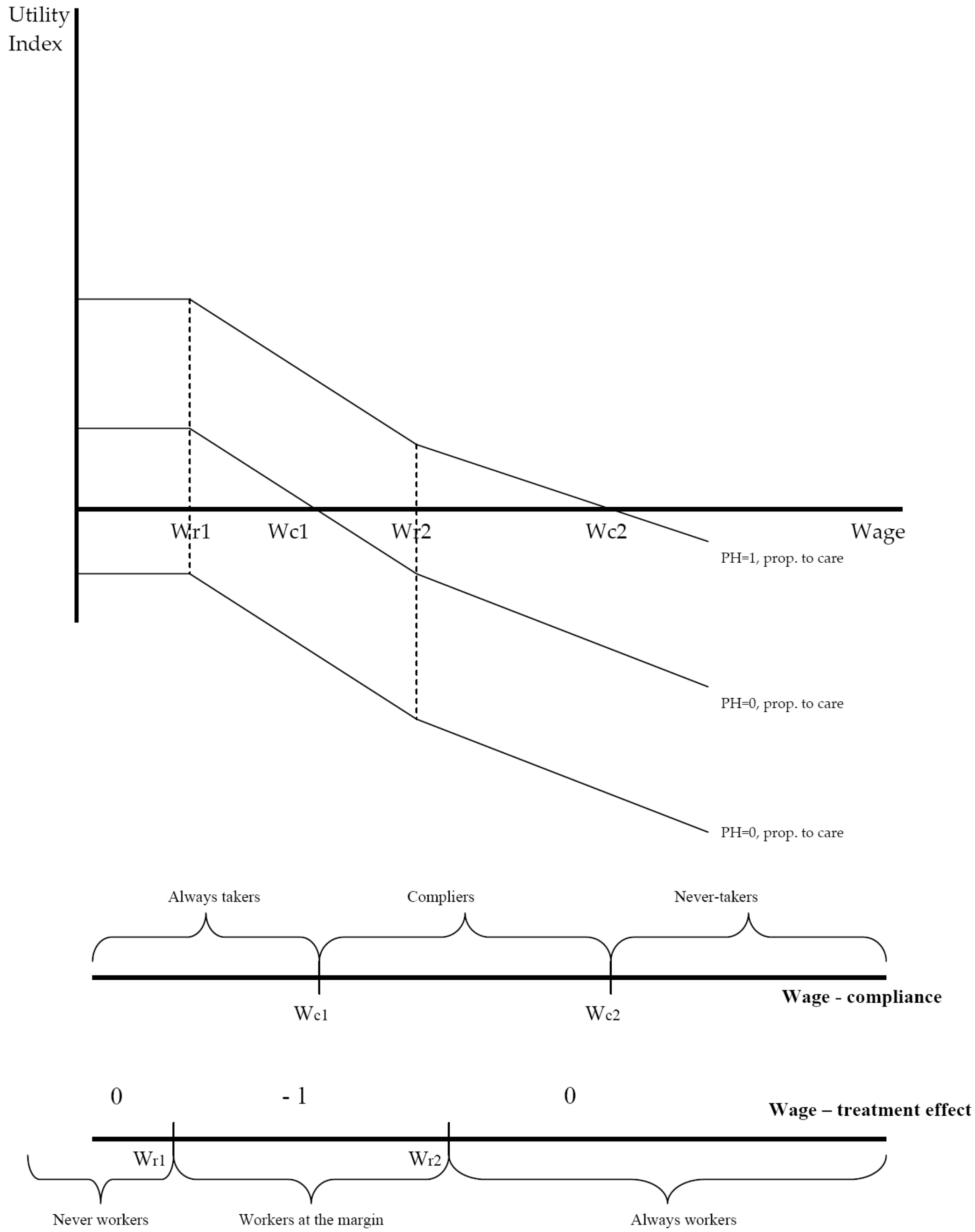


Figure 3: Propensity-to-care index, treatment effects and compliance types

TABLES

Table 1. Descriptive Statistics for Caregiving Variables in the Sample (Wave 2)

	NC	*CC*	*SC*
Caregiver	0.4584	0.3286	0.2598
Intensive Caregiver (Daily/Weekly)	0.1956	0.2238	0.2117
Intensive Caregiver (Daily)	0.0248	0.0708	0.1168
Sample Size	685	1059	685
Sample of Caregivers			
	NC	*CC*	*SC*
Frequency of Caregiving			
Intensive (Daily/Weekly)	0.4267	0.6810	0.8146
Intensive (Daily)	0.0541	0.2155	0.4494
Sample Size	314	348	178
Sample of Daily/Weekly Caregivers			
Intensive (Daily)	0.1269	0.3164	0.5517
Sample Size	134	237	145

Table 2. employment & Daily Caregiving (%)

	NC		CC		SC	
	IC	Non-IC	IC	Non-IC	IC	Non-IC
LP	76.47	82.93	57.33	71.24	38.75	46.28
Sample Size	17	668	75	984	80	605

Table 3. Descriptive Statistics for the Overall Sample (Wave 2)

	NC	*CC*	*SC*
Labour Participant	0.8277	0.7025	0.4540
Daily Caregivers	0.0248	0.0708	0.1168
Age	54.5401	54.6742	54.3985
	(2.9583)	(2.7779)	(2.8929)
Married/Partnership	0.7839	0.7583	0.8525
Education			
Educ1	0.0365	0.0925	0.3226
Educ2	0.2029	0.1511	0.2131
Educ3	0.2861	0.4627	0.2569
Educ4	0.4744	0.2937	0.2073
Health			
Excellent/Very Good	0.5635	0.4353	0.3854
Good	0.2934	0.4079	0.4
Fair	0.1255	0.1350	0.1810
Poor	0.0175	0.0217	0.0336
Non-wage Income(2)	19.940	21.0363	14.9879
	(19.7832)	(24.7994)	(16.9971)
Children	2.3182	2.0651	1.9781
	(1.2534)	(1.1701)	(0.9902)
Parental Health ($\overline{PH}$)	0.2219	0.1879	0.2073
Brothers	1.2686	0.9254	1.1474
	(1.2499)	(1.1655)	(1.1187)
Sisters	1.2204	0.9792	1.1474
	(1.3463)	(1.2226)	(1.2956)
Sample Size	685	1059	685

Note: Means of the variables considered in the analysis and standard deviations in parentheses. (2) Net annual non-wage income is expressed in thousands of ppp-adjusted euros.

Table 4. Descriptive Statistics by Daily Caregiving Status (Wave 2)

	IC			Non-IC		
	NC	*CC*	*SC*	*NC*	*CC*	*SC*
Labour Participant	0.7647	0.5733	0.3875	0.8293	0.7124	0.4628
Age	54.2941	55.5867	54.65	5.455	54.6047	54.3653
	(3.057)	(2.8996)	(2.8777)	(2.958)	(2.7576)	(2.896)
Married/Partnership	0.9412	0.8133	0.9	0.7799	0.7541	0.8463
Education						
Educ1	0.0588	0.0933	0.375	0.0359	0.0925	0.3157
Educ2	0.2941	0.16	0.2625	0.2006	0.1504	0.2066
Educ3	0.5294	0.4667	0.225	0.2799	0.4624	0.2611
Educ4	0.1176	0.28	0.1375	0.4835	0.2947	0.2165
Health						
Excellent/Very Good	0.7647	0.3333	0.375	0.5584	0.4431	0.3868
Good	0.1176	0.5067	0.45	0.2979	0.4004	0.3934
Fair	0.1176	0.1467	0.15	0.1257	0.1341	0.1851
Poor	0	0.0133	0.025	0.0180	0.0223	0.0347
Non-wage Income$^{(2)}$	32.5044	19.9946	16.0510	19.6204	21.1157	14.8473
	(22.5383)	(20.7495)	(16.7507)	(19.6224)	(25.0885)	(17.0381)
Children	2.5294	1.86667	1.95	2.3129	2.0803	1.9818
	(0.7174)	(1.0310)	(0.8554)	(1.2640)	(1.1791)	(1.0072)
Parental Health ($\overline{PH}$)	0.5294	0.3333	0.45	0.2141	0.1768	0.1752
Brothers	1.7059	0.8267	0.875	1.2575	0.9329	1.1835
	(1.5718)	(1.0574)	(1.0835)	(1.2401)	(1.1734)	(1.1192)
Sisters	0.94118	0.48	1.175	1.2275	1.0173	1.1438
	(1.5601)	(0.6649)	(1.2404)	(1.3410)	(1.2470)	(1.3036)
Sample Size	17	75	80	668	984	605

Note: Means of the variables considered in the analysis and standard deviations in parentheses. (2) Net annual non-wage income is expressed in thousands of ppp-adjusted euros.

Table 5. Descriptive Statistics by Parents' Health Status (Wave 2)

	With PH=1			With PH=0		
	NC	*CC*	*SC*	*NC*	*CC*	*SC*
Labour Participant	0.7960	0.6533	0.3803	0.8368	0.7139	0.4733
Daily Caregivers	0.0592	0.1256	0.2535	0.0150	0.0581	0.0810
Age	54.7632	55.0703	54.5563	54.476	54.5826	54.3573
	(3.1386)	(2.8222)	(3.1590)	(2.905)	(2.7612)	(2.8208)
Married/Partnership	0.7631	0.7085	0.8873	0.7899	0.7698	0.8435
Education						
Educ1	0.0263	0.0955	0.3591	0.0394	0.0919	0.3131
Educ2	0.1842	0.1457	0.2887	0.2082	0.1523	0.1934
Educ3	0.2960	0.4774	0.1901	0.2833	0.4593	0.2744
Educ4	0.4934	0.2814	0.1620	0.4690	0.2965	0.2191
Health						
Excellent/Very Good	0.4474	0.3266	0.2183	0.5966	0.4605	0.4300
Good	0.2763	0.4322	0.4155	0.2983	0.4023	0.3959
Fair	0.2237	0.1909	0.2887	0.0976	0.1221	0.1528
Poor	0.0526	0.0502	0.0775	0.0075	0.0151	0.0221
Non-wage Income$^{(2)}$	19.0091	21.5819	14.5269	20.2057	20.9100	15.1084
	(19.1675)	(25.5900)	(16.2079)	(19.9650)	(24.6264)	(17.2099)
Children	2.3618	2.0553	2.1056	2.3058	2.0674	1.9447
	(1.5074)	(1.1468)	(1.0699)	(1.1723)	(1.1760)	(0.9665)
Brothers	1.1974	0.8241	1.2324	1.2889	0.9488	1.1252
	(1.2182)	(1.1389)	(1.0763)	(1.2592)	(1.17093)	(1.1295)
Sisters	1.2171	0.8141	1.4789	1.2214	1.0174	1.0608
	(1.2283)	(1.1593)	(1.6361)	(1.3791)	(1.2343)	(1.1771)
Co-resident	0	0.0151	0.021	0	0.008	0.033
Less than 5km	0.256	0.281407	0.570	0.304	0.399	0.521
Sample Size	152	199	142	533	860	543

Note: Means of the variables considered in the analysis and standard deviations in parentheses. (2) Net annual non-wage income is expressed in thousands of ppp-adjusted euros.

Table 6. Evidence from cross-sectional variation in parental health (Daughters-sample)

	Numerator			Denominator (compliers)			LATE
	NC	CC	SC	NC	CC	SC	SC
Cross-section 2006							
No controls	-0.041	-0.061	-0.093**	0.044**	0.067***	0.172***	
	(0.036)	(0.037)	(0.046)	(0.020)	(0.025)	(0.038)	
Controls (weighting)$^{(1)}$	-0.047	-0.035	-0.031	0.044**	0.058**	0.203***	-0.153
	(0.037)	(0.036)	(0.044)	(0.019)	(0.024)	(0.039)	(0.25,8.25,68.75,22.75)
Sample Size	685	1059	685	685	1059	685	685
Longitudinal							
No controls	0.016	-0.038	-0.141**	0.067**	0.089**	0.126**	
	(0.046)	(0.047)	(0.063)	(0.031)	(0.035)	(0.051)	
Controls (biprobit)$^{(1)}$	-0.002	-0.044	-0.090	0.075**	0.069**	0.125**	
	(0.048)	(0.047)	(0.058)	(0.034)	(0.034)	(0.055)	
Controls (biprobit)$^{(2)}$	-0.0008	-0.017	-0.028	0.076	0.066**	0.120 **	-0.230
		(0.029)	(0.039)		(0.033)	(0.052)	(4.75,16.75,59.75,18.75)
Sample Size	366	633	400	366	633	400	400

Note: Bootstrapped standard errors in parenthesis for the numerator and the denominator. Bootstrapped empirical distributions for conditional LATE in parenthesis. Density distributed in four intervals: $(-\infty,-1),[-1,-0.5),[-0.5,0),[0,\infty)$. (*) Significant at 10%. (**) Significant at 5%. (***) Significant at 1%. (1) Estimation which includes three controls: age, a dummy variable for less than upper secondary education and the number of sisters. (2) Estimation which includes four controls: age, a dummy variable for less than upper secondary education, the number of sisters and first lag for LP. Standard errors for the NC under specification (2) are not provided because of convergence problems.

Table 7. Evidence from longitudinal variation in parental health (Daughters sample)

	Numerator			Denominator (compliers)			LATE
	NC	CC	SC	NC	CC	SC	SC
Longitudinal $PH_{t-1} = 0$							
No controls	-0.031	-0.011	-0.263***	0.095*	0.103**	0.220**	
	(0.072)	(0.066)	(0.086)	(0.055)	(0.052)	(0.090)	
Controls (biprobit)$^{(1)}$	-0.047	-0.024	-0.227**	0.104	0.083*	0.248***	
		(0.065)	(0.088)		(0.050)	(0.105)	
Controls (biprobit)$^{(2)}$	-0.017	-0.027	-0.118*	0.101	0.084*	0.238**	-0.496**
		(0.042)	(0.065)		(0.048)	(0.102)	(16.75,29.25,53.25,0.75)
Sample Size	290	482	303	290	482	303	303
		Numerator-SC			Denominator-SC		LATE-SC
Subgroups (controls, biprobit)							
Low educated, no sisters, lagged participants (N=12)$^{(2)}$		-0.229*			0.260**		-0.881**
		(0.138)			(0.122)		(45.75,27.25,26.75,0.25)
Low educated, no sisters, no lagged participants (N=32)$^{(2)}$		-0.039**			0.308**		-0.127**
		(0.019)			(0.122)		(2.00,2.75,93.75,1.50)
High educated, sisters, lagged participants (N=61)$^{(2)}$		-0.167			0.188*		-0.890**
		(0.115)			(0.097)		(46.50,25.75,26.25,1.50)
Low educated, sisters, lagged participants (N=28)$^{(2)}$		-0.207			0.215**		-0.962**
		(0.128)			(0.108)		(49.75,28.00,20.00,2.25)
	NC	CC	SC				
Prevalence of $\overline{PH} = 1$ (%)	12.07	11.20	8.91				
Sample Size	290	482	303				

Note: Bootstrapped standard errors in parenthesis for the numerator and the denominator. Bootstrapped empirical distributions for conditional LATE in parenthesis. Density distributed in four intervals: $(-\infty, -1), [-1, -0.5), [-0.5, 0), [0, \infty)$. (*) Significant at 10%. (**) Significant at 5%. (***) Significant at 1%. (1) Estimation which includes three controls: age, a dummy variable for less than upper secondary education and the number of sisters. (2) Estimation which includes four controls: age, a dummy variable for less than upper secondary education, the number of sisters and first lag for LP. Standard errors for the NC under specificationS (1) and (2) are not provided because of convergence problems.

Table 8. Evidence from multiple measures of parental disability (Parents sample. Cross-section)

Panel A: Compliance types, by instrument	Always Takers			Compliers			Never Takers			Prevalence (%)		
biprobit	NC	CC	SC	NC	CC	SC	NC	CC	SC	NC	CC	SC
$PH = 0$ is $Z = (0, ., ., .)^{(1)}$	0.012	0.030***	0.097***	0.0005	0.077***	0.185***	0.988	0.892***	0.718***	88.94	83.56	76.63
		(0.006)	(0.014)		(0.027)	(0.040)		(0.026)	(0.0371876)			
$PH = 0$ is $Z = (0, 0, 0, .)^{(1)}$	0.007	.011**	0.048***	0.013	0.089***	0.208***	0.980	0.900***	0.744***	72.12	64.77	56.34
		(0.004)	(0.013)		(0.017)	(0.031)		(0.016)	(0.028)			
$PH = 0$ is $Z = (0, 0, 0, 0)^{(1)}$	0.006	0.005	0.016	0.010	0.068***	0.167***	0.983	0.927***	0.817***	53.10	45.75	26.09
		(0.004)	(0.010)		(0.013)	(0.021)		(0.012)	(0.019)			
Sample Size	678	894	552	678	894	552	678	894	552	678	894	552
Panel B: Employment, caregiving and	Numerator			Denominator (compliers)				LATE		Prevalence (%)		
parental disability (biprobit)	NC	CC	SC	NC	CC	SC		SC		NC	CC	SC
Adl: $\widetilde{Z} = (0, 1, 0, .)^{(1)}$	-0.018	-0.015	-0.008	0.019	0.078***	0.155***		-0.052		13.86	16.22	16.49
		(0.040)	(0.064)		(0.022)	(0.038)		(1.75, 16.50, 36.00, 45.75)				
Poor: $\widetilde{Z} = ((1, 0, 0, .), (1, 1, 0, .))^{(2)}$	0.020	0.024	-0.144**	0.008	0.056**	0.236***		-0.609**		9.44	14.09	18.66
		(0.040)	(0.063)		(0.023)	(0.043)		(10.25, 51.75, 36.50, 1.50)				
Dementia: $\widetilde{Z} = (., ., 1, .)^{(1)}$	-0.020	-0.001	-0.169**	0.010	0.253***	0.417***		-0.405*		4.57	4.92	8.51
		(0.065)	(0.083)		(0.070)	(0.072)		(1.75, 31.50, 65.75, 1.00)				
Mobility: $\widetilde{Z} = (0, 0, 0, 1)^{(1)}$	-0.012	-0.016	-0.101*	0.003	0.017	0.061***		-1.684*		19.03	19.02	30.25
		(0.037)	(0.057)		(0.012)	(0.022)		(74.75, 14.50, 5.25, 5.50)				
Summary$^{(1)}$	-0.008	-0.003	-0.099**	0.010	0.068***	0.167***		-0.598**		46.90	54.25	73.91
		(0.029)	(0.043)		(0.013)	(0.021)		(7.50, 58.25, 33.25, 1.00)				
Panel C: Other care	Numerator			Denominator (compliers)				LATE				
biprobit	NC	CC	SC	NC	CC	SC		SC				
Summary$^{(2)}$	-0.008	0.008	-0.103**	0.012	0.068***	0.173***		-0.599**				
		(0.030)	(0.047)		(0.013)	(0.021)		(6.00,56.50,36.75,0.75)				
Summary$^{(3)}$	0.024	0.017	-0.110**	0.014	0.058***	0.172***		-0.641**				
		(0.029)	(0.049)		(0.011)	(0.022)		(11.75,55.75,31.75,0.75)				
Sample Size	678	894	552	678	894	552		552		678	894	552

Note: Bootstrapped standard errors in parenthesis for the numerator and the denominator. Bootstrapped empirical distributions for conditional LATE in parenthesis. Density distributed in four intervals: $(-\infty, -1), [-1, -0.5), [-0.5, 0), [0, \infty)$. (*),(**),(***) Significant at 10%, 5% and 1% respectively. (1) Three controls: age, a dummy variable for less than upper secondary education and the number of sisters. (2) Three controls: age, a dummy variable for less than upper secondary education and a dummy at least one parent receiving informal care from other people. (3) Four controls: same as (2) plus a dummy for at least one parent receiving formal care. Standard errors for the NC are not provided because of convergence problems.

Appendix A1. Description of variables and some statistics

1. Description of variables: We enumerate the list of variables used in the analysis and the codes of the variables of SHARE used for their construction:

1.1. "Daughters-sample": The main variables of interest are defined according to the section 2. The variable LP is equal to 1 if the woman reports in variable ep005 to be employed or self-employed (including working for family business). The variable IC equals to one if the woman reports to have provided care to at least one elderly parent on a daily basis in the last 12 months. This indicator is constructed using the variables sp008, sp009, sp011, and sp019. The variable $\overline{PH}$ equals to 1 if at least one parent is in poor health and is contructed using the variables dn033. Regarding covariates, we use information on the daughter's age, current marital status, education, health, income, children, living parents and siblings. The variable *Age* is constructed based on dn003. The dummy variable *Married/Partnership* is equal to one if the woman is married or engaged in a registered partnership (dn014). Education is measured by four dummy variables (*Educ1, Educ2, Educ3, and Educ4*) generated from the highest level of education completed according to the ISCED-97 code (isced_r).[37] The first dummy corresponds to no schooling, still in school or primary education (ISCED-97 code 1), the second one refers to lower secondary education (ISCED-97 code 2), the third corresponds to (upper) secondary education (ISCED-97 code 3) and, the last one reflects graduate, undergraduate or second level of professional studies (post-secondary, non-tertiary, first stage of tertiary and second stage of tertiary. ISCED-97 code 4-6). Health is measured by the respondent's self-perceived health status (ph003) with one generated dummy variable for each of the categories (*Excellent/Very Good, Good, Fair, and Poor*). Non-wage income (*Non-wage Income*) is computed as the difference between the gross annual total household income (YhhP) and the gross annual individual income derived from employment (YindP) and self-employment (YdipP), expressed in thousands of ppp-adjusted euros.[38] We also consider in the analysis variables reflecting other family responsibilities as the number of living children (*Children*) based on variable ch001,[39] and alternative potential sources of informal care for elderly parents as the number of the respondent's siblings (*Brothers, Sisters*) based on variables dn036 and dn037.

2.1. "Parents-sample": The variable LP is equal to 1 if the family respondent reports in ch016 that the daughter is full-time employed, part-time employed or self-employed or working working for own family business. The variable IC equals to one if at least one parent reports to have received care from the daughter on a daily basis in the last 12 months. This indicator is constructed using the variables sp002, sp003, sp005, and sp021. The multiple measures of parental disability POOR, ADL, DEMENTIA and MOBILITY are described in detail in section 4.3. Specifically, these variables are constructed based on the variables ph003, ph049, ph048 and ph006, respectively. Regarding covariates, we

[37] ISCED stands for International Standard Classification of Education.

[38] The amounts of euros have been corrected for PPP to control for the differences in the price levels among countries.

[39] It is important to remark that these children could be natural, fostered, adopted or stepchildren. For couples, they could be from one member of the couple or from both of them.

use information on the daughter's age, current marital status, education, children, living parents and siblings reported by the family respondent. The variable *Age* is constructed based on ch006. The dummy variable *Married/Partnership* is equal to one if the woman is married or engaged in a registered partnership (ch012). Education is measured by four dummy variables (*Educ1, Educ2, Educ3, and Educ4*) generated from the highest level of education completed according to the ISCED-97 code (generated from the information in variables ch017 and ch018). The first dummy corresponds to no schooling, still in school or primary education (ISCED-97 code 1), the second one refers to lower secondary education (ISCED-97 code 2), the third corresponds to (upper) secondary education (ISCED-97 code 3) and, the last one reflects graduate, undergraduate or second level of professional studies (post-secondary, non-tertiary, first stage of tertiary and second stage of tertiary. ISCED-97 code 4-6). The number of children of each daughter (*Children*) is constructed from the variable ch019 and the number of the daughter's siblings (*Brothers, Sisters*) are computed based on the gender of each child reported by the family respondent in variable ch005. The variable *Fcare* is constructed using the variables hc029, hc032 and cv178. Finally, the variable *Icare_other* is based on the variables sp003 and sp021. In order to identify the four children selected by the program in the case of respondents with more than four children (see footnote 5), we use the children identification variables chselch1-chselch4.

2. Some statistics:

2.1. Summary statistics and kernel density estimates of the distribution of weekly hours of care and hours of work ("Daughters-sample"):

Table A1.1. Weekly Hours of Care and Work
Conditional on Participation[1]

	Hours of Care			Hours of Work		
	NC	CC	SC	NC	CC	SC
P5th	7	7	7	11	10	3
P25th	10.5	7	14	25	22.25	26
P50th (median)	14	14	21	35	35	40
P75th	31	21	28	40	40	40
P95th	84	56	63	49	52	60
Mean	25.75	19.98	25.40	32.35	32.39	35.05
Std.Dev.	23.77	17.70	23.48	10.83	12.65	14.48
Sample Size	16	64	61	560	736	298

Note: (1) Weekly hours of work conditional on being employed and weekly hours of care conditional on being a daily caregiver of a person outside the household.

Figure A1.1: Kernel density estimates of the distribution of weekly hours of work conditional on participation across country pools.

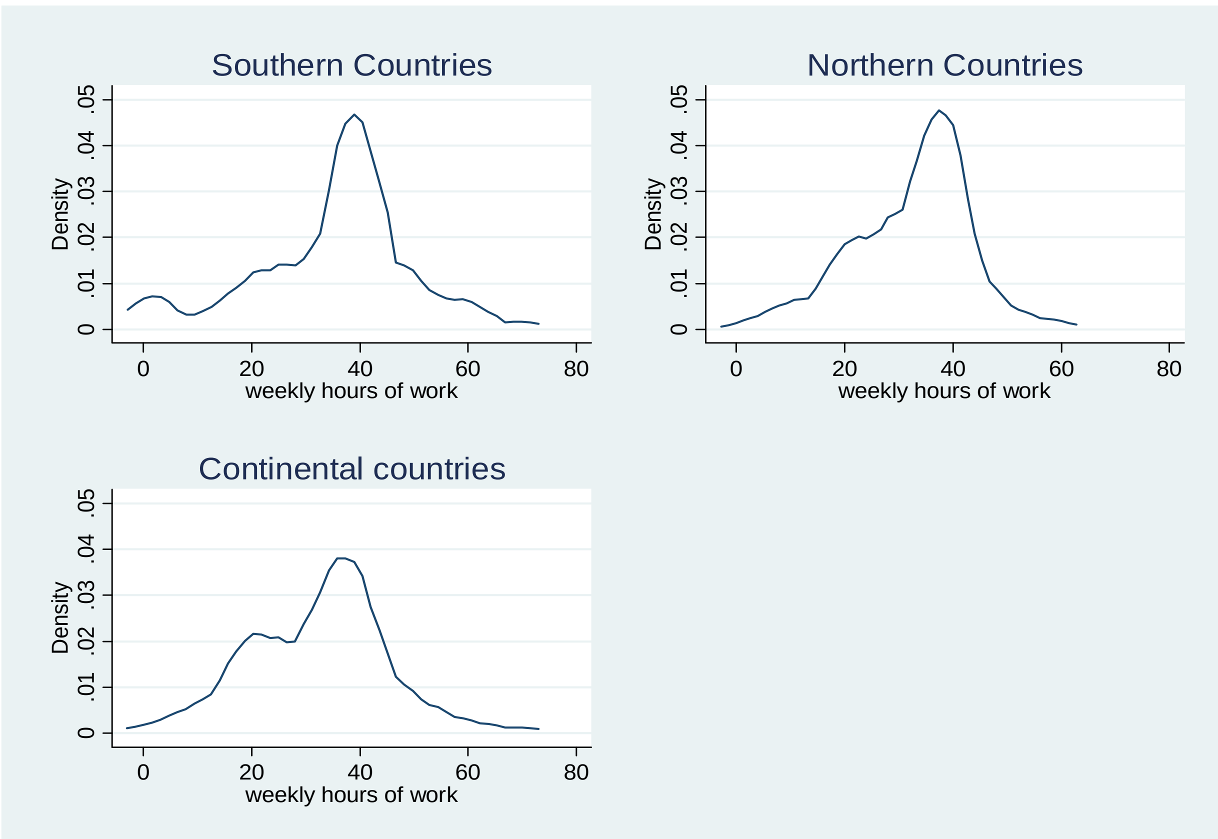
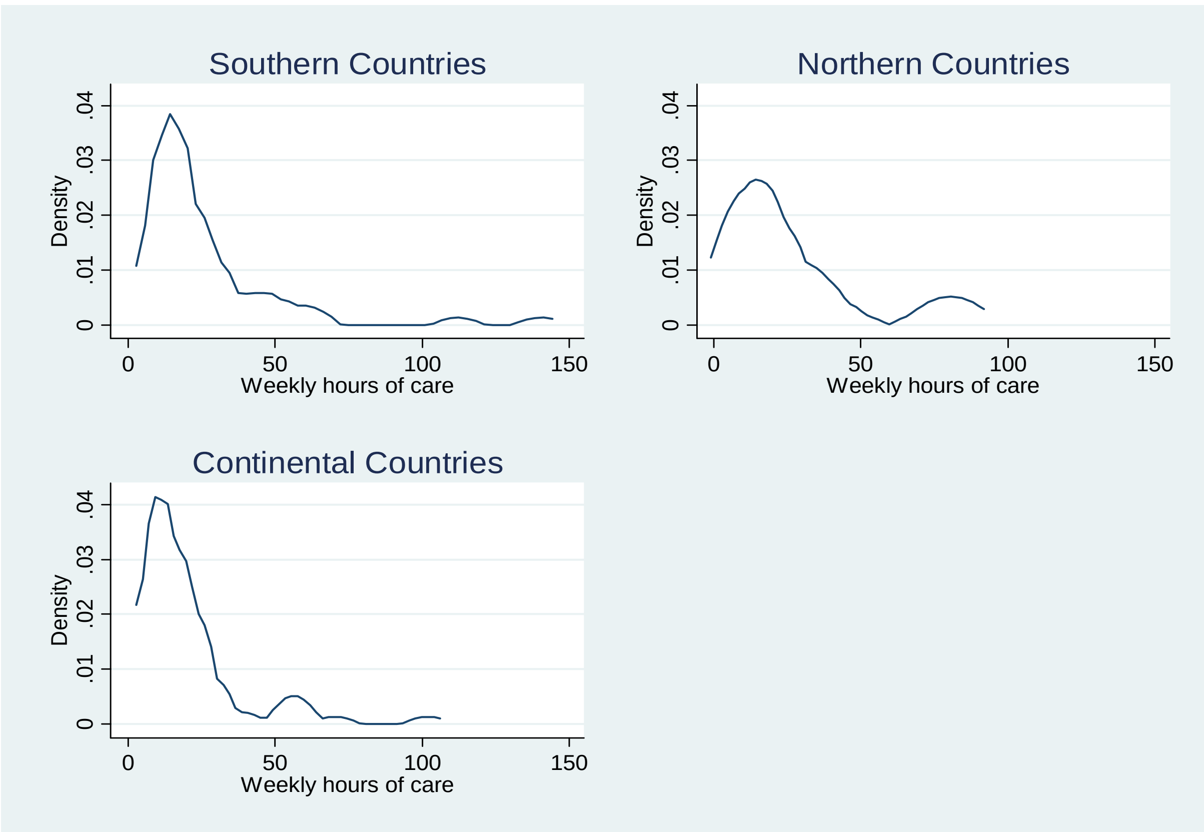


Figure A1.2: Kernel density estimates of the distribution of weekly hours of care conditional on being daily caregiver of someone outside the household.



2.2. Descriptive statistics for the "Parents-sample":

Table A1.2. Descriptive Statistics for the "Parents-Sample" (Wave 2)

	NC	CC	SC
Labour Participant	0.8672	0.7662	0.5199
Daily Caregivers	0.0118	0.0436	0.1413
Age	53.9115	53.6600	53.7663
	(3.007)	(2.8557)	2.9518
Married/Partnership	0.7640	0.7069	0.8279
Education			
Educ1	0.0236	0.0537	0.2627
Educ2	0.2198	0.0995	0.2935
Educ3	0.3053	0.3613	0.2518
Educ4	0.4513	0.4854	0.1920
Children	2.1018	1.7808	1.9203
	(1.2976)	(1.1645)	(1.1097)
Parental Health ($\overline{PH}$)	0.1106	0.1644	0.2337
Parental Health (ADL)	0.2286	0.2863	0.3243
Parental Health (MENTAL)	0.0457	0.0492	0.0851
Parental Health (MOBILITY)	0.3982	0.4888	0.7083
Other Care (Icare_other)	0.2537	0.2640	0.3061
Formal Care (Fcare)	0.3289	0.3009	0.1304
Brothers	1.1106	0.9922	1.0652
	(0.9909)	(0.9716)	(1.0362)
Sisters	1.0383	1.0414	0.9728
	(1.0560)	(1.1086)	(1.0765)
Sample Size	678	894	552

Note: Means of the variables considered in the analysis and standard deviations in parentheses.

Appendix A2. Analysis of the discrete choice model of time allocation.

Unless otherwise stated, we assume that the exclusion and monotonicity restrictions hold. Then the utility difference giving the propensity to work conditional on treatment IC is

$$\Delta U_L(IC, w; .) = w\overline{h} - \alpha_{12}\left[(w\overline{h})^2 + 2C_0 w\overline{h}\right] - \alpha_{31}\overline{h} - \alpha_{32}\left(\overline{h}^2 - 2\overline{h}\widetilde{h}_0\right) + \alpha_5\left[w\overline{h}(T - \overline{h} - IC) - C_0\overline{h}\right] \tag{1}$$

where $C_0 = y + \beta_1 IC - \beta_2 FC(IC)$ is consumption when the daughter does not work and $\widetilde{h}_0 = T - IC$ is her leisure when she does not work. Therefore, the effect of informal care on the propensity-to-work index is

$$\Delta^2 U(w, .) = -2\alpha_{32}\overline{hIC} - 2\alpha_{12}(w\overline{h})\Delta C_0 - \alpha_5\overline{h}(w\overline{IC} + \Delta C_0)$$

which will (almost certainly) be negative if utility is concave in consumption and leisure. Note that $\Delta C_0 = \beta_1\overline{IC} - \beta_2(FC_1 - FC_0)$ is the difference in the 'no work' level of consumption when caring and not caring, and we assume $\Delta C_0 > 0$.[A suffcient condition is that spending on formal care is lower when the daughter provides daily informal care.] Finally, the treatment effect is

$$TE(.) = I(\Delta U_L(0, w; .) + \Delta^2 U(w, .) > 0) - I(\Delta U_L(0, w; .) > 0)$$

which has to be 0 or -1 if $\Delta^2 U(w, .) < 0$.

We now derive the utility difference which determines the caregiving choice. Let $U(0, 0; PH)$ be the baseline utility when the daughter does not work and does not provide informal care. Let $\Delta U_C(PH)$ denote the utility difference or 'propensity to care' for a daughter who does not work. That is,

$$\begin{aligned}\Delta U_C(PH) &= \Delta C_0 - \alpha_{12}\left[(\Delta C_0)^2 - 2y(\Delta C_0)\right] + \alpha_5\left[(\Delta C_0)(T - \overline{IC}) - y\overline{IC}\right] + \\ &\quad + \alpha_2\frac{\partial W_p}{\partial IC}() - \left[\alpha_{31}\overline{IC} + \alpha_{32}(\overline{IC}^2 - 2T\overline{IC})\right]\end{aligned}$$

where the last two terms are the marginal utility of parental welfare and the (leisure) disutility of care, and the terms in the first line are cross terms and the marginal utility of consumption changes brought about by caring.

Recall from section 3.2 that the daughter provides care iff $\max[U(1, 1), U(0, 1)] - \max[U(1, 0), U(0, 0)] > 0$. We can now write this "propensity to care" utility index in terms of a baseline utility and three utility differences as follows:

$$\begin{aligned}IC^*(w, PH; .) &= \max[U(0, 0; PH) + \Delta U_L(0; .) + \Delta^2 U(.) + \Delta U_C(PH), U(0, 0; PH) + \Delta U_C(PH)] \\ &\quad - \max[U(0, 0; PH) + \Delta U_L(0; .), U(0, 0; PH)]\end{aligned}$$

where $U(0, 0; PH)$ is baseline utility, $\Delta U_C(PH)$ is propensity to care when not working, $\Delta U_L(0; .)$ is propensity to work when not caring, and $\Delta^2 U(.)$ is "treatment effect" of care on the propensity to work. Note that the first two terms depend on PH but not on the wage, whereas the third and fourth term depend on the wage but not on PH.

Remark 1: The treatment effects thresholds of Result 1 in section 3.2 are the reservation wages w_{r1} and w_{r2} that satisfy $\Delta U_L(0, w_{r1}; .) = 0$, and $\Delta U_L(0, w_{r2}; .) + \Delta^2 U(w_{r2}, .) = 0$ respectively. When $w < w_{r1}$ the wage is so low that the daughter will never work and the "treatment" effect of daily caregiving on employment is 0. When $w_{r1} < w < w_{r2}$ the daughter works if she is not caring but does not work if she provides care so the treatment effect is -1. When $w > w_{r2}$ the daughter always works and the treatment effect is zero. We know that $w_{r2} > w_{r1}$ because $\Delta^2 U()$ is negative, and this also implies that the treatment effect cannot be positive. [If we relax the exclusion restrictions then the two reservation wages may depend on PH.]

Remark 2: For fixed PH, consider the IC^* index as a function of the wage. For $w < w_{r1}$ it is a constant. For $w_{r1} < w < w_{r2}$ it is clearly decreasing in w. For $w > w_{r2}$ we have $IC^*(w, .) = \Delta^2 U(w.) + \Delta U_C(PH)$. So the slope of the index is the slope of $\Delta^2 U(w.)$. The leading terms of $\Delta^2 U(w.)$ are $-2\alpha_{32}\overline{hIC} - \alpha_5 \overline{hIC} w$ because ΔC_0 is likely to be negligible. If consumption and leisure are complements in utility then $IC^*()$ is still decreasing in this range. [Even if it is not, the slope should be small - it seem very implausible that increasing the wage would increase the propensity to provide daily care of a daughter who works.] To conclude, the IC^* index for both values of PH is flat for low wages, and decreasing therafter.

Remark 3: For any fixed w, the IC^* is strictly greater for $PH = 1$ than for $PH = 0$ as long as $f_{21} > 0$. This implies monotonicity of the instrument, and it is straightforward to prove from the expression for IC^* above, which is based on the Exclusion and Separability restrictions. [The result follows because $\Delta U_C(1) > \Delta U_C(0)$, and the difference between the IC^* indexes for $PH = 1$ and $PH = 0$ is constant in the wage.]

Remark 4: Result 2 establishing the existence of "compliance type" thresholds w_{c1} and w_{c2} follows from Remarks 2 and 3. The thresholds w_{c1} and w_{c2} are defined by the conditions $IC^*(w_{c1}, 0) = 0$, and $IC^*(w_{c2}, 1) = 0$ respectively (see Figure 3). For $w < w_{c1}$ daughters are always takers, for $w_{c1} < w < w_{c2}$ daughters are compliers and for $w > w_{c2}$ daughters are never takers.

Remark 5: The first part of Result C1 is trivial from the expression for $IC^*()$. The second part follows from the plausible assumption that $f_{24} < 0$, because this implies that $\frac{\partial \Delta U_C}{\partial OC}() < 0$ for both values of PH.

Appendix A3. Formulas for the compliance types and the LATE parameter based on a biprobit model.

In our context, an alternative specification to the matching estimator considered in (6) is given by a parametric estimator based on the following model:

$$\begin{aligned} LP &= I(\gamma_0 + \gamma_1 PH + \phi' X + \varepsilon \geq 0), \\ IC &= I(\pi_0 + \pi_1 PH + \delta' X + \nu \geq 0). \end{aligned}$$

where $I(\cdot)$ is the indicator function that is equal to one if the condition in parenthesis holds and zero otherwise and $(\varepsilon, \nu)'$ is the vector of unobservable characteristics of the daughters or parents that could also potentially influence their choices. These error terms are assumed to be *iid* and follow a bivariate normal distribution

$$\begin{pmatrix} \varepsilon \\ \nu \end{pmatrix} \mid PH, X \sim N \left[0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]$$

where ρ is the correlation of the errors and the variances are normalized to 1. Given this specification, we can decompose the population in compliers, always takers and never takers using the fact that the index model assumption for IC is equivalent to the monotonicity assumption on the relationship between IC and PH. In particular, the two potential outcomes for IC are the following

$$\begin{aligned} IC_1 &= I(\pi_0 + \pi_1 + \delta' X + v \geq 0) \\ IC_0 &= I(\pi_0 + \delta' X + v \geq 0) \end{aligned}$$

If we assume that $\pi_1 \geq 0$, the compliance type depends on the individual's value of v as follows:

Compliers: Units with $IC_1 = 1, IC_0 = 0 \Longrightarrow v \geq -\pi_0 - \pi_1 - \delta' X$ but $v < -\pi_0 - \delta' X$.

Always takers: Units with $IC_1 = 1, IC_0 = 1 \Longrightarrow v \geq -\pi_0 - \delta' X$.

Never takers Units with $IC_1 = 0, IC_0 = 0 \Longrightarrow v < -\pi_0 - \pi_1 - \delta' X$.

Therefore, the mass of each group is given by

$$\int [\Phi(\pi_0 + \pi_1 + \delta' X) - \Phi(\pi_0 + \delta' X)] dF(X) \text{ for compliers}$$

$$\int [\Phi(\pi_0 + \delta' X)] dF(X) \text{ for always takers}$$

$$\int [1 - \Phi(\pi_0 + \pi_1 + \delta' X)] dF(X) \text{ for never takers}$$

where Φ denotes the standard normal distribution.

And the average treatment effect for the compliers or LATE based on this parametric specification has the following expression

$$\beta = \frac{\int [\Phi(\gamma_0 + \gamma_1 + \phi' X) - \Phi(\gamma_0 + \phi' X)] dF(X)}{\int [\Phi(\pi_0 + \pi_1 + \delta' X) - \Phi(\pi_0 + \delta' X)] dF(X)}.$$

Appendix A4. Lack of Precision and Large Standard Errors.

As it is stated in the text, standard errors for LATE turn out to be very large for all estimators and therefore this effect is very imprecisely estimated, specially for northern countries. To explain this, we explored the analytical formula for the estimated asymptotic variance of the Wald estimate (IV for the linear model $y_i = \alpha + \beta x_i + u_i$ where both the endogenous regressor and the instrument z_i are binary dummy variables)[40] and how it depends on its determinants. In particular, the asymptotic distribution of this estimator is given by

$$\sqrt{N}(\widehat{\beta}^*_{IV} - \beta^*) \longrightarrow \ N(0,\ \sigma^2 E(Z_i X_i')^{-1} E(Z_i Z_i')(E(Z_i X_i')^{-1})')$$

where $Z_i = (1, z_i)'$, $X_i = (1, x_i)'$ and $\beta^* = (\alpha, \beta)'$. Therefore, the estimated asymptotic variance of $\widehat{\beta}$ is

$$\widehat{Var}(\widehat{\beta}_{IV}) = \frac{\widehat{\sigma}^2}{N} \frac{(\overline{x}^2 + \overline{z} - 2\overline{xz})}{(\overline{z}(1-\overline{z})(\overline{x}_1 - \overline{x}_0))^2}$$

where $\overline{x}$ and $\overline{z}$ are the sample means of x, and z respectively, and $\overline{x}_1$ and $\overline{x}_0$ the sample mean of x for those individuals with $z = 1$ and with $z = 0$ respectively. Notice that we have assumed for simplicity the homoskedasticity of u_i ($E(u_i|z_i) = \sigma^2$).

From this expression, we can analyse how each element of the formula affects the estimated variance. On the one hand, standard errors depend negatively on the mass of compliers $(\overline{x}_1 - \overline{x}_0)$ and the sample variance of the instrument $(\overline{z}(1 - \overline{z}))$. On the other hand, the numerator depends negatively on $\overline{x}$

$$\frac{\partial((\overline{x}^2 + \overline{z} - 2\overline{xz})}{\partial \overline{x}} = 2(\overline{x} - \overline{z}) < 0$$

given that for all three groups of countries the sample mean of daily caregiving is smaller than the sample mean of the instrument having at least one parent in bad health. Therefore, this shows why we also find a gradient North-South when computing the standard errors of our Wald estimate. For northern countries, the estimated proportion of compliers and $\overline{x}$ are much lower than for southern countries whereas $\overline{z}$ is similar in all the three groups of countries. In fact, these two elements are rather small for all groups, which may explain the imprecision of our estimates.[41]

[40] In this notation y, x and z stand for LP, IC and PH, respectively.

[41] This formula and discussion applies only to the standard errors for the Wald estimate. We have not derived the corresponding formulas for the conditional estimators (matching and biprobit) but we think that similar arguments may apply to explain the large standard errors.