Ca' Foscari
University
of Venice

Department
of Economics

Working Paper

Luis Aranda

**Doubling Up: A Gift or a
Shame? Multigenerational
Households and Parental
Depression of Older
Europeans**

ISSN: 1827-3580
No. 29/WP/2013

Doubling Up: A Gift or a Shame? **Multigenerational Households and Parental Depression** **of Older Europeans**

Luis Aranda

*Advanced School of Economics
Ca' Foscari University*

10 November 2013

Abstract: The Great Recession has brought along a rearrangement of living patterns both in the U.S. and in Europe. This study seeks to identify the consequences of the “doubling up” of two or more generations of adults into the same household. In particular, a difference-in-difference (DID) propensity score matching approach is employed to target the causal effect of a change in geographical closeness of respondents and their children —either moving together (doubling up) or apart (splitting up)— on the well-being of the older generation, proxied by their depression score. We find that, although heterogeneous across European regions, in no case does doubling up pose a negative effect to the quality of life of older Europeans. The opposite is true for central and southern Europe, where a double up seems to be followed by a significant reduction in the depression level of the older generation. Our results highlight that, although a negative connotation has usually been attached to multigenerational living arrangements in the post-WWII era, its benefits are evident and, in a time marked by increasing demographic aging, can lead to significant improvements in the quality of life of older Europeans.

Keywords: Doubling up; Depression; Aging; Difference-in-differences; Matching estimator

JEL Codes: J140; I310

Address for correspondence:

Luis Aranda
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
Phone: (+39) 345 311 1707
e-mail: luis.aranda@unive.it

This Working Paper is published under the auspices of the Department of Economics of the Ca' Foscari University of Venice. Opinions expressed herein are those of the authors and not those of the Department. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional character.

The Working Paper Series
is available only on line
(http://www.unive.it/nqcontent.cfm?a_id=86302)
For editorial correspondence, please
contact: wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

Doubling Up: A Gift or a Shame?

Multigenerational Households and Parental Depression of Older Europeans

Luis Aranda

Advanced School of Economics
Ca' Foscari University of Venice

October 2013

Abstract: The Great Recession has brought along a rearrangement of living patterns both in the U.S. and in Europe. This study seeks to identify the consequences of the “doubling up” of two or more generations of adults into the same household. In particular, a difference-in-difference (DID) propensity score matching approach is employed to target the causal effect of a change in geographical closeness of respondents and their children —either moving together (doubling up) or apart (splitting up)— on the well-being of the older generation, proxied by their depression score. We find that, although heterogeneous across European regions, in no case does doubling up pose a negative effect to the quality of life of older Europeans. The opposite is true for central and southern Europe, where a double up seems to be followed by a significant reduction in the depression level of the older generation. Our results highlight that, although a negative connotation has usually been attached to multigenerational living arrangements in the post-WWII era, its benefits are evident and, in a time marked by increasing demographic aging, can lead to significant improvements in the quality of life of older Europeans.

Keywords: Doubling up; Depression; Aging; Difference-in-differences; Matching estimator

JEL classification: J140; I310

1 Introduction: Setting the Stage

Multigenerational homes, which gather together adults from three or more generations, have been on the rise for the last few decades. The post-WWII trend of independent living seems to have been stopped and even slightly reverted, giving way to an increased number of multigenerational households. In recent years, more and more families are deciding to live together under one roof. There has been a sharp increase in the number of households where more than one generation of adults live together, a phenomenon of increasing social and economic importance known as “doubling up.” In particular, an increasing number of young adults (aged 25-34) are moving back into their parents’ home –arguably a strategic and protective response to the economic hardships and high unemployment rates brought about by the Great Recession (Mykyta, 2012; Kaplan, 2012).¹

In this respect, a paradox has been brought to light by ongoing research showing an association between economic downturns, health improvements, and reduced death rates among the elderly (Ruhm, 2000, 2003, 2005; Miller et al., 2009; Hoynes et al., 2012). In searching for explanations to this puzzle, we turn to coresidence trends –and doubling up in particular– in the hope of making useful contributions to the discussion. This hypothesis is based upon the following logic: recessions and economic difficulties may increase the chances that the elderly share a household with their children, who in turn provide support, social interaction and companionship to their parents. By decreasing loneliness and solo-living, this fulfillment of family roles potentially results in an increase in life-satisfaction and emotional well-being of the older generation.²

The causes of multigenerational household arrangements are without a

¹The increasing demographic importance of the group of young adults moving back to their parents’ home has led sociologists to describe them by coining the term “boomerang” generation.

²A positive association between loneliness and depression in old age is consistent in the literature (Green et al., 1992; Singh and Misra, 2009).

doubt of great interest and importance. However, in this paper we shift our attention to the analysis and measurement of its effects (if any) which have been greatly disregarded in the literature, especially so for individuals belonging to the 50+ age group. So far, much has been said and done about the effects of leaving or returning to the parents' household on the younger generation (Taylor et al., 2012; Parker, 2012; Mykyta, 2012; Wiemers, 2012; Kaplan, 2012), but how do older adults fare when such changes take place? This paper therefore takes a closer look at the effects of doubling up on the quality of life of older Europeans, seeking in particular to untangle its effects on their psychological health as proxied by a self-reported depression index. Specifically, should we expect an increase in depressive symptoms deriving from increasing loneliness following a child's move out of the household? Or do older Europeans accept it as nature's course? Is there an adaptation period? In any case, are the effects contingent upon institutional arrangements and cultural factors pervading the region of residence? These questions motivate the present work.

In what follows, the transnational nature of the "double up" phenomenon is documented by exploring its presence and development in the United States and as well as in Europe and providing evidence of its increasing importance as a relevant social occurrence in the last decades. Subsequently, in Section 2 the theoretical foundations of our analysis are laid, spanning from the presentation of the common problem of properly evaluating a given intervention, to the introduction of the difference-in-differences estimator and the subsequent exposition of matching methods, which culminates in the DID matching estimator. Section 3 constructs the empirical methodology from which the results are obtained. In it, the data are described, the variables explained and the problem at hand modeled. In Section 3 the results are presented and justified through a series of balancing tests. The analysis is conducted sequentially, starting with the whole sample but proceeding steadily into a macro-regional dissection where northern and southern Europe are compared

in an attempt to capture any possible latent cultural factors. Finally, the sensitivity and robustness of our results are tested in Sections 5 and 6, followed by a brief discussion and concluding remarks in Section 7.

1.1 The American case

In the twentieth century, a revolutionary change in living arrangements occurred in the United States, characterized by an astronomical increase in solo-living by the elderly.³ This cultural trend toward more autonomy might have been the result of increasing affluence and choice in American society, giving rise to the (increasingly contested) popular belief that privacy is a normal good. However, according to a 2010 Pew Research Center report based on U.S. Census Bureau data, in the United States a growing percentage of adult children living with their parents has been observed in recent decades. As shown in Figure 1, such trend admits a higher proportion of multigenerational households now than at any time since the 1950s, having increased significantly in the last decade (*e.g.* 13% of parents in 2009 reported adult children moving back home in the last 12 months, while 24% of 18 to 34 year-olds say they have moved back in with parents in recent years because of economic conditions (Kochhar and Cohn, 2011; Taylor et al., 2012)). Although the trend grew solidly starting in the 1970s, it has recently received more attention as a result of the economic hardships imposed on families by the Great Recession.

More than a new phenomenon, it looks as though multigenerational households were in the process of making a staggering comeback. The economic safety net that such living arrangements represent go without saying and are backed up by evidence: sharing cost-of-living expenses and other household needs, savings in rent, childcare and old-age care giving (*e.g.* for

³In the beginning of the century, 75% of adults aged 45-64 lived with their children, other relatives, or both, and only one in seven lived alone or in a childless couple. In the end of the century, however, half of the individuals in this age category lived alone (Fischer and Hout, 2006).

Figure 1: Rising share (%) of young adults (25-34) living in multigenerational households in the U.S. *Source:* Pew Research Center analysis of U.S. Decennial Census data, 1940-2000 and 2010 American Community Survey (IPMUS).

grandchildren by grandparents and for grandparents by the rest of the family, respectively), among others. In many cases doubling up shields younger adults from going into poverty: in 2010, 9.8% of Americans aged 25-34 living in a multigenerational household lived below the poverty line, as opposed to 17.4% of those who lived in another type of household (Kochhar and Cohn, 2011). Additionally, according to Figure 2 the relational effect of these arrangements is not all that bad either: only one out of four young adults reports a deterioration in her relationship with her parents resulting from doubling up.⁴

⁴This is accompanied by a paradoxical shift in attitudes against independent living. After being asked “As you know, many older people share a home with their grown children. Do you think this is generally a good idea or a bad idea?” two-thirds of Americans in 1957 considered it a “bad idea,” in a time where only half of the elderly lived independently. However, support for independent living fell all the way down to one-third of Americans in 2000, even though solo-living among the elderly had increased to 70% (Fischer and Hout, 2006).

Figure 2: Percentage of “boomerang children” in the U.S. saying that living with their parents at this stage of life has been “good,” “bad,” or “has made no difference” for relationship. *Note:* based on 25-34-year-olds, $n=121$. “Don’t know/Refused” responses shown but not labeled. *Source:* Parker (2012)

1.2 The European case

Dramatic recent increases in doubling up trends have not only been documented in the United States but also in Europe, where the multigenerational household phenomenon is more pronounced. As reported in 2010 by statistical office of the European Union (Eurostat), 19.6% of European women and 32% of European men aged 25-34 lived with their parents in 2008, as opposed to 18% and 22% of their American counterparts in the same year. In Europe, moreover, the proportion of multigenerational living arrangements encounters great cross-country variation, being lower in Nordic countries and higher in South and East Europe (Table 1).

Parallel to what documented in Section 1.1 for the United States and even though only data from the last decade are available, Figure 3 corroborates the positive trend in the share of young European adults living with their parents. Overall, in the last decade this type of household has seen an average increase of 3.5% in the EU-19 area countries.

Such demographic shift has been underscored by varied cultural, demographic, and economic factors, among which the financial crisis and the following Great Recession are of prominent importance. Arguably, similarly to what evidenced for the United States, the European recession has driven

Table 1: Share of young adults living with their parents by age group and country, averaged over the period 2005-2011.

	Aged 18-34	Aged 25-34
Denmark	16.6	2.1
Sweden	23.2	3.7
Norway	18.5	3.8
Finland	21.0	5.1
Netherlands	33.8	9.0
France	32.2	10.1
Iceland	32.3	11.1
Switzerland ^a	40.8	12.1
United Kingdom	36.4	15.0
Belgium	41.0	15.7
Germany	43.5	16.5
Luxembourg	46.2	22.6
Estonia	45.4	22.8
Austria	46.4	23.8
Ireland	48.7	23.6
Cyprus	54.3	28.8
Lithuania	53.6	30.0
Czech Republic	51.4	31.3
Latvia	55.6	36.4
Hungary	55.2	36.5
Spain	54.8	38.3
Romania ^a	58.3	39.7
Poland	58.7	39.7
Portugal	59.5	43.4
Italy	61.0	44.6
Malta	66.6	46.0
Greece	59.4	49.2
Slovenia	66.8	48.7
Bulgaria	62.0	48.7
Slovakia	70.3	50.5

Notes:

^aData for these countries are for the period 2007-2011.

Source: Eurostat online database (EU-SILC) (ilc_lvps08).

Figure 3: Percentage of young adults aged 18-34 living with their parents in EU-19 area by country (*Source:* Eurostat EU-SILC. Last updated on July 7, 2013; extracted on September 6, 2013.)

families to double up in multigenerational homes.

2 Theoretical Methodology

We use a non-parametric difference-in-difference (DID) propensity score matching approach to assess the causal effect of a child moving in or outside the household (*treatment*) on the depression level of elder Europeans. In other words, we would like to see if a change in proximity between the *closer* child and the respondent/parent has any impact on the psychological health of the respondent/parent.

2.1 The Evaluation Problem

Measuring the effect of a given intervention is usually straight forward. However, for a credible evaluation of the effect of an intervention, one must also be able to make an inference about the outcomes that would have been observed had the treated individuals not undergone treatment, that is, had the participants not participated.⁵ This information is commonly referred to as the *missing counterfactual*. More specifically, the impact of participating in a program is given by:

$$\Delta = Y_1 - Y_0$$

where Y_1 and Y_0 are the outcomes conditional on participation and non-participation in the program, respectively. The evaluation problem is basically a missing data problem: it emerges with the fact that either Y_1 or Y_0 are observed for each participant, but not both (and thus Δ is never observed). This makes it impossible to compute individual treatment effects; consequently, the main parameter of interest in the program evaluation lit-

⁵“Treatment,” “intervention” and “participation” are used interchangeably throughout this paper.

erature is the *mean impact of treatment on the treated* (see *e.g.*, LaLonde, 1986; Heckman, 1990), which is given by:

$$\begin{aligned} ATT &= E(\Delta|X, D = 1) = E(Y_1 - Y_0|X, D = 1) \\ &= E(Y_1|X, D = 1) - E(Y_0|X, D = 1) \end{aligned}$$

where $D = 1$ ($D = 0$) denotes an individual's (non-)participation in the program, for whom outcome Y_1 (Y_0) is observed, and X is a vector of observable individual characteristics used as control variables. Under this notation, the last term, $E(Y_0|X, D = 1)$, is the missing counterfactual.

Under randomized experiments, the missing counterfactual mean can be directly estimated through the use of a control group. Such direct estimates are not available, however, for non-experimental or observational studies, for which the alternative is to adjust the outcomes of the untreated individuals econometrically and use them as proxies for what would have been the counterfactual (Smith and Todd, 2005a). Consequently, selection (or evaluation) bias emerges as the difference between such adjusted outcomes and the true (but unobserved) counterfactual mean. Failure to properly take into account such bias will ultimately lead the program evaluator into making distorted conclusions.

In the present study, we employ a non-parametric difference-in-differences (DID) propensity score matching approach to estimate the missing counterfactual. It does so by combining two popular econometric techniques, namely difference-in-differences and matching, to achieve a contestably more appropriate estimation method than a simple instrumental variable approach, particularly because strong exclusion restrictions are not required (*e.g.*, Blundell and Dias, 2000; Girma and Goerg, 2007). This offers a clear advantage for our purposes, given the dense endogeneity cloud blurring the already intricate relationship between family issues and psychological health (*e.g.* depression),

and the consequent difficulty in finding valid instruments directly correlated to the earlier but not the latter.

2.2 Difference-In-Difference Estimator

The standard DID estimator can be described as follows: outcomes are observed for two groups for two time periods, $t = 0$ and $t = 1$. In the second period, one of the groups is exposed to a treatment, while the other group is not. No units are exposed to the treatment in the first period. Denote $D_{it} = 1$ if individual i has been exposed to the treatment, and $D_{it} = 0$ otherwise. We refer to individuals with $D_{it} = 1$ as treated, and those with $D_{it} = 0$ as controls or untreated.

In panel data settings where the same units within a group are observed in each time period, the average gain over time in the control group is subtracted from the average gain over time in the treatment group. The purpose of this double differencing is to smooth between-group comparisons by a) removing biases resulting from permanent differences between the control and the treatment groups, and b) removing time trends unrelated to the treatment in the treatment group, which could result in biases over time (Imbens and Wooldridge, 2009).

Following Smith and Todd (2005a) framework, we start by defining the possible outcomes for individual i as

$$\begin{aligned} Y_{1it} &= f_1(X_{it}) + U_{1it} \text{ if treated, and} \\ Y_{0it} &= f_0(X_{it}) + U_{0it} \text{ otherwise.} \end{aligned} \tag{1}$$

Here, Y_{it} is the outcome for individual i at time t ; $f(X_{it})$ is a function of the vector of observed individual characteristics X_{it} ; and the residuals U_{it} (or unobservable characteristics): a) are normalized to have mean zero; b) are assumed to be independent of the group indicator; and c) have the same

distribution over time (Smith and Todd, 2005a).

The observed outcome for i is then given by

$$Y_{it} = D_i Y_{1it} + (1 - D_i) Y_{0it} \quad (2)$$

By plugging (1) into (2) and rearranging terms, we have

$$Y_{it} = f_0(X_{it}) + D_i \delta^*(X_{it}) + U_{0it} \quad (3)$$

where treatment impact is given by $\delta^*(X_{it}) = f_1(X_{it}) - f_0(X_{it}) + U_{1it} - U_{0it}$. To ease up on notation, assume δ^* is constant across individuals. Now, denoting t and t' respectively the time periods before and after the intervention and recalling that $D_i \delta^* = 0$ at t for all i , we make use of pre- and post-treatment data to construct the standard DID estimating equation⁶

$$Y_{it'} - Y_{it} = f(X_{it'}) - f(X_{it}) + D_i \delta^* + U_{it'} - U_{it} \quad (4)$$

Simple least-square (OLS) methods can be used to estimate the treatment impact δ^* ⁷, which corresponds to the sample counterpart of the DID population estimand $\hat{\delta}_{DID}$, given by:

$$\begin{aligned} \hat{\delta}_{DID} &= E[Y_{i1}|X_{i1}, D_{i1} = 1] - E[Y_{i1}|X_{i1}, D_{i1} = 0] \\ &= (E[Y_{i1}|X_{i1}, D_{i1} = 1] - E[Y_{i0}|X_{i0}, D_{i0} = 1]) \\ &\quad - (E[Y_{i1}|X_{i1}, D_{i1} = 0] - E[Y_{i0}|X_{i0}, D_{i0} = 0]) \end{aligned} \quad (5)$$

In this equation, we assume that X_i , the vector of individual characteristics,

⁶An implicit restriction in standard DID models is assuming a coefficient equal to one for the lagged outcome Y_{it} (Smith and Todd, 2005a, p. 312).

⁷Allowing for time-specific intercepts that are common across groups, the DID estimator appropriately prevents confounding problems between, for instance, δ^* and a given time-specific intercept (Smith and Todd, 2005a, p. 312).

is predetermined at t (or $t=0$)⁸. This calls for a short note on the treatment of the covariates in DID models. In an ideal DID model, covariates should be treated in a non-parametric way, as in equation (5), in order to avoid any potential inconsistency arising from a functional form misspecification (Abadie, 2005). To illustrate, introducing the covariates linearly into the parametric model given in (4) will help account for heterogeneity in outcome dynamics, but may not be appropriate in the presence of heterogeneity in treatment effects, *i.e.*, if treatment exerts different influences on different groups (Meyer, 1995; Abadie, 2005).

Nevertheless, even a non-parametric treatment of covariates may be jeopardized when the number of covariates required to attain identification is large. To obtain interpretable results in such cases, researchers are forced to make use of other available econometric tools that allow for some sort of integration procedure over X_i . Propensity score matching techniques (Rosenbaum and Rubin, 1983) allow us to do just that, and are introduced following a brief review of matching methods.

2.3 Matching Methods

In measuring the counterfactual, one of the crucial steps is to construct an appropriate control group. We do this through the use of matching techniques. Matching estimators assess the effect of a program by pairing participants with observably similar non-participants and comparing their outcomes (Rosenbaum and Rubin, 1983; Smith and Todd, 2005a). Two assumptions justify the use of matching estimators: the unconfoundedness and the

⁸Abadie et al. (2010, p. 494) make a claim against conditioning on after-treatment characteristics: "...researchers [should] decide on study design without knowing how those decisions will affect the conclusions of their studies. (Rubin, 2001) and others have advocated that the ability to make decisions on research design while remaining blind to how each particular decision affects the conclusions of the study is an important device for promoting research honesty in observational studies." The same argument can and should be extended to non-experimental studies.

overlap assumptions.⁹ The earlier assumes that outcomes are independent of treatment conditional on a set of observable individual characteristics X_i , *i.e.*, there are no unobserved characteristics (beyond the observed covariates) associated both with the potential outcome and the treatment:

$$[Y_i(0), Y_i(1)] \perp D | X \quad (6)$$

Although controversial, this sort of assumption is neither new nor unique to matching methods, for it is also commonly made when conducting multiple regression analyses.¹⁰

On the other hand, the overlap assumption states that for all X_i there is a positive probability of either receiving treatment ($D = 1$) or not ($D = 0$):

$$0 < Pr(D = 1 | X) < 1 \quad (7)$$

which implies that the support of the conditional distribution of X_i overlaps for the $D = 0$ and $D = 1$ groups or, put in a more intuitive way, that a match can be found for all treated individuals. However, if there are regions where both distributions do not overlap, then matching makes sense only if done over the *region of common support* (Smith and Todd, 2005a; Imbens and Wooldridge, 2009) .

Since matching involves comparing the treatment and control units across a number of observable pre-treatment characteristics, it may be hard to implement when the number of such characteristics is large. In that case, the so-called “curse of dimensionality” in econometrics may considerably slow down the convergence of the matching estimator, as the number of observations required increases very rapidly with the dimension of X . This problem

⁹Using Rosenbaum and Rubin (1983) terminology, the combination of unconfoundedness and overlap ensure identification by inferring “strong ignorability:” when both assumptions are satisfied, treatment assignment is “strongly ignorable” given X_i .

¹⁰Given a multiple regression, $Y_i = \alpha + \gamma W_i + \beta' X_i + \varepsilon_i$, the unconfoundedness assumption is equivalent to assuming that ε_i and W_i are independent conditional on X_i (Imbens and Wooldridge, 2009).

can be solved by reducing the dimensionality of the conditioning problem in order to match solely on the basis of a single index which collects the information from all control variables: the univariate propensity score constitutes such an index (Rosenbaum and Rubin, 1983).

2.3.1 Propensity Scores

Simply stated, the propensity score is the probability that a given individual gets exposed to treatment conditional on her characteristics. As shown by Rosenbaum and Rubin (1983), when outcomes are independent of treatment given covariates X , they are also independent of treatment given the propensity score $Pr(D = 1|X)$.¹¹ In their own words, given strong ignorability of treatment assignment, “adjustment for a balancing [*e.g.* propensity] score is sufficient to produce unbiased estimates of the average treatment effect” (p. 45) and, moreover, “units with the same value of the balancing [*e.g.* propensity] score but different treatments can act as controls for each other, in the sense that the expected difference in their responses equals the average treatment effect” (p. 46). Therefore, matching individuals on the single propensity score effectively reduces the dimensionality of the problem and breaks the dimensionality curse. The propensity score can be estimated using parametric methods, such as a logit or a probit model.

¹¹In equation form, Rosenbaum and Rubin (1983) show that $E(D|Y, Pr(D = 1|X)) = E(E(D|Y, X)|Y, X)|Y, Pr(D = 1|X))$, so that if $E(D|Y, X) = E(D|X) = Pr(D = 1|X)$, then $E(D|Y, Pr(D = 1|X)) = E(D|Pr(D = 1|X))$, which is the desired result.

2.3.2 Matching Estimators (Part I: Kernel Matching)

Although different estimators can be used to perform matching methods, most of them are variations of the more general form given by¹²

$$\hat{\delta}_M = \frac{1}{n_{1P}} \sum_{i \in I_1 \cap S_P} [Y_{1i} - \hat{E}(Y_{0i} | D_i = 1, P_i)] \quad (8)$$

where

$$\hat{E}(Y_{0i} | D_i = 1, P_i) = \sum_{j \in I_0} W(i, j) Y_{0j}$$

Under this notation, I_1 and I_0 are the sets of the treated and untreated individuals, respectively; S_P is the common support region; and n_{1P} is the number of treated individuals whose propensity score lies under the region of common support, *i.e.*, those belonging to the set $I_1 \cap S_P$. Weighted averages of the outcomes of the untreated individuals are constructed in order to match the treated as closely as possible; the weights $W(i, j)$ depend on the distance between the propensity scores P_i and P_j (Smith and Todd, 2005a). Additionally, a neighborhood region $R(P_i)$ for each treated individual i can be fabricated with a number of untreated individuals j whose propensity score falls inside i 's neighborhood. Stated differently, nonparticipants for whom $P_j \in R(P_i)$ are matched to participant i and serve as its counterfactual reference for comparison.

Different matching estimators can be constructed depending on how neighborhood regions are defined and how weights $W(i, j)$ are established. Among the most widely used are *nearest-neighbor* matching which uses m closest comparison units; *caliper* matching which builds a radius around a point to create the control group; *stratification* matching which breaks the sample into several intervals and estimates the effect of treatment separately in each region; and *kernel* and *local linear* matching which put some sort of

¹²Our notation follows that of Smith and Todd (2005a). As commonly done in the literature, for ease of notation let $P = Pr(D = 1 | X)$.

distribution around each treated individual so that closer control units receive a higher weight than those which are farther away in the distribution. In small samples, different matching methods may give diverging results, since there exists a trade-off between variance and bias depending on the choice of the algorithm (Heckman et al., 1997).¹³ In our specific case, the limited sample size plays a crucial role in choosing a matching algorithm. Consequently, the likely unsatisfactory number of control units at our disposal limits our choice to only those algorithms where matching with replacement is allowed. To make the most of the control units that we do have, however, using information from more than one nearest neighbor will likely increase the precision of the estimates by achieving a lower variance (Caliendo and Kopeinig, 2008). Moreover, arbitrary choices that have to be made when using caliper and stratification matching (*e.g* the tolerance level and the number of strata, respectively) are not required when using kernel matching.¹⁴ Hence, due to its advantages in smaller samples and the lower variance achieved by using more information from a broader control base, the kernel matching estimator with replacement proves itself the better option for the present study.

Kernel estimators use a kernel-weighted average over multiple individuals in the comparison group to construct a match for each treated person. Following Heckman et al. (1997); Heckman, Ichimura and Todd (1998); Heckman, Ichimura, Smith and Todd (1998), the kernel matching estimator is

$$\hat{\delta}_{KM} = \frac{1}{n_{1P}} \sum_{i \in I_1} \left(Y_{1i} - \frac{\sum_{j \in I_0} Y_{0j} G\left(\frac{P_j - P_i}{b_n}\right)}{\sum_{k \in I_0} G\left(\frac{P_k - P_i}{b_n}\right)} \right) \quad (9)$$

¹³No such problem exists asymptotically for large samples, since all matching estimators become closer to finding perfectly equal matches as the sample size grows (Smith, 2000; Caliendo and Kopeinig, 2008).

¹⁴As per DiNardo and Tobias (2001) and Caliendo and Kopeinig (2008), although different kernel functions can be chosen when performing kernel matching, in practice the choice of these functions appears to be relatively unimportant.

where the kernel function is given by $G(\cdot)$ and b_n is its bandwidth parameter. The weights $W(i, j)$ are given by the term

$$\frac{G(\frac{P_j - P_i}{b_n})}{\sum_{k \in I_0} G(\frac{P_k - P_i}{b_n})}$$

and depend on the distance between each untreated unit in the comparison group and the treated individual for whom the counterfactual is being built up. Additionally, different kernel functions require different specifications of the neighborhood region $R(P_i)$. In estimation, I use the Epanechnikov kernel function, given by $G(s) = (3/4)(1 - s^2)$ for $|s| \leq 1$, else $G(s) = 0$. Under standard conditions (see Smith and Todd, 2005a, footnote 16), the second term in brackets in (9), namely $\sum_{j \in I_0} Y_{0j} W(i, j)$, consistently estimates the counterfactual term in equation (8), $E(Y_{0i} | D = 1, P_i)$.

2.3.3 Matching Estimators (Part II: DID Matching)

Conditioning on observable characteristics may not, however, be enough to completely identify our matching estimators. The assumption that mean outcomes are conditionally mean independent of treatment may be violated and systematic differences between the treated and the untreated may persist even after conditioning on observables. For instance, selection into treatment may be influenced by unobservable or unmeasured characteristics; outcomes may be measured in different ways for treated and untreated individuals, as when different survey tools are used to collect the data; or there may exist fixed differences in outcome levels over time between the treatment and control groups. In the presence of any of those aforementioned factors, the identification conditions required for matching are violated.

If longitudinal data are available, some of the deficiencies observed in matching methods can be ameliorated by incorporating them into a difference-in-differences framework. This results in what Heckman et al.

(1997); Heckman, Ichimura, Smith and Todd (1998) defined as a difference-in-differences matching strategy, which has the additional advantage of removing unobserved time-invariant differences in outcomes between treated and untreated individuals that standard matching estimators fail to eradicate. According to Blundell and Dias (2000, p. 438), the combination of matching estimators and DID methodology has the potential to “improve the quality of non-experimental evaluations results significantly.”

Although analogous to the standard DID regressor estimator defined in (4), the DID matching estimator is convenient in that it does not impose a linear functional form restriction to estimate the outcome’s conditional expectation, and uses the weighting function defined by the matching estimator to re-weight the observations (Smith and Todd, 2005a). The general DID propensity score matching estimator for panel data¹⁵ is given by

$$\hat{\delta}_{DIDM} = \frac{1}{n_{1P}} \sum_{i \in I_1 \cap S_P} \left[(Y_{1it'} - Y_{0it}) - \sum_{j \in I_0 \cap S_P} W(i, j) (Y_{0jt'} - Y_{0jt}) \right] \quad (10)$$

where the weights depend on the type of matching estimator adopted. Beside the overlap or support condition (7), the following equality must also be satisfied for the DID matching estimator to work:

$$E(Y_{0t'} - Y_{0t} | D = 1, P) = E(Y_{0t'} - Y_{0t} | D = 0, P)$$

In what follows, the empirical part of this study is materialized, starting with the description of the data and evolving through the methodology until the estimators are computed, presented, and appraised. The aforementioned DID matching estimator for panel data is implemented throughout as the empirical backbone of this work.

¹⁵The specification varies if cross-sectional instead of longitudinal data is used. However, as reported in Smith and Todd (2005a), the panel data DID matching estimator is more robust than its cross-sectional counterpart.

3 Empirical Methodology

3.1 Description of the Data

We use data from the first (2004), second (2006) and fourth (2010) waves of the Survey on Health, Aging and Retirement in Europe (SHARE)¹⁶, which surveys people aged 50 and over in 19 European countries (and Israel). Since we exploit the longitudinal nature of the data, only respondents present in all three waves mentioned above will be considered, limiting our study to the 10 countries present in those waves, which are: Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Spain, Sweden, and Switzerland. SHARE is a multidisciplinary and cross-national database which provides detailed information on physical and mental health, socio-economic status, and social and family networks of respondents and their households. International comparisons are allowed by the inter-country standardization of all questions.

The current sample is made up of 10,107 individuals participating in wave 1, 2 and 4 (56% females) who make for 30,321 observations, of which 12,463 contain complete information on depression levels, geographical proximity with children, and the usual controls (*e.g.* socioeconomic and health indicators).

¹⁶This article uses data from SHARE 2004, 2006 and 2008, Wave 1, 2 and 4, release 2.5.0 for wave 1 and 2 and release 1.1.1 for wave 4. The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N 211909, SHARE-LEAP, N 227822 and SHARE M4, N 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions).

3.1.1 Independent Variable

The purpose of this paper is to analyze how levels of depression of older Europeans are affected by changes in their geographical closeness with their children. Children proximity in SHARE is measured by asking respondents about their children’s living arrangements, *i.e.* “Where does child [*child name*] live?” with nine possible answers: 1) in the same household; 2) in the same building; 3) less than 1 kilometer away; 4) between 1 and 5 kilometers away; 5) between 5 and 25 kilometers away; 6) between 25 and 100 kilometers away; 7) between 100 and 500 kilometers away; 8) more than 500 kilometers away; and 9) more than 500 kilometers away in another country. Since 97.2% of respondents in our dataset mention having at most five children, our analysis is limited to the geographical proximity of first five children of each respondent.

Given the longitudinal nature of the dataset, we are able to identify changes in family proximity between the respondent and her children from one wave to the next, as well as to observe the respondent’s depression level before and after the move. The analysis is split up into two periods, depending on when the move took place: short term (between wave 1 in 2004 and wave 2 in 2006) and medium term (between wave 2 in 2006 and wave 4 in 2010). These short and medium term schemes seem adequate in capturing the real and lasting effects of a shock in family proximity, since depression is known to be a recurrent condition with symptoms and effects that can take some time to develop and even longer dissipate, continuing for months and even years.¹⁷

One of the drawbacks of the data is that we cannot follow children in time, which, among other things, hinders our ability to make conclusions regarding the gender and age of children. Moreover, although after a move we are unable to identify who the actual mover is, for our purposes the factor of

¹⁷See the description of “depression” provided by the American Psychiatric Association in their website <http://www.psychiatry.org/depression> (accessed on July 23, 2013).

interest remains family closeness *per se* irrespective of which family member is the protagonist of such move (be it child *A*, child *B*, *etc.* or the respondent herself). Thus, given such data limitations, family closeness in our analysis is a measure of geographical proximity between the respondent and her *closest* child. Thus, family proximity, the independent or explanatory variable of this study, is constructed as a binary “in household” indicator, equal to one if the respondent and *at least* one of her children live in the same household and zero otherwise.

3.1.2 Dependent Variable

Our dependent variable of interest is depression, defined on the basis of a symptom-oriented measure known as the EURO-D scale. The EURO-D is made up of 5 depression measures¹⁸ which are harmonized to produce a 12-item scale comprising the following symptoms: pessimism, depressed mood, suicidal thoughts, guilt, sleep, interest, irritability, fatigue, concentration, appetite, enjoyment, and tearfulness (Prince et al., 1999). The index was developed in an effort to compare depression symptoms in fourteen European centers and is constructed by summing the binary items, ranging from a score of 0 (no symptoms) to 12 (all symptoms). According to Prince *et al.*, the EURO-D is a valid and internally consistent scale which correlates well with other well-known mental health measures, providing for a valid comparison of risk factor associations in mental health between units. The EURO-D variable in our sample is distributed as shown in Figure 4.

3.1.3 Matching Covariates (Controls)

The propensity score is built upon a number of observables which may prove themselves important factors influencing the outcome of interest, which in

¹⁸Geriatric Mental State-AGECAT (GMS-AGECAT), SHORT-CARE, Centre for Epidemiological Studies Depression scale (CES-D), Zung Self-Rating Depression Scale (ZSDS), and Comprehensive Psychopathological Rating Scale (CPRS).

Figure 4: Distribution of the EURO-D depression score variable in our sample.

this case is the depression level of the respondent. Such observables include: a) sociodemographic characteristics of the respondent, such as age, gender, marital status, years of education, area of residence (either urban or rural), and health (number of chronic conditions as well as difficulties in performing activities of daily living); and b) her economic situation, measured by her household income (split into quintiles by country), her ability to “make ends meet” financially (*i.e.* financial distress), and her employment status (employed, unemployed, or retired). We also use the available information on the respondent’s offspring to match on the number of children who are still alive per respondent and their average age.¹⁹

Table 2 presents summary statistics of the main variables in our sample. The mean age at survey entry (*i.e.* in the year 2004) is 63.5 and 64.1 years for men and women, respectively. 70% of respondents are married and living

¹⁹Additional variables, such as the number of grandchildren, were included to assess the robustness of the model at hand, but were then dropped for lack of explanatory power. As per Bryson et al. (2002) and Caliendo and Kopeinig (2008), including extraneous variables in the model may aggravate the common support problem, and hence researchers should refrain from over-parameterizing their models.

together with their spouse, while about a fifth of them are either widowed or divorced. In 2004, respondents' children were, on average, 34.9 years old. Approximately one third of them report being financially constrained. When it comes to health, almost half suffer from two or more chronic conditions (a list of 17 items going from asthma and high cholesterol and blood pressure to cancer and Alzheimer's disease), while 10% them report at least 1 limitations in performing activities of daily living (ADLs). More than half of the respondents are retired, and only about 3% of them report being unemployed. One in three live in a rural area. The average respondent declares having between 2 or 3 children.

Table 2: Summary statistics of the main variables in our sample.

Variable	Mean	Median	S.D.	Min	Max	N
<i>EURO-D</i>	2.27	2	2.14	0	11	12463
<i>Age</i>	63.82	63	9.16	50	94	12463
<i>Female</i>	0.57	1	0.49	0	1	12463
<i>Married</i> ^a	0.58	1	0.49	0	1	12463
<i>Single</i> ^b	0.42	0	0.49	0	1	12463
<i>Education (years)</i>	10.75	11	4.42	0	25	12463
<i>Financial distress</i>	0.31	0	0.46	0	1	12463
<i>Income quintile</i>	2.90	3	1.40	1	5	12463
<i>Employed</i>	0.09	0	0.29	0	1	12463
<i>Unemployed</i>	0.03	0	0.16	0	1	12463
<i>Retired</i>	0.57	1	0.49	0	1	12463
<i>Chronic conditions (> 1)</i>	0.45	0	0.50	0	1	12463
<i>ADLs (≥ 1)</i>	0.09	0	0.29	0	1	12463
<i>Urban arean</i>	0.45	0	0.50	0	1	12463
<i>Rural area</i>	0.55	1	0.50	0	1	12463
<i>Number of children</i>	2.46	2	1.32	0	12	12463
<i>Age of children</i>	34.93	34	10.06	2	70	12463

Notes:

^a Married and living together or in a registered partnership.

^b Separated, never married, divorced, or widowed.

3.2 Modeling the Problem

Modeling the problem empirically, we classify individual respondents depending on whether a change in the geographical proximity to their closest child was registered in between two waves, independent of the direction of the change (*e.g.* more or less proximity). Following the theory developed in Section 2, let $D_{it} \in 0, 1$ be an indicator of change of proximity from one wave to the next. If such change is registered, respondents are classified as treated and are assigned a value of $D_{it} = 1$. Otherwise, respondents are left untreated ($D_{it} = 0$) if parental proximity with the closest child remains unchanged between waves. Moreover, let $Y_{1i(t+s)}$ be the respondent's depression score at time $t + s$, $s \geq 0$, following a change in proximity, while $Y_{0i(t+s)}$ represents the depression score had there not been such a change. The causal effect of a change in family proximity for respondent i at time $t + s$ is then given by $Y_{1i(t+s)} - Y_{0i(t+s)}$, which leads to the estimation of the average treatment effect on the treated:

$$E[Y_{1i(t+s)} - Y_{0i(t+s)} | D_{it} = 1] = E[Y_{1i(t+s)} | D_{it} = 1] - \underbrace{E[Y_{0i(t+s)} | D_{it} = 1]}_{\theta} \quad (11)$$

The last term in (11), labeled θ for short, represents the expected outcome that treated respondents would have experienced had there not been a change in proximity, *i.e.* the counterfactual. Given that $Y_{0i(t+s)}$ is unobserved for respondents undergoing a change in proximity, causal inference relies on its proper approximation through a suitable construction of the counterfactual.

Provided the absence of confounding factors which might influence treatment assignment (and thus fill our results with biases and endogeneity), a valid approximation to $E[Y_{0i(t+s)} | D_{it} = 1]$ would be given by the average depression level of respondents not experiencing a change in family proximity, namely $E[Y_{0i(t+s)} | D_{it} = 0]$. This is not always the case, however, and so the need arises to resort to matching techniques in order to minimize estima-

tion bias by constructing an appropriate control group. By building up the probability of receiving treatment (*i.e.* propensity score) for each respondent on the basis of observable characteristics, we are able to pair each individual undergoing a inter-wave change in children proximity with a comparable respondent for whom such move is nonexistent. In this manner, individual differences in observable attributes are minimized and depression score dynamics for the untreated group can be used to construct the counterfactual for the treated individuals. A probit model is used to calculate the propensity score $P(D_{it} = 1) = f(X_{i(t-1)})$, where, as mentioned in Section 2, $X_{i(t-1)}$ constitutes a vector of observed pre-treatment individual characteristics.

The average treatment effect on the treated is then estimated by inserting the propensity score into the DID-matching estimator, as defined in (11). Although this estimator still requires the selection on observables assumption inherent to matching, performing DID effectively eliminates all unobserved time invariant characteristics of respondents which would otherwise bias the results (Girma and Goerg, 2007).

Finally, the assumption of independence conditional on observables given by (6) implies that the treatment and control groups should be balanced made similar in terms of the pre-treatment characteristics. To ensure that no significant differences exist between treatment and control groups, such balance must be achieved. By guaranteeing comparability, a set of properly balanced covariates will make a case for the validity of our results. Another way to diagnose the quality of the propensity score matching estimates is to perform sensitivity and robustness checks by exposing the model to minor changes, which is done in Section 5. The findings presented in the following section are chaperoned by several balancing tests displayed in both numerical and graphical format.

4 Results

The analysis is split into two subsections: the first one takes the whole sample pooling all countries together to perform the DID matching estimates, while the second splits the sample up into two European macro blocks (northern and southern Europe) in an attempt to individuate any potential large scale cultural effects.

4.1 Whole Sample

Table 3 displays the DID matching estimates²⁰ of the doubling up effect when the ten European countries in our sample are pooled together.²¹ As described in Section 2.3.3, such estimates differ from traditional cross-sectional matching in that they remove the time-invariant factors in both the treatment and comparison groups conditional on the propensity score $P(X)$. Surprisingly, the estimates indicate that neither the doubling up nor the separation of a child-parent household between waves has a significant effect on the levels of depression of the parents. This is true both in the short and the medium term when the whole sample is taken into account.

The validity of our estimates is justified by providing graphical and numerical evidence on the successful balancing of the treatment and control groups, which satisfies the crucial assumption that, conditional on the propensity score, the treatment and comparison groups are in fact compara-

²⁰The *diff* and *psmatch2* user-written Stata commands and their subcommands were employed for estimation (Villa (2011) and Leuven and Sianesi (2003), respectively). In particular and as mentioned in Section 2.3.2, matching results throughout this study are based on the Epanechnikov kernel distribution.

²¹Bootstrapping methods with 200 repetitions are used to estimate the standard errors (variance) of the DID matching estimates throughout this study. In spite of being extremely time-consuming and computationally demanding, bootstrapping aims at obtaining a distribution of average treatment effects that approximates the sampling distribution of the population mean (Caliendo and Kopeinig, 2008). Bootstrapping in a matching context has been commonly done in the literature (Lechner, 2002; Heckman et al., 1997; Black and Smith, 2004) although some controversy exists (Imbens, 2004; Abadie and Imbens, 2008).

Table 3: “Double up” and “split up” DID matching estimates for the whole sample.

	More Proximity (Double up)		Less Proximity (Split up)	
	S.T.	M.T.	S.T.	M.T.
<i>Matching estimate</i>	0.191	-0.300	-0.051	-0.225
<i>p-value</i>	0.615	0.161	0.825	0.315
<i>Std. Error</i>	0.380	0.214	0.232	0.224
<i>Number treated</i>	69	215	282	281
<i>Number control</i>	3017	2946	791	573

Notes:

S.T.=Short term (2 years); **M.T.**=Medium term (4 years).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

ble. Table 4 shows that matching is effective in removing differences in observable characteristics between individuals undergoing **treatment (either doubling up with parents or splitting up into different households)** and their control counterparts. It presents the standardized difference²² (or bias) for all the covariates used in the propensity score estimation (Rosenbaum and Rubin, 1983). For instance, the

standardized bias for a given covariate is defined as the difference in means between the treated sample and the matched control group used as comparison, divided by the square root of the average of the variances of such covariate in the original pre-matching samples (Smith and Todd, 2005b).²³

Intuitively, the standardized difference can be described as the difference in means of a covariate divided by the pooled standard deviation. The lower the standardized bias, the more balanced (or comparable) the treatment and control groups in terms of the covariate under scrutiny. Given the lack of

²²Standardized differences of means is commonly preferred to t -tests as a measure of balance. Given the assumptions required for t -tests to work (*e.g.* normal distribution of variables) and its high sensitivity to sample size, post-matching comparisons based on t -tests are greatly disputable.

²³Formally, the standardized bias for a given covariate X_k is calculated as follows (*e.g.*

formal criteria for evaluating the size of the bias, we follow Rosenbaum and Rubin (1983) in assuming that a value of 20 is “large.” Additionally, we abide by the common practice in the literature of estimating the standardized bias for all the covariates included in the matching, since bias reduction for individual covariates might be misleading (*e.g.* if the means are already so close together in the unmatched sample that matching methods have basically no room for improvement, it would seem as if matching was not able to reduce the bias). Notably, adopting the propensity score matching approach reduces the median absolute bias of the models substantially by a factor ranging from 65% to 87%. The standardized differences between the treatment and control samples are all less than 3%. Matching also effectively removes any explanatory power of the covariates in the model, as indicated by a pseudo R-squared close to zero.

The second balancing assessment is given in Figure 5, where the standardized percentage bias across individual covariates before and after matching can be compared. The reduction in standardized percentage bias for each of the models is evident, indicating that the chosen matching technique is effective in removing a substantial part of the standardized percentage biases present in the covariates before the treated and control individuals are matched.

Histograms of the propensity scores of the treatment and control groups are plotted in Figure 6 in order to assess if enough overlap exists to make reasonable comparisons between them. The treated and control cases are shown in green and violet on top and bottom, respectively. The charts are

Rosenbaum and Rubin, 1985; Smith and Todd, 2005b):

$$SDIFF(X_k) = \frac{100 \frac{1}{n_{1P}} \sum_{i \in I_1} \left(X_{ik} - \sum_{j \in I_0} W(i, j) X_{jk} \right)}{\sqrt{\frac{var_{i \in I_1}(X_{ik}) + var_{j \in I_0}(X_{jk})}{2}}}$$

Table 4: Balancing statistics for the whole sample.

	Pseudo R^2		LR χ^2		$p > \chi^2$		Median bias		% reduction in bias
	Raw	Matched	Raw	Matched	Raw	Matched	Raw	Matched	
More Proximity									
<i>Short term</i>	0.067	0.010	89.57	4.09	0.000	0.990	23.2	2.9	87.5
<i>Medium term</i>	0.037	0.001	115.08	1.67	0.000	0.999	6.4	1.4	78.1
Less Proximity									
<i>Short term</i>	0.045	0.003	112.12	3.86	0.000	0.993	7.6	2.6	65.8
<i>Medium term</i>	0.028	0.003	60.92	4.65	0.000	0.982	5.9	2.0	66.1

(a) Double up (short term)

(b) Double up (medium term)

(c) Split up (short term)

(d) Split up (medium term)

Figure 5: Standardized percentage bias across covariates before and after matching for the whole sample. *Note:* the variables in alphabetical order are limitations in activities of daily living (*adl*); age of respondent (*age*); age of children (*agech*); area of residence (*areabldg*); two or more chronic conditions (*chronic2*); financial distress (*findis*); gender (*gender*); income quintile (*incometile*); marital status (*marstat*); number of children (*nchild*); retired (*retired*); unemployed (*unemp*); and years of education (*yrseduc*).

Figure 6: Propensity score histogram for the treatment and control groups for the whole sample (20 bins).

satisfactory in that they show similar propensity score distributions for both groups; put differently, there are enough control cases to serve as analogue counterparts for the treated cases.²⁴ In Figure 7, a more detailed picture of the propensity score distribution is given by increasing the number of bins from 20 to 100.

The balancing tests presented above indicate that the balancing conditions are satisfied and suggest that the chosen matching specification effec-

²⁴Nevertheless, the fact that matching with replacement is being used implies that comparisons would still be possible even if only few control cases existed for a given propensity score, since such cases would be used over and over again as matches to treated cases.

(a) Double up (short term)

(b) Double up (medium term)

(c) Split up (short term)

(d) Split up (medium term)

Figure 7: Propensity score histogram for the treatment and control groups for the whole sample (100 bins).

tively accounts for factors that determine selection into treatment.

The foreseeable question in view of the lack of significance in our findings is whether there exists a cultural factor which may determine the impact of household arrangements on parental depression levels and which would be eclipsed by the pooling of all ten European countries in our data. For instance, it is well known that in Mediterranean countries such as Italy or Spain adult children leave the household (if ever) at a much older age than they would do, say, in Sweden or the Netherlands. Such cultural differences are explored in more detail in what follows.

4.2 North *vs.* South

It is possible that the impact of treatment differs across the treated depending, for instance, on cultural factors, making the estimation of the average treatment effect on the treated ambiguous and inconclusive. Thus, in an attempt to account for such cultural differences, we split the analysis into two geographical macro regions: North (Sweden, Denmark, Germany, the Netherlands, and Switzerland) and South (France, Belgium, Austria, Italy, and Spain).^{25,26} DID matching estimates are then calculated for each of these regions separately and then compared among them to assess the extent to which regional differences play a determinant role in the doubling up phenomenon.

²⁵This arrangement follows Inglehart and Welzel (2005) “cultural map of the world,” a chart that positions countries according to their people’s values and not their geographical location. Using data from the World Values Survey (WVS), they find that two major dimensions of cross-cultural variations –namely, (1) traditional *vs.* secular-rational and (2) survival *vs.* self-expression values– are enough to explain more than 70% of the cross-national variance in a factor analysis of ten indicators. In this way, western Europe is split into two culturally and historically-diverse categories, Protestant and Catholic, which in this work will be referred to as “North” and “South,” respectively.

²⁶In a further re-categorization, the sample is split up into three regions, namely northern, central, and southern Europe. This is presented in Section 5 and aims at assessing the robustness of our results. Although it is clear that the analysis would be best at a country or even city level, small sample sizes make this option statistically unviable.

Table 5 shows the DID matching estimates split up by region.²⁷ One significant effect emerges: in southern Europe, doubling up significantly reduces parental depression in the medium term, making it approximately 19% lower than what it would have been had the double up not taken place;²⁸ on the other hand, no significant effect of doubling up on parental depression levels is observed in northern Europe.

Several reasons make this a noteworthy finding:

- Doubling up never increases depression significantly, neither in the South nor in the North, in the short nor in the medium term. Nor does a split up –or dissolution– of intergenerational households.
- An apparent cultural divide emerges in the case of increased intergenerational proximity, as the effects of a double up in northern and southern Europe seem to move in opposite directions. In the short term, for instance, more proximity decreases parental depression in the North but leads it to an increase in the South (although in neither case significantly so).
- In the medium term, however, more intergenerational proximity increases parental depression in the North but proves itself significantly beneficial in the South. This finding could be attributed to the more family-oriented culture in the South where the so-called welfare State may be lacking in comparison to the North. A weaker welfare State combined with a stronger family tradition in southern Europe may result in a scenario where elderly parents rely mainly on the help, assistance, and companionship of their children. Contrastingly, while family unity and intergenerational support might be the rule in the

²⁷42.6% of respondents live in the North, 57.4% in the South.

²⁸The initial average number of depression symptoms in the South were 2.57. Following the trend in the control group, had the treatment not taken place the number of symptoms would have increased by 0.25, to reach a level of 2.83. According to our results, however, after a double up the depression score of southern Europeans drops to 2.29.

Table 5: “Double up” and “split up” DID matching estimates by regions (northern *vs.* southern Europe).

	Northern Europe				Southern Europe			
	Double up		Split up		Double up		Split up	
	S.T.	M.T.	S.T.	M.T.	S.T.	M.T.	S.T.	M.T.
<i>Matching estimate</i>	-0.543	0.108	-0.078	-0.226	0.559	-0.540	-0.077	-0.220
<i>p-value</i>	0.434	0.731	0.802	0.478	0.243	0.049**	0.804	0.406
<i>Std. Error</i>	0.694	0.313	0.311	0.318	0.478	0.275	0.311	0.264
<i>Number treated</i>	24	82	114	100	43	133	168	182
<i>Number control</i>	1273	1456	187	133	1481	1492	577	432

Notes:

Northern Europe=Sweden, Denmark, the Netherlands, Germany, and Switzerland.

Southern Europe=Belgium, France, Austria, Italy, and Spain.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

South, Northern values might instead encourage professional success, personal independence, and mobility.²⁹

- The fact that doubling up increases depression in the short term but drastically decreases it in the medium term³⁰ in southern European countries may be an indicator of the existence of a parental **adaptability period**, characterized by an initial negative connotation attached to a double up (*e.g.* signaling a problematic situation, such as a lack of independence, unemployment, etc.) but which either turns out beneficial or is just accepted as a fact in the long term. Blatantly put, when it comes to doubling up southern European elders learn to literally “live with it” –and they seem to enjoy it.
- Throughout Europe, the dissolution of intergenerational households might be regarded as “nature’s course” or the “appropriate” life-cycle path, as it reduces depression levels in both short and medium terms for both southern and northern countries, although never significantly so.

The balancing tests that follow (Table 6 and Figures 8, 9, 10, 11, 12, and 13) assert the overall validity of our findings. All eight regional DID matching models presented pass the balancing tests by which treatment and control units can be effectively regarded as comparable. However (and although the estimates are not statistically significant), caution is advised when interpreting the effect of a northern European split up in the medium term, given the relatively small (18.8%) post-matching bias reduction and the fact that some of the matched covariates come close to the 20% standardized bias threshold (Figure 8d).³¹ Comparability among the treatment and comparison groups may be challenged when such threshold is surpassed.

²⁹See citetIsengard2012 for supportive evidence of this intra-European cultural divide.

³⁰Not significantly and significantly so, respectively.

³¹This might be due to the comparatively smaller control group for such case (see Table 5).

Table 6: Balancing statistics by region (northern *vs.* southern Europe).

	Pseudo R^2		LR χ^2		$p > \chi^2$		Median bias		% reduction in bias
	Raw	Matched	Raw	Matched	Raw	Matched	Raw	Matched	
a) Northern Europe									
More Proximity									
Short term	0.095	0.024	49.47	3.00	0.000	0.998	20.1	7.8	61.2
Medium term	0.039	0.007	49.64	3.02	0.000	0.998	12.8	2.7	78.9
Less Proximity									
Short term	0.107	0.031	90.00	16.66	0.000	0.215	16.5	7.6	53.9
Medium term	0.029	0.025	18.64	12.28	0.135	0.505	9.6	7.8	18.8
b) Southern Europe									
More Proximity									
Short term	0.071	0.013	56.80	3.00	0.000	0.998	22.9	6.1	73.4
Medium term	0.037	0.002	68.25	1.56	0.000	0.999	7.7	1.9	75.3
Less Proximity									
Short term	0.034	0.004	55.85	3.81	0.000	0.993	8.0	2.7	66.3
Medium term	0.029	0.007	44.75	6.40	0.000	0.930	5.5	3.1	43.6

Notes:

Northern Europe=Sweden, Denmark, the Netherlands, Germany, and Switzerland.

Southern Europe=Belgium, France, Austria, Italy, and Spain.

(a) Double up (short term)

(b) Double up (medium term)

(c) Split up (short term)

(d) Split up (medium term)

Figure 8: Standardized percentage bias across covariates before and after matching for northern Europe. *Note:* the variables in alphabetical order are limitations in activities of daily living (*adl*); age of respondent (*age*); age of children (*agech*); area of residence (*areabldg*); two or more chronic conditions (*chronic2*); financial distress (*findis*); gender (*gender*); income quintile (*incometile*); marital status (*marstat*); number of children (*nchild*); retired (*retired*); unemployed (*unemp*); and years of education (*yrseduc*).

(a) Double up (short term)

(b) Double up (medium term)

(c) Split up (short term)

(d) Split up (medium term)

Figure 9: Standardized percentage bias across covariates before and after matching for southern Europe. *Note:* the variables in alphabetical order are limitations in activities of daily living (*adl*); age of respondent (*age*); age of children (*agech*); area of residence (*areabldg*); two or more chronic conditions (*chronic2*); financial distress (*findis*); gender (*gender*); income quintile (*incometile*); marital status (*marstat*); number of children (*nchild*); retired (*retired*); unemployed (*unemp*); and years of education (*yrseduc*).

(a) Double up (short term)

(b) Double up (medium term)

(c) Split up (short term)

(d) Split up (medium term)

Figure 10: Propensity score histogram for the treatment and control groups for northern Europe (20 bins).

(a) Double up (short term)

(b) Double up (medium term)

(c) Split up (short term)

(d) Split up (medium term)

Figure 11: Propensity score histogram for the treatment and control groups for southern Europe (20 bins).

(a) Double up (short term)

(b) Double up (medium term)

(c) Split up (short term)

(d) Split up (medium term)

Figure 12: Propensity score histogram for the treatment and control groups for northern Europe (100 bins).

(a) Double up (short term)

(b) Double up (medium term)

(c) Split up (short term)

(d) Split up (medium term)

Figure 13: Propensity score histogram for the treatment and control groups for southern Europe (100 bins).

5 Sensitivity Analysis

In this section, the model is subjected to minor changes in order to assess the reliability of our results. The model can be deemed more solid if the estimates are not sensitive to such changes. In particular, further regional re-categorizations are proposed to test the sensitivity of our estimates to the geographical specification at hand.

Eurostat data³² were used to rank the countries according to the average percentage of young adults (aged 24-35) living in cohabitation with their parents from 2005 to 2011 (see Table 1). This information was used to split our sample into three groupings according to the average coresidence rate: low (less than 10%), medium (more than 10% but less than 20%), and high (more than 20%). A geographical pattern emerges: low coresidence in northern Europe (Denmark, Sweden, and the Netherlands), medium in central (France, Switzerland, Belgium, and Germany), and high in southern Europe (Austria, Spain, Italy).

DID matching estimates for this re-categorization exercise are given in Table 7 and confirm the robustness of our results. Contrary to what would be otherwise expected, the effect of doubling up in countries with the highest percentage of coresidence (Austria, Italy, and Spain) loses conventional significance levels. In this respect, doubling up seems to be more beneficial in countries with medium rates of coresidence (France, Switzerland, Belgium, Germany) than in their (usually referred to as) more family-oriented southern counterparts. This is evidenced by a p-value of 3.6% and a negative matching estimate of -0.700 in the former as opposed to 23.9% and -0.302 in the latter. However, when both groups are combined the effect of doubling up is magnified to a statistical significance level of 1.4%.³³ The fact that the

³²All data can be publicly accessed under the link: http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/data/database.

³³This might be a reflection of the increasing number of treatment and control individuals in the sample.

effect of a double up in northern Europe runs in the opposite direction in both short and medium terms to that in southern Europe is worthy of note.³⁴

Although sample sizes suffer tremendously with the choice to split the sample into more than two macro regions, our findings seem robust in that they advocate for a strong cultural difference present across Europe, hence confirming our cross-regional analysis approach as appropriate in correctly assessing the effect of the phenomenon at hand and enhancing the interpretability of results.

6 Robustness Checks

As discussed in Section 2, in order for the econometric methods employed in this paper to hold, three assumptions were made: the overlap condition and the unconfoundedness assumption (both indispensable to the matching estimator) as well as the parallel trends assumption (crucial to the DID estimator). This section aims at assessing the plausibility of these assumptions under our current framework and data limitations.

First, the robustness of the results with respect to deviations from the common support condition is discussed. The common support condition was imposed in all the aforementioned results: no comparable control units exist for treated individuals that fall outside the common support region, which are then disregarded and hence their treatment effects not calculated. However, if the number of individuals dropped out of the sample is large, the estimates obtained for the remaining units may neither be representative nor consistent for the whole sample (Bryson et al., 2002; Caliendo and Kopeinig, 2008). Therefore, although the common support condition ensures comparability, in some cases its application may be misleading. In response, all the DID matching estimates aforementioned were re-calculated without im-

³⁴Results from the corresponding balancing tests are available from the author upon request.

Table 7: “Double up” and “split up” DID matching estimates by region (northern, central, and southern Europe).

		N	C	S	N+C	S+C	
Double up	S.T.	Estimate	-0.221	-0.153	0.776	-0.171	0.313
		p-value	0.810	0.733	0.269	0.707	0.435
		Std. Error	0.920	0.449	0.701	0.454	0.401
		Treated	13	32	22	46	55
		Control	397	1369	587	2184	1845
	M.T.	Estimate	0.428	-0.700**	-0.302	-0.257	-0.574**
		p-value	0.342	0.036	0.239	0.364	0.014
		Std. Error	0.451	0.334	0.256	0.283	0.235
		Treated	55	82	77	137	159
		Control	1002	1400	543	2380	1916
Split up	S.T.	Estimate	-0.352	-0.072	-0.008	-0.156	0.016
		p-value	0.345	0.817	0.985	0.519	0.950
		Std. Error	0.373	0.313	0.426	0.242	0.249
		Treated	69	113	94	186	210
		Control	111	291	359	404	663
	M.T.	Estimate	-0.185	-0.472	0.205	-0.384	-0.215
		p-value	0.656	0.114	0.652	0.169	0.411
		Std. Error	0.414	0.298	0.454	0.279	0.261
		Treated	65	138	74	201	214
		Control	65	199	289	276	496

Notes:

North (N)=Denmark, Sweden, and the Netherlands.

Center (C)=France, Switzerland, Belgium, and Germany.

South (S)=Austria, Spain, and Italy.

S.T.=Short term (2 years); **M.T.**=Medium term.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

posing the overlap assumption this time around. While preserving a larger sample size, no change in results is observed when the common support condition is not imposed (Table 8). This is an indication that, for the most part, treatment effects are not heterogeneous across individuals inside and outside the common support region.

Second, the unconfoundedness assumption critical to our identification strategy is discussed and examined. Although this assumption is not directly testable, a number of indirect ways of assessing its validity exist and are often used in practice. For the most part, they consist on estimating a causal effect that is known to be zero. An estimated effect different from zero is an indication that the treated and control units are different in terms of this particular covariate conditional on the others, rendering the unconfoundedness assumption less plausible. Otherwise, unconfoundedness would be supported by an estimated effect close to zero. Moreover, the power of this proxy test is higher when the variables used are closely related to outcome of interest and, in consequence, to the unobservable factors likely affecting it.³⁵

To illustrate, consider the number of chronic illnesses diagnosed a priori to a given individual. It is well known that health and depression are closely correlated, while it is at the same time unlikely that the number of long-term illnesses (such as asthma, arthritis, cancer, and Parkinson’s disease) will be affected by a double up. Hence, the causal effect of a double up on the number of long-term illnesses must be zero; otherwise, if significant differences in chronic conditions are found before and after treatment, we would have evidence against the validity of the unconfoundedness assumption. The same case could be made for other variables, such as education and marital status.³⁶

³⁵See Lee (2008) and Battistin et al. (2009) for an application of these methods to a regression discontinuity design.

³⁶Higher educational levels as well as marriage have been shown to have a protective effect against depression (Bjelland et al. (2008) and Ross and Mirowsky (2006); and Simon (2002) and Kim and McKenry (2002), respectively).

Table 8: “Double up” and “split up” DID matching estimates without imposing the common support condition.

		Pooled	North	South	
Double up	S.T.	<i>Estimate</i>	0.235	-0.191	0.549
		<i>p-value</i>	0.469	0.676	0.165
		<i>Std. Error</i>	0.324	0.456	0.395
		<i>Treated</i>	115	52	73
		<i>Control</i>	4089	2505	2098
	M.T.	<i>Estimate</i>	-0.309	-0.095	-0.495**
		<i>p-value</i>	0.101	0.751	0.035
		<i>Std. Error</i>	0.188	0.297	0.235
		<i>Treated</i>	301	134	188
		<i>Control</i>	4178	2591	2148
Split up	S.T.	<i>Estimate</i>	0.034	-0.077	0.050
		<i>p-value</i>	0.859	0.758	0.831
		<i>Std. Error</i>	0.193	0.252	0.236
		<i>Treated</i>	413	208	246
		<i>Control</i>	1115	377	840
	M.T.	<i>Estimate</i>	-0.144	-0.317	-0.085
		<i>p-value</i>	0.451	0.256	0.731
		<i>Std. Error</i>	0.191	0.279	0.246
		<i>Treated</i>	385	183	252
		<i>Control</i>	869	275	665

Notes:

North=Sweden, Denmark, the Netherlands, Germany, and Switzerland.

South=France, Belgium, Austria, Italy, and Spain.

S.T.=Short term (2 years); **M.T.**=Medium term.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In Table 9 we present the estimates of the effect of a double up in southern Europe in the medium term on several outcomes likely to be determined prior to the treatment itself. They include dummies for educational level, marital status, unemployment, and financial distress, as well as the number of children, ADLs, and chronic conditions. As the results indicate, all the considered cases are consistent with our identification restriction of unconfoundedness.

Table 9: Overidentification tests (using variables likely determined before double up).

Variable	Estimate	Std. Error	z	p -value
Primary school	0.007	0.056	0.120	0.905
High school diploma	-0.005	0.062	-0.080	0.939
College degree	-0.002	0.048	-0.040	0.968
Graduate degree	-0.001	0.031	-0.020	0.983
Number of chronic illnesses	0.015	0.063	0.240	0.812
Number of ADLs	-0.049	0.094	-0.520	0.601
Married	-0.008	0.066	-0.120	0.906
Single	0.008	0.065	0.120	0.905
Unemployed	0.011	0.011	0.930	0.355
Financially distressed	-0.028	0.066	-0.043	0.669
Number of children	0.072	0.156	0.460	0.645

An alternative way of assessing the validity of the unconfoundedness assumption is by the use of lagged outcome variables. In this regard, results remain unchanged and significance is unaffected when the lagged depression variable from the first wave was included as a matching covariate (p -value= 0.046).

Finally, two remarks are at hand when it comes to the parallel trends assumption. The first is the finding that the rate of older adults with depressive symptoms tends to increase with advancing age (U.S. Department of Health and Human Services, 1999). This is attributed not only to negative late-life events (such as the loss of friends, widowhood, chronic disorders, declining

health, etc.) but also to a natural process of physiological change that comes with age.³⁷ If such an increasing trend in depressive symptoms with age is accurate and holds true in developed countries and given that our sample is, in fact, representative of the population, there would be no reason to expect the treated and control units in our sample to follow different trends in the absence of treatment.

In a second exercise to test the parallel trends assumption, we conduct a placebo test by exploiting the panel nature of the SHARE dataset. The test consisted in moving the treatment back one period (before it actually happened) to assess whether the difference in trends between the treated and controlled individuals had been, in fact, historically persistent and thus unrelated to the treatment. Focusing in the particular on the case for which the effects of a double up were significantly different between northern and southern Europe, the placebo test delivers no significant difference in pre-treatment depression trends between the treated and controlled individuals ($p=0.50$). Although a more sequential and chronological analysis of trends is not possible with the current data, the results of the placebo test based on the three waves of available data are unable to contest the parallel trends assumption.

7 Discussion and Final Remarks

In the last decade or so, multigenerational households in the developed world have been on the rise, reverting a post-WWII trend of independent living. As a consequence, an increasing number of parents and their adult children are doubling up into the same household: more young adults are returning to

³⁷According to the American Psychiatric Association, the risk of experiencing depression is augmented by natural physiological changes associated with aging. In particular, there is evidence that lower concentrations of folate in the blood and nervous system may contribute to depression, mental impairment, and dementia (<http://www.apa.org/helpcenter/aging-depression.aspx>).

and remaining in the parental household, equivalently to the growing number of older individuals who are moving in with their adult children. Among the main causes behind this social occurrence is, arguably, the harsh economic situation brought about in the last decade by the “Great Recession.”

Regardless of whether they arise out of economic necessity, cultural preferences, or strong family ties, this study is concerned not with the causes themselves (for they have already received considerable attention) but rather with the effects of the doubling up phenomenon, particularly its effects on the psychological health of older Europeans.

Although strongly culture dependent, not once in our study do the levels of depression of older Europeans increase after a double up. On the contrary: after a short adaptation period, in central and southern Europe a double up seems to be accompanied by a greater peace of mind of the older generation. Obtained by means of a painstaking analysis of the SHARE longitudinal data through a difference-in-difference propensity score matching approach, these findings go in defiance of the much negative stigma that has been attached to multigenerational households in recent decades (which is based on the popular view that privacy is a normal good).

Several data limitations give room for improvement. For instance, more waves would allow for a stronger support to the parallel trends assumption, as well as an analysis of the phenomenon in the long term. Additionally, it is not yet possible identify who the closest child is in a given wave nor to discern whether the child moved back to the parents’ home or vice-versa. The motives behind the coresidence decision are unknown as well. Data improvements in all these areas would permit a better identification of important dynamics in intergenerational living.

In an increasingly aging European society where the mental health of older individuals is of growing concern, this finding is not to be taken lightly—especially so in the current climate of slow economic recovery where double ups and rapid shocks to living arrangements are expected to continue. Impor-

tant policy implications in sectors such as health, housing, and employment are evident. Particularly, relevant policies should give multigenerational living its proper weight as a favorable social force that brings about not only economic advantages but proves itself beneficial also in terms of psychological well-being for older generations.

References

- Abadie, A. (2005). Semiparametric difference-in-differences estimators, *Review of Economic Studies* **72**: 1–19.
- Abadie, A., Diamond, A. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: estimating the effect of California’s tobacco control program, *Journal of the American Statistical Association* **105(490)**: 493–505.
- Abadie, A. and Imbens, G. (2008). On the failure of the bootstrap for matching estimators, *Econometrica* **76(6)**: 1537–1557.
- Battistin, E., Brugiavini, A., Rettore, E. and Weber, G. (2009). The Retirement Consumption Puzzle: Evidence from a Regression Discontinuity Approach, *American Economic Review* **99(5)**: 2209–2226.
- Bjelland, I., Krokstad, S., Mykletun, A., Dahl, A., Tell, G. and Tambsf, K. (2008). Does a higher educational level protect against anxiety and depression? The HUNT study, *Social Science and Medicine* **66**: 1334–1345.
- Black, D. and Smith, J. (2004). How robust is the evidence on the effects of the college quality? evidence from matching, *Journal of Econometrics* **121(1)**: 99–124.
- Blundell, R. and Dias, M. C. (2000). Evaluation methods for non-experimental data, *Fiscal Studies* **21**: 427–468.
- Bryson, A., Dorsett, R. and Purdon, S. (2002). The use of propensity score matching in the evaluation of labour market policies, *Department for Work and Pensions Working Paper No. 4*.

- Caliendo, M. and Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching, *Journal of Economic Surveys* **22(1)**: 31–72.
- Choroszewicz, M. and Wolff, P. (2010). 51 million young adults lived with their parent(s) in 2008. Statistics in Focus 50/2010.
URL: http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-10-050/EN/KS-SF-10-050-EN.PDF
- Corselli-Nordblad, L. (2010). Young adults in the EU27 in 2008, Eurostat Press Office. Newsrelease 149/2010.
URL: http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/3-08102010-AP/EN/3-08102010-AP-EN.PDF
- DiNardo, J. and Tobias, J. (2001). Nonparametric density and regression estimation, *Journal of Economic Perspectives* **15(4)**: 11–28.
- Fischer, C. and Hout, M. (2006). *Century of Difference: How America Changed in the Last One Hundred Years*, Russell Sage, New York.
- Girma, S. and Goerg, H. (2007). Evaluating the foreign ownership wage premium using a difference-in-differences matching approach, *Journal of International Economics* **72**: 97–112.
- Green, B., Copeland, J., Dewey, M., Shamra, V., Saunders, P., Davidson, I., Sullivan, C. and McWilliam, C. (1992). Risk factors for depression in elderly people: A prospective study, *Acta Psychiatrica Scandinavica* **86(3)**: 213–217.
- Heckman, J. (1990). Varieties of selection bias, *American Economic Review* **80(2)**: 313–318.
- Heckman, J., Ichimura, H., Smith, J. and Todd, P. (1998). Characterising selection bias using experimental data, *Econometrica* **66**: 5.

- Heckman, J., Ichimura, H. and Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme, *Review of Economic Studies* **64**: 605–654.
- Heckman, J., Ichimura, H. and Todd, P. (1998). Matching as an econometric evaluation estimator, *Review of Economic Studies* **65**: 261–294.
- Hoynes, H., Miller, D. and Schaller, J. (2012). Who Suffers during Recessions?, *Journal of Economic Perspectives, American Economic Association* **26(3)**: 27–48.
- Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: a review, *Review of Economics and Statistics* **86(1)**: 4–29.
- Imbens, G. and Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation, *Journal of Economic Literature* **47(1)**: 5–86.
- Inglehart, R. and Welzel, C. (2005). *Modernization, Cultural Change and Democracy*, New York: Cambridge University Press.
- Kaplan, G. (2012). Moving Back Home: Insurance against Labor Market Risk, *Journal of Political Economy* **120(3)**: 446–512.
- Kim, H. and McKenry, P. (2002). The Relationship Between Marriage and Psychological Well-Being, *Journal of Family Issues* **23(8)**: 885–911.
- Kochhar, R. and Cohn, D. (2011). Fighting poverty in a tough economy, Americans move in with relatives, Pew Research Center, Washington, D.C. Accessed on May 23, 2013.
URL: <http://www.pewsocialtrends.org/files/2011/10/Multigenerational-Households-Final1.pdf>
- LaLonde, R. (1986). Evaluating the econometric evaluations of training programs with experimental data, *American Economic Review* **76**: 604–620.

- Lechner, M. (2002). Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods, *Journal of the Royal Statistical Society* **A165**: 59–82.
- Lee, D. (2008). Randomized Experiments from Non-Random Selection in the U.S. House Elections, *Journal of Econometrics* **142(2)**: 675–97.
- Leuven, E. and Sianesi, B. (2003). Psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. This version 4.0.6 17may2012.
URL: <http://ideas.repec.org/c/boc/bocode/s432001.html>
- Meyer, B. D. (1995). Natural and quasi-experiments in economics, *Journal of Business & Economic Statistics* **13**: 151–161.
- Miller, D., Page, M., Stevens, A. and Filipsky, M. (2009). Why Are Recessions Good for your Health, *American Economic Review: Papers and Proceedings* **99**: 122–127.
- Mykyta, L. (2012). Economic Downturns and the Failure to Launch: The Living Arrangements of Young Adults in the U.S. 1995-2011, *U.S. Census Bureau SEHSD Working Paper* (2012-24).
- Parker, K. (2012). The boomerang generation: Feeling ok about living with mom and dad, Pew Research Center, Washington, D.C.
URL: <http://www.pewsocialtrends.org/2012/03/15/the-boomerang-generation/>
- Prince, M., Reischies, F., Beekman, A., Fuhrer, R., Jonker, C. and Kivela, S. (1999). Development of the EURO-D scale—a European Union initiative to compare symptoms of depression in 14 European centres, *British Journal of Psychiatry* **174**: 330–338.
- Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects, *Biometrika* **70**: 41–55.

- Rosenbaum, P. and Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *The American Statistician* **39(1)**: 33–38.
- Ross, C. and Mirowsky, J. (2006). Sex differences in the effect of education on depression: Resource multiplication or resource substitution?, *Social Science and Medicine* **63**: 1400–1413.
- Rubin, D. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation, *Health Services and Outcomes Research Methodology* **1**: 169–188.
- Ruhm, C. (2000). Are recessions good for your health?, *Quarterly Journal of Economics* **115**: 617–650.
- Ruhm, C. (2003). Good times make you sick, *Journal of Health Economics* **22**: 637–658.
- Ruhm, C. (2005). Healthy living in hard times, *Journal of Health Economics* **24**: 341–363.
- Simon, R. (2002). Revisiting the Relationships Among Gender, Marital Status, and Mental Health, *American Journal of Sociology* **4**: 1065–1096.
- Singh, A. and Misra, N. (2009). Loneliness, depression and sociability in old age, *Industrial Psychiatry Journal* **18(1)**: 51–55.
- Smith, J. (2000). A critical survey of empirical methods for evaluating active labor market policies, *Schweizerische Zeitschrift fuer Volkswirtschaft und Statistik* **136(3)**: 1–22.
- Smith, J. and Todd, P. (2005a). Does matching overcome LaLondes critique of nonexperimental estimators, *Journal of Econometrics* **125**: 305–353.
- Smith, J. and Todd, P. (2005b). Rejoinder, *Journal of Econometrics* **125**: 365–375.

- Taylor, P., Parker, K., Kochhar, R., Fry, R. and Funk, C. (2012). Young, underemployed and optimistic: Coming of age, slowly, in a tough economy, Pew Research Center, Washington, D.C. Accessed on June 5, 2013.
URL: <http://www.pewsocialtrends.org/files/2012/02/SDT-Youth-and-Economy.pdf>
- Taylor, P., Passel, J., Fry, R. and Morin, R. (2010). The return of the multi-generational family household, Pew Research Center, Washington, D.C. Accessed on June 5, 2013.
URL: <http://pewsocialtrends.org/2010/03/18/the-return-of-the-multi-generational-familyhousehold/>
- U.S. Department of Health and Human Services (1999). Older Adults and Mental Health. In: Mental Health: A Report of the Surgeon General. Accessed June 20, 2013.
URL: <http://www.surgeongeneral.gov/library/mentalhealth/chapter5/sec1.html>.
- Villa, J. (2011). Diff: Stata module to perform differences in differences estimation, *Statistical Software Components* . Boston College Department of Economics.
- Wiemers, E. (2012). The Effect of Unemployment on Household Composition and Doubling Up, *National Poverty Center* . Working Paper.