FP7 MULTILINKS Work package 5: **Short report on methodology for analysing intergenerational relationships in a comparative perspective**

Deliverable 5.1, Project number 217523

September, 2009

Arnstein Aassve, Bruno Arpino, Nicolas Robette

Department of Decision Sciences and "Carlo F. Dondena" Research Centre for Research on Social Dynamics, Bocconi University, Via Roentgen 1, 20136 Milano

The MULTILINKS research project touches on several dimensions of social science, including sociology, demography, social policy and economics. At the same time, it requires use of large-scale data analysis, both cross sectional and longitudinal. The key issue in the MULTILINKS project is to consider intergenerational linkages, measured in various ways, and relate them to certain outcomes of interest. As a result, appropriate use of statistical techniques becomes crucial. This short report, the first deliverable of Work Package 5, gives a summary of the issues involved and serves as a guideline for appropriate use of empirical methods.

Data sources

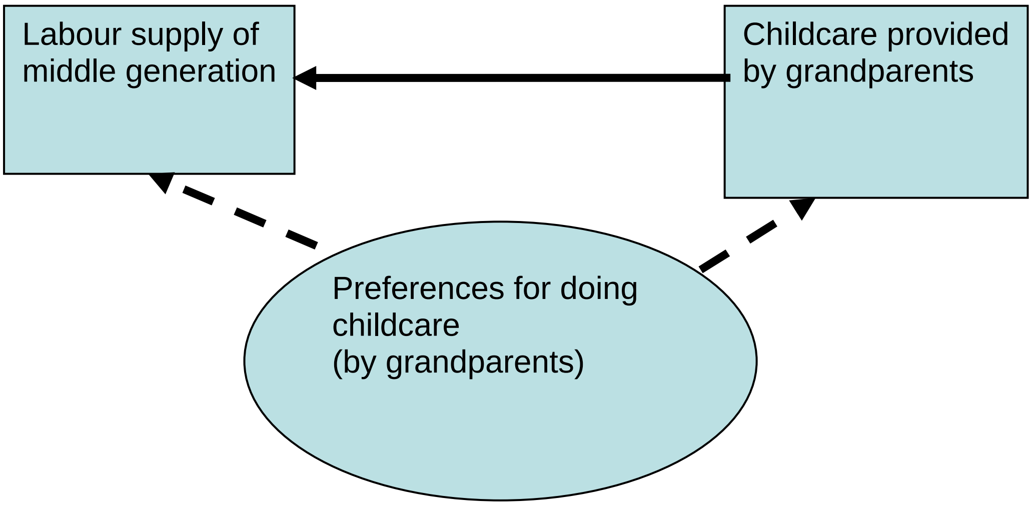
There are several relevant data sources available for this project. Three sources appear particularly useful for the purposes of MULTILINKS, and their design has an impact on the statistical modelling. The datasets are: Gender and Generations Surveys (GGP), Survey of Health and Retirement of Europe (SHARE), and the European Social Survey (ESS). They have features in common but also differ in important ways. All of them contain retrospective information on important family events, whereas only the first two have longitudinal elements. Moreover, they differ in terms of the age of respondents. The ESS has limited information on inter-generational relationships, whereas the first two have such information.

Endogeneity and selection.

The chosen statistical technique depends obviously on the nature of the research question. However, in so far we are considering the effect of behaviours of a certain generation on another, the issue of endogeneity bias becomes relevant (Holland 1986). The problem lies in the fact that generations do not act independently of each other, whereas these interactions are not necessarily observed in the surveys. Moreover, they might have unobserved preferences which affect their behaviour and hence the outcomes of interest. Take grandparents' efforts in childcare for their grandchildren as an example, and consider the effect it may have on labour supply of the middle generation (i.e. mothers' employment decision). Most likely we find a positive relationship. That is, the more childcare grandparents provide, the more mothers work (i.e. the children of the grandparents). We might therefore conclude that having grandparents providing childcare increases labour supply and with simple regression techniques we get an estimate of this effect. If we are considering hours worked as the outcome we can do the estimation with Ordinary Least Squares (OLS), or a Tobit specification (if we want to take into account the left hand side truncation manifested by many women working zero hours), or a probit specification if estimating labour force participation (i.e. the dependent variable is binary and measures whether the respondent work or not). The problem of course, is that because the middle generation works, they also make an imposition on the grandparents – making them more likely to provide childcare. In this light, it is the work decision of the middle generation that drives childcare provision of grandparents. What is not observed in our surveys is how the generations may interact and negotiate in order to reach the labour supply / childcare outcomes (which are observed). Similarly, preferences for work (of the middle generation) and preferences for childcare (of the grandparents) are also unobserved. In so far such unobservables are important for the outcomes we are faced with a classic omitted variable bias, meaning that our parameter estimates do not reflect the true effect of grand parents' childcare effort on labour supply of the middle generation. The endogeneity issue in this setting is not well studied, though techniques are available to deal with it.

Before discussing the statistical techniques available, it is useful to compare this endogeneity problem with a very classic example, namely that of returns to education. That is, what is the effect of educational attainment on earnings? Here the issue is that unobserved ability drives both educational choice (i.e. years spent in education) and earnings. In a naïve regression we find a strong positive association. Once adjusting for the ability of the individual, which is positively correlated with both earnings and education, the effect of education becomes smaller (but still positive). Thus, the naïve estimator (such as OLS) will overestimate the effect of education on earnings. In our setting the endogeneity issue becomes more complicated and there are two reasons for this. First, keeping in line with the example outlined above, the outcome is defined over the middle generation (e.g. the amount of work a woman provides to the market), whereas the endogenous variable (childcare) is defined over the grandparents (i.e. the parents of the middle generation). Thus, the unobserved component is not limited to the characteristics of one individual, but rather the unobserved interaction between two related individuals, as well as their unobserved characteristics, which may include preferences. There is clearly an important need based on theoretical considerations, to better understand how such interaction may impact the outcome of interest. The second issue is of course a need for having access to information about all generations involved. If the middle generation has high education, she is more likely to work, all else equal. If the child in receipt of childcare is demanding, and assuming the grandparents have a general supportive attitude, this may impose more childcare provision (they see a need to help out the middle generation). On the contrary, grandparents may be more reluctant to provide childcare if one is dealing with a difficult child. Similar arguments go for characteristics of the grandparents themselves. If they are in good health, they might be willing to provide more childcare – less if they are in poor health. Also, their age matters. If grandparents are relatively young, then they might still be in gainful employment, making it more difficult to provide childcare. If they are relatively old, they will be less able to provide childcare because of frailty. As always, it would also be important to control for the characteristics of the partner, which undoubtedly will have an impact on women's labour supply decisions and hence childcare outcomes of the grand parents. The key is to better understand the unobserved components that may have an impact on the bargaining process between generations and therefore impact on the intergenerational support. This can be derived from theoretical perspectives but also empirically. The fact that we have access to comparative data from different countries – with different institutional settings – will inform us about the direction of endogeneity bias. At this point it is useful to present a stylized example. Consider the following figure:

Figure 1: Observed and unobserved effects on Labour supply



Here the solid line implies a positive effect, whereas the dotted lines represent negative effects. Hence, we have assumed here that access to childcare increases the labour supply of the middle generation. We also assume that grandparents have a negative preference for doing childcare, and again, we assume that their negative preference for doing childcare has a negative impact on labour supply on the middle generation. Bear in mind that these are assumptions imposed to ease the exposition of the endogeneity problem. There might certainly be contexts where these effects are not negative – or they may have very little impact on the outcome of interest (and the explanatory variable). Still let us assume for the time being that this is the relationship between labour supply, childcare provided by grandparents, and, their preferences for doing childcare. As already stated, a naïve estimator is most likely to give a positive relationship between labour supply and childcare. However, if the stated assumptions in Figure 1 hold, the naïve estimator will in fact underestimate the effect of childcare on labour supply. A correction of endogenity of grandparents' preferences, will consequently increase the effect of childcare on labour supply. One important lesson one can gain from comparative work is that these preference will differ for different countries. For instance, grandparents may have a much stronger negative preference for doing childcare in Scandinavian countries where generally speaking, childcare is provided by the state and not by grandparents. This is in contrast to Mediterranean countries where it is much more common for grandparents to provide childcare. Due to the prevalent social norms, which may of course be driven by the institutional setting and structure of the society, their preference towards providing childcare may be much less negative than what would be observed in Scandinavian countries. If the assumptions hold, we should observe a stronger bias for the Scandinavian countries than in the Mediterranean ones. In fact, in some countries, the preferences towards doing childcare might be neutral or even positive, in which case an endogeneity correction would provide a smaller estimate of the effect of childcare on labour supply.

There are several statistical approaches to deal with these endogeneity issues. In general, it is important to bear in mind that the techniques discussed in the following section are concerned with *controlling* for endogeneity that arises from complex intergenerational links. That is, the methods are not suitable if the purpose is to model the various links, in which case one may want to consider social network analysis (Freeman, 2004). The underlying principle of the following methods, is that due to various complexities, omitted variable biases may arise, whereas the key statistical idea is to control for the bias. Thus, implementing these statistical techniques cannot tell us directly which linkages that may drive the bias.

The first statistical tool to use Instrumental Variable (IV) techniques (Wooldridge 2004). In simple terms, the technique makes a correction for the bias induced by the unobserved and omitted variable. The key requirement for implementing the IV estimator is to have access to variables that are correlated with the endogenous variable of interest (in our example grandparents childcare provision), but uncorrelated with the error term (again in this example – the error term in the regression for women's labour supply). If this requirement is met, this approach provides unbiased estimates and the technique is easily implemented in available statistical software packages. As always, the key problem lies in finding valid and relevant variables that serves as instruments.

To demonstrate the technique in more detail, assume that we are interested in an outcome that can be defined over a binary state. This could for instance whether a person works or not, a highly relevant example when consider to what extent mothers' employment depends on intergenerational linkages. Keeping line with the example outlined above, assume the intergenerational linkages is measured in terms of how much childcare grandparents provide to working mothers. A simple, but naïve, econometric approach consists in specifying a probit model where labour market participation is the dependent variable. The outcome is a binary variable, which indicates the participation in the labour market:

$$W = \begin{cases} 1 = \text{"work"} & \textit{if} \;\; W^* > 0 \\ 0 = \text{"no work"} & \textit{otherwise} \end{cases}$$

Assume that W* is a continuous latent variable representing the propensity to participate in the labour market. If this propensity is sufficiently high to overcome a given threshold, which in probit models is usually set to 0, the woman decides to work (W = 1); otherwise, the woman decides not to participate to the labour market (W = 0). The propensity to participate, W*, is modelled as a linear function of a set of covariates X^w and the childcare help received by grandparents, H (a dummy variable taking value 1 if woman receives childcare help and 0 otherwise):

$$W^* = X^W \beta^W + \delta H + \varepsilon^W \tag{1}$$

The usual assumption in probit models is that the error term follows a standard normal distribution: $\varepsilon^w \sim N(0,1)$. The appropriateness of the simple probit approach described above relies on the assumption of selection on observables; namely, in order to guarantee that δ consistently estimate the effect of grandparents childcare help (H), we have to rule out the possibility of unobservable characteristics that influence both H and W. However, there are several reasons to believe that the assumption of non-existence of selection over observable is not satisfied, i.e. childcare help received from grandparents is endogenous to the woman's labour supply decision. This implies a correlation between the error term ε^w and the variable of interest H, and ultimately that the estimate of δ is biased. With the method of Instrumental Variables, instruments, Z, are variables associated with the endogenous covariate (H) and are supposed to influence the outcome (W) only through the effect on the grandparents help; that is, they should not have a direct effect on the outcome. The IV approach is usually implemented in a two-step procedure (two stages least square, 2SLS). At the first stage H is regressed on Z together with other exogenous variables, whereas at the second stage W is regressed on the exogenous variables and on the predicted value of the endogenous variable, grandparents' childcare help obtained from the previous stage. When dealing with a dichotomous outcome, the 2SLS estimator has been proved to be inconsistent (Foster, 1997). An alternative approach is to specify a joint model combining treatment and outcome and estimate this structural model by Full Information Maximum Likelihood (FIML). Since in our case both the outcome W and the endogenous H variables are binary, we can use a bivariate probit model (see e.g.; Maddala, 1983; Hardin, 1996; Greene, 1997). Other models are necessary in case the outcome measures are not binary.

Model (1) will be therefore estimated together with a second probit where the outcome is the help received by grandparents. Again we can think of this model in terms of the underlying latent variable. We observe the decision to receive (H = 1) or not help (H = 0) as the result of an underlying unobservable propensity to receive grandparents help, H*:

$$H = \begin{cases} 1 = \text{"help"} & if \;\; H^* > 0 \\ 0 = \text{"no help"} & otherwise \end{cases}$$

The propensity to receive help depends on a set of covariate X^H, which may (or may not) coincide with those affecting the decision to work. Since we want to identify the impact of actually receiving grandparents' childcare help on women labour supply, rather than the impact of the propensity to receive grandchild care, we adopt the type II specification in the terminology of Blundell and Smith (1993). That is, we adopt a recursive model in which grandparents' childcare is assumed to influence the probability that a woman works:

$$\begin{cases} W^* = X^W \beta^W + \delta\, H + \varepsilon^W \\ H^* = X^H \beta^H \qquad\quad + \varepsilon^H \end{cases} \tag{2}$$

The error terms of the two equations are allowed to be freely correlated in order to account for the possibility that some unobserved factors influence both decisions to work and receive grandparents help. More precisely, in the bivariate probit model the error terms in the two equations follow a bivariate normal distribution, with zero averages and a variance–covariance matrix which has values of 1 on the leading diagonal while the off-diagonal elements are to be estimated.

$$(\varepsilon^W, \varepsilon^H) \sim BN\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right]$$

We estimate the bivariate probit models using the command biprobit in Stata. As proved by Wilde (2000), it is only necessary to have variation in the set of exogenous regressors to avoid identification problems in this recursive bivariate probit model and exclusion restriction are not required: this is known as "identification by functional form". However, this identification could nevertheless be weak because it relies on the bivariate normality of the error terms and it is common practice to impose some exclusion restriction to improve identification. Therefore, the two sets of covariates X^W and X^H generally speaking differ. In the example pursued here, two possible instruments to estimate the probability that the mother received care for the grandchild appear promising. The first (Z_1) is a dummy variable taking value 1 if the woman's mother is alive at the time of the interview, 0 otherwise and (Z_2) the number of siblings the woman has. Both variables can be a priori considered good instruments since they are expected to be associated with the childcare received but not with the labour supply decision (after help and other covariates are controlled for). In particular, Z_1 is expected to be positively (and strongly) correlated with "Help".

Another strategy is to use propensity score matching techniques (Rosenbaum and Rubin 1983). In contrast to the two approaches listed above, this estimation technique is non-parametric and does not impose any assumptions about functional forms. The matching approach involves comparing pairs of observational units that are similar in all possible respects apart from the "treatment", which may in our case be a change in demographic status. In line with our exemple above, treatment would be defined over grandparents providing childcare and the outcome of interest is the labour supply decision of the mother. Having performed the matching, the difference in outcome is then attributed to the treatment (Dehejia and Wahba 1999, 2002). The matching approach is, however, based on observed variables, whereas selection might be driven by unobserved factors, as discussed above when presenting the IV approach. One solution to this problem is to ensure that both the explanatory and the outcome variable are measured in differences. That is, rather than using variables measured in levels, one constructs variables that represents the change in these variables. The idea is that in so far unobserved heterogeneity, which in regression models would be represented by the error term, remains fixed over time, they (the error term) will be cancelled from the regression function. The requirement is however, that the variables are observed at different points in time, either through longitudinal or retrospective information. As a result, these techniques are more relevant once the longitudinal elements of the GGS and SHARE are available (Aassve et al 2007; Heckman et al 1998), and it is not particularly relevant for the ESS, which is cross-sectional[1]. With two waves the Difference-in-Difference estimator is essentially a fixed effect estimator where treatment (explanatory variables) and outcomes represent a change over two time periods. In general, it is desirable to use more than one approach to assess the sensitivity and consistency of our results. It is well known that IV estimation is sensitive to the choice of instruments and in the case of weak instruments (i.e. exogenous variables are only weakly correlated with the endogenous variable) estimation results may become misleading. For further details on the Propensity Score Matching approach and how it compares with the Instrumental Variable approach, see Arpino and Aassve (2008).

[1] Of course, the ESS does include retrospective information, which can be used to device variables that measures changes over time.

Retrospective versus longitudinal data.

It is clear that the structure of the data influence the extent to which we can establish causal effects for potentially complex inter-linkages. Clearly longitudinal data is desirable as a means to establish these causal effects, especially when the outcomes are not well suited to retrospective questions (Heckman et al 1999). The retrospective information in the GGS and SHARE involves recordings of several trajectories of demographic and labour market behaviour. The longitudinal dimension of GGS (with the introduction of the second and third waves) facilitates panel data regression techniques – as does the SHARE data. Both will improve the causality analysis. The key here is that the panel provides repeated observations for variables that normally cannot be easily measured respectively, or at least not in a reliable way. Typical examples include income and other measures of wellbeing such as happiness or loneliness. In the SHARE data we can assess changes in grandparents' support to the middle generation and hence assess changes in the responses of the middle generation. As already mentioned, the key benefit lies in the possibility of excluding the error term from the regression functions, which, if correlated with explanatory variables, produce biased estimates.

Modelling life course trajectories.

The methods listed above are relevant for analysing the effects and relationships between generations, when there is a need to control for endogeneity issues arising from unobserved complexities. It may also be of interest to consider how certain life course *trajectories* are related across generations. The idea is to better understand if those individuals who undertake an "orderly" transition to adulthood (i.e. completion of education – obtaining stable employment – marriage – childbearing) is related to the support grandparents provide. Alternatively, if data is available, one can compare life course trajectories for the different generations. The analysis of life course trajectories as conceptual units represents a considerable challenge, even when focussing on a specific section of the life course such as the young adult years (Giele and Elder 1998). This happens because one needs to simultaneously take into account the timing, sequencing, and quantum of life course events. However, a promising approach is to use sequence analysis (Abbott and Tsay 2000; Elsinga 2003). The benefit of this approach is that it enables us to study a complex set of life course trajectories as they actually take place, providing ideal-types of trajectories that can be interpreted and analysed in a meaningful way both in terms of theoretical perspectives and policy implications (Aassve et al 2007; Billari and Piccarreta 2006). Given a sequence representation of individuals' life trajectories, one can apply standard clustering algorithms to identify groups with similar life course trajectories (Abbott and Tsay 2000), and importantly, hold it up against the characteristics of the older generation and their life course behaviour. The comparative aspect brings further insight to this issue. It might for instance be the case that some transition patterns appear disadvantageous in some countries, but not in others. Using information on country specific policies and regulations, we can make inference on why such differences arise.

The need for multi-level modelling

The comparative perspective is important for our understanding of why intergenerational relationships differ across different socio-economic and socio-political landscapes. Typically comparative analysis is undertaken by comparing regression coefficients across countries (Aassve et al 2007). In other cases, countries are pooled together in one sample and country differences are estimated through a set of binary variables (Uunk 2004), thereby creating country level fixed effects. Observed differences are then often explained in light of welfare regime theory (Esping-Andersen 1990; Trifiletti 1999). Whereas this is informative, it is nevertheless a crude approach to comparative analysis. There are two main objections to this tradition of comparative analysis. First, controlling for country differences through binary variables, possibly supplemented by some interactions, produces estimates that have little direct meaning. That is, apart from indicating that countries differ in the intercept of a regression line, we cannot say much about what lies behind these differences. Second, behaviour is heterogeneous within countries (Testa and Grilli 2007). For instance, demographic behaviour and hence their linkages between generations, might be very different in South of Europe than what is the case in North Europe. Moreover, observed behaviour in certain regions in one country might be very similar to what is observed in some regions in other countries, whereas at the same time overall country differences remain significant. In essence, this means that behaviour may not only be clustered by countries, but also by regions within countries. Such features of the surveys call for appropriate multilevel modelling (Skrondal and Rabe-Hesketh 2004). In general, these models facilitate a much more rigorous analysis of the underlying causes for the outcomes of interest. A multilevel can be expressed in the following way:

$$Y_{irc} = \alpha_0 + \beta X_{irc} + \mu S_{rc} + \gamma D_c + u_{0c} + \eta_{0rc} + \varepsilon_{irc}$$

Here Y_{irc} represents the outcome of interest measured for individual i in region r of country c, X_{irc} is a vector of individual characteristics, S_{rc} represents covariates measured at the regional level and D_c are variables measured at the country level. Importantly, the error term is decomposed into three different components, the first at the individual-level ε_{irc}, regional-level η_{0rc} and country-level u_{0c}. The country specific error term represent unobserved country characteristics that determine the differences in the outcome across countries. Similarly, η_{0rc} represents unobserved difference across regions within countries. One important issue here is that the decomposition will capture the fact that observations within countries and regions may not be independent of each other. Thus, in case of hierarchical data structure, failing to decompose the error term in the proper way, will produce biased estimates of the standard errors, generally leading to an overestimate of the statistical significance (Goldstein, 2003).

The multilevel structure also enables us to distinguish the variance of the response variable at country level, i.e. the variance across countries, the variance across regions and variability within countries and regions. Through the intra-class correlation coefficient ρ one can compute which is the proportion of total variance accounted for between-country variation. The coefficient is given by:

$$\rho = \frac{Var(u_{0c})}{Var(u_{0c}) + Var(\eta_{0rc}) + Var(\varepsilon_{irc})}$$

where $Var(u_{0c})$ is the variance across countries, $Var(\eta_{0rc})$ across regions in country c and $Var(\varepsilon_{irc})$ among individuals in region r and country c. The typical approach in applied analysis, is to start by estimating a model where no explanatory variables are included (also called a null model). From this one get estimates of variance of the error terms. Thus, one can assess to what extent the total variance in the outcome is factored into country, region and individual levels. Importantly, one can now include country and region level variables to understand how country and region characteristics determine the outcome of interest. Moreover, given the decomposition of the error term we can be sure to have correct estimates of the standard errors, thus providing correct significance levels for the estimated coefficients. As variables are included at country, region and individual levels, one can assess to what extent the observed variables can explain variations of the error term – all along comparing them when estimated in the null model (when no explanatory variables were included).

There are however, some practical limitations to the use of multilevel models. For instance, with the GGS, information from seven countries is available. In general, this is a too small sample to estimate country effects safely. The situation is much better for the ESS where one might have access to 25 countries (depending on which round one is using). The SHARE surveys include 10 countries, which is also a small number if considering country effects in a multilevel setting. However, sample size is generally large enough at the region level. Thus, in so far one has information measured at the regional level this can be easily embedded in the analysis. However, so far our empirical analysis seem to suggest that for many phenomena of interest, variation appear to take place predominantly at the country and individual levels. Another issue concerns interdependencies at the family level. Given the multilevel approach, and the interest in intergenerational links, one might argue to introduce a fourth level – the family. This would be important given the previous discussion about endogeneity of grandparents' behaviour and actions with regard to other family members (e.g. the mother's employment decision). However, in this case the multilevel model does not solve the endogeneity problem. Whereas grandparents' behaviour is not independent of the mother's employment and vice versa, only one of these family members are actually recorded as respondents in the survey. Moreover, the outcome of interest is only defined over one of the family members – in our example the mother. In other words, the endogeneity problem needs to taken seriously also in the multilevel setting.

References

Aassve A, FC Billari and R. Piccarreta (2007) “Strings of Adulthood: A Sequence Analysis of Young British Women’s Work-Family trajectories”, *European Journal of Population*

Aassve A, G Betti, S Mazzuco and L Mencarini (2007) “Marital Disruption and Economic Wellbeing: A comparative analysis”, *Journal of the Royal Statistical Society Series A* 170(3): 781-799

Arpino B and A Aassve (2008) Estimating the causal effect of fertility on economic wellbeing: Data requirements, identifying assumptions and estimation methods, DONDENA Working Paper no 13.

Abbott, A and A Tsay (2000) Sequence Analysis and Optimal Matching Methods in Sociology, *Sociological Methods & Research*, 29, 3-33.

Billari FC and R Piccarreta (2005) Analysing Demographic Life Courses through Sequence Analysis, *Mathematical Population Studies*, 12, 81-106.

Blundell, R.W., Smith, R.J., (1993) Simultaneous microeconometric models with censored or qualitative dependent variables. In: Maddala, G.S., Rao, C.R., Vinoc, H.D. (Eds.), Handbook of Statistics, Vol. II. Elsevier, Amsterdam.

Dehejia R and S Wahba (1999) Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *Journal of the American Statistical Association*, 94 (448): 1053-1062.

Dehejia R and S Wahba (2002) Propensity Score Matching Methods for Nonexperimental Causal Studies, *The Review of Economics and Statistics*, 84(1), 151-161.

Elzinga CH (2003) Sequence Similarity – A Non-Aligning Technique, *Sociological Methods & Research*, 31, 3-29.

Esping-Andersen G (1990) *The Three Worlds of Welfare Capitalism*. Princeton, NJ: Princeton University Press.

Fishbein M and I Ajzen (1975) *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley

Foster, E.M. (1997) IV for logistic regression: an illustration. *Social Science Research*, 26, 487-504.

Freeman LC (2004) *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver: Empirical Press

Gender and Generations Program, UNECE, at: http://www.unece.org/pau/ggp/Welcome.html

Greene, W. (1997) Econometric Analysis, Prentice Hall: Upper Saddle River, New Jersey, USA.

Giele JZ and GH Elder Jr. (Eds.) (1998) *Methods of Life Course Research. Qualitative and Quantitative Approaches*, Thousand Oaks: Sage.

Hardin, J.W. (1996) sg61: Bivariate probit models. *Stata Tech. Bull*., 33, 15-20.

Heckman, J. (1974), Effects of child-care programs on women's work effort, Journal of Political Economy, volume 82: pp. s136-s163

Heckman JJ, H Ichimura, J Smith and PE Todd (1998) Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies* 64, 605 654

Holland PW (1986) Statistics and Causal Inference. *Journal of the American Statistical Association* 81, 945 970.

Maddala G.S., Phillips, P.C.B., Srinivasa, T.N. eds. (1995) Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C.R. Rao, Blackwell, Oxford UK and Cambridge, USA.

Pagani, L., Marenzi, A. (2008), The Labour Market Participation of Sandwich Generation Italian Women, *Journal of Family and Economic Issues*, volume 29: pp. 427-444

Rosenbaum PT and DB Rubin (1983) The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, 70(1), 41 55.

Skrondal A and S Rabe-Hesketh (2004) *Generalized Latent Variable Modelling: Multilevel, Longitudinal and Structural Equation Models*, Boca Raton, FL: Chapman & Hall/CRC.

Smock PJ, WD Manning and S Gupta (1999) The Effect of Marriage and Divorce on Women's Economic Well-Being, *American Sociological Review*, 64, 794-812.

Trifiletti R (1999) Southern European Welfare Regimes and the Worsening Position of Women, *Journal of European Social Policy*, 9(1), 49-64.

Wilde, J. (2000) Identification of Multiple Equation Probit Models with Endogenous Dummy Regressors. *Economics Letters*, 69: 309-312.

Uunk W (2004) The Economic Consequences of Divorce for Women in the European Union: The Impact of Welfare State Arrangements. *European Journal of Population*, 20, 251 285.