Sociology Working Papers

Paper Number 2012-01

Calibrating time-use estimates for the British Household Panel Survey

Cristina Borra
(University of Seville)

Almudena Sevilla-Sanz
(University of Oxford)

Jonathan Gershuny
(University of Oxford)

Department of Sociology

University of Oxford

Manor Road

Oxford OX1 3UQ

www.sociology.ox.ac.uk/swp.html

Calibrating time-use estimates for the British Household Panel Survey*

Cristina Borra
(University of Seville)

Almudena Sevilla-Sanz
(University of Oxford)

and
Jonathan Gershuny
(University of Oxford)

Cristina Borra is Associate Professor at the Department of Economics and Economic History, University of Seville (Spain). Her research focuses on Labour and Demographic Economics, including immigration, job satisfaction, work-life balance, and childcare issues. She has recently published in the Journal of Family and Economic Issues.

Almudena Sevilla-Sanz is Senior Research Officer at the Department of Economics and Centre for Time Use Research (Department of Sociology), University of Oxford. Her research interests include Household and Population Economics, Time Use Research, Consumption Theory and Labour Economics. She has recently published papers in Feminist Economics, Demography, and Oxford Economic Papers.

Jonathan Gershuny is Director of the Center for Time Use Research and Professor in the Department of Sociology at the University of Oxford. His research interests include the analysis of narrative datasets (life and work histories, time use diaries) and interconnections between household organisation, labour force participation, and household formation/dissolution. He has recently published *Changing Times: Work and Leisure in Postindustrial Society*. (Oxford, UK: Oxford University Press. 2000).

* This paper has benefited from comments provided by Man Yee Kan. Any remaining errors are our own. This paper was partly prepared while prof. Borra was an Academic Visitor in the Centre for Time Use Research at the University of Oxford (summer 2009). The authors would like to express their thanks for the financial support provided by the Spanish Ministry of Education and Science (ProjectECO2008-01297), by the Andalusian Government (Convocatoria IAC 2009 2), and by the Economic and Social Research Council (Grant Number RES-060-25-0037).

Correspondence to Almudena Sevilla-Sanz. University of Oxford. Department of Economics, Manor Road Building, Manor Road, OX 02115. Phone: +44 01865 286178 Email: *almudena.sevilla@economics.ox.ac.uk*.

Calibrating time-use estimates for the British Household Panel Survey

Abstract

This paper proposes an innovative statistical matching method to combine the advantages of large national surveys and time diary data. We use data from two UK datasets that share stylised time-use information, crucial for the matching process. In particular, time-diary information of an individual from the Home On-line Study, our donor data set, is imputed to a similar individual from the British Household Panel Survey, our recipient dataset. Propensity score methods are used in conjunction with Mahalanobis matching to increase matching quality.

Keywords

Statistical matching, propensity score, Mahalanobis distance, childcare time

1. Introduction

This paper suggests an innovative approach to increase the power of stylized time-use data usually found in surveys. Current research distinguishes two methods of measuring time-use: direct, stylized questioning and time-diary methods.[1]Stylised questions are typically incorporated into national surveys. However, this kind of data usually does not cover all types of daily activities (e.g., Kan and Pudney 2008; Kan and Gershuny 2009). On the contrary, diaries are seen as the most reliable and comprehensive sources of time budgets (Juster et al. 2003). However, keeping a diary is more complex, more expensive, and more time-consuming than obtaining stylised time-use information and therefore diaries are not usually included in most surveys (Bonke 2005;Kan and Gershuny 2009; Schulz and Grunow 2011). As stated by Kan and Pudney (2008), researchers are frequently faced with the dilemma of opting for detailed and presumably more reliable time-use data at the cost of severe constraints on the type of research that can be done or accepting poorer quality, less wide-ranging time-use data, to give them greater research scope.

In this paper we propose an innovative statistical matching method to solve the dilemma by combining the advantages of large national surveys and time diary data. To do this, we use data from two UK datasets: the Home On-line Study (HOL) and the British Household Panel Survey (BHPS). The 1999–2001 HOL is a unique longitudinal data set, which contains both stylized and diary estimates of time devoted to a wide range of activities.It has a smaller sample and covers a shorter period than the BHPS, though. The BHPS gives only stylized estimates of time use

[1]Recently, researchers also consider experience sampling methods whereby respondents record what they are doing at randomly selected moments of time (Juster et al.2003; Gershuny 2004).

devoted to a few activities. Nonetheless, this information can be used to calibrate time-use estimates for BHPS respondents.

The proposed statistical matching procedure uses variables common to both data sets to identify similar individuals in order to generate a new synthetic dataset (Kum and Masterson 2008). In particular, information of an individual from the HOL, our donor data set, is imputed to a similar individual from the BHPS, our recipient dataset. A special feature of our method is that the set of common variables used in the matching process includes the stylised time-use variables present in both surveys.

The practical application of this paper focuses on the important issue of childcare time, although the methods proposed here can be used for other activities such as leisure, personal care, or education time. Unlike housework time and market work time which are more easily obtained from stylised questions, the BHPS does not contain information on childcare time. Our method provides an alternative way to relying in indirect proxies like working status or number of children for accounting for mothers' time investments in children. This technique can be used to calibrate time devoted to different activities for other longitudinal surveys, such as the PSID and the SOEP, using the auxiliary time use data sets for these countries.

Unlike traditional data imputation techniques which merge evidence across datasets using variables unrelated to the variables of interest, our study contributes to the literature by using a wide range of stylised time use variables (about paid work, childcare responsibilities, and housework) present in both surveys. A previous study by Kan and Gershuny's (2009) also make use of this information. Our study improves on this and other previous studies (Sutherland et al. 2002; Bloemen et al. 2010) by using propensity score matching techniques to impute the missing information. These techniques are preferred over OLS regression prediction because they preserve the variability of the data (Peichl and Schaefer 2009). In addition the use of a propensity score overcomes the dimensionality problem that arises if many covariates are used for matching (Rosenbaum and Rubin 1983). We use recent developments in matching theory to increase matching quality by using Mahalanobis matching on key covariates together with propensity score matching techniques (Rubin and Thomas 2000; Zhao 2004; Stuart and Rubin 2007).

This paper is organized as follows. Section 2 describes the data and the methodology used in the analysis. Section 3 presents the results and assesses the quality of the matching. Section 4 performs the robustness check. Section 5 concludes.

2. Data and Methods

HOL is a three-wave household panel data conducted annually in 1999, 2000, and 2001, and containing about 1,000 households drawn from a national random sample (with an over-sample of computer users). As stated by Kan (2008), this study has two main distinctive advantages: first, it contains both stylised estimates and diary-based estimates of time spent on market work and housework; and second, it collected 7-day diaries from respondents, while other studies usually collect only 1- or 2- day diaries. HOL collected around 2,300 weekly diaries, covering 16,100 diary days.

The BHPS is an annual survey that interviews all members of a random selection of about 5,000 households and 10,000 individuals. From 1994 (Wave 4) onwards, the BHPS asked respondents a series of stylised time-use questions, including usual weekly paid work hours and housework hours, and the distribution of various domestic and childcare tasks within the households.

In both datasets, samples were selected including mothers aged 18 to 64 with children present in the household. In our main study sample, only waves 9 to 11 (1999-2001) of the BHPS are used, but we include an analysis for the whole 1994-2006 BHPS dataset in the robustness check section. Our sample is thus composed of 404 observations from HOL and 7,265 observations from BHPS, in the time-restricted case, and 27,538, in the unrestricted case.

Our main aim is to impute values for a variable that is missing in the recipient dataset, the BHPS, from the donor dataset, the HOL. In our study the HOL is the donor dataset and the BHPS is the recipient dataset. Both datasets have a collection of common variables that are labelled *X*. We want to add one variable, childcare time *C*, from the HOL to the BHPS.

We may think of different methods for imputing the missing variable in the recipient dataset. One is the regression method. In this approach, the specific variable from the donor dataset *C* is regressed on the vector of common variables *X*:

$$C = X\beta + v \tag{1.}$$

The estimated coefficients β from the donor dataset are then used to predict the values of *C* in the recipient dataset (Peichl and Schaefer 2009).[2] In this case, the imputed measures are not values observed on a "similar" individual who participates in the survey, but are simply estimates. The main advantage of this strategy is its simplicity. Connelly and Kimmel (2009) recognize that the main disadvantage is that the variance in the imputed variable is lost since it is a predicted value based on estimated coefficients.

Statistical matching is the other broad option. Matching involves pairing units from different datasets. It uses variables common to both datasets to identify similar records that can be linked

[2]Kan and Gershuny (2009) and Connelly and Kimmel (2009) include examples of this method.

in order to generate a new synthetic data set that allows more flexible analysis than would be possible with the two discrete data sets. In particular, the associations between variables never jointly observed are often the main motivation for interest in such a complete, but synthetic, dataset (Kum and Materson 2008). As stated by Judson and Popoff (2004), the problem of imputing values for a variable that is missing may be thought to be analogous to constructing a pseudo control group for an experimental design study when a random assignment between treatment and control groups is not possible. The main advantage of these matching methods is that the variation in the imputed variable that occurs in the donor dataset is simulated as closely as possible, given that a unique donor amount can be found for each recipient record (Connelly and Kimmel 2009).

Exact matching is only possible for a reduced number of common variables. However, when dealing with multiple covariates it becomes very difficult to find matches with close or exact values of all covariates. Different methods have been devised to summarize the information in the covariates in just one scalar. The two most popular are the propensity score and the Mahalanobis metric (Stuart and Rubin 2007).

To implement propensity score methods, both datasets are combined and a dummy variable is constructed taking value one if the observation belongs to the recipient file BHPS and value 0 if the observation belongs to the donor file HOL. The propensity score is formally defined as the probability of belonging to the recipient database conditional on the common observed covariates:

$$p(X_i) = \Pr(i \epsilon \mathbf{BHPS} | X = x) \quad (2.)$$

Rosenbaum and Rubin (1983) show that at each value of the propensity score, the distribution of the covariates *X* defining the propensity score is the same in the recipient and donor groups. In other words, conditioning on covariates and conditioning on the propensity score will both make the distribution of the covariates in the recipient group the same as the distribution of the covariates in the donor group.

The Mahalanobis metric is a measure of dissimilarity between observations. The Mahalanobis metric measures the distance between units *i* from the recipient dataset BHPS and *j* from the donor dataset HOL weighting each coordinate of *X* in inverse proportion to the variance of that coordinate (Zhao 2004):

$$d_M = (X_i - X_j)D^{-1}(X_i - X_j)' \qquad (3.)$$

Where *D* is the variance-covariance matrix of *X*.

Gu and Rosenbaum (1993) and Rubin and Thomas (2000) compare the performance of matching methods based on Mahalanobis metric matching and propensity score matching, and they find that the two distance measures perform similarly when there are a relatively small number of covariates, but that propensity score matching works better than Mahalanobis metric matching with large numbers of covariates (greater than 5). Nonetheless Zhao (2004) reports that when the sample size is too small, propensity score matching does not perform well compared with Mahalanobis matching, which is relatively robust to specification error.

One possible solution to these complexities considered in the literature is combining both distance measures in the matching algorithm (Lechner 2002; Zhao 2004; Stuart and Rubin 2008). This can be especially illuminating for cases where there are some key covariates on which particularly close matches are desired (Rubin and Thomas 2000). Mahalanobis matching on the key covariates can be combined with propensity score matching to improve matching quality (Stuart and Rubin 2008).

This is the approach followed in this paper. Previous studies on the determinants of mothers' childcare time have underlined the importance of market work and the age of the child in deciding mothers' time spent with their children (Zick and Bryant 1996; Baydar et al. 1999; Bittman, 2004; Joesch and Spiess 2006). Thus, in performing the matching we would like to obtain very high quality pairings with respect to these two variables that are believed to be particularly predictive of the variable of interest childcare time. In order to do so we follow the subsequent matching protocol (Rosenbaum and Rubin 1985; Rubin and Thomas 2000; Lechner 2002): We first specify and estimate a binomial probit model of the probability of belonging to the BHPS sample, that is, we obtain the propensity score. Second, we restrict the BHPS sample to observations whose estimated propensity score lies within the ranges of estimated propensity scores of the HOL sample, that is, we impose the common support condition. Third, all HOL subjects within intervals surrounding each BHPS subject's estimated propensity score are identified as potential matches. And finally, Mahalanobis metric matching is applied on the propensity score and the key covariates considered: usual working hours and age of youngest child. Rubin and Thomas (2000) underline that Mahalanobis metric matching on key covariates within relatively coarse propensity score calipers is an effective method for ensuring matching quality while allowing researchers to use information about the relative prognostic value of different covariates.

3. Results

Consistent with our matching strategy we first estimate the propensity score. We run a probit regression of the binary indicator taking value 1 for observations in the BHPS sample (and 0 for the HOL sample) over the set of common variables. In particular we consider demographic and personal characteristics of the mother (a quadratic in age, highest educational level in three categories, civil status), household characteristics (number of children in different age brackets), reported time-use behaviour (average weekly working hours, a quadratic in usual housework hours, and, for those living in couples, whether childcare is a shared responsibility with their partners). We also include year indicators and a control for computer use to take account of over sampling in the HOL survey.

Table 1 displays the results from the probit model of the likelihood of belonging to the BHPS sample. Many of the variables considered significantly explain sample membership; interestingly, all the variables computed with stylised time-use information.

Table 1. Propensity Score Coefficient Estimates

	Label	Coef.	Std. Err.	
Demographic and personal variables				
age	Age	-0,1811	0,068	***
agesq	Age squared	0,0013	0,001	
married		0,1185	0,158	
Education levels				
degree_further	=1 degree	0,3286	0,165	**
alev	=1 A-level	-0,7139	0,178	***
olev	=1 O-level	-0,6156	0,154	***
less	=1 less than O-level	ref		
Household characteristics				
nch02	number children 0-2 years	-0,3853	0,146	***
nch34	number children 3-4 years	-0,4186	0,133	***
nch511	number children 5-11 years	0,0039	0,072	
Time use and childcare behaviour				
howlng	housework time	0,0365	0,011	***
howlngsq	housework time squared	-0,0003	0,000	*
tothrs	working hours	0,0115	0,004	***
joint_childcare	=1 childcare joint responsability	-0,3454	0,209	*
	interaction			
tot_joint	tothrs*joint_childcare	-0,0159	0,007	**
Design variables				
wave10	=1 year 2000	0,7029	0,131	***
wave11	=1 year 2001	0,6282	0,129	***
cd8use	=1 computer use	-0,5440	0,132	***
_cons		7,5264	1,290	***

Notes:
This table shows the probit regression of treatment status on available covariates. The samples include all mothers 18-64, where mother is defined as having a child under the age of 15 in the house.
* significant at 10% ** significant at 5%; *** significant at 1%.

After imposing the common support assumption[3], Mahalanobis matching within balanced propensity score intervals is performed. Within each block, we pair each recipient unit with that donor for which the Mahalanobis distance metric is lower. We consider three variables in the computation of this metric: the propensity score, the average weekly working hours, and the age of the youngest child.

[3]597 BHPS observations had to be discarded because they had no HOL close matches available.

In order to analyze the quality of the matching Rosenbaum and Rubin (1983) suggest that we check whether significant differences between the average values of the variables for both groups exist after matching. Before matching we expect differences, yet after matching the variables should be balanced in both groups and significant differences should not persist. We followed the test of stratification suggested by Dehejia and Wahba (2002). We divided the observations into blocks based on the estimated propensity score and we later checked whether within each block significant differences in the distribution of each of the explanatory variables persisted. In the application of this test to our sample we obtained nine sectors. We checked whether significant differences in the distribution of observed variables persisted after matching for each sector and obtained that in general the distribution of the variables was balanced.[4]

As a final assessment of the virtues of this method, the distribution of the imputed variable is compared to the distribution of the original variable and the distribution of an OLS regression estimation of it. Following Rubin and Thomas (2000), to increase the reliability of OLS estimates, the regression is performed on the matched sample, that is, imposing the common support assumption. Table 3 presents descriptive statistics of the three variables. The distribution of the Childcare variable estimated by the matching method resembles the distribution of the original HOL variable better than the OLS prediction. In particular, the variation in the original variable is closely simulated, as measured by the estimated standard deviation. In addition, the matching method is utterly superior to the OLS regression in that no negative childcare hours are predicted and in that the estimated proportion of mothers reporting zero hours of childcare is closer to the original variable, as evidenced by the almost negligible proportion mothers with OLS-estimated zero (or negative) childcare hours.

---

[4] We used the procedure developed by Leuven and Sianesi (2003) for STATA (revised Nov. 2010). A summary of the results is presented in the Appendix.

Table 3 Descriptive Statistics of Original and Estimated Variables. Restricted sample

Variable	No. Obs	Estimated No. Obs	Mean	Std. Dev.	Min	Max	Proportion zero childcare
Original		398					
Childcare	404		15.18	19.61	0.00	128.50	0.141
Matching		3711					
Childcare	6668		17.45	18.21	0.00	97.00	0.097
OLS Childcare	6668	3711	17.44	10.79	-3.86	51.52	0.002

Notes:
The samples include all mothers 18-64, where mother is defined as having a child under the age of 15 in the house. All estimates are adjusted to take into account sample weights.

4. Robustness check

In order to test the sensitivity of our results with respect to the sample selection criteria being used, we perform the matching process using the unrestricted BHPS sample (1994-2006) together with the HOL sample. In this case, the calibration involves assuming that time-use patterns among different social groups have not changed significantly over the years, as well supported by past research (Gershuny, 2000).

In the estimation of the propensity score, the year dummies were obviously not included. Additional interaction variables were used as explanatory variables in order to reach convergence in Dehjia and Wahba's (2003) proposed algorithm. Nonetheless, results were quite similar to those reported in Table 1 for the restricted sample.[5]

Table 4 reports the distribution of the estimated matched variable, together with the distribution of the original variable and an OLS regression estimation of it. Again the matched estimates

[5] The additional variables were interaction terms between total working hours and age, civil status, education categories, number of children, childcare responsibility, and computer use. Results are available upon request.

resemble the original variable more closely with regard to variability of the data and proportion of zero outcomes.

Table 4 Descriptive Statistics of Original and Estimated Variables. Unrestricted sample.

Variable	No. Obs	Estimated No. Obs	Mean	Std. Dev.	Min	Max	Proportion zero childcare
Original		398					
Childcare	404		15.18	19.61	0.00	128.50	0.141
Matching		18514					
Childcare	25,206		17.97	19.69	0.00	128.50	0.077
OLS Childcare	25,206	18514	17.58	11.06	-7.69	57.40	0.012

Notes:
The samples include all mothers 18-64, where mother is defined as having a child under the age of 15 in the house. All estimates are adjusted to take into account sample weights.

5. Conclusions

This paper proposes an innovative statistical matching method to combine the advantages of large national surveys and time diary data. In particular we use stylized time use data, usually found in large national datasets, to match similar individuals from the national survey and the diary data. The proposed method involves performing Mahalanobis matching on special key variables within relatively coarse intervals of the estimated propensity score.

The method is tested using two UK datasets, the BHPS and the HOL, that share stylized time use information. We focus on the important issue of childcare time, estimating time devoted to that activity for all mothers in the BHPS sample. In fact matching estimates are utterly superior to OLS estimates with respect to variability of the data and proportion of zero outcomes.

The methods proposed here can be expanded to calibrate diary time devoted to different activities such as leisure, personal care, or education. It also opens new avenues for imputing time use information for other longitudinal surveys, such as the PSID and the SOEP, using the auxiliary time use data sets for these countries.

Appendix

Table A.1. Matching quality

	Unmatched sample					Matched sample											
Variable	Recipient	Donor	%bias	t-test		t-test.4th block		t-test 5th block		t-test 6th block		t-test 7th block		t-test 8th block		t-test 9th block	
Personal and demographic variables																	
age	35,43	38,34	-41	-7,58	***	-4.70	***	5.27	***	-0.06		4.32	***	2.83	***	0.29	
agesq	1309,40	1516,10	-39,4	-7,54	***	-4.54	***	4.52	***	0.01		4.40	***	3.65	***	0.17	
married	0,79	0,85	-14,9	-2,70	***	4.63	***	0.63		-3.57	***	2.31	**	-0.26		1.1	
Education levels																	
degree_further	0,37	0,26	22,5	4,13	***	1.00		1.19		-3.62	***	2.51	**	1.11		-4.3	***
Alev	0,12	0,19	-18,9	-3,93	***	-2.49	**	-0.13		3.04	***	1.20		2.09	**	0.38	
Olev	0,26	0,35	-21,3	-4,25	***	6.62	***	-3.39	***	1.5		1.73		-8.69	***	4.49	***
Household characteristics																	
nch02	0,24	0,19	11,5	2,08	**	3.01	***	-4.3	***	-2.64	***	-6.37	***	-0.68		-7.61	***
nch34	0,22	0,22	1	0,20		3.89	***	-3.57	***	0.13		2.14	**	-10.72	***	6.93	***
nch511	0,81	0,78	3,7	0,68		-2.14	**	3.08	***	4.04	***	-2.87	***	3.28	***	4.37	***
age_young	6,04	7,87	-39,5	-7,47	**	-	***	-		-0.4		-0.15		-1.05		0.74	

est					*	5.50		0.37									
Time-use and childcare behaviour																	
howlng	19,31	17,29	16,8	3,11	**	3.41	***	0.22		-0.99		-1.09		-3.35	***	14.52	***
howlngsq	526,83	433,13	12,3	2,20	**	1.87	*	0.58		-1.18		-2.55	**	-4.21	***	13.47	***
tothrs	18,28	19,87	-9,1	-1,75	*	-0.87		0.76		0.6		0.15		0.40		1.07	
joint_childcare	0,22	0,33	-26,4	-5,36	***	1.50		-2.24	**	0.61		1.55		7.23	***	7.19	***
tot_joint	5.27	9.14	-26.7	-5.69	***	4.50	***	-1.69	*	-0.8		0.49		7.28	***	-0.22	
Design variables																	
wave9	0,33	0,46	-27,8	-5,45	***	-1.74	*	-0.6		-0.46		-4.34	***	-3.98	***	-9.35	***
wave10	0,34	0,26	18,7	3,45	***	1.76	*	3.67	***	3.68	***	0.46		-4.91	***	-3.4	***
wave11	0,33	0,28	10,7	2,00	**	0.83		-2.14	**	-2.57	*	4.40	***	9.49	***	12.72	***
cd8use	0,64	0,76	-27,4	-4,97	***	1.04		1.56		-6.33	***	-2.67	***	3.78	***	-18.46	***

References

Baydar, N., Greek, A., & Gritz, M. R. (1999). “Young mothers’ time spent at work and time spent caring for children”. *Journal of Family and Economic Issues*, 20, 61–84.

Becker, S.. O. & Ichino, A.(2002). "Estimation of average treatment effects based on propensity scores".*Stata Journal*, 2(4), 358-377.

Bittman, M. (2004) “Parenting and employment What time-use surveys show” in Michael Bittman; Nancy Folbre (Ed) *Family Time: The Social Organization of Care*, 152 – 170.

Bloemen, H. & Pasqua, S. & Stancanelli, E. (2010). "An empirical analysis of the time allocation of Italian couples: are they responsive?,".*Review of Economics of the Household*, 8(3), 345-369.

Bonke, J. (2005). Paid work and unpaid work. Diary information versus questionnaire information. *Social Indicators Research*, 70 , 349–368.

Connelly, R. & Kimmel, J. (2009). "Spousal influences on parents’ non-market time choices".*Review of Economics of the Household*, 7(4), 361-394.

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

Gershuny, J. (2000). Changing times: Work and leisure in postindustrial society. Oxford: Oxford University Press.

Gu, X., and Rosenbaum, P. (1993), "Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms".*Journal of Complutational and Graphical Statistics*, 2, 405-420.

Joesch, Jutta M., & Spiess, K. (2006). “European Mothers’ Time Spent Looking after Children—Differences and Similarities across Nine Countries.” *Electronic International Journal of Time Use Research,* 3(1): 1–27.

Judson, D.H. and Poppoff, C. L. (2004) “Selected General Methods” in Jacob S. Siegel and David Swanson (Eds.) *The methods and materials of demography*. San Diego, CA: Elsevier. pp.667-732

Juster, F. T., Ono, H. and Stafford, F. P. (2003). “An assessment of alternative measures of time use”. *Sociological Methodology,* 33, 19–54.

Kan, M. Y. (2008) “Measuring Housework Participation: the gap between "stylised" questionnaire estimates and diary-based estimates”. *Social Indicators Research,* 86(3), 381-400.

Kan, M. J. & Gershuny, J. (2009) “Calibrating Stylised Time Estimates Using UK Diary Data” *Social Indicators Research*, 93, 239–243.

Kan, M. Y., & Pudney, S. (2008). Measurement error in stylized and diary data on time use. *SociologicalMethodology*. 38, 101–132.

Lechner, Michael, (2002) “Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies,” *Review of Economics and Statistics*, 84, 205–220.

Leuven, E. & Sianesi, B.(2003). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing," Statistical Software ComponentsS432001, Boston College Department of Economics, revised 11 Nov 2010.

Kum, H. & Masterson, T. (2008). "Statistical Matching Using Propensity Scores: Theory and Application to the Levy Institute Measure of Economic Wellbeing," Economics Working Paper Archive wp_535, Levy Economics Institute.

Peichl, A. & Schaefer, T. (2009). "FiFoSiM- An integrated tax benefit microsimulation and CGE model for Germany".*International Journal of Microsimulation*, 2(1) 1-15.

Rosenbaum, P. R. and Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician* 39, 33–38.

Rubin, D. B. and Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association* 95, 573–585.

Schulz, F. & Grunow, D. (forthcoming): “Comparing Diary and Survey Estimates on Time Use.” *European Sociological Review*. First published online April 22, 2011 doi:10.1093/esr/jcr030.

Sutherland, H., Taylor, R. and Gomulka, J. (2002) “Combining household income and expenditure data in Policy Simulations” *Review of Income and Wealth* 48, Number 4, 517-536.

Stuart, E. A. and Rubin, D. B. (2008) “Best Practices in Quasi-Experimental Designs: Matching Methods for Causal Inference” in Jason W. Osborne (Ed.) *Best Practices in Quantitative Methods*, 155-176.

Zhao, Z. (2004). Using matching to estimate treatment effects: Data requirements, matching metrics, and monte carlo evidence. *Review of Economics and Statistics* 86, 1, 91–107.

Zick, C. D., & Bryant, W. K. (1996). “A new look at parents’ time spent in child care: Primary and secondary time use”. *Social Science Research 25,* 1-21