The Impact of Complex Sampling Designs in the Analysis of Longitudinal Survey Data

Marcel de Toledo Vieira
Departamento de Estatística, Universidade Federal de Juiz de Fora, Brazil
and
Chris Skinner
Division of Social Statistics, University of Southampton, UK

Summary. This paper is motivated by the use of data from the British Household Panel Survey (BHPS) to study attitudes to gender roles and their relation to demographic and economic variables. Such household surveys often use a complex sampling design to select the sample to be followed up over time. It is well known that complex sampling schemes may inflate the variances of estimators, especially as a result of clustering. The design effect measures the inflation of the sampling variance of an estimator as a result of the use of a complex sampling scheme. There is some empirical evidence that this impact may be less the more complex the analysis and this may sometimes be used to justify ignoring the complex sampling scheme in analysis. The aim of this paper is to show that design effects for longitudinal analyses can be greater than for corresponding cross-sectional analyses, implying that more caution is required before ignoring the complex design in standard error estimation. A possible theoretical explanation is provided.

Keywords: clustering; design effect; longitudinal analysis; random effects model

1. Introduction

This paper develops methodology for the analysis of complex survey data (Skinner, Holt and Smith, 1989) to address longitudinal aspects of regression analyses of British Household Panel Survey (BHPS) data on attitudes to gender roles and their relation to demographic and economic variables. The general question of interest in this paper is: is the impact of the complex sampling design on variance estimation for analyses of these longitudinal data greater or less than for corresponding cross-sectional analyses? Kish and Frankel (1974) presented empirical work which suggested that the impacts of complex designs on variances are reduced for more complex analytical statistics and so one might conjecture that the impact on longitudinal analyses might also be reduced. We shall provide evidence in the opposite direction that, at least for the specific analyses considered, the impact on longitudinal analyses tends to be greater. Given that an impact does exist, the second question addressed is how to undertake variance estimation. We shall focus in the paper on the clustering impact of the sampling design by adopting survey sampling variance estimation procedures (Skinner et al, 1989).

When asking how an analysis should take account of complex sampling, it is natural first to ask whether the parameters of interest should depend on the design, via the population structure underlying the sampling (Skinner et al., 1989). In this paper we shall assume this is not the case, since the primary sampling units in the BHPS are postcode sectors, determined by the needs of the British postal system and assumed here not to be relevant to the definition of parameters of scientific interest. A second question which might be asked is how the sampling impacts on point estimation, e.g. via the use of sampling weights. We shall refer to this question briefly, but we shall largely suppose that point estimation is unaffected by the design. Our main focus will be on the impact of the design on variance estimation.

The impact on variance estimation will be measured here by the 'misspecification effect', denoted *meff* (Skinner, 1989a), which is the variance of a point estimator divided by the expectation of the variance estimator. This is a measure of relative bias of the variance estimator, analogous to the 'design effect' or *deff* of Kish (1965), which measures the impact of a design on a variance, defined as the variance of the point estimator under the given design divided by its variance under simple random sampling with the same sample size. In the application in this paper, estimated *meffs* may be treated as equivalent to estimated *deffs* when the variance estimator ignores the complex design.

One reason for studying *meffs* for variance estimators which ignore the design is that analysts of longitudinal survey data face many difficult methodological challenges and they may be tempted to view the impact of complex sampling on standard errors as a relatively minor issue which, if ignored, is unlikely to lead to misleading inferences. Indeed, in cases where the survey documentation indicates that the design effect of the mean of the analyst's outcome variable of interest is not much larger than one, the analyst might justify ignoring the design when estimating standard errors by appealing to the observation of Kish and Frankel (1974, p.13) that "design effects for complex statistics tend to be less than those for means of the same variables".

The paper is motivated by a regression analysis of five waves of BHPS data, based upon work of Berrington (2002) and described in Section 2. After a description of models and estimation methods in Section 3, the paper proceeds in Section 4 to provide evidence that *meffs* for longitudinal analyses can be greater than for corresponding cross-sectional analyses, implying that more caution is required before ignoring the complex design in standard error estimation.

We ignore the effects of stratification and weighting in the empirical work in section 4 in order to isolate the source of potential misspecification effects and to avoid introducing the more complex weighting issues arising with multilevel models (Pfeffermann et al., 1998). We make brief remarks on these effects in the concluding discussion in Section 5.

2 The motivating application to BHPS data

Recent decades have witnessed major changes in the roles of men and women in the family in many countries. Social scientists are interested in the relation between changing attitudes to gender roles and changes in behaviour, such as parenthood and labour force participation (e.g. Morgan and Waite, 1987; Fan and Marini, 2000). A variety of forms of statistical analysis are used to provide evidence about these relationships. In this paper we consider a longitudinal regression analysis, based upon a model considered by Berrington (2002), with a measure of gender role attitude as the dependent variable. We also consider some simpler versions of this analysis to facilitate understanding of the methodological issues outlined in Section 1. The models will be set out formally in Section 3.

The data come from waves 1, 3, 5, 7 and 9 (collected in 1991, 1993, 1995, 1997, and 1999 respectively) of the BHPS, when respondents were asked whether they 'strongly agreed', 'agreed', 'neither agreed nor disagreed', 'disagreed' or 'strongly disagreed' with a series of statements concerning the family, women's roles, and work out of the household. Responses were scored from 1 to 5. Factor analysis was used to assess which statements could be combined into a gender role attitude measure. The attitude score considered here is the total score for six selected statements. Higher scores signify more egalitarian gender role attitudes. Berrington (2002) provides further discussion of this variable.

Covariates for the regression analysis were selected on the basis of discussion in Berrington (2002) but reduced in number to facilitate a focus on the methodological issues of interest. The covariate of primary scientific interest is economic activity, which distinguishes in particular between women who are at home looking after children (denoted 'family care') and women following other forms of activity in relation to the labour market. Variables reflecting age and education are also included since these have often been found to be strongly related to gender role attitudes (e.g. Fan and Marini, 2000). All these covariates may change values between waves. A year variable is also included. This may reflect both historical change and the general ageing of the women in the sample.

The BHPS is a household panel survey of individuals in private domiciles in Great Britain (Taylor *et al.*, 2001). Given the interest in whether women's primary labour market activity is 'caring for a family', we define our study population as women aged 16-39 in 1991. This results in a subset of data on n = 1340 women. This subset consists of those women in the eligible age range for whom full interview outcomes (complete records) were obtained in all the five waves. We comment further on the treatment of nonresponse in section 3.

The initial (wave one) sample of the BHPS in 1991 was selected by a stratified multistage design in which households had approximately equal probabilities of inclusion. As primary sampling units (PSUs), 250 postcode sectors were selected, with replacement and with probability of selection proportional to size using a systematic procedure. Addresses were selected as secondary sampling units, with the adoption of an analogous systematic procedure. In addresses with up to 3 households present, all households were included, and in those with more than 3 households, a random selection procedure, using a Kish grid, was used for the selection of 3 households. Then, all resident household members aged 16 or over were selected. All adults selected at wave one, were followed from wave two and beyond. A consequence of this design is that inclusion probabilities of adults vary little. The impact of weighting is considered briefly in section 5. The 1340 women represented in the data are spread fairly evenly across the 250 postcode sectors. The small average sample size of around five per postcode sector combined with the relatively low intra-postcode sector correlation for the attitude variable of interest leads to relatively small impacts of the design, as measured by *meffs*. Since our aims are methodological ones, to compare *meffs* for different analyses, we have chosen to group the postcode sectors into 47 geographically contiguous clusters, to create sharper comparisons, less blurred by sampling errors which can be appreciable in variance estimation. The *meffs* in the tables we present therefore tend to be greater than they are for the actual design. The latter results tend to follow similar patterns, although the patterns are less clear-cut as a result of sampling error.

3. Regression model and inference procedures

Let y_{it} denote the value of the attitude score for woman i at wave t (coded t = 1, … , T = 5 to correspond to 1991, 1993, …,1999) and let $y_i = (y_{i1},..., y_{iT})'$ be the vector of repeated measures. We consider linear models of the following form to represent the expectation of y_i given the values of covariates:

$$E(y_i) = x_i\beta \,, \qquad (1)$$

where $x_i = (x_{i1}',..., x_{iT}')'$, x_{it} is a $1 \times q$ vector of specified values of covariates for woman i at wave t, β is the $q \times 1$ vector of regression coefficients and the expectation is with respect to a superpopulation model (Goldstein, 2003, p. 164). A more sophisticated analysis might include a measurement error model for attitudes (e.g. Fan and Marini, 2000), with each of the five-point responses to the six statements treated as ordinal variables. Here, we adopt a simpler approach, treating the aggregate score y_{it} and the associated coefficient vector β as scientifically interesting, with the measurement error included in the error term of the model.

We consider estimation of β based on data from the 'longitudinal sample', s_T, i.e. the sample for which observations are available for each of $t = 1,...,T$. We did not attempt to use data observed only at a subset of the five waves, partly for simplicity but also because our primary interest is clustering and we did not wish differences in clustering effects over time to be confounded with differences in incomplete data effects. A concern with the use of the longitudinal sample s_T is that the underlying attrition process may lead to biased estimation of β. One possible way of attempting to correct for this potential biasing effect is via the use of longitudinal survey weights, $w_{iT}, i \in s$ (Lepkowski, 1986).

The most general estimator of β we consider is

$$\hat{\beta} = \left(\sum_{i \in s_T} w_{iT} x_i' V^{-1} x_i \right)^{-1} \sum_{i \in s_T} w_{iT} x_i' V^{-1} y_i \,, \qquad (2)$$

where V is a 'working' variance matrix of y_i (Diggle et al. 2002, p.70), taken as the exchangeable variance matrix with diagonal elements σ^2 and off-diagonal elements $\hat{\rho}\sigma^2$, and $\hat{\rho}$ is an estimator of the intra-individual correlation, obtained by iterating between generalised least squares estimation of β and survey-weighted moment-based estimation of the intra-individual correlation (Liang and Zeger, 1986; Shah et al., 1997). Note that σ^2 cancels out in (2) and hence does not need to be estimated for $\hat{\beta}$.

This variance matrix, V, would arise if y_{it} obeyed the multilevel (mixed linear) model:

$$y_{it} = x_{it}\beta + u_i + v_{it} \,, \qquad (3)$$

with independent random effects u_i and v_{it} with variances $\sigma_u^2 = \rho\sigma^2$ and $\sigma_v^2 = (1-\rho)\sigma^2$ respectively. We find that this model provides a first approximation to the variance structure for the regression models fitted in section 4. For illustration, we find $\hat{\rho} = 0.59$ in the most elaborate regression model implying a fairly substantial between-woman component in the attitude scores unexplained by the chosen covariates. It is not necessary, however, for the error structure to follow the specific model in (3) exactly for $\hat{\beta}$ to be consistent.

To estimate the covariance matrix of $\hat{\beta}$ allowing for the complex sampling design, we may use the linearization estimator (Skinner, 1989b, p.78):

$$v(\hat{\beta}) = \left[\sum_{i \in s_T} w_{iT} x_i' V^{-1} x_i \right]^{-1} \left[\sum_h n_h/(n_h - 1) \sum_a (z_{ha} - \bar{z}_h)^2 \right] \left[\sum_{i \in s_T} w_{iT} x_i' V^{-1} x_i \right]^{-1}, \qquad (4)$$

where h denotes stratum, a denotes area (primary sampling unit, PSU), n_h is the number of PSUs in stratum h, $z_{ha} = \sum_i w_{iT} x_i' V^{-1} e_i$, $\bar{z}_h = \sum_a z_{ha} / n_h$ and $e_i = y_i - x_i\hat{\beta}$. Note that this variance estimator requires use of the stratum and primary sampling unit identifiers. See Lavange et al. (1996) and Lavange et al. (2001) for applications of a similar approach to allowing for complex sampling designs in regression analyses of repeated measures data from different longitudinal studies.

In order to assess the impact of the complex design on variance estimation, we also consider a linearization variance estimator which ignores the complex design, denoted $v_0(\hat{\beta})$, given by expression (4) where the PSUs become the same as women so that z_{ha} is replaced by $w_{iT} x_i' V^{-1} e_i$ and there is only a single stratum so that $n_h = n$ is the overall sample size and the term $\bar{z}_h$ disappears. Ignoring the weights and the term $n/(n-1)$, this is the 'robust' variance estimator presented by Liang and Zeger (1986) as consistent when (1) holds, but where the working variance matrix, V, may not reflect the true variance structure. See also Diggle et al. (2002, section 4.6).

Following Skinner (1989a, p.24), we refer to $v(\hat{\beta}_k) / v_0(\hat{\beta}_k)$, the ratio of these two variance estimators for the k^{th} element of $\hat{\beta}$, as an estimated misspecification effect and denote it *meff*. This ratio may be viewed as an estimator of the misspecification effect, defined as $\mathrm{var}(\hat{\beta}_k) / E[v_0(\hat{\beta}_k)]$, on the assumption that $v(\hat{\beta})$ is a consistent estimator of $\mathrm{var}(\hat{\beta})$. This quantity is a measure of the relative bias of the 'incorrectly specified' variance estimator $v_0(\hat{\beta}_k)$ as an estimator of $\mathrm{var}(\hat{\beta}_k)$. This concept is closely related to that of the design effect of Kish (1965) which is more relevant to the choice of design than to the choice of standard error estimator.

In general, *meffs* will reflect the impact of weighting, clustering and stratification. In order to disentangle these effects, we shall first in section 4 only consider the impact of clustering. We thus treat the weights as constant and ignore stratification.

4. Misspecification effects: the impact of ignoring clustering in longitudinal analyses

In this section we explore the impact of ignoring clustering in standard error estimation for various longitudinal analyses. To provide theoretical motivation for the kind of impact we may expect, consider converting the two-level model in (3) into a simple three-level model (Goldstein, 2003) as:

$$y_{ait} = x_{ait}\beta + \eta_a + u_{ai} + v_{ait} \,, \qquad (5)$$

where an additional subscript a has been added to denote area (cluster) and an additional random term η_a with variance σ_η^2 represents the area effect (assumed independent of u_{ai} and v_{ait}). We now let σ_u^2 and σ_v^2 denote the variances of u_{ai} and v_{ait} respectively. Let us use this model to consider first the expected nature of misspecification effects in the case of cross-sectional analyses, where t is kept fixed as $t=1$. In this case, if we suppose for simplicity that $x_{ait} \equiv 1$ and β is the mean of y_{ait} in (5) and that there is a common sample size m per cluster, the misspecification effect is approximately equal to $1+(m-1)\tau$, where $\tau=\sigma_\eta^2/(\sigma_\eta^2+\sigma_u^2+\sigma_v^2)$ is the intracluster correlation (Skinner, 1989b, p. 38). If the sample sizes per cluster are unequal a common approximation is to replace m in this formula by $\bar{m}$, the average sample size per cluster.

Turning to the longitudinal case, where again $x_{ait}\equiv 1$ and now β is a longitudinal mean of y_{ait} for $t=1,\ldots,T$, the same theory for misspecification effects will apply, but where τ is now the intracluster correlation for η_a and $u_{ai}+v_{ait}$ averaged over the waves., i.e. $\tau=\sigma_\eta^2/(\sigma_\eta^2+\sigma_u^2+\sigma_v^2/T)$. Hence, under this model, the misspecification effect increases as T increases, if $\sigma_v^2>0$.

Let us now compare this expected theoretical pattern with the empirical findings. Using data from just the first wave and setting $x_{ait}\equiv 1$, the *meff* for this cross-sectional mean is given in Table 1 as about 1.5. This value is plausible since the average sample size per cluster is $\bar{m}=1340/47\approx 29$ and using the $1+(\bar{m}-1)\tau$ formula, the implied value of τ is about 0.02 and such a small value is in line with other estimated values of τ found for attitudinal variables in British surveys (Lynn and Lievesley, 1991, App. D).

To assess the impact of the longitudinal aspect of the data, we re-estimate the *meff* using data for waves *1,…,t* for *t*=2, 3, … 5. Table 1 suggests a tendency for the *meff* to increase with the number of waves, as anticipated from the theoretical reasoning. These *meffs* are certainly subject to sampling error and there appears to be some genuine variation in misspecification effects for cross-sectional estimates at different waves but this variation does not appear to be sufficient to explain this trend.

To pursue the theoretical rationale for this finding further, note that model (5) is likely to be an oversimplification because the area effects are likely to display some variation over time, in which case we write η_{at} rather than η_a. In this case, τ becomes $\tau=\mathrm{var}(\bar{\eta}_a)/[\mathrm{var}(\bar{\eta}_a)+\mathrm{var}(\bar{u}_a+\bar{v}_a)]$, where $\bar{\eta}_a=\sum_t\eta_{at}/T$ and $\bar{u}_a+\bar{v}_a=\sum_t(u_{ai}+v_{ait})/T$. Now, it seems plausible that the average level of egalitarian attitudes in an area will vary less from year to year than the attitude scores of individual women, since the latter will be affected both by measurement error and genuine changes in attitudes, so that $\mathrm{var}(\bar{\eta}_a)$ may be expected to decline more slowly with T than $\mathrm{var}(\bar{u}_a+\bar{v}_a)$. We may therefore expect τ, and consequently the *meff,* to increase as T increases, as we observe in Table 1.

We next elaborate the analysis by including indicator variables for economic activity as covariates. The resulting regression model has an intercept term and four covariates representing contrasts between women who are employed full-time and women in other categories of economic activity. The *meffs* are presented in Table 2. The intercept term is a domain mean and standard theory for a *meff* of a mean in a domain cutting across clusters (Skinner, 1989b, p.60) suggests that it will be somewhat less than the *meff* for the mean in the whole sample, as indeed is observed with the *meff* for the cross-section domain mean of 1.13 in Table 2 being less than the value 1.51 in Table 1. As before, there is some evidence in Table 2 of tendency for the *meff* to increase, from 1.13 with one wave to 1.50 with five waves, albeit with lower values of the *meffs* than in Table 1. The *meffs* for the contrasts in Table 2 vary in size, some greater than and some less than one. These *meffs* may be viewed as a combination of the traditional variance inflating effect of clustering in surveys together with the familiar variance reducing effect of blocking in an experiment. The main feature of these results of interest here is that there is again no tendency for the *meffs* to converge to one as the number of waves increases. If there is a trend, it is in the opposite direction. For the contrast of particular scientific interest, that between women who are full-time employed and those who are ‘at home caring for a family’, the *meff* is consistently well below one.

We next elaborate the model further by including, as additional covariates, age group, year and qualifications. The results for *meffs* are given in Table 3. The *meffs* for the economic activity covariates again vary, some being above one and some below one. There is again some evidence of a tendency for these *meffs* to diverge away from one as the number of waves increases. A comparison of Tables 1 and 3 confirms the observation of Kish and Frankel (1974) that *meffs* for regression coefficients tend not to be greater than *meffs* for the means of the dependent variable.

5. Discussion

We have presented some theoretical arguments and empirical evidence that the impact of ignoring clustering in standard error estimation for certain longitudinal analyses can tend to be larger than for corresponding cross-sectional analyses. The implication is that it is, in general, at least as important to allow for clustering in standard error estimation for longitudinal analyses as for cross-sectional analyses. Thus, the expectation from the finding of Kish and Frankel (1974) that complex sampling has less of an impact on variances for more complex analytical statistics was not borne out in this case.

The longitudinal analyses considered in this paper are of a certain kind and we should emphasise that the patterns observed for *meffs* in these kinds of analyses may well not extend to other kinds of longitudinal analyses. To speculate about the class of models and estimators for which the patterns observed in this paper might apply, we conjecture that increased *meffs* for longitudinal analyses will arise when the longitudinal design enables temporal ‘random’ variation in individual responses to be extracted from betweenperson differences and hence to reduce the component of standard errors due to these differences, but provides less 'explanation' of between cluster differences, so that the relative importance of this component of standard errors becomes greater.

The empirical work presented in this paper has also been restricted to the impact of clustering. We have undertaken corresponding work allowing for weighting and stratification and found broadly similar findings. Stratification tends to have a smaller effect than clustering. The sample selection probabilities in the BHPS do not vary greatly and the impact of weighting by the reciprocals of these probabilities on both point and variance estimates tends not to be large. There is rather greater variation among the longitudinal weights, w_{iT}, which are provided with BHPS data for analyses of sets of individuals who have responded at each wave up to and including a given year, *T*. The impact of these weights on point and variance estimates is somewhat greater. As *T* increases and further attrition occurs, the weights, w_{iT}, tend to become more variable and lead to greater inflation of variances. This tends to compound the effect we have described of *meffs* increasing with *T*.

Acknowledgments

The research of the first author was supported by grant 20.0286/01.3 from the Brazilian National Council for Scientific and Technological Development (CNPq).

References

Berrington, A. (2002) Exploring Relationships Between Entry Into parenthood and Gender Role Attitudes: Evidence from the British Household Panel Study. In Lesthaeghe, R. ed. *Meaning and Choice: Value Orientations and Life Course Decisions*. Brussels, NIDI.

Diggle, P. J., Heagerty, P., Liang, K. & Zeger, S. L. (2002) *Analysis of Longitudinal Data*. 2nd ed. Oxford: Oxford University Press.

Fan, P.-L. and Marini, M.M. (2000) Influences on gender-role attitudes during the transition to adulthood. *Social Science Research*, **29**, 258-283.

Goldstein, H. (2003) *Multilevel Statistical Models*, 3rd. Ed. London: Arnold.

Kish, L. (1965) *Survey Sampling*. New York: Wiley.

Kish, L. and Frankel, M. R. (1974) Inference from complex samples. *J. R. Statist. Soc.* B, **36**, 1-37.

Lavange, L.M., Koch, G.G. and Schwartz, T.A. (2001) Applying sample survey methods to clinical trial data. *Statistics in Medicine*, **29**, 2609-23.

Lavange, L.M., Stearns, S.C., Lafata, J.E., Koch, G.G. and Shah, B.V. (1996) Innovative strategies using SUDAAN for analysis of health surveys with complex samples. *Statistical Methods in Medical Research*, **5**, 311-329.

Lepkowski, J.M. (1986) Treatment of wave nonresponse in panel surveys. In Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P. eds. *Panel Surveys*. New York: Wiley.

Liang, K.Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

Lynn, P. and Lievesley, D. (1991) *Drawing General Population Samples in Great Britain*. London: Social and Community Planning Research.

Morgan, S.P. and Waite, L.J. (1987) Parenthood and the attitudes of young adults. *Am. Sociological Review,* **52**, 541-547.

Pfeffermann, D., Skinner, C.J., Holmes, D.J., Goldstein, H. and Rasbash, J. (1998) Weighting for unequal selection probabilities in multilevel models. *J. R. Statist. Soc.*B, 60, , 23-56.

Shah, B. V., Barnwell, B.G. and Bieler, G.S. (1997) *SUDAAN User's Manual, Release 7.5*. Research Triangle Park, NC: Research Triangle Institute.

Skinner, C. J. (1989a) Introduction to Part A. In Skinner, C. J., Holt, D. and Smith, T. M. F. eds. *Analysis of Complex Surveys*. Chichester: Wiley, pp.23-58.

Skinner, C. J. (1989b) Domain means, regression and multivariate analysis. In Skinner, C. J., Holt, D. and Smith, T. M. F. eds. *Analysis of Complex Surveys*. Chichester: Wiley, pp. 59-87.

Skinner, C. J., Holt, D. and Smith, T. M. F. eds. (1989) *Analysis of Complex Surveys*. Chichester: Wiley.

Taylor, M. F. ed, with Brice J., Buck, N. and Prentice-Lane E. (2001) *British Household Panel Survey - User Manual - Volume A: Introduction, Technical Report and Appendices*. Colchester, University of Essex.

Table 1 – Estimates of Longitudinal Means

	point estimate	s.e.	*meffs*				
Waves	1-9	1-9	1	1,3	1,3,5	1-7	1-9
	19.83	0.12	1.51	1.50	1.68	1.81	1.84

Table 2 – Estimates by Economic Activity

	$\hat{\beta}$	s.e.	*meffs*				
Waves	1-9	1-9	1	1,3	1,3,5	1-7	1-9
Intercept	20.58	0.11	1.13	1.01	1.09	1.38	1.50
Contrasts for							
PT employed	-1.03	0.10	0.93	0.91	0.93	1.00	0.89
Other inactive	-0.80	0.15	0.60	0.96	0.68	0.76	0.81
FT student	0.41	0.24	1.10	1.32	1.14	1.48	1.44
Family care	-2.18	0.10	0.72	0.49	0.58	0.66	0.60

Note: intercept is mean for women full-time employed
contrasts are for other categories of economic activity relative to full-time employed

Table 3 – Estimates of Regression Coefficients

	$\hat{\beta}$	s.e.	*meffs*				
Waves	1-9	1-9	1	1,3	1,3,5	1-7	1-9
Intercept	20.20	0.30	0.95	0.87	0.87	1.04	1.07
Year, t	-0.04	0.01	-	0.86	0.69	0.59	0.96
Age Group							
16-21	0.00	-					
22-27	-0.71	0.25	1.22	1.37	1.44	1.73	1.64
28-33	-0.89	0.27	1.38	1.40	1.46	1.68	1.59
34+	-1.03	0.27	0.94	1.10	1.13	1.26	1.34
Econ. Activity							
FT employed	0.00	-					
PT employed	-0.93	0.10	0.97	0.95	0.96	1.06	0.91
Other inactive	-0.75	0.15	0.60	0.96	0.68	0.77	0.81
FT student	0.17	0.24	0.93	1.32	1.23	1.39	1.32
Family care	-2.09	0.10	0.77	0.59	0.70	0.78	0.67
Qualification							
Degree	0.00	-					
QF	-0.52	0.21	0.77	0.64	0.75	0.87	0.85
A-level	-0.61	0.24	0.98	0.87	0.94	0.94	1.01
O-level	-0.44	0.20	0.62	0.62	0.59	0.69	0.73
Other	-1.16	0.22	0.83	0.83	0.78	0.80	0.82