DE GRUYTER
OPEN

Journal of Official Statistics, Vol. 32, No. 2, 2016, pp. 487–505, http://dx.doi.org/10.1515/JOS-2016-0025

Misspecification Effects in the Analysis of Panel Data

Marcel de Toledo Vieira[1], *Peter W.F. Smith*[2], *and Maria de Fátima Salgueiro*[3]

Misspecification effects (*meffs*) measure the effect on the sampling variance of an estimator of incorrect specification of both the sampling scheme and the model considered. We assess the effect of various features of complex sampling schemes on the inferences drawn from models for panel data using *meffs*. Many longitudinal social survey designs employ multistage sampling, leading to some clustering, which tends to lead to *meffs* greater than unity. An empirical study using data from the British Household Panel Survey is conducted, and a simulation study is performed. Our results suggest that clustering impacts are stronger for longitudinal studies than for cross-sectional studies, and that *meffs* for the regression coefficients increase with the number of waves analysed. Hence, estimated standard errors in the analysis of panel data can be misleading if any clustering is ignored.

Key words: Longitudinal survey; sampling variance; multistage sampling; stratification; weighting.

1. Introduction

Interest in fitting models to longitudinal complex survey data has grown in the last decade. Longitudinal surveys often make use of complex sampling procedures, such as unequal selection probabilities, stratification and multistage sampling, to select the initial panel sample at the first wave in order to best use the available resources (e.g., Smith et al. 2009). Nevertheless, to our knowledge, insufficient attention is still paid to the impacts of sampling complexities on the regression analysis of panel data in the survey-sampling literature.

Researchers and other users of panel data often make use of standard statistical techniques, which in most of the cases do not take account of the complex sample designs. These techniques may assume that the data are (after conditioning on some covariates) realizations of independent and identically distributed random vectors, which is rare in practice. The standard formulation of inference methods is often not valid when analysing data collected using a complex sampling scheme. According to Chambers and Skinner

[1] Departamento de Estatística e Programa de Pós-Graduação em Economia, Universidade Federal de Juiz de Fora (UFJF), Rua José Lourenço Kelmer, s/n, Campus Universitário, Bairro São Pedro, 36036-900, Juiz de Fora, MG, Brazil. Email: marcel.vieira@ice.ufjf.br
[2] Southampton Statistical Sciences Research Institute (S3RI), University of Southampton, Southampton, SO17 1BJ, United Kingdom. Email: P.W.Smith@soton.ac.uk
[3] Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit, Av. Forças Armadas, 1649-026, Lisbon, Portugal. Email: fatima.salgueiro@iscte.pt
Acknowledgments: This work was supported by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais [grant numbers CEX-APQ-01799-10, CEX-APQ-02071-12 to the first author]; and the Fundação para a Ciência e a Tecnologia [grant number PTDC/GES/72784/2006 to the third author]. Visits to Southampton were partially supported by the Southampton Statistical Sciences Research Institute for the first and third authors.

(2003), even when the sampling design is considered ignorable, the standard inferential procedures may not satisfactorily reproduce the population complexities underlying the sampling mechanism. For a discussion on design-based and model-based methods for estimating model parameters under both ignorable and nonignorable sampling designs, see also Binder and Roberts (2003).

Moreover, complex sampling schemes may induce a correlation structure among observations, as elements in the same cluster are likely to be more similar than elements in different clusters. Therefore, when a sample is selected by complex sampling at Wave 1, a correlation structure among the observations, additional to the longitudinal correlation, may be induced. Under this situation, the use of standard statistical techniques with complex sampling data may lead to seriously biased point and standard-error estimates (see e.g., Nathan and Holt 1980). Ignoring clustering and weighting effects, for example, tends to lead to the underestimation of standard errors, and therefore to narrowed confidence intervals and to the incorrect rejection of null hypotheses. Stratification normally affects the analysis in an opposite direction. Thus ignoring clustering, weighting and stratification effects may lead to inappropriate statistical inference.

There is a well substantiated literature on methods for taking account of complex sampling schemes in the analysis of survey data. Skinner et al. (1989), Chambers and Skinner (2003), and Pfeffermann (2011), for example, provide further information and references. For cross-sectional data, Kish and Frankel (1974), Holt and Scott (1981), Scott and Holt (1982), Skinner (1986, 1989a, b), and Feder (2011), for example, have considered the effects of complex sampling on regression model parameters estimation.

Furthermore, Feder et al. (2000) proposed combining multilevel modelling, time-series modelling and survey-sampling methods for panel data analysis; Sutradhar and Kovacevic (2000) developed a generalised estimating equations approach by considering an autocorrelation structure in multivariate polytomous panel data models. In addition, Skinner and Holmes (2003) studied two approaches for dealing with sampling effects, either by taking the repeated observations as multivariate outcomes and utilising weighted estimators that account for the correlation structure, or by considering a two-level longitudinal model.

Skinner and Vieira (2007) presented some empirical and theoretical evidence that the variance-inflating impacts of clustering may be higher for longitudinal analyses than for corresponding cross-sectional analyses and that those effects may increase with the number of waves considered in some types of analysis. Moreover, Vieira and Skinner (2008) considered parametric models for panel data and have proposed methods of estimating model parameters that allow for complex schemes by incorporating survey weights into alternative point estimation procedures and using linearisation methods for variance estimation (see Vieira 2009 for further references).

Large-scale longitudinal studies usually involve the selection of a probability sample from a population at the time the panel starts. Weighting in the panel data context has three main aims. If we consider, for example, a survey with two waves, then the longitudinal weight at Wave 2 would: (i) account for unequal selection probabilities at Wave 1, (ii) adjust for unit nonresponse which may occur at Waves 1 and 2, and (iii) adjust (via poststratification, raking or calibration) so that weighted sample estimates for certain auxiliary variables match their respective known population parameters. Longitudinal weights, therefore, allow for different selection probabilities and nonresponse at Wave 1 and attrition, and are adjusted, at each wave, to take account of previous wave respondents' absence through refusal at the current wave or through some other way of sample attrition. Longitudinal weights are calculated in order to guarantee the property that weighted sample moments are consistent for population moments with respect to the joint sampling/nonresponse probability distribution.

The current article further examines the impacts of clustering in panel data analysis, previously investigated by Skinner and Vieira (2007). Moreover, the impacts of survey weighting and stratification are studied by comparing these with the impact on corresponding cross-sectional analyses and by examining how these effects behave with increases in the number of survey waves considered in the analysis. Misspecification effects (*meffs*) for parameter estimates in regression models for (i) the logarithm of household income and (ii) a material satisfaction score are used to evaluate the impact of various features of complex designs on inference. The data are taken from Waves 12 to 15 of the British Household Panel Study (BHPS). To validate the conclusions from an empirical study, a simulation study is also performed, where the use of the *meffs* as a measure of incorrect specification of the model considered is also extensively explored in the longitudinal data analysis context.

The contribution of the current article, when compared to Skinner and Vieira (2007), is (i) the investigation of the impacts of survey weighting and stratification, (ii) the consideration of alternative *meff* measures, (iii) the undertaking of a detailed simulation study, and (iv) the use of the *meffs* as a measure of the impact of incorrect specification of longitudinal models.

This article is organised in six sections. In Section 2 we introduce the panel data under analysis. Section 3 introduces the models, point and variance estimation procedures, and describes the various *meffs*. In Section 4 we present our motivating application and empirical results obtained from real panel data. In Section 5, the simulation study conducted is described and its results are presented. The concluding discussion is presented in Section 6.

2. Data and Sampling Design

The empirical evidence presented in this article is based upon data from the BHPS, which was a large nationally representative household panel survey of individuals in private domiciles in Great Britain (see Taylor et al. 2010). This survey had the main objective of providing information about social and economic change at the individual and household levels.

The BHPS is a longitudinal survey and adopts a complex multistage sampling scheme for collecting data. In addition, it has a multiple-cohort prospective panel design. At Wave 1, in 1991, the survey design involved (i) a multistage stratified clustered probability design with systematic sampling and (ii) approximately equal probability selection of households. As primary sampling units (PSUs or clusters), 250 postcode sectors were selected, with replacement, and with probability of selection proportional to size, using a systematic sampling procedure. The final strata are the result of several stratification stages, which may be summarised as follows:

(a) First, the population was divided into 18 implicit regional strata (regions).
(b) Within each region, PSUs were ranked and then split into major strata of approximately equal size based on the proportion of heads of households in professional or managerial positions.
(c) Within major strata, PSUs were reranked by the proportion of their population in pensionable age.
(d) Major strata were then split into two minor strata: a nonmetropolitan area, with PSUs sorted by their proportion of employed population in agriculture; and a metropolitan area, with PSUs sorted by their population both under pensionable age and living in single-person households. For further details on the BHPS sampling design, see Taylor et al. (2010).

Our analyses are based upon a subset of 2,255 men and women aged 16 or more, clustered in 234 PSUs, who were original sample members, who gave a full interview in Waves 12 to 15 (collected from 2002 until 2005), and who were employed throughout the period. This results in a balanced panel. Note that we study the same subsample considered by Salgueiro et al. (2013), which does not include the BHPS extension samples selected from Scotland, Wales, and Northern Ireland. Therefore, $T = 4$, where T is the number of waves considered. BHPS respondents were asked to answer several questions related to sociodemographic, economic, and attitudinal characteristics. The following variables are considered in our analysis: gender, age category, number of children in the household, education level, social class, marital status, health status, hours normally worked per week, and the logarithm of the household income.

The BHPS data set includes longitudinal weights w_i, which are provided for individual cases that have responded at each wave up to and including the latest wave (Wave 15 in our analysis). The longitudinal weight at any wave generally accounts for losses between each immediate pair of waves up to that point and for the initial sampling design. For information regarding how the weights are defined for the BHPS, see Taylor et al. (2010), where further details about the sampling design of the BHPS are also given.

We have also included a material satisfaction score variable in our data set. Factor analysis, undertaken by Salgueiro et al. (2013), was used to assess which BHPS measures of subjective wellbeing could be combined into a measure of satisfaction with material dimensions of life. A material satisfaction score has subsequently been calculated for each respondent as the total sum of the responses to the following three satisfaction variables: (i) satisfaction with household income, (ii) satisfaction with house/flat, and (iii) satisfaction with job. These three variables were originally measured on a scale from 1 (not satisfied) to 7 (completely satisfied).

In our sample, the relative frequency for males and females is approximately 50%. The distribution of the age category variable is negatively skewed, as the frequencies for the older categories are larger. Most of the respondents were either married or living as a couple in 2002. Approximately 80% of the respondents considered themselves in either a good or excellent health condition. Furthermore, over 75% of the individuals worked at least 30 hours per week. About 55% of the individuals had a high level of education, and only 16% of them occupied a partly skilled or an unskilled position in their last job. Almost 62% of the respondents had no children in the household where they live.

Moreover, in 2002, the average household income of the sample members was approximately 3,365 British pounds in the month before the interview was made.

3. Models, Estimation Procedures, and Misspecification Effects

Regression models have found a wide range of useful applications with panel data (e.g., Diggle et al. 2002; Fitzmaurice et al. 2004). Such data consist of repeated observations on the same variables for the same individuals across equally spaced waves of data collection. The models considered here are concerned with representing the relationship between one of the variables, treated as dependent, and several other variables, treated as covariates. We shall adopt i to denote an individual and t to denote time. We denote the survey variable of interest as y_{it} for individual i at time t. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ be the vector of repeated measures. For the population, we consider linear models of the following form to represent the expectation of $\mathbf{y}_i$ given the values of covariates:

$$E(\mathbf{y}_i) = \mathbf{x}_i \boldsymbol{\beta}, \tag{1}$$

where $\mathbf{x}_i = (\mathrm{x}_{i1}, \ldots, \mathrm{x}_{iT})'$ is a $T \times q$ matrix, $\mathbf{x}_{it}$ is a vector of specified values of q covariates for individual i at time t, $\boldsymbol{\beta}$ is a $q \times 1$ vector of regression coefficients, and the expectation is with respect to the model.

The estimation of $\boldsymbol{\beta}$ is based on data from the 'longitudinal sample', s, (i.e., the sample for which observations are available for each $t = 1, \ldots, T$). Following the pseudolikelihood approach (Skinner 1989b), the most general estimator of $\boldsymbol{\beta}$ considered in this article is (Skinner and Vieira 2007)

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i \epsilon s} w_i \mathbf{x}_i{}' \mathbf{V}^{-1} \mathbf{x}_i \right)^{-1} \sum_{i \epsilon s} w_i \mathbf{x}_i{}' \mathbf{V}^{-1} \mathbf{y}_i, \tag{2}$$

where w_i is a longitudinal survey weight, $\mathbf{V}$ is a $T \times T$ estimated 'working' variance matrix of $\mathbf{y}_i$ given $\mathbf{x}_i$ (Diggle et al. 2002), taken as the exchangeable variance matrix with diagonal elements $\hat{\sigma}^2$ and off-diagonal elements $\hat{\rho}\hat{\sigma}^2$ and $(\hat{\rho}, \hat{\sigma}^2)$ is an estimator of $(\rho, \hat{\sigma}^2)$. Further details on the pseudolikelihood approach may be found in Vieira (2009). The parameter ρ is the intra-individual correlation and σ^2 is the variance of y_{it}. Further discussion on the estimation of $\boldsymbol{\beta}$ and ρ is presented in Skinner and Vieira (2007). Notice that $\hat{\boldsymbol{\beta}}$ would be fully efficient when the underlying working model holds. Furthermore, under (1), $\hat{\boldsymbol{\beta}}$ is approximately unbiased with respect to the model and to the survey design, and may still be expected to associate both within and between individual information in a reasonably efficient manner, even if the working model for the error structure does not hold exactly (Skinner and Vieira 2007). Without the weight terms and survey-sampling considerations, the form of $\hat{\boldsymbol{\beta}}$, given by (2), is motivated by the generalised estimating Equations (GEE) approach of Liang and Zeger (1986). We shall denote this *unweighted* version by $\hat{\boldsymbol{\beta}}^{\mathrm{u}}$. The following estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$ allows for a stratified multistage sampling scheme and it is based upon the classical method of linearisation (Binder 1983; Skinner 1989b; Skinner and Vieira 2007)

$$\mathbf{v}(\hat{\boldsymbol{\beta}}) = \left[\sum_{i \epsilon s} w_i \mathbf{x}_i{}' \mathbf{V}^{-1} \mathbf{x}_i\right]^{-1} \left[\sum_h n_h/(n_h - 1) \sum_a (\mathbf{z}_{ha} - \bar{\mathbf{z}}_h)(\mathbf{z}_{ha} - \bar{\mathbf{z}}_h)'\right] \left[\sum_{i \epsilon s} w_i \mathbf{x}_i{}' \mathbf{V}^{-1} \mathbf{x}_i\right]^{-1},$$

where h denotes stratum, a denotes PSU, n_h is the number of PSUs in stratum h,

$\mathbf{z}_{ha}=\sum_i w_i \mathbf{x}_i{}'\mathbf{V}^{-1}\mathbf{e}_i$, $\bar{\mathbf{z}}_h=\sum_a \mathbf{z}_{ha}/n_h$ and $\mathbf{e}_i = \mathbf{y}_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$. If the weights, the sampling scheme and the difference between $n/(n-1)$ and 1 are ignored, this estimator reduces to the 'robust' variance estimator presented by Liang and Zeger (1986), which is as consistent when (1) holds, even when the working variance matrix $\mathbf{V}$ does not reflect the true variance structure (Diggle et al. 2002). We shall consider three further alternatives for estimating the covariance matrix of $\hat{\boldsymbol{\beta}}$: (i) $\mathbf{v}_a(\hat{\boldsymbol{\beta}})$, which considers that the population consists of only one stratum ($h = 1$), and therefore ignores stratification but takes *area* clustering into account; (ii) $\mathbf{v}_h(\hat{\boldsymbol{\beta}})$, which considers that each individual i is a PSU, and therefore ignores clustering but takes stratification into account; and (iii) the *naive* $\mathbf{v}_n(\hat{\boldsymbol{\beta}})$, which considers that $h = 1$ and that each individual is a PSU, and therefore ignores both stratification and clustering. We shall also perform variance estimation for $\hat{\boldsymbol{\beta}}^{\mathrm{u}}$, which is the point estimator that ignores the weights and stratification, and considers each individual as a PSU.

We shall be concerned with the potential bias of $\mathbf{v}_a(\hat{\boldsymbol{\beta}})$, $\mathbf{v}_h(\hat{\boldsymbol{\beta}})$ and $\mathbf{v}_n(\hat{\boldsymbol{\beta}})$ when in fact the design is complex. Skinner (1989a) has proposed the *misspecification effect* (*meff*), which is designed to measure the effects of incorrect specification of both (i) all the features of the sampling scheme and (ii) the model considered. The effect of the complex sampling scheme on $\mathbf{v}_a(\hat{\boldsymbol{\beta}})$, $\mathbf{v}_h(\hat{\boldsymbol{\beta}})$ and $\mathbf{v}_n(\hat{\boldsymbol{\beta}})$ can be evaluated by considering alternative *meffs* estimators, such as

$$meff_a\left[\hat{\beta}_k, v_a(\hat{\beta}_k)\right] = v(\hat{\beta}_k)/v_a(\hat{\beta}_k),\ meff_h\left[\hat{\beta}_k, v_h(\hat{\beta}_k)\right]$$
$$= v(\hat{\beta}_k)/v_h(\hat{\beta}_k),\ \text{and}\ meff_n\left[\hat{\beta}_k, v_n(\hat{\beta}_k)\right] = v(\hat{\beta}_k)/v_n(\hat{\beta}_k),$$

where $\hat{\beta}_k$ denotes the kth element of $\hat{\boldsymbol{\beta}}$. The $meff_a$, $meff_h$, and $meff_n$ separately estimate the impacts of stratification, clustering, and both stratification and clustering, respectively, and therefore are particular cases of the original *meff* of Skinner (1989a). We shall also calculate all the versions of the *meff* measure considered for $\hat{\boldsymbol{\beta}}^u$. Furthermore, a general *meff*,

$$meff_g = v(\hat{\beta}_k)/v_n\left(\hat{\beta}_k^u\right),$$

with $\hat{\beta}_k^u$ denoting the kth element of $\hat{\boldsymbol{\beta}}^u$, defined above, shall be calculated in order to access the bias caused by ignoring all the sampling-scheme features.

4. Applications

We consider two applications of regression analysis for four waves of the BHPS data, which include (i) the logarithm of the household income and (ii) a material satisfaction score as the dependent variables. Covariates were selected on the basis of the discussion in Salgueiro et al. (2013) and include time, gender, age category, marital status, number of children in the household, education level, social class, health status, and number of hours normally worked per week. We first estimate *meffs* for the linearisation estimator, considering $\hat{\boldsymbol{\beta}}$, as discussed in Section 3. By using data from just the first wave and setting $x_i = 1$, the estimated $meff_n$ for this cross-sectional mean is given in Table 1 and equals 1.343. In order to evaluate the impact of the longitudinal aspect of the data, we estimated a sequence of each of the *meffs* discussed above, using data for time $1, \ldots, t$, for $t = 2, 3, 4$. It is important to note that the estimation of cross-sectional and longitudinal means is often

Table 1. Meff estimates for estimated longitudinal means.

Dependent Variable	*Meff*	Waves			
		12	12 and 13	12 to 14	12 to 15
Log of the household income	$meff_a[\hat{\beta}_k, v_a(\hat{\beta}_k)]$	0.971	0.965	0.965	0.963
	$meff_h[\hat{\beta}_k, v_h(\hat{\beta}_k)]$	1.490	1.653	1.699	1.695
	$meff_n[\hat{\beta}_k, v_n(\hat{\beta}_k)]$	1.282	1.431	1.474	1.458
	$meff_a\left[\hat{\beta}_k^u, v_a\left(\hat{\beta}_k^u\right)\right]$	0.969	0.963	0.961	0.960
	$meff_h\left[\hat{\beta}_k^u, v_h\left(\hat{\beta}_k^u\right)\right]$	1.572	1.795	1.830	1.870
	$meff_n\left[\hat{\beta}_k^u, v_n\left(\hat{\beta}_k^u\right)\right]$	1.343	1.504	1.575	1.653
	$meff_g$	1.494	1.598	1.778	1.706
Material satisfaction score	$meff_a[\hat{\beta}_k, v_a(\hat{\beta}_k)]$	0.994	0.997	0.993	0.889
	$meff_h[\hat{\beta}_k, v_h(\hat{\beta}_k)]$	1.075	1.125	1.190	1.197
	$meff_n[\hat{\beta}_k, v_n(\hat{\beta}_k)]$	1.087	1.104	1.135	1.132
	$meff_a\left[\hat{\beta}_k^u, v_a\left(\hat{\beta}_k^u\right)\right]$	1.000	1.000	0.996	0.996
	$meff_h\left[\hat{\beta}_k^u, v_h\left(\hat{\beta}_k^u\right)\right]$	1.079	1.113	1.182	1.199
	$meff_n\left[\hat{\beta}_k^u, v_n\left(\hat{\beta}_k^u\right)\right]$	1.119	1.155	1.207	1.203
	$meff_g$	1.306	1.309	1.328	1.297

the aim of official statistics agencies, and therefore we consider the following analysis to be of particular relevance.

Although these estimated *meffs* are subject to sampling error, there seems to be some evidence from Table 1 of a tendency for $meff_h$, $meff_n$, and $meff_g$ to increase with the number of waves. It therefore seems like it becomes more important to allow for clustering and for the complex sampling design in general when the number of waves in the analysis increases. This result agrees with Skinner and Vieira (2007). Furthermore, the stratification effects appear ($meff_a$) to remain constant as the number of waves increases.

The models with logarithm of the household income as the dependent variable appear to have larger values for $meff_h$, $meff_n$, and $meff_g$ than the models with a material satisfaction score as the dependent variable. This result was expected, as attitudinal variables tend to have small estimated intracluster (intra-postcode) correlations for variables in British surveys (Lynn and Lievesley 1991; Vieira and Skinner 2008).

We have elaborated the analysis by including educational level as a covariate and we present in Table 2 only *meff* estimates for the estimated coefficients for the constant term of the longitudinal models.

The main feature of these results is that, as before, there is some evidence that $meff_h$, $meff_n$, and $meff_g$ increase with the number of waves. The intercept term may be seen as a domain mean, and standard survey-sampling theory for a *meff* of a mean in a domain cutting across clusters (Skinner 1989b; Skinner and Vieira 2007) implies that it will be somewhat less than a *meff* for the mean in the whole sample, as we have generally observed when comparing the results in Table 2 with those from Table 1. Moreover, such a comparison also confirms the observation of Kish and Frankel (1974) and Skinner and Vieira (2007) that *meffs* for regression coefficients tend not to be greater than *meffs* for the

Table 2. Meff estimates for the estimated constant terms in the longitudinal models (with one education covariate).

		Waves			
Dependent Variable	*Meff*	12	12 and 13	12 to 14	12 to 15
Log of the household income	$meff_a[\hat{\beta}_k, v_a(\hat{\beta}_k)]$	1.000	1.000	1.000	0.980
	$meff_h[\hat{\beta}_k, v_h(\hat{\beta}_k)]$	1.000	1.127	1.179	1.230
	$meff_n[\hat{\beta}_k, v_n(\hat{\beta}_k)]$	1.016	1.108	1.118	1.143
	$meff_a\left[\hat{\beta}_k^u, v_a\left(\hat{\beta}_k^u\right)\right]$	0.983	0.982	0.980	0.980
	$meff_h\left[\hat{\beta}_k^u, v_h\left(\hat{\beta}_k^u\right)\right]$	1.104	1.117	1.274	1.330
	$meff_n\left[\hat{\beta}_k^u, v_n\left(\hat{\beta}_k^u\right)\right]$	1.051	1.131	1.208	1.237
	$meff_g$	1.195	1.190	1.208	1.214
Material satisfaction score	$meff_a[\hat{\beta}_k, v_a(\hat{\beta}_k)]$	0.996	0.998	0.998	1.000
	$meff_h[\hat{\beta}_k, v_h(\hat{\beta}_k)]$	1.038	1.052	1.111	1.065
	$meff_n[\hat{\beta}_k, v_n(\hat{\beta}_k)]$	0.972	1.046	1.128	1.087
	$meff_a\left[\hat{\beta}_k^u, v_a\left(\hat{\beta}_k^u\right)\right]$	0.993	0.995	0.998	1.002
	$meff_h\left[\hat{\beta}_k^u, v_h\left(\hat{\beta}_k^u\right)\right]$	1.127	1.172	1.137	1.120
	$meff_n\left[\hat{\beta}_k^u, v_n\left(\hat{\beta}_k^u\right)\right]$	1.069	1.176	1.180	1.174
	$meff_g$	1.247	1.268	1.406	1.323

means of the dependent variable. Again the stratification effects appear to be constant with increases in the number of waves.

Although we have chosen not to present the *meffs* for the contrasts (coefficients for the education level covariate considered in the model), we have observed that they have varied in size and generally do not show any tendency to converge to one as the number of waves analysed increases, which would indicate no misspecification. As observed by Skinner and Vieira (2007), a *meff* for a contrast may be considered a combination of the traditional variance-inflating effect of clustering in surveys together with the variance-reducing effect of blocking in an experiment. Such variance reduction may be observed when the domains being contrasted share a common cluster effect that tends to cancel out in the contrasts, and therefore may imply that the actual variance of the contrast is lower than the expectation of the variance estimator which assumes independence between domains (Skinner and Vieira 2007).

The models have been further refined by the inclusion of additional covariates:

- time
- gender (g1 ¼ male, reference category; and g2 ¼ female)
- age category (ac1 ¼ 16 to 21 years, reference category; ac2 ¼ 22 to 29 years; ac3 ¼ 30 to 39 years; ac4 ¼ 40 to 49 years; and ac5 ¼ 50 years or older)
- number of children in the household
- education level (el1 ¼ first or higher degree, reference category; el2 ¼ other higher qualification; el3 ¼ nursing or A-levels; el4 ¼ other levels; and el5 ¼ no post-school qualification)

- social class (sc1 $\frac{1}{4}$ professional occupation, reference category; sc2 $\frac{1}{4}$ managerial or technical; sc3 $\frac{1}{4}$ skilled; and sc4 $\frac{1}{4}$ partly skilled or unskilled),
- health status (hs1 $\frac{1}{4}$ excellent, reference category; hs2 $\frac{1}{4}$ good; hs3 $\frac{1}{4}$ fair; and hs4 $\frac{1}{4}$ poor), numbers of hours normally worked per week (nh1 $\frac{1}{4}$ less than 16 hours, reference category; nh2 $\frac{1}{4}$ 16 to 29 hours; nh3 $\frac{1}{4}$ 30 to 40 hours; and nh4 $\frac{1}{4}$ more than 40 hours)
- and marital status (ms1 $\frac{1}{4}$ married or living as a couple, reference category, and ms2 $\frac{1}{4}$ widowed, divorced, separated or never married).

For the model with a material satisfaction score as the dependent variable, we have also added the logarithm of the household income as a covariate. As before, in Table 3 we present *meff* estimates only for the estimated coefficients for the constant term of the further elaborated longitudinal models.

There is some evidence of a tendency in the *meffs* for the constant to diverge from unity as the number of waves increases, especially for the model with a material satisfaction score as the dependent variable. Although we have not presented the *meffs* for the covariates, we have observed that $meff_h$, $meff_n$, and $meff_g$ generally have not shown any tendency to converge to one, for the same reasons as we have argued above. In general, when comparing the results in Table 3 with those in Tables 1 and 2, we have also confirmed the observation of Kish and Frankel (1974) and Skinner and Vieira (2007) that *meffs* for regression coefficients tend not to be greater than *meffs* for the means of the dependent variable, except for the estimated *meffs* for the constant term of the model with a material satisfaction score as dependent variable, which has presented surprisingly high *meffs* for the more elaborate model.

Table 3. Meff estimates for the estimated constant terms in the longitudinal models (with several covariates).

		Waves			
Dependent Variable	*Meff*	12	12 and 13	12 to 14	12 to 15
Log of the household income	$meff_a[\hat{\beta}_k, v_a(\hat{\beta}_k)]$	0.982	1.000	1.000	1.001
	$meff_h[\hat{\beta}_k, v_h(\hat{\beta}_k)]$	1.000	0.948	0.994	0.944
	$meff_n[\hat{\beta}_k, v_n(\hat{\beta}_k)]$	0.829	0.938	1.000	0.970
	$meff_a\left[\hat{\beta}_k^u, v_a\left(\hat{\beta}_k^u\right)\right]$	1.000	0.994	1.000	1.002
	$meff_h\left[\hat{\beta}_k^u, v_h\left(\hat{\beta}_k^u\right)\right]$	0.980	0.966	0.981	0.916
	$meff_n\left[\hat{\beta}_k^u, v_n\left(\hat{\beta}_k^u\right)\right]$	0.849	0.955	0.994	0.947
	$meff_g$	1.000	1.124	1.175	1.138
Material satisfaction score	$meff_a[\hat{\beta}_k, v_a(\hat{\beta}_k)]$	0.992	0.996	0.992	1.000
	$meff_h[\hat{\beta}_k, v_h(\hat{\beta}_k)]$	1.211	1.273	1.311	1.112
	$meff_n[\hat{\beta}_k, v_n(\hat{\beta}_k)]$	1.184	1.278	1.349	1.205
	$meff_a\left[\hat{\beta}_k^u, v_a\left(\hat{\beta}_k^u\right)\right]$	0.993	0.996	0.991	1.000
	$meff_h\left[\hat{\beta}_k^u, v_h\left(\hat{\beta}_k^u\right)\right]$	1.176	1.225	1.369	1.200
	$meff_n\left[\hat{\beta}_k^u, v_n\left(\hat{\beta}_k^u\right)\right]$	1.155	1.250	1.432	1.306
	$meff_g$	1.413	1.573	1.628	1.446

Table 4 presents coefficient, standard error ($\mathbf{se}(\hat{\boldsymbol{\beta}}) = \sqrt{\mathbf{v}(\hat{\boldsymbol{\beta}})}$ and $\mathbf{se}_n(\hat{\boldsymbol{\beta}}) = \sqrt{\mathbf{v}_n(\hat{\boldsymbol{\beta}})}$) and *meff* estimates for the model for Waves 12 to 15 with logarithm of the household income as the dependent variable and several covariates. The differences observed when we compare the point estimates produced by the standard Liang and Zeger (1986) estimator ($\hat{\boldsymbol{\beta}}_n$, given by Equation (2) without the weight terms) and the weighted pseudolikelihood estimator ($\hat{\boldsymbol{\beta}}$, given by Equation (2)) suggest that using standard statistical techniques with complex sampling data may lead to biased point estimates. Note the differences in the estimated coefficients for gender, age category, health status, and numbers of hours normally worked, confirming Nathan and Holt's (1980) results produced in a cross-sectional context. Moreover, the results in Table 4 also suggest that, in general, the BHPS complex sampling effects, if not taken into account in the estimation procedure, tend to lead to an underestimation of standard errors (compare columns labelled (1) and (2) and columns labelled (3) and (4)), and therefore to narrowed confidence intervals and possibly to the incorrect rejection of null hypotheses. In our application, complex sampling effects may lead to inappropriate statistical conclusions. This is confirmed by the estimated *meffs*, which are generally above one and even above two for gender. The $meff_n$ for $\hat{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}$ are similar, suggesting the impact of the complex sampling is the same irrespective of whether or not weights are used. However, $meff_g$ is nearly always larger than both these $meff_n$, suggesting that the effect of weighting is to further increase the estimated standard errors.

Figure 1 includes confidence intervals for both $\hat{\boldsymbol{\beta}}^u$ and $\hat{\boldsymbol{\beta}}$, considering both $\mathbf{se}_n(.)$ and $\mathbf{se}(.)$, for coefficients of covariates which had at least one $meff_g > 1.5$. Horizontal lines are represented both at $\beta = 0$ and $\hat{\boldsymbol{\beta}}$ for the plots on the left-hand side, and only $\beta = 0$ for the right-hand ones. Four different confidence intervals were calculated for each coefficient, labelled as: (a) confidence interval for $\hat{\boldsymbol{\beta}}^u$ based on $\mathbf{se}_n(.)$, (b) confidence interval for $\hat{\boldsymbol{\beta}}^u$ based on $\mathbf{se}(.)$, (c) confidence interval for $\hat{\boldsymbol{\beta}}$ based on $\mathbf{se}_n(.)$, and (d) confidence interval for $\hat{\boldsymbol{\beta}}$ based on $\mathbf{se}(.)$. Note, therefore, that: (a) does not allow for any sampling design features, (b) allows for clustering and stratification, (c) allows for weighting, and (d) allows for clustering, stratification, and weighting. The comparison of (a), (b), (c), and (d) helps us to evaluate the different sampling misspecification effects. Our plots demonstrate that different coefficients show different types of effects. The plot for the variable number of children, for example, shows a common situation faced by data analysts. Note the coefficients are considered significant when the sampling design is not considered in (a). Moving from (a) to (d), sampling design features are gradually being considered, leading to the coefficient not being significant in (d). Plots for social class and gender show weighting and stratification effects in the standard-error estimation. The plot for age category illustrates the effects of weighting, and the possibility of bias, in the point estimates. Plots for time and health status show different patterns for the evaluated effects depending on which point estimator is being considered.

5. Simulation Study

As the results reported in Section 4 are subject to sampling error, we conducted a simulation study to evaluate the behaviour of the *meff* measures. Each of the $d = 1, \ldots, D$ replicate samples is based on the BHPS data subset described above, which is

Table 4. Estimates for the four-waves model with logarithm of the household income as the dependent variable and several covariates.

Covariates		$\hat{\beta}_k^u$	$se_n(\hat{\beta}_k^u)$ (1)	$se(\hat{\beta}_k^u)$ (2)	$meff_n$ $(2)^2/(1)^2$	$\hat{\beta}_k$	$se_n(\hat{\beta}_k)$ (3)	$se(\hat{\beta}_k)$ (4)	$meff_n$ $(4)^2/(3)^2$	$meff_g$ $(4)^2/(1)^2$
Constant		3.66	0.029	0.028	0.947	3.64	0.031	0.031	0.970	1.138
Time		0.01	0.001	0.002	1.299	0.01	0.002	0.002	1.235	1.689
Gender	*g2*	−0.01	0.007	0.006	0.691	−0.02	0.012	0.010	0.680	2.085
Age category	*ac2*	−0.10	0.023	0.024	1.053	−0.08	0.025	0.028	1.172	1.433
	ac3	−0.09	0.022	0.023	1.095	−0.08	0.027	0.029	1.189	1.701
	ac4	−0.04	0.023	0.024	1.073	−0.03	0.025	0.026	1.145	1.310
	ac5	−0.10	0.022	0.023	1.076	−0.08	0.024	0.025	1.131	1.283
No. children		−0.01	0.004	0.005	1.179	−0.01	0.005	0.006	1.079	1.758
No. of hours	*nh2*	0.04	0.015	0.015	0.936	0.03	0.014	0.014	0.963	0.861
	nh3	0.06	0.012	0.011	0.915	0.06	0.013	0.013	0.938	1.126
	nh4	0.08	0.015	0.015	1.070	0.07	0.015	0.015	1.040	1.051
Education level	*el2*	−0.08	0.010	0.011	1.269	−0.08	0.011	0.012	1.056	1.371
	el3	−0.09	0.013	0.014	1.166	−0.09	0.015	0.016	1.067	1.359
	el4	−0.13	0.012	0.013	1.199	−0.13	0.014	0.014	1.102	1.378
	el5	−0.16	0.014	0.015	1.261	−0.15	0.015	0.016	1.208	1.396
Social class	*sc2*	−0.02	0.010	0.010	0.913	−0.02	0.015	0.014	0.958	1.920
	sc3	−0.05	0.011	0.011	1.073	−0.05	0.014	0.014	1.063	1.696
	sc4	−0.07	0.013	0.013	1.070	−0.07	0.016	0.017	1.134	1.681
Health status	*hs2*	−0.02	0.005	0.005	1.037	−0.01	0.004	0.004	1.117	0.536
	hs3	−0.03	0.006	0.006	1.216	−0.03	0.007	0.008	1.231	1.883
	hs4	−0.03	0.009	0.009	0.994	−0.03	0.011	0.012	1.104	1.806
Marital status	*ms2*	−0.07	0.011	0.011	0.967	−0.07	0.012	0.012	1.020	1.184

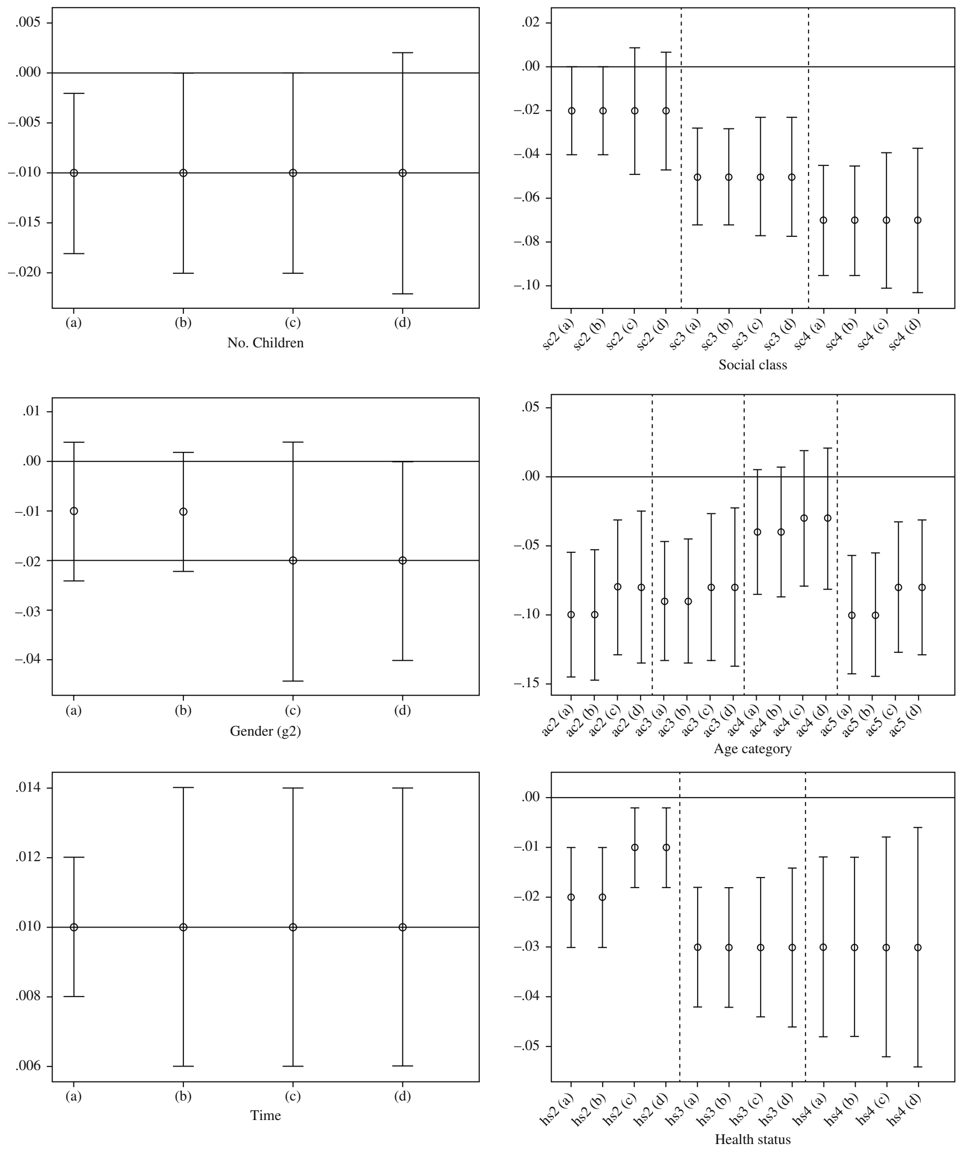

Fig. 1. Confidence intervals for coefficients of covariates with $meff_g > 1.5$.

considered as the ‘target population’. We evaluated the properties of variance estimators for unweighted point estimators and assessed only the impacts of clustering. We studied the *meff* when the number of waves in the analysis is increased. Note that we do not assess the impact of stratification, unequal probability sampling, nonresponse, and attrition.

Let y_{iat} be the value for the study variable for unit $i = 1, 2, \ldots, n_a^{sim}$, in PSU $a = 1, \ldots, m^{sim}$, at Wave t of the survey, where n_a^{sim} and m^{sim} are the sample sizes and the number of PSUs for the replicate sample d. To generate the values of y_{iat} for the simulation study, we used the following uniform correlation model, which includes a clustering effect:

$$y_{iat} = \mathbf{x}_{iat}\boldsymbol{\beta} + \eta_a + u_{ia} + v_{iat}, \quad (3)$$

where η_a represents the PSU (postcode area) random effects, u_{ia} denotes individual-level random effects (or unobservable individual specific factors), and v_{iat} are residuals, with $\eta_a \sim N\left(0, \sigma_\eta^2\right), u_{ia} \sim N\left(0, \sigma_u^2\right)$, and $v_{iat} \sim N\left(0, \sigma_v^2\right)$. We consider the logarithm of the household income and a material satisfaction score as dependent variables and the remaining variables listed and described in Section 2 as covariates. We have held the values of the covariates fixed.

The values adopted for $\boldsymbol{\beta}$, σ_η, σ^u, and σ_v were based on maximum-likelihood estimates for the model fitted to the 'target population', which were 0.16 and 2.10 for σ_u, and 0.11 and 1.88 for σ_v, respectively, for the models with the logarithm of the household income and with a material satisfaction score as dependent variables. In order to evaluate the effects of different impacts of clustering on the variance estimation procedures considered, we used the following realistic choices for σ_η: (i) $\sigma_\eta = 0.06$ (actual value estimated from fitting Model (3) to the data), $\sigma_\eta = 0.12$ and $\sigma_\eta = 0.18$, for the model with the logarithm of the household income as dependent variable; and (ii) $\sigma_\eta = 0.35$ (actual value estimated from fitting (3)), $\sigma_\eta = 0.70$, and $\sigma_\eta = 1.05$ for the model with a material satisfaction score as the dependent variable.

Let

$$\hat{E}(m\hat{e}ff) = \frac{1}{D}\sum_{d=1}^{D} m\hat{e}ff^{(d)}$$

be the mean of our *meff* of interest estimated over repeated simulation,

$$var(m\hat{e}ff) = \frac{1}{D-1}\sum_{d=1}^{D}\left[m\hat{e}ff^{(d)} - \hat{E}(m\hat{e}ff)\right]^2$$

be a simulation estimator of VAR($m\hat{e}ff$), the population variance of the misspecification-effect measure, and

$$se[\hat{E}(m\hat{e}ff)] = \sqrt{var(m\hat{e}ff)/D}$$

be the simulation standard error of $\hat{E}(m\hat{e}ff)$.

We initially set $x_i = 1$ in the models fitted to each generated replicate sample and therefore studied the behaviour of the *meff* for longitudinal means. We set n_a^{sim} equal the sample size for PSU a in our BHPS subsample and $m^{sim} = 234$, which is equal to the number of PSUs in our BHPS subsample. Therefore, Table 5 presents simulation results for four scenarios, including one that considers $\sigma_\eta = 0.00$ (i.e., no clustering effect), when $D = 1{,}000$.

The simulation results also provide evidence that the *meffs* increase as the number of waves in the analysis increases, at least for longitudinal means. This increase seems to be stronger for larger intracluster correlation. We also observe an increase in the *meff* when the intracluster correlation increases, as expected from the survey-sampling literature (Kish and Frankel 1974; Holt and Scott 1981; Scott and Holt 1982; Skinner 1986; and Skinner 1989a).

Table 5. $\hat{E}(m\hat{e}ff)$ and $se[\hat{E}(m\hat{e}ff)]$ (in brackets), for four scenarios (for longitudinal means).

Dependent Variable	σ_η	Waves			
		12	12 and 13	12 to 14	12 to 15
Log of the household income	0.00	1.1448	1.1561	1.1589	1.1597
		(0.0035)	(0.0035)	(0.0036)	(0.0036)
	0.06	1.1862	1.2018	1.2078	1.2107
		(0.0042)	(0.0043)	(0.0043)	(0.0043)
	0.12	1.2697	1.2940	1.3019	1.3073
		(0.0053)	(0.0055)	(0.0056)	(0.0056)
	0.18	1.3774	1.4061	1.4190	1.4255
		(0.0068)	(0.0070)	(0.0070)	(0.0071)
Material satisfaction score	0.00	1.0826	1.0986	1.1017	1.1030
		(0.0033)	(0.0033)	(0.0033)	(0.0033)
	0.35	1.0890	1.1063	1.1105	1.1129
		(0.0033)	(0.0034)	(0.0035)	(0.0035)
	0.70	1.1086	1.1363	1.1428	1.1462
		(0.0035)	(0.0037)	(0.0037)	(0.0037)
	1.05	1.1498	1.1806	1.1889	1.1936
		(0.0044)	(0.0046)	(0.0047)	(0.0048)

We also notice that the *meffs* are greater than one even when $\sigma_\eta = 0.00$. We believe that this is due to the model that is being fitted (with no covariates), which is different from the true model (with several covariates) that was used to generate the data. Therefore, this is a good example of the use of the *meff* to measure the effects of incorrect specification of both the sampling scheme and the model considered.

Following the same strategy considered in Section 4, we have elaborated the analysis by including educational level as a covariate. Tables 6 and 7 present the results for the constant term and one of the contrasts (one category) of the educational level covariate, for the logarithm of the household income and material satisfaction models respectively, using the same four scenarios as before.

The simulation results with the logarithm of the household income as the dependent variable and the educational level as the covariate also generally show a tendency for the *meffs* to increase as the number of waves in the analysis increases, more clearly for the constant (domain mean) but also for the contrasts (including those contrasts that were not presented in Table 6). This increase seems, again, to be stronger for larger clustering impacts. Furthermore, we notice once again that the *meffs* are greater than one even when $\sigma_\eta = 0.00$, but not as much as we observed in Table 5, as the model that is now being fitted (with one covariate) is slightly closer to the true model.

The simulation results with the material satisfaction score as the dependent variable and the educational level as the covariate, presented in Table 7, lead to very similar conclusions to those drawn from Table 6. In fact the increase in the *meff* is now even clearer. Moreover, when comparing the results from Tables 6 and 7 to those presented in Table 5, we confirm our results from Section 4, and the observation of Kish and Frankel (1974) and Skinner and Vieira (2007) that *meffs* for regression coefficients tend not to be greater than *meffs* for the means of the dependent variable.

Table 6. $\hat{E}(\hat{meff})$ and $se[\hat{E}(\hat{meff})]$ (in brackets), considering four scenarios for the logarithm of the household-income model with one education covariate.

Dependent Variable	σ_η	Coefficient	Waves			
			12	12 and 13	12 to 14	12 to 15
Log of the household income	0.00	Constant	1.0454 (0.0043)	1.0441 (0.0043)	1.0427 (0.0043)	1.0435 (0.0042)
		el5	1.0478 (0.0036)	1.0473 (0.0036)	1.0493 (0.0036)	1.0473 (0.0037)
	0.06	Constant	1.0872 (0.0044)	1.0908 (0.0044)	1.0921 (0.0043)	1.0933 (0.0043)
		el5	1.0864 (0.0037)	1.0897 (0.0037)	1.0881 (0.0038)	1.0827 (0.0037)
	0.12	Constant	1.1671 (0.0055)	1.1892 (0.0057)	1.1920 (0.0056)	1.1971 (0.0056)
		el5	1.1683 (0.0048)	1.1835 (0.0049)	1.1791 (0.0048)	1.1709 (0.0048)
	0.18	Constant	1.2760 (0.0069)	1.2950 (0.0068)	1.2926 (0.0067)	1.2976 (0.0067)
		el5	1.2644 (0.0057)	1.2786 (0.0058)	1.2607 (0.0055)	1.2458 (0.0053)

We included the following additional covariates: time, gender, age category, marital status, number of children in the household, education level, social class, health status, and numbers of hours normally worked. As the simulation results presented in Tables 5, 6, and 7 suggested very similar conclusions drawn from the models with the two different

Table 7. $\hat{E}(\hat{meff})$ and $se[\hat{E}(\hat{meff})]$ (in brackets), considering four scenarios for the material satisfaction score model with one education covariate.

Dependent Variable	σ_η	Coefficient	Waves			
			12	12 and 13	12 to 14	12 to 15
Material satisfaction score	0.00	Constant	1.0604 (0.0043)	1.0667 (0.0042)	1.0721 (0.0041)	1.0794 (0.0041)
		el5	1.0488 (0.0034)	1.0513 (0.0034)	1.0551 (0.0034)	1.0570 (0.0034)
	0.35	Constant	1.0672 (0.0043)	1.0786 (0.0043)	1.0843 (0.0043)	1.0897 (0.0043)
		el5	1.0503 (0.0037)	1.0585 (0.0036)	1.0644 (0.0035)	1.0662 (0.0035)
	0.70	Constant	1.0886 (0.0043)	1.0986 (0.0044)	1.1075 (0.0045)	1.1148 (0.0045)
		el5	1.0752 (0.0036)	1.0837 (0.0037)	1.0895 (0.0037)	1.0913 (0.0037)
	1.05	Constant	1.1106 (0.0048)	1.1300 (0.0047)	1.1406 (0.0048)	1.1507 (0.0047)
		el5	1.0924 (0.0038)	1.1094 (0.0039)	1.1151 (0.0040)	1.1188 (0.0040)

Table 8. $\hat{E}(m\hat{e}ff)$ *and* $se[\hat{E}(m\hat{e}ff)]$ *(in brackets), considering four scenarios for the logarithm of the household-income model with several covariates.*

Dependent Variable	σ_η	Coefficient	Waves			
			12	12 and 13	12 to 14	12 to 15
Log of the household income	0.00	Constant	0.9855 (0.0038)	0.9903 (0.0037)	0.9925 (0.0037)	0.9926 (0.0039)
		el5	0.9884 (0.0034)	0.9914 (0.0033)	0.9933 (0.0033)	0.9955 (0.0033)
	0.06	Constant	0.9911 (0.0037)	1.0105 (0.0037)	1.0158 (0.0038)	1.0193 (0.0040)
		el5	0.9834 (0.0033)	1.0087 (0.0033)	1.0196 (0.0033)	1.0222 (0.0034)
	0.12	Constant	0.9938 (0.0039)	1.0655 (0.0042)	1.0879 (0.0047)	1.0961 (0.0049)
		el5	0.9869 (0.0032)	1.0572 (0.0037)	1.0870 (0.0040)	1.1003 (0.0041)
	0.18	Constant	0.9793 (0.0037)	1.1109 (0.0050)	1.1607 (0.0057)	1.1766 (0.0059)
		el5	0.9877 (0.0033)	1.1138 (0.0043)	1.1626 (0.0049)	1.1814 (0.0052)

dependent variables considered, we have chosen to present results for the logarithm of the household-income models for the more complex model with several covariates. Table 8 presents results for the constant term and for the same contrast that was included in Tables 6 and 7.

Table 8 also shows an increase in the *meff* as the number of waves in the analysis increases. We may draw very similar conclusions to those regarding Tables 6 and 7. Furthermore, we now notice that the *meffs* are much closer to one when $\sigma_\eta = 0.00$, especially for the situation where we consider four waves, as the model that is being fitted in that case (with several covariates) is the true model and no clustering effect is induced. We believe these *meff* results are not significantly different to one as their 95% simulation confidence intervals include one for the four-waves model for most of the estimated coefficients.

6. Discussion

We have presented evidence that the impact of clustering may be stronger for longitudinal studies than for cross-sectional studies, and that *meffs* for the regression coefficients increase with the number of waves considered in the analysis, which confirms previous theoretical results by Skinner and Vieira (2007; Expression (11)). Longitudinal household surveys tend to have a long life in most countries (e.g., Panel Study of Income Dynamics in the United States; German Social Economic Panel in Germany) and therefore a large number of waves, and in such cases our conclusions are particularly relevant. Moreover, we have also observed that *meffs* for regression coefficients tend not to be greater than *meffs* for the means of the dependent variable. In fact, lower *meffs* are expected for models with increasing complexity (with more covariates) or for models that are closer to the true population model, which has been observed in our results. However, as previously stated, official statistical agencies often wish to estimate domain means, which correspond to simple models with, for example, a single covariate, and again in such cases our conclusions are particularly relevant.

Furthermore, our application results suggest that stratification effects remain constant with increases in the number of waves. This conclusion does not seem to be dependent upon the complexity of the model (i.e., number of covariates) that is being considered.

The main implication of our findings is that standard errors estimated in the analysis of panel data may be misleading if the initial sample was clustered and if this clustering is ignored in the analysis, more strongly so in situations where descriptive statistics (such as means) are being estimated or when the model that is being fitted is not well specified. Our results also suggest that longitudinal weighting has implications on both point and standard-error estimation. The analysis of longitudinal data collected by surveys that adopt unequal probability selection procedures, unit-nonresponse weighting adjustments for protection against attrition, and other weighting adjustments requires allowances for such features. Therefore, the types of misspecification that investigators need to protect against are those related to clustering and weighting. We believe that by taking our findings into account, analysts of longitudinal data will be able to produce better inferential results for panel surveys.

Possible future work could include investigating the impacts of various sampling design features in the analysis of panel data based on estimating marginal models for a binary response, such as the ones considered by Roberts et al. (2009). Moreover, the effects of item nonresponse and of the use of imputation in variance estimation in the longitudinal data context, which has not been dealt with here, could also be investigated in future work.

7. References

Binder, D.A. 1983. "On the Variances of Asymptotically Normal Estimators from Complex Surveys." *International Statistical Review* 51: 279–292. Doi: http://dx.doi.org/10.2307/1402588.

Binder, D.A. and G.R. Roberts. 2003. "Design-Based and Model-Based Methods for Estimating Model Parameters." In *Analysis of Survey Data*, edited by R.L. Chambers and C.J. Skinner. 29–48. Chichester: Wiley.

Chambers, R.L. and C.J. Skinner. 2003. *Analysis of Survey Data*. Chichester: Wiley.

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

Feder, M., "Fitting Regression Models to Complex Survey Data – Gelman's Estimator Revisited." International Statistical Institute, Proceedings of the 58th World Statistical Congress, Session CPS028, 21–26 August 2011, Dublin, Ireland. Available at: http://2011.isiproceedings.org/ (accessed January 2016).

Feder, M., G. Nathan, and D. Pfeffermann. 2000. "Multilevel Modelling of Complex Survey Longitudinal Data with Time Varying Random Effects." *Survey Methodology* 26: 53–65.

Fitzmaurice, G.M., M.N. Laird, and J.H. Ware. 2004. *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley.

Holt, D. and A.J. Scott. 1981. "Regression Analysis Using Survey Data." *The Statistician* 30: 169–178. Doi: http://dx.doi.org/10.2307/2988047.

Kish, L. and M.R. Frankel. 1974. "Inference from Complex Samples." *Journal of the Royal Statistical Society, Series B* 36: 1–37.

Liang, K. and S.L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73: 13–22. Doi: http://dx.doi.org/10.1093/biomet/73.1.13.

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

Nathan, G. and D. Holt. 1980. "The Effect of Survey Design on Regression Analysis." *Journal of the Royal Statistical Society, Series B* 42: 377–386.

Pfeffermann, D. 2011. "Modelling of Complex Survey Data: Why is it a Problem? How Should We Approach It?" *Survey Methodology* 37: 115–136.

Roberts, G., Q. Ren, and J.N.K. Rao. 2009. "Using Marginal Mean Models for Data from Longitudinal Surveys with a Complex Design: Some Advances in Methods." In *Methodology of Longitudinal Surveys*, edited by P. Lynn. 351–366. Chichester: Wiley.

Salgueiro, M.F., P.W.F. Smith, and M.D.T. Vieira. 2013. "A Multi-Process Second-Order Latent Growth Curve Model for Subjective Well-Being." *Quality and Quantity* 47: 735–752. Doi: http://dx.doi.org/10.1007/s11135-011-9541-y.

Scott, A.J. and D. Holt. 1982. "The Effect of Two-Stage Sampling on Ordinary Least Squares Methods." *Journal of the American Statistical Association* 77: 848–854. Doi: http://dx.doi.org/10.1080/01621459.1982.10477897.

Skinner, C.J. 1986. "Design Effects of Two-Stage Sampling." *Journal of the Royal Statististical Society, Series B* 48: 89–99.

Skinner, C.J. 1989a. "Introduction to Part A." In *Analysis of Complex Surveys*, edited by C.J. Skinner, D. Holt, and T.M.F. Smith. 23–58. Chichester: Wiley.

Skinner, C.J. 1989b. "Domain Means, Regression and Multivariate Analysis." In *Analysis of Complex Surveys*, edited by C.J. Skinner, D. Holt, and T.M.F. Smith. 59–87. Chichester: Wiley.

Skinner, C.J. and D. Holmes. 2003. "Random Effects Models for Longitudinal Survey Data." In *Analysis of Survey Data*, edited by R.L. Chambers and C.J. Skinner. 205–219. Chichester: Wiley.

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

Skinner, C. and M.D.T. Vieira. 2007. "Variance Estimation in the Analysis of Clustered Longitudinal Survey Data." *Survey Methodology* 33: 3–12.

Smith, P., P. Lynn, and D. Elliot. 2009. "Sample Design for Longitudinal Surveys." In *Methodology of Longitudinal Surveys*, edited by P. Lynn. 21–33. Chichester: Wiley.

Sutradhar, B.C. and M. Kovacevic. 2000. "Analysing Ordinal Longitudinal Survey Data: Generalised Estimating Equations Approach." *Biometrik* 87: 837–848. Doi: http://dx.doi.org/10.1093/biomet/87.4.837.

Taylor, M.F. (ed.), J. Brice, N. Buck, and E. Prentice-Lane. 2010. *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. Colchester: University of Essex.

UK Data Service. *UK Data Service*. Available at: http://ukdataservice.ac.uk (accessed January 2016).

Vieira, M.D.T. 2009. *Analysis of Longitudinal Survey Data*, 1st ed. Saarbrücken: VDM Verlag.

Vieira, M.D.T. and C.J. Skinner. 2008. "Estimating Models for Panel Survey Data under Complex Sampling." *Journal of Official Statistics* 24: 343–364.

Received July 2014
Revised December 2015
Accepted January 2016