

Article

Multilevel modelling of complex survey longitudinal data with time varying random effects

by Moshe Feder, Gad Nathan and Danny Pfeffermann

June 2000



Multilevel modelling of complex survey longitudinal data with time varying random effects

Moshe Feder, Gad Nathan and Danny Pfeffermann¹

Abstract

Longitudinal observations consist of repeated measurements on the same units over a number of occasions, with fixed or varying time spells between the occasions. Each vector observation can be viewed therefore as a time series, usually of short length. Analyzing the measurements for all the units permits the fitting of low-order time series models, despite the short lengths of the individual series. We illustrate this paradigm using simulated data that follow the rotation scheme of the Israel Labor Force Survey (LFS). This survey employs a rotating panel sampling scheme of two quarters in the sample, two quarters out of the sample then two quarters in again. The model consists of two-level linear models for single time points that are connected by allowing the second level effects (corresponding to households) and the first level residuals (corresponding to individuals) to evolve stochastically over time. The likelihood of the model is easily constructed by employing the time series properties of the combined model. However, in view of the large number of unknown parameters, direct maximization of the likelihood could yield unstable estimators. Therefore, a two-stage procedure is adopted. At the first stage, a separate two-level model is fitted for each time point, thus yielding estimators for the fixed effects and the variances. At the second stage, the time series likelihood is maximized only with respect to the time series model parameters. This two-stage procedure has the further advantage of permitting appropriate first and second level weighting to account for possible informative sampling effects. Empirical results when fitting the model to data collected by the Israel LFS are also presented.

Key Words: Informative sampling; Probability weighted IGLS; Rotating panel schemes; State-space models.

1. Introduction

1.1 Background and objectives

In recent years there has been a growing interest in fitting models to data collected from longitudinal surveys that use complex sampling designs. This interest reflects expansion in requirements by policy makers and social scientists for in-depth studies of social processes over time, rather than of one-time “snap-shots” provided by cross-sectional analyses. A familiar example is the estimation of gross flows between social and demographic states such as employment states or health and education levels. For discussions of these issues and the problems they raise with respect to the design and analysis of longitudinal surveys, see Duncan and Kalton (1987) and Binder (1998).

Examples of surveys we wish to consider in this paper are of three types:

1. Rotating panel surveys such as labor force surveys carried out in many countries. These surveys were often designed originally for cross-sectional analysis of household and individual data, so as to study labor force and other socio-economic characteristics on a current basis. Complex rotating sampling schemes have later been introduced in order to improve comparisons over time. For example, the quarterly Israel Labor Force Survey (LFS) employs a rotating panel sampling scheme whereby each unit in the sample is interviewed for two consecutive quarters; it is left out of the

sample for the next two quarters and then is interviewed again for two more consecutive quarters. In the U.S.A. and Brazil, a more complicated sampling scheme of 4 months in the sample, 8 months out of the sample and then 4 months in again is used. Australia, Canada and the U.K. employ sampling schemes by which sampled units are interviewed over a succession of months or quarters before being dropped from the sample. These kinds of surveys are increasingly used for short-term longitudinal analysis, such as the estimation of gross flows between labor force states or studies of social mobility. This has not always proved simple due to the complexity of the survey designs, difficulties in matching and response errors.

2. Medium term panel surveys, such as the U.S. Survey of Income and Programme Participation (SIPP, Herriot and Kasprzyk 1984), the U.S. Panel Study of Income Dynamics (PSID, Survey Research Center 1984) and the Canadian Survey of Labor and Income Dynamics (SLID, Webber 1994). These surveys differ from labor force surveys in being specially designed for longitudinal analysis of economic and social characteristics of households and individuals. For example, SIPP includes an intensive investigation in the form of a full retrospective interview every 4 months. It provides a complete work history for the

1. Moshe Feder, Department of Social Statistics, University of Southampton, Southampton, S017 1BJ, U.K.; Gad Nathan and Danny Pfeffermann, Department of Statistics, Hebrew University, Jerusalem, 91905, Israel.

survey period (30-48 months) by combining the continuous retrospective four-month recall data with a reconciliation of data provided for longer periods.

3. Longitudinal cohort studies characterized by the follow-up of a cohort sample over a long time period. For example, in the British Household Panel Survey, starting from a sample of addresses selected in 1991, data have been collected on the same households in subsequent annual waves for over seven years. A wide range of data is collected on labor force characteristics, economic resources and health and education, with emphasis on longitudinal aspects. In this survey all members of the originally selected households were followed and the sample was supplemented by the addition of entrants to the sample households, including children born to sample household members. Other longitudinal cohort studies such as the British National Child Development Study and the British Cohort Study have surveyed a cohort of births over periods of up to 40 years. See Nathan (1999) for description and discussion of the latter three studies.

Most of the studies associated with these surveys require longitudinal analysis for populations that have a complex hierarchical structure, based on data collected from complex sampling designs. Standard analysis of longitudinal survey data often fails to account for the complex nature of the sampling design such as the use of unequal selection probabilities, clustering, post-stratification and other kinds of weighting used for the treatment of non-response. The effect of sampling on the analysis is due to the fact that the models in use typically do not incorporate all the design variables determining the sample selection, either because there may be too many of them or because they are not of substantive interest. However, if the design is “informative” in the sense that the outcome variable is correlated with the design variables not included in the model, even after conditioning on the model covariates, standard estimates of the model parameters can be severely biased, leading possibly to false inference. Pfeffermann (1993, 1996) reviews many examples reported in the literature that illustrate the effects of ignoring the sampling process when fitting models to survey data and discusses methods that have been proposed to deal with this problem. See also the book edited by Skinner, Holt, and Smith (1989) and the more recent paper by Pfeffermann, Skinner, Goldstein, Holmes, and Rasbash (1998) to which we refer in more detail below. It should be emphasized that standard inference may be biased even when the original sample design is simple random within design strata, due to non-response, attrition, and imperfect frames that result in de facto a posteriori differential inclusion probabilities. Special features of longitudinal studies, such as late additions of individuals who join panel households, can also lead to de facto unequal inclusion probabilities.

In this paper we propose to deal with the problems arising from the hierarchical nature of the target population, the longitudinal aspect of the analysis and the effects of complex sampling designs by combining three separate statistical methodologies. These are multilevel modelling (MLM), time series modelling and methods of analysis under complex informative sampling. Multilevel models are used to deal with the hierarchical structure of many human populations like persons within households, pupils within classes, classes within schools and so forth. The models, extensively employed by social scientists especially in the field of education, account for the effects of observed covariates at the lower and higher levels of the structure, with fixed or random coefficients. Common unobservable random effects within the higher levels capture further unexplained variations. The method of Iterative Generalized Least Squares (IGLS) is commonly used for estimating the model parameters, Goldstein (1986, 1995).

Simple state-space time series models are used to combine the multilevel models operating at different time points via a set of linear transition equations that account for the time series relationships of the random covariate coefficients and the higher level random effects. The Kalman filter is used for estimating the model parameters and predict the random effects for current and future time points. Smoothing algorithms can be used for updating past predictions, Harvey (1989). Methods of model fitting under informative sampling are employed to control the bias resulting from the sample selection process. Such methods have been investigated in recent years in the context of analytic inference from complex sample surveys, mostly for cross-sectional analysis of single-level models, *cf.* Skinner *et al.* (1989). In the present paper we utilize the methodology of sample weighting for multilevel modelling as developed by Pfeffermann *et al.* (1998).

The aims of the present study are then to develop models and methods of estimation for longitudinal analysis of hierarchically structured data, taking unequal sample selection probabilities into account. The main feature of our approach is that the model is fitted at the individual level but it contains common higher level random effects that change stochastically over time. The model enables to predict the higher and lower level random effects (like household and individual person effects in the present application), using the data for all the time points with observations. This should enhance model-based inference from complex survey data since it permits a better understanding of the structure and correlation pattern of the longitudinal measurements. In particular, it is bound to improve the prediction of individual measurements compared to the use of aggregate time series models, which by their nature fail to separate the individual (person) effects from the common higher level (household) effects. These advantages are partly illustrated in the example of section 6 and more so in a related paper by Pfeffermann and Nathan (forthcoming) which focuses on the imputation of missing data. It is

important to emphasize in this regard that although the length of each individual longitudinal record is often very short (4 measurements for each individual in our application), the number of records is usually sufficiently large to warrant the application of classical time series estimation and model diagnostic procedures. In this article we only consider parameter estimation under a given model but the use of test statistics and diagnostic procedures that employ the empirical innovations for model identification follows through with minor modifications by virtue of the use of maximum likelihood estimation methods and the consistency of the parameter estimators.

In section 2 we overview the main features of the aforementioned statistical methodologies that are employed in subsequent sections. In section 3 we propose a model that addresses the longitudinal aspects discussed above. Estimation procedures are discussed in section 4. Section 5 contains the results of a simulation study carried out for assessing the performance of the various estimators under different sampling scenarios. Results obtained when fitting the model to real data collected by the Israel LFS are presented in section 6, followed by a brief summary in section 7 of possible model extensions and applications.

1.2 Literature review

Previous work in this area deals mostly with longitudinal data in a non-survey context and does not consider hierarchically structured populations. In particular, none of the studies that we have come across permits the second level effects (common household effects in our application) to evolve over time. For example, Goldstein, Healy and Rasbash (1994) consider the analysis of repeated measurements using a two-level model with individuals as second levels and the repeated measurements as the first levels. The model extends the standard two-level model by permitting the first level measurements to be correlated over time. The authors consider several possibilities of modelling the auto-correlation structure, which include autoregressive models when the measurements are taken at equally spaced time points and autocorrelation functions when the observations are taken at unequal time intervals. In the latter case the autocorrelation function is linearized for estimation purposes.

Several authors study the application of time series models for the analysis of longitudinal data. In a series of papers by Jones and his co-authors (Jones and Ackerson 1990, Jones and Boadi-Boating 1991, Jones and Vecchia 1993) and the book by Jones (1993), the authors consider observations taken at unequally spaced time gaps. The observations referring to the same subject are allowed to be serially correlated by postulating continuous autoregressive moving average models. These models contain fixed and random effects, but do not have a hierarchical population structure. Weighted least squares and state space modelling combined with the Kalman filter are used for calculating the likelihood function.

Continuous time autoregressive models for irregularly spaced longitudinal data are considered also by Belcher, Hampton and Tunnicliffe (1984), using linear stochastic differential equations for describing the process generating the data. An Empirical Bayes approach is proposed by Bryant and Day (1991) for the simultaneous analysis of a system of mixed linear models, having linked and serially correlated random effects. Chi and Reinsel (1989) consider a score test for autocorrelation between individual errors under a “conditional independence” random effects model. The authors derive a maximum likelihood estimation procedure and use the estimators for predicting the random effects by application of Empirical Bayes.

Diggle, Liang and Zeger (1994) propose the use of generalized linear models for the analysis of longitudinal data. They consider a transition (Markov) model by considering past values as additional predictor variables. Transitional extensions of the GLM are used for maximum likelihood estimation under linear link functions, whereas for non-linear link functions the estimation is based on conditional score functions. Lawless (1999) uses an event history approach for the analysis of longitudinal data. By this approach, the dependent variable is the number of occurrences of a particular event up to a given time point t , with the limiting transitional probabilities being modelled as functions of the previous history and covariates. Zimmerman and Nunez-Anton (1997) propose a structured antedependence model for longitudinal data, primarily in the context of growth analysis. Neither of the above studies considers a hierarchical structure or a complex sampling design.

Finally, Skinner and Holmes (1999) consider a model for longitudinal observations that consists of a “permanent” random effect at the individual level and autocorrelated transitory random effects corresponding to different waves of investigation. The authors study two approaches for the estimation of the unknown model parameters with both approaches accounting for sampling effects and “non-informative” attritions. The first approach treats the repeated observations as correlated multivariate outcomes and derives probability-weighted estimators that account for the correlation structure. The second approach considers the model as a two-level model with “individuals” as the second level units and the repeated measurements as first level units. Estimation of the unknown parameters under this approach is carried out by a modification of the PWIGLS method of Pfeiffermann *et al.* (1998, see section 2.2).

2. Statistical methodologies underlying the proposed approach

2.1 Multilevel models

In what follows we consider a two-level model for the response variable y in a population consisting of

$i = 1, \dots, M$ second level units (household, schools, ...) and $j = 1, \dots, N_i$ individuals within second level unit i . The model is,

$$y_{ij} = x'_{ij} \beta + z'_{ij} u_i + z_{0ij} e_{ij}, \quad i = 1, \dots, M; \quad j=1, \dots, N_i, \quad (2.1)$$

where x_{ij} , z_{ij} , and z_{0ij} are known covariate values of dimensions p , q and 1 respectively, β is a fixed parameter vector of dimension p and $u_i \sim N(0, \Omega)$ and $e_{ij} \sim N(0, \sigma^2)$ are independent random second level effects and first level residuals of orders p and 1 respectively.

The inclusion of the multipliers z_{0ij} allows for first level heteroscedasticity whereas the common second level effects u_i explain the (interclass) correlations between individual measurements corresponding to the same second level unit. In the simple case of the "random intercept model", $y_{ij} = x'_{ij} \beta + u_i + e_{ij}$, these correlations take the familiar form, $\text{Corr}(y_{ij}, y_{ik}) = \sigma_u^2 / (\sigma_u^2 + \sigma^2)$. The random intercept model is often applied for small area estimation (see below).

As stated in the introduction, models like (2.1) are widely used by social scientists for studying the effects of the covariate variables and the interrelationships between observations corresponding to the same higher level unit. In such cases, primary interest is in the estimation of the vector coefficient β and the vector θ of the distinct elements of Ω and σ^2 . Another, well-known application of the two-level model is for "small area estimation", in which case the second levels are geographical areas or other domains of study. In small area estimation, the target of the analysis is the prediction of the second level (area) means $\bar{X}'_i \beta + \bar{Z}'_i u_i$, where \bar{X}_i and \bar{Z}_i are the true area covariate means, and the estimation of the model parameters is only an intermediate step. See Rao (1999) for a recent review.

Estimation of the unknown model parameters is carried out most conveniently by use of the Iterative Generalized Least Squares (IGLS) algorithm (Goldstein 1986, 1995). For a random sample of m second level units and n_i first level units within second level unit i , the model holding for the sample data is first written in matrix form as

$$y_i = X_i \beta + d_i, \quad i = 1 \dots m \quad (2.2)$$

where $y_i = [y_{i1}, \dots, y_{in_i}]'$, $X_i = [x_{i1}, \dots, x_{in_i}]'$ and $d_i = [d_{i1}, \dots, d_{in_i}]'$ with $d_{ij} = (z'_{ij} u_i + z_{0ij} e_{ij})$. Then, $d_i \sim N(0, V_i)$, where $V_i = Z_i \Omega Z_i' + \sigma^2 Z_{0i}^2 = V_i(\theta)$; $Z_i = [z_{i1} \dots z_{in_i}]'$ and $Z_{0i} = \text{diag}[z_{0i1} \dots z_{0in_i}]$. The IGLS algorithm iterates between the estimation of β , with θ considered known, and the estimation of θ , with β considered known. At each iteration, the estimate obtained for the other vector parameter on the previous iteration is used as the "known" parameter. This process is a special case of the EM algorithm and it converges to the corresponding maximum likelihood estimators (MLE) under the stated normality assumptions. It is known to provide consistent estimators under more general conditions.

2.2 MLM estimation under informative sampling

The IGLS algorithm described in section 2.1 assumes that the model defined by (2.2) holds for the sample data. This would be the case if selection of the first and second level units is carried out by simple random sampling. However, as discussed in the introduction, the selection of the sample could be informative so that the model holding for the sample units differs from the model holding in the population. For example, in an educational survey, schools in poor areas could be sampled with higher probabilities. In a household survey, higher selection probabilities could be assigned to households in areas characterized by high proportions of minorities or to persons that are unemployed. All illustrated by Pfeffermann *et al.* (1998) and also in section 5 of the present paper, the use of the IGLS algorithm in such cases could yield severely biased estimators for all the parameters. The authors propose therefore a probability weighted IGLS (PWIGLS) algorithm that protects against informative sampling.

The algorithm is an adaptation of the pseudo-MLE method (Binder 1983, Skinner *et al.* 1989, Pfeffermann 1993). Suppose that the two-level model defined by (2.1) holds for the target population. Had all the population values been observed, the IGLS would converge at the end of the iterative process to the census estimators, $(\hat{\beta}_c, \hat{\theta}_c)$. At each iteration, the intermediate estimators $(\hat{\beta}_{(i)}, \hat{\theta}_{(i)})$ are products of matrices with elements that are functions of sums of the population values. When the IGLS is applied to sample data, the population sums are substituted by the corresponding sample sums. The PWIGLS consists of further replacing the unweighted sample sums by weighted sums. Denote by $\pi_i = \Pr(i \in s)$ the second level sample inclusion probabilities and by $\pi_{j|i} = \Pr(j \in s | i \in s)$ the conditional first level inclusion probabilities. The PWIGLS estimators are obtained by, 1-replacing each second level sample sum of the general form $\sum_{i=1}^n g_i$ by the weighted sum $\sum_{i=1}^n w_i g_i$, where $w_i = \pi_i^{-1}$ and 2-replacing each first level sample sum $\sum_{j=1}^{n_i} g_{ij}$ by the weighted sum $\sum_{j=1}^{n_i} w_{j|i} g_{ij}$ with $w_{j|i} = \pi_{j|i}^{-1}$. Note that the weighting process requires the knowledge of the inclusion probabilities at both stages of the selection process and not just the final overall inclusion probabilities $\pi_{ij} = \pi_{j|i} \times \pi_i$.

As established by Pfeffermann *et al.* (1998), the PWIGLS estimators are consistent for the model parameters when both the first and second level sample sizes increase, but the estimators of the variances are not consistent if the first level sample sizes are bounded. For this case, the authors propose appropriate scaling of the weights $w_{j|i}$ that eliminates the bias, provided that the sample selection within the second level units is noninformative. It is important to emphasize that standard weighting of the sample measurements by the weights $w_{ij} = \pi_{ij}^{-1}$, which is routinely applied for single level models yields consistent estimators only for β .

2.3 State-space models

State-space models as considered here consist of two sets of equations:

1. The measurement (observations) equations:

$$y_t = X_t \beta_t + L_t \alpha_t + \varepsilon_t; E(\varepsilon_t) = 0,$$

$$E(\varepsilon_t \varepsilon'_{t-k}) = \delta_k H_t, t = 1, \dots, T \tag{2.3}$$

2. The transition (system) equation:

$$\alpha_t = G_t \alpha_{t-1} + \eta_t; E(\eta_t) = 0,$$

$$E(\eta_t, \eta'_{t-k}) = \delta_k Q_t, t = 1, \dots, T \tag{2.4}$$

where $\delta_k = 1$ for $k = 0$ and $\delta_k = 0$ otherwise. We also assume $E(\varepsilon_t \eta'_s) = 0$ for all t and s . Note that both y_t and α_t can be multivariate. The measurement equations relate the observations y_t at any given time point to covariate values X_t with fixed (nonstochastic) vector coefficients β_t , and linear functions L_t of an unobservable state vector α_t . The transition equations describe the time series relationships between the components of the state vector. The matrices X_t , L_t and G_t are assumed to be nonstochastic although they may change over time, as is the case with the vector coefficients β_t . Notice that the latter vectors can be included as part of the state vectors by taking their transition matrix to be the zero matrix of corresponding order and defining the corresponding residual variances in Q_t to be very large. See Sallas and Harville (1981) for details.

Although not written here in its most general form, the state-space model defined by (2.3) and (2.4) is known to include as special cases many of the time series and mixed linear models in common use. As important examples we mention the family of ARIMA models and models with random regression coefficients. The MLM defined by (2.1) can also be easily structured in a state-space form. To see this, replace the index i by t and define $L_t = [X_t, Z_t]$, $\alpha_t = [\beta'_t, u'_t]'$, $H_t = \sigma^2 Z_{0t}$ and $G_t = [I_p, 0_q]$ where I_p and 0_q define the identity matrix and the zero matrix of the appropriate orders. (The matrices Z_t and X_t are defined below (2.2).) The vector coefficient β_t is added for convenience to the state vector. The covariance matrix Q_t is block diagonal with 0_p and $Z_t \Omega Z'_t$ as the two blocks. The use of the zeroes matrix 0_p for the covariance of $(\beta_t - \beta_{t-1})$ guarantees that the β -coefficients are fixed over time, in accordance with (2.1). (The representation of the MLM in a state-space form is not unique.)

For given covariance matrices $\{H_t, Q_t\}$ and assuming that β_t , L_t and G_t are known for all t , the best linear unbiased predictor (BLUP) of the state vector at any given time t , based on all the data accumulated until that time, is conveniently obtained by means of the Kalman filter. Let $\hat{\alpha}_{t-1}$ define the BLUP of α_{t-1} based on the observations

until time $(t - 1)$, with covariance matrix $P_{t-1} = \text{Cov}(\hat{\alpha}_{t-1} - \alpha_{t-1})$. The BLUP of α_t at time $(t - 1)$ is then, $\hat{\alpha}_{t|t-1} = G_t \hat{\alpha}_{t-1}$ with covariance matrix $P_{t|t-1} = \text{Cov}(\hat{\alpha}_{t|t-1} - \alpha_t) = \text{Cov}(\hat{\alpha}_{t|t-1} - \alpha_t) = G_t P_{t-1} G'_t + Q_t$. When new observations y_t become available, the predictor $\hat{\alpha}_{t|t-1}$ and the corresponding covariance matrix are updated as

$$\hat{\alpha}_t = \hat{\alpha}_{t|t-1} + P_{t|t-1} L'_t F_t^{-1} (y_t - X_t \beta_t - L_t \hat{\alpha}_{t|t-1})$$

$$P_t = P_{t|t-1} - P_{t|t-1} L'_t F_t^{-1} L_t P_{t|t-1} \tag{2.5}$$

where $F_t = L_t P_{t|t-1} L'_t + H_t = \text{Var}(y_t - \hat{y}_{t|t-1})$ with $\hat{y}_{t|t-1} = X_t \beta_t + L_t \hat{\alpha}_{t|t-1}$ defining the BLUP of y_t at time $(t - 1)$. The actual application of the Kalman filter requires a proper initialization for $\hat{\alpha}_{1|0}$ and $P_{1|0}$ which depends on the model under study. See section 4 for the initialization under the model proposed in this paper.

The unknown model parameters (β_t , elements of H_t , Q_t and possibly L_t and G_t) are ordinarily estimated by MLE with the likelihood conveniently constructed by use of the “prediction error decomposition”. Assuming that $\dim(y_t) = n$, the log-likelihood takes the general form,

$$\log(L) = - \left\{ T \frac{n}{2} \log(2\pi) + \frac{1}{2} \sum_{t=1}^T \log |F_t| + \frac{1}{2} (Y_t - \hat{Y}_{t|t-1})' F_t^{-1} (Y_t - \hat{Y}_{t|t-1}) \right\}. \tag{2.6}$$

For a thorough discussion of state-space models and their applications, see Harvey (1989).

3. Model for hierarchical longitudinal data

In this section we propose a time series multilevel model which combines separate cross-sectional two-level models by modelling the evolution of the first and second level random effects over time. Let S_t define the sample available at time t , composed of m_t level 2 units with n_h level 1 units in level 2 unit h . The formulation of the overall sample in terms of the subsets S_t covers situations where the longitudinal observations are collected at different time periods. The proposed model allows also for the rotation patterns mentioned previously and for wave non-response. Note that the samples observed at different time points are generally not disjoint and that the assumption that n_h is fixed over time is not restrictive. Pfeffermann and Nathan (forthcoming) consider the case of temporal missing data for which this supposition does not hold. As long as the missing data are missing completely at random, generalization of the present methodology to this case is straight-forward. We assume the following two-level model to hold for the sample S_t :

$$y_{hjt} = x'_{hjt} \gamma_t + z'_{ht} v_t + z'_{ht} u_{ht} + e_{hjt},$$

$$h = 1, \dots, m_t, j = 1, \dots, n_h, \tag{3.1}$$

where y_{hjt} is the outcome for first level unit j in second level unit h , x_{hjt} and z_{ht} are fixed known covariate vectors of dimensions p and q respectively, γ_t and v_t are fixed (unknown) vector coefficients and u_{ht} and e_{hjt} are independent second level and first level random effects. For given time t , the model defined by (3.1) is basically the same as the MLM model defined by (2.1), except that we assume $z_{hjt} = z_{ht}$ for all j and t , thus distinguishing between first level covariates and second level covariates. We assume also for convenience $z_{0hjt} \equiv 1$. The model is quite general in that all the covariate variables, the fixed vector coefficients and the random effects are allowed to vary over time in ways defined below. Notice that by assuming that (3.1) holds for the sample data, it is implicitly assumed that the sampling design is noninformative. See the discussion in section 2.2 and also section 4 below.

As in (2.2), the model defined by (3.1) can be formulated in matrix form as,

$$Y_{ht} = X_{ht} \gamma_t + Z_{ht} v_t + Z_{ht} \mu_{ht} + I_{n_h} e_{ht}, \quad (3.2)$$

where $Y_{ht} = [y_{h1t}, \dots, y_{hm_t t}]'$, $X_{ht} = [x_{h1t}, \dots, x_{hm_t t}]'$, $Z_{ht} = 1 \otimes z_{ht}$ and $e_{ht} = [e_{h1t}, \dots, e_{hm_t t}]'$ with \otimes defining the Kronecker product. The matrix representation (3.2) can be written concisely as,

$$Y_{ht} = \tilde{X}_{ht} \beta_t + \tilde{Z}_{ht} \alpha_{ht}, \quad (3.3)$$

where $\tilde{X}_{ht} = [X_{ht}, Z_{ht}]$; $\tilde{Z}_{ht} = [Z_{ht}, I_{n_h}]$; $\beta_t = [\gamma_t', v_t']'$; $\alpha_{ht} = [u_{ht}', e_{ht}']'$.

Next we model the time series relationships of the vector coefficients and the random effects. We assume that the vectors β_t , $t = 1, 2, \dots$ are fixed without specifying the way they evolve over time. This assumption is generally not restrictive because in practical applications the overall sample size in any given time point is usually sufficiently large to allow accurate estimation of the vector coefficients without having to borrow information across time. For the random second and first level effects we postulate first order autoregressive [AR(1)] relationships of the form,

$$u_{ht} = A u_{h,t-1} + \delta_{ht}; \quad e_{ht} = \rho e_{h,t-1} + \varepsilon_{ht} \quad (3.4)$$

where A is a $(q \times q)$ matrix of fixed coefficients, ρ is a fixed scalar and $\delta_{ht} \sim N(0_q, \Delta)$; $\varepsilon_{ht} \sim N(0_{n_h}, \sigma_\varepsilon^2 I_{n_h})$ are independent white noise series. The model defined by (3.4) is rather simple and as a further simplification we assume that A and Δ are diagonal, implying that the second level random effects are independent. It is assumed also that $|\rho| < 1$ and $|A_{kk}| < 1$ for all k to guarantee stationarity. More complex models can be considered in principle but it should be emphasized that unlike in classical (aggregate) time series analysis, longitudinal observations may only be taken over a very short time period in which case the use of models that incorporate lagged values of high order may no longer be operational. For example, in the quarterly Israel LFS described in the introduction, individuals are in the

sample for a total of 4 quarters over a time period of 6 quarters which clearly limits the class of time series models that can be postulated for the random effects.

The AR(1) models defined by (3.4) can be written concisely as

$$\alpha_{ht} = G_h \alpha_{h,t-1} + \eta_{ht}, \quad h = 1, \dots, m_t \quad (3.5)$$

where,

$$G_h = \begin{bmatrix} A & 0 \\ 0 & \rho I_{n_h} \end{bmatrix}, \quad \eta_{ht} = \begin{bmatrix} \delta_{ht} \\ \varepsilon_{ht} \end{bmatrix},$$

$$\eta_{ht} \sim N(0, Q_h), \quad Q_h = \begin{bmatrix} \Delta & 0 \\ 0 & \sigma_\varepsilon^2 I_{n_h} \end{bmatrix}. \quad (3.6)$$

By writing the proposed model using the equations (3.3), (3.5) and (3.6) and setting $\tilde{Z}_{ht} = L_{ht}$, $H_{ht} = 0$, it is easily seen to belong to the class of state-space models presented in section 2.3, with no residual errors in the measurement equation. The model is defined for distinct second level units h but unlike in classical time series analysis where the data consist of a single long series, the data in our case consist of many independent short (longitudinal) series that could be observed over different time periods. Note that the transition matrix, G_h and the covariance matrix, Q_h depend on h through the second level size n_h but they are time invariant. In situations where the second level sizes are not fixed over time (for example, because of missing data), these matrices also change accordingly.

4. Estimation of the model parameters

In principle, the likelihood function holding for the model defined by (3.3), (3.5) and (3.6) can be maximized to obtain the maximum likelihood estimators (MLE) of all the unknown model parameters. However, the number of estimated parameters would usually be very large, which can intensify the computations and result in statistically unstable estimators. For instance, even for $p = q = 2$ and $T = 10$ there are already 46 unknown parameters. We propose therefore a two-stage estimation procedure that employs MLM estimation for the ‘‘cross-sectional parameters’’ and state-space model estimation for the ‘‘time series parameters’’. The use of this procedure has the further advantage of accommodating appropriate weighting to protect against informative sampling.

The procedure starts off by fitting the MLM defined by (3.1) to each sample S_t separately, to obtain IGLS estimates of the time-dependent fixed effects $\beta_t = [\gamma_t', v_t']'$ and the variances of the random effects u_{ht} and e_{hjt} . Notice that by (3.4),

$$\begin{aligned} \text{Var}(u_{ht}) &= \Delta^* = (I - A^2)^{-1}\Delta; \\ \text{Var}(e_{hjt}) &= \sigma_e^2 = (1 - \rho^2) \sigma_\varepsilon^2 \end{aligned} \quad (4.1)$$

using familiar relationships holding for AR(1) models. The use of this step yields estimates $\{\hat{\beta}_t, \hat{\Delta}_t, \hat{\sigma}_{et}^2\}$ for $\{\beta_t, \Delta^*, \sigma_e^2\}$ respectively. Under the model, the true variances (Δ^*, σ_e^2) are fixed over time and assuming that the sample sizes at the various time points are fairly constant, the estimates $\hat{\Delta}_t^*$ and $\hat{\sigma}_{et}^2$ can be averaged to yield single estimates

$$\bar{\Delta}^* = \sum_{t=1}^T \hat{\Delta}_t^* / T; \bar{\sigma}_e^2 = \sum_{t=1}^T \hat{\sigma}_{et}^2 / T. \quad (4.2)$$

In the second stage the remaining parameters are estimated by maximizing the likelihood of the combined model defined by (3.3) (3.5) and (3.6), with the parameters estimated in the first stage held fixed at their estimated values. Since observations on different second level units are independent, the log-likelihood has the form $\log(L) = \sum_h \log(L_h)$ where L_h , the contribution to the likelihood from second level unit h , is defined by (2.6) with the index h added to all the components thus distinguishing between different second level units. As pointed out before, the number of time points for which the second level units are observed and the time periods over which the observations are taken may differ between units so that the notation T in (2.6) for the number of time points needs also to be changed to T_h .

When fitting the model to data obtained from rotating panel sampling designs as in the empirical study of the present paper, a further modification is required to account for the intermediate periods without observations. For example, for the Israel LFS described in the introduction, with rotation pattern of two quarters in the sample, two quarters out of the sample and two quarters in again, $T_h = 4$ but the transition equations from $t = 2$ to $t = 3$ (the next quarter with observations) have to be changed to account for the two quarters with missing observations. Repeated substitutions in (3.5) yield the following relationships:

$$\begin{aligned} \alpha_{h3} &= G_h^3 \alpha_{h2} + \eta_{h3}^*; \eta_{h3}^* \sim N(0, Q_{h3}^*), \\ Q_{h3}^* &= \begin{bmatrix} (A^4 + A^2 + I)\Delta & 0 \\ 0 & (\rho^4 + \rho^2 + 1)\sigma_\varepsilon^2 I_{n_h} \end{bmatrix}. \end{aligned} \quad (4.3)$$

In order to apply the Kalman filter and compute the likelihood, it is needed to set initial values for $\alpha_{1|0}$ and $P_{1|0}$. This is simple under the present model as $\alpha_{ht} = [u'_{ht}, e'_{ht}]'$ is stationary with zero mean and covariance matrix defined by (4.1). Thus, the filter is started by setting,

$$\begin{aligned} \alpha_{h1|0} &= E(u'_{h1}, e'_{h1}) = 0; \\ P_{h1|0} &= \text{Var}[u'_{h1}, e'_{h1}] \\ &= \text{diag}\{(I - A^2)^{-1}\Delta, \sigma_\varepsilon^2(1 - \rho^2)^{-1}I_{n_h}\}. \end{aligned} \quad (4.4)$$

In the empirical study described in the next two sections we compare two methods regarding the set of parameters estimated in the second stage.

Method 1: The parameters estimated in Stage 2 are the three AR coefficients ρ, A_{11}, A_{22} and the corresponding residual variances $\sigma_\varepsilon^2 = \text{Var}(\varepsilon_{hjt})$ and $\Delta = \text{Var}(\delta_{ht})$, (equation 3.6, three variances in total). Note that under this method the only estimates utilized from Stage 1 are the fixed parameter estimates $\{\hat{\beta}_t = [\hat{\gamma}'_t, \hat{v}'_t]'\}$. By (4.1), the variances $\Delta^* = \text{Var}(u_{ht})$ and $\sigma_e^2 = \text{Var}(e_{hjt})$ are estimated as

$$\hat{\Delta}^* = (I - \hat{A}^2)^{-1}\hat{\Delta}; \hat{\sigma}_e^2 = (1 - \hat{\rho}^2)^{-1}\hat{\sigma}_\varepsilon^2. \quad (4.5)$$

Method 2: The only parameters estimated in Stage 2 are the AR coefficients ρ, A_{11}, A_{22} (Equation 3.4). Note that with this method the variances Δ and σ_ε^2 are set in the likelihood as, $\Delta = (I - A^2)\bar{\Delta}^*$ and $\sigma_\varepsilon^2 = (1 - \rho^2)\bar{\sigma}_e^2$ utilizing (4.1), where $\bar{\sigma}_e^2$ and $\bar{\Delta}^*$ are defined by (4.2).

The estimation procedures described so far assume implicitly noninformative sampling. As discussed in the introduction, complex sample surveys often involve selection with unequal probabilities that could be correlated with the values of the response variable. When this is the case, the model holding for the sample data may differ from the model holding in the population. A further advantage of the proposed two-stage estimation method is that it can be adapted to protect against informative sampling. This is done by applying the weighting procedure described in section 2.2 in the first stage, replacing the iterative IGLS algorithm by the PWIGLS procedure. Thus, for each sample S_t , PWIGLS is used for estimating the MLM model parameters instead of using the IGLS.

Comment 1: Informative selection of the first and second level units does not affect the conditional distributions of the random effects as defined by (3.4). Thus, although the distribution of u_{h1} and e_{h1} could be largely distorted because of the sample selection at time $t = 1$, this has no effect on the distributions of $u_{h2}|u_{h1}$, or $e_{h2}|e_{h1}$. The implication of this property is that the computation of the likelihood in the second stage remains the same, but care should be taken of a proper initialization of the Kalman filter. As defined by (4.4), the filter is initialized by the unconditional means and variances of the random effects under the model, but at time $t = 1$ the moments holding for units in the sample can be different because of the sampling effects. As is well known, for long enough series and under

some regularity conditions, the estimates derived from maximization of the likelihood are not sensitive to the initialization procedure but with short series, improper initialization under informative sampling could distort the estimation process. Nonetheless, as illustrated in section 5, having a moderate number of longitudinal observations even of very short length (at most 4 observations in our application) and weighting the likelihood contributions by the inverse of the sample inclusion probabilities (application of the pseudo likelihood approach) yields approximately unbiased estimators for all the time series model parameters.

5. Simulation results

In this section we report the results of a Monte Carlo study carried out for assessing the performance of the various estimation procedures described in section 4 under noninformative and informative rotating sampling schemes.

5.1 Description of simulation study

A) Generation of population data and sample rotation scheme

Population values have been generated for individuals (first level units) within households (second level units), using the model defined by (3.1) and (3.4) (see below). The number of persons n_h observed within household h was selected at random with possible values of 2, 3 or 4. A new panel of households has been generated in each of 11 quarters and a sample of these households has been observed following the Israel Labor Force Survey rotation scheme of two quarters in the sample, two quarters out of the sample and two quarters in again. As easily checked, this process yields a complete sample of four panels in each of the quarters 6-11, with one panel in each quarter observed for the first time, one panel observed for the second time, one for the third time and one for the fourth and last time. (In the first quarter there is only one panel, in the next three quarters there are two panels and in the fifth quarter there are 3 panels.) In what follows we only consider the data observed for quarters 6-11.

B) Population model

The model used for generating the y -values for a given household h is defined by (3.1) and (3.4) with $x'_{hjt} \equiv (x_{hj1}, x_{hj2})$ and $z'_{ht} \equiv (3, z_{h2})$, such that the covariate values are fixed over time. The x -values were generated independently from the uniform distribution $U[1, 2]$. Values z_{h2} were generated from the uniform distribution $U[1, 5]$. In order to simplify the presentation and evaluation of the results, we also set the model coefficients to be time invariant such that $\gamma_t = \gamma = (6, -2)'$ and $v_t = v = (1, 2)'$. The random error terms were generated independently

between households using the model (3.4) with $A = \text{diag}[0.5, 0.7]$, $A = \text{diag}[0.8, 0.5]$, $\rho = 0.4$ and $\sigma_e^2 = 0.25$. Notice from (4.1) that $\text{Var}(u_{ht}) = \Delta^* = \text{diag}[1.067, 0.980]$ and $\text{Var}(e_{hjt}) = \sigma_e^2 = 0.298$.

C) Sample selection

We consider two separate sampling schemes.

C1) Noninformative sampling

Population values have been generated for panels of 30 households, with all the households belonging to a given panel selected to the sample and observed following the sample rotation scheme described in A above. The total number of sampled households in each of the quarters 6-11 is therefore $m = 120$. All the individuals belonging to a given household have been observed, yielding an expected sample size of $n = 360$ individuals for each of the quarters. This sampling scheme corresponds to simple random sampling of households and individuals within the selected households.

C2) Informative sampling

Population values have been generated for panels of 55 households. Households with random effects $u_{h1,1} < 0$ (the value of the first random effect at the first time point) have been sampled with probability 1, households with random effects $u_{h1,1} > 0$ have been sampled independently (Poisson sampling) with probability 0.1. All the individuals belonging to a sampled household have been observed. This sampling scheme yields an expected sample size of approximately 30 households per panel and expected sample sizes of approximately $m = 120$ households and $n = 360$ individuals per quarter, similarly to the sampling scheme C1.

Comment 2: It should be emphasized that even though there are 4 panels observed in each of the quarters 6-11, there are only 11 separate panels that are used for estimation of the model parameters. Moreover, out of the 11 panels, only the panel entering the sample in quarter 6 for the first time is observed in 4 quarters, only 2 panels are observed in 3 quarters, 6 panels are observed in 2 quarters and 2 panels are observed in only one quarter. This implies a total of 13 panels transitions, with about 390 household transitions observed for estimation of the time series parameters. (By a panel transition we mean that the same panel is observed on two occasions. For 3 of these panel transitions there is a time gap of 2 quarters between the two observations). We refer to this sample structure when assessing the estimation of the time series model parameters.

The whole process of generating population values and selecting the sample has been repeated 100 times for each of the two sampling schemes C1 and C2, with one sample selected from each population. For each sample we applied

the two estimation procedures described in section 4. The simulations were run using the Gauss software package. Maximization of the likelihood has been carried out using the numerical optimization procedure, OPTMUM.

5.2 Results

The results of the simulation study are summarized in Tables 1-4 as averages over the 100 samples selected under the two sampling schemes. Each table contains the mean estimates of the model parameters, the empirical standard deviations (SD) of the estimators and the conventional *t*-statistics obtained by dividing the difference between the mean estimates and the true parameter values by the standard errors (SE), computed as SD/10. Notice that the estimates of the fixed vector coefficients $\beta_t = (\gamma_t', v_t)'$ are the same under the two estimation methods.

Perhaps the most important outcome of this study, revealed from Table 1, is that under noninformative sampling it is indeed possible to fit successfully simple but nontrivial time series models to very short longitudinal series, provided that the number of observed series is sufficiently large. (The model is not trivial because even after subtracting the fixed effects, the dependent response variable is the sum of three AR(1) processes.) This conclusion is further strengthened by the fact that 8 out of the 11 panels have been observed for at most 2 times, yielding a total of 13 panel transitions, three of which with a gap of 2 quarters. See Comment 2 at the end of section 5.1.

Table 1
Means, Standard Deviations (SD) and *t*-statistics of estimators under two estimation methods. Noninformative sampling

Parameter	True Value	Method 1			Method 2		
		Mean	SD	<i>t</i> -statistic	Mean	SD	<i>t</i> -statistic
γ_1	6.000	6.002	0.03	0.677	6.002	0.03	0.677
γ_2	-2.000	-2.000	0.03	0.078	-2.000	0.03	0.078
v_1	1.000	0.989	0.08	-1.357	0.989	0.08	-1.357
v_2	2.000	2.008	0.08	0.997	2.008	0.08	0.997
A_{11}	0.500	0.497	0.07	-0.391	0.491	0.07	-1.271
A_{22}	0.700	0.696	0.07	-0.532	0.695	0.07	-0.820
Δ_{11}^*	1.067	1.054	0.08	-1.668	1.045	0.08	-2.677
Δ_{22}^*	0.980	0.991	0.10	1.042	0.990	0.11	0.906
ρ	0.400	0.398	0.02	-0.937	0.397	0.02	-1.637
σ_e^2	0.298	0.298	0.01	-0.062	0.297	0.01	-1.382

Evaluation of the performance of the two sets of estimators in Table 1 shows that all the estimators under Method 1 are highly insignificant based on the conventional *t*-statistics and only the estimator of Δ_{11}^* is significant under Method 2. Note that even in that case the absolute relative bias is about 2% and considering that MLE of time series parameters are generally not strictly unbiased, such a small bias in one of 10 parameters is expected. Notice also that the standard errors of the mean estimators under the two methods are very similar, a result observed also in the other tables.

Next we consider the case of informative sampling. Table 2 shows the results obtained when ignoring the informative sampling process, using the same estimation procedures as used for the noninformative case. As indicated very clearly by this table, some of the parameter estimates are highly significant, particularly the estimators of the parameters indexing the time series model of the random effects u_{hit} that define the sample selection probabilities. Thus, we find that the absolute relative bias in estimating v_1 is about 27%, and large absolute relative biases are also observed for the estimators of A_{11} and Δ_{11}^* . (The model defined by (3.1) can be rewritten as $y_{hjt} = x'_{hjt} \gamma_t + z'_{ht} u_{ht}^* + e_{hjt}$ where $u_{ht}^* = u_{ht} + v_t$, such that for $v_t \equiv v$ as under the simulation model, $v_1 = E(u_{hit}^*)$). Note that the three biases are negative, which is explained by the fact that the selection mechanism utilized for this study oversamples individuals with observations that contain negative random effects u_{hit} . In this case again, the two estimation methods perform very similarly.

Table 2
Means, Standard Deviations (SD) and *t*-statistics of estimators under two estimation methods. Informative sampling, unweighted estimators

Parameter	True Value	Method 1			Method 2		
		Mean	SD	<i>t</i> -statistic	Mean	SD	<i>t</i> -statistic
γ_1	6.000	5.998	0.02	-0.768	5.998	0.02	-0.768
γ_2	-2.000	-2.000	0.03	0.104	-2.000	0.03	0.104
v_1	1.000	0.728	0.09	-34.385	0.728	0.09	-34.385
v_2	2.000	2.005	0.09	0.564	2.005	0.09	0.564
A_{11}	0.500	0.438	0.09	-6.742	0.434	0.09	-7.453
A_{22}	0.700	0.738	0.09	4.078	0.735	0.09	3.941
Δ_{11}^*	1.067	0.995	0.09	-7.766	0.994	0.09	-7.883
Δ_{22}^*	0.980	1.003	0.10	2.352	0.987	0.10	0.698
ρ	0.400	0.407	0.02	3.184	0.405	0.02	2.218
σ_e^2	0.298	0.298	0.01	0.644	0.296	0.01	-1.800

Table 3 shows the results obtained when using the PWIGLS algorithm for the estimation of the MLM parameters (section 2.2) and weighting the time series likelihood contributions $\log(L_h) = -\{1/2 T_h n_h \log(2\pi) + 1/2 \sum_{t=1}^{T_h} \log|F_{ht}| + 1/2 (Y_{ht} - \hat{Y}_{ht|t-1})' F_{ht}^{-1} (Y_{ht} - \hat{Y}_{ht|t-1})\}$ by the household sampling weights $w_h = 1/\text{Pr}(h \in s)$, using the same 100 samples as used for Table 2. Weighting the likelihood contributions by the inverse of the sample inclusion probabilities is an application of the pseudo likelihood approach that is often recommended for fitting single level models to cross-sectional data, see, e.g., Binder (1983), Skinner *et al.* (1989) and Pfeiffermann (1993). As revealed from this table, the use of the PWIGLS algorithm and weighting the likelihood eliminates the large biases observed in Table 2, despite the improper initialization of the Kalman filter with very short series. (See the discussion in Comment 1 at the end of section 4.) Here again, the two estimation methods perform quite similarly, yielding one

biased estimator in each case but with both biases being relatively very small.

It is important to mention that the SD's of the weighted estimators shown in Table 3 are always larger than the corresponding SD's of the unweighted estimators displayed in Table 2. As pointed out by one of the referees, this implies that the empirical root mean square errors (RMSE's) of the unweighted estimators in Table 2 are in fact larger than the empirical RMSE's of the corresponding estimators in Table 3. This outcome, however, is due to the relatively small sample sizes employed in this study. For larger samples (larger numbers of households and individuals within the households) the RMSE is dominated by the bias which, unlike the variance, is not reduced as the sample size increases. Thus, it is clear that as the sample size increases the RMSE's of the weighted estimators become smaller than the RMSE's of the unweighted estimators. The fact that probability weighted estimators have larger variances than the corresponding unweighted estimators is well known from many other studies, see Pfeffermann (1993) for discussion and references.

Table 3
Means, Standard Deviations (SD) and t -statistics of estimators under two estimation methods. Informative sampling, weighted estimators

Parameter	True Value	Method 1			Method 2		
		Mean	SD	t -statistic	Mean	SD	t -statistic
γ_1	6.000	5.997	0.04	-0.607	5.997	0.04	-0.607
γ_2	-2.000	-2.000	0.05	-0.007	-2.000	0.05	-0.007
v_1	1.000	0.978	0.14	-1.518	0.978	0.14	-1.518
v_2	2.000	2.019	0.14	1.330	2.019	0.14	1.330
A_{11}	0.500	0.490	0.15	-0.695	0.477	0.14	-1.611
A_{22}	0.700	0.699	0.17	-0.066	0.709	0.16	0.545
Δ_{11}^*	1.067	1.055	0.17	-0.664	1.040	0.17	-1.560
Δ_{22}^*	0.980	1.023	0.19	2.199	1.010	0.19	1.571
ρ	0.400	0.401	0.04	0.135	0.397	0.04	-0.813
σ_e^2	0.298	0.297	0.01	-0.486	0.294	0.01	-3.340

As discussed in Comment 1 at the end of section 4, informative sampling distorts the cross-sectional distribution of the sample observations and the initialization of the Kalman filter, but does not affect the conditional distributions of the first and second level random effects defined by (3.4). Thus, it is interesting to test whether the use of the PWIGLS algorithm for estimating the cross-sectional model parameters but without weighting the time series likelihood likewise controls the bias. Table 4 shows there results obtained for this case with the same samples as used for Tables 2 and 3. The estimators of the fixed vector coefficients $\beta_t = (\gamma_t', v_t')'$ are the same as in Table 3 and hence are not shown again. Notice that the estimators of Δ_{11}^* , Δ_{22}^* and σ_e^2 under Method 2 are also the same as the corresponding estimators in Table 3.

The interesting result revealed from Table 4 is that the estimators of A_{11} and A_{22} have now a non-negligible bias,

unlike the corresponding estimators in Table 3. This result can be explained as follows. Under the informative sampling scheme, the expectation of the random effects $u_{h,1}$ corresponding to households h in the sample is below zero, $E(u_{h,1}|h \in s) < 0$, and hence the initialization of the Kalman filter by the population expectation ($Eu_{h,1} = 0$, Equation 4.4) yields biased estimators. On the other hand, by weighting the likelihood contributions L_h by the inverse of the sample selection probabilities, the proportions of likelihoods L_h corresponding to random effects that are below and above the model expectation is balanced to the population proportions and thus the use of the model expectation for the initialization process does not bias the estimation process. As noticed for the previous tables, the SD's of the unweighted estimators in Table 4 are much smaller than the SD's of the corresponding weighted estimators in Table 3.

Table 4
Means, Standard Deviations (SD) and t -statistics of estimators under two estimation methods. Informative sampling, weighted MLM, unweighted likelihood

Parameter	True Value	Method 1			Method 2		
		Mean	SD	t -statistic	Mean	SD	t -statistic
A_{11}	0.500	0.468	0.09	-3.477	0.453	0.10	-4.569
A_{22}	0.700	0.742	0.11	3.948	0.737	0.11	3.197
Δ_{11}^*	1.067	1.060	0.11	-0.598	1.040	0.17	-1.560
Δ_{22}^*	0.980	1.008	0.11	2.449	1.010	0.19	1.571
ρ	0.400	0.407	0.02	3.021	0.402	0.02	0.894
σ_e^2	0.298	0.298	0.01	1.013	0.294	0.01	-3.340

6. Application of the model to LFS data

We fitted the model defined by (3.1) and (3.4) to an empirical data set extracted from data collected by the Israel LFS for Jerusalem during the years 1990-1994. The data contain complete records for 567 individuals in 475 households, with each individual observed in four quarters according to the rotation pattern described before and used for the simulation study. Out of the 475 households, 385 have one individual record, 88 have 2 individual records and only 2 households have 3 individual records. The outcome variable is y = number of hours worked during the week preceding the interview, ($\bar{y} = 39.8$, $sd(y) = 14.8$; calculated over all individuals and all the quarters). The individual level auxiliary variables are x_1 = years of education, ($\bar{x}_1 = 13.4$, $sd(x_1) = 4.8$) and x_2 = gender, (41% females). The household level auxiliary variables are $z_1 = 1$ and z_2 = number of employed persons in the household ($\bar{z}_2 = 1.48$, $sd(z_2) = 0.56$).

We estimated the model parameters using the two methods described in section 4. The sampling weights attached to these data are very similar across households and individuals so that we only computed the unweighted

estimators. The GLS algorithm produced negative variance estimates for Δ_{22}^* in some of the quarters and these estimates have been set to zero when averaging the variance estimates under Method 2. The quarterly estimates of the fixed model coefficients have not been averaged as they change significantly over the five years period.

The estimates computed by the two methods for the variances and autoregression coefficients are shown in Table 5 using the same notation as in the previous tables. The two sets of estimates are not very far except for the estimator of Δ_{22}^* which, already mentioned was found to be negative in some of the separate IGLS runs. Note in this respect that for most of the households there is only a single individual record (see above), and that for almost all of these households $z_2 = 1$. This complicates the estimation process since for such households it is impossible to distinguish the first (individual) level effect from the two household effects, which are likewise confounded. (Note that the sum of the latter two variances is similar under the two methods.) As discussed below, the estimators in Table 5 are dominated by the observations obtained for households with two individual records.

Table 5

Estimates of variances and autoregression coefficients under two estimation methods. LFS data

Parameter	A_{11}	A_{22}	Δ_{11}^*	Δ_{22}^*	ρ	σ_e^2
Method 1	0.915	-0.606	73.88	2.541	0.242	102.306
Method 2	0.976	-0.548	56.88	14.753	0.448	101.001

Under the Israel LFS sampling design, each individual record consists of 4 observations taken in quarters 1, 2, 5 and 6, with quarter 1 defining the first calendar quarter t that the individual is in the sample. In order to assess the prediction power of the model, we computed for every individual record (h, j) the empirical innovations when predicting the adjusted values $r_{hj,q} = (y_{hj,q} - x'_{hj,q} \hat{\gamma}_q - z'_{hq} \hat{v}_q)$ using the household data observed for the preceding quarters that the individual has been in the sample. Note that by subtracting the fixed effects from the original observations, the distribution of the adjusted values no longer depends on the calendar quarters. The innovation for quarter q is the corresponding prediction error which, by (3.1) is computed as $d_{hj,q} = (r_{hj,q} - z'_{hq} \hat{u}_{hq|q-m} - \hat{e}_{hj,q|q-m}) = r_{hj,q} - (z'_{hq}, 1)' \hat{\alpha}_{q|q-m}, q = 2, 5, 6$ where $\hat{\alpha}_{q|q-m}$ is the predictor of the state vector $\alpha_q = (u'_{hq}, e_{hj,q})'$ using the data observed until quarter $q - m$, with $m = q - 1$ for $q = 2, 6$ and $m = 3$ for $q = 5$. The predictor $\hat{\alpha}_{q|q-m}$ is obtained by application of the Kalman filter with the corresponding estimated parameters (see section 2.3 and Equations 3.5 and 4.3).

Table 6 shows the roots of the means of the square innovations (RMSI) by quarter and the number of household (HH) records, as obtained under the two estimation methods (using the parameter values displayed in Table 5).

For comparison, we also show the RMSI's of the innovations obtained by predicting the adjusted value for quarter q by the adjusted value in the preceding quarter. The "naive" predictor $\hat{r}_{hj,q} = r_{hj,q-m}$ can be interpreted as being the optimal predictor under the simple random walk model $r_{hj,q} = r_{hj,q-m} + \text{error}$. The means of the innovations ($r_{hj,q} = r_{hj,q-m}$) for $q = (2, 5, 6)$ are (0.68, 0.24, 0.301) for households with one record, (1.24, -1.20, 0.60) for households with two records and (4.02, -5.82, 7.68) for households with 3 records but recall that the latter means are based on only 2 households. The corresponding means of the empirical innovations computed under the model are smaller in absolute value in all the cases.

Table 6

Root mean square of innovations by number of household records and quarter under two estimation methods and naive prediction. LFS data

HH Records	1			2			3		
Quarter	2	5	6	2	5	6	2	5	6
Method 1	11.54	11.16	11.62	12.26	11.71	10.88	9.61	9.98	8.94
Method 2	11.71	11.16	11.49	12.10	11.48	10.91	9.30	9.78	7.90
Naive Pred.	14.00	11.92	13.60	14.71	15.12	13.47	7.50	13.32	11.29

The data analyzed in this section behave much more erratically than the data used for the simulation study generated under the model and we cannot claim that the model employed yields the best possible fit (see also below). Nonetheless, the values displayed in Table 6 illustrate some important features of the model. We mention first the generally much better performance of the model predictors compared to the naive predictor $\hat{r}_{hj,q} = r_{hj,q-m}$ with the two estimation methods yielding similar RMSI's. The superiority of the model is explained by the fact that whereas the first order autocorrelations of the two random household effects used for the model predictions are high in absolute value (very high for the first component), the autocorrelations of the adjusted values (the "total" errors) are only of moderate size. The first order autocorrelations of the random components are the corresponding autoregression coefficients, see Table 5. The empirical autocorrelations of the adjusted values, $\text{Corr}(\hat{r}_{hj,q}, \hat{r}_{hj,q-m}), q = 2, 5, 6; m = 1$ for $q = 2, 6; m = 3$ for $q = 5$ are correspondingly (0.46, 0.59, 0.51) for one record households, (0.48, 0.36, 0.45) for two record households and (0.92, 0.43, 0.63) for three record households (based on 6 individual records).

As already noted, the fact that most households have only one individual record introduces identifiability problems since for such households it is impossible to distinguish between the three random effects. Computation of the correlations $\text{Corr}(\hat{r}_{hj,q}, \hat{r}_{hj,q-m})$, under the model using the parameter estimates in Table 5 shows a good fit to the correlations computed for two record households. This in turn illustrates that the estimators in Table 5 are

dominated by these observations and we conclude that the model fits best the observations obtained for the households with two records. Note, however, that the RMSI's obtained for the other household sizes are not higher than the RMSI's computed for the two record households (see also below). It is important to mention in this regard that if the data had been aggregated over all the individuals observed in a given calendar quarter, it would have been impossible to account for the random household effects, resulting in inferior predictions of the individual observations. See the discussion in the introduction. (Modelling the aggregate data is rather complicated in this case since the sample in each calendar quarter consists of 4 different panels as defined by the number of times that individuals are in the sample. This implies that the models holding for these panels are different, depending on the number of observations available for each panel.)

Other interesting results noted in table 6 are that the RMSI's under the model are generally lower for $q = 6$ than for $q = 2$, as explained by the use of more observed data for the same individual in the prediction process (more observed data for estimating the random effects in the preceding quarter). Also, for $q = 6$ the RMSI's decrease as the number of household records increases, as explained by the use of data observed for other household members. Finally, the RMSI's for households with 3 records are much lower by use of the model than the RMSI's obtained for households with 1 and 2 records but we mention again that there are only 2 households with three records. The unexpected results in Table 6 are that for households with one record the RMSI's are somewhat larger for $q = 6$ than for $q = 5$ (note the relatively high and unexplained correlation of 0.59 between the adjusted values 3 quarters apart computed for these households), and that for $q = 2$ and $q = 5$ the RMSI's for households with 2 records are larger than the corresponding RMSI's for households with 1 record. With empirical data of relatively small size such anomalies are not unusual and they show up even more prominently with the naive predictor. (The fact that for a given number of household records the RMSI's by use of the model for $q = 5$ are of similar magnitude to the other RMSI's is reassuring given that the predictions in this case are 3 quarters ahead.)

7. Conclusions and model extensions

The results of this paper illustrate that it is possible to fit time series models to longitudinal series of very short length and with missing observations. The model used in the present study is an extension of the standard two level linear model by which both the first and second level random effects evolve stochastically over time. This kind of model is suitable for modelling longitudinal measurements that are taken for hierarchical populations. Application of the PWIGLS algorithm combined with standard probability

weighting of the time series likelihood is shown to protect against the effects of informative sampling.

Multilevel models are often fitted to discrete data, in which case the models contain nonlinear components. In principle, the two-stage estimation method proposed in this paper can be applied in this case as well, although with very short longitudinal series the range of models that can be fitted is obviously limited. Moreover, a common procedure for estimating the unknown model parameters in the discrete case consists of linearizing the nonlinear components on each iteration of the IGLS around estimates obtained on the previous iteration, and then applying the standard IGLS for computing the revised estimates. See Goldstein (1995) for details. Thus, it seems feasible to extend the PWIGLS algorithm to the discrete case without major difficulties.

In this paper we have not considered variance estimation. This is no problem under the standard IGLS and Pfeffermann *et al.* (1988) propose simple variance estimators for the PWIGLS procedure. However, estimation of the variances of estimators obtained from maximization of the time series likelihood is more problematic because of two reasons. First, the possibly short length of the longitudinal series may no longer justify the use of the information matrix or permit stable estimation thereof, even with large number of second level units. Second, the MLM estimators are held fixed when maximizing the likelihood, implying that the MLE abstract from the sampling errors in the estimation of the MLM parameters. A possible solution to this problem is the use of re-sampling methods that allow to account for all sources of variation in the estimation process.

Finally, we mention an important application of the proposed model for the imputation of missing data. In a recent article, Pfeffermann and Nathan (forthcoming) illustrate the large reductions in the imputation variance that can be achieved under the model compared to the use of more standard imputation methods that ignore the common household effects.

References

- Belcher, J., Hampton, J.S., and Tunnicliffe Wilson, G. (1994). Parameterization of continuous time autoregressive models for irregularly sampled time series data. *Journal of the Royal Statistical Society, Series B*, 56, 141-155.
- Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51, 279-292.
- Binder, D.A. (1998). Longitudinal surveys: Why are these surveys different from all other surveys? *Survey Methodology*, 24, 101-108.
- Bryant, J., and Day, R. (1991). Empirical Bayes analysis for systems of mixed models with linked autocorrelated random effects. *Journal of the American Statistical Association*, 86, 1007-1012.

- Chi, E.M., and Reinsel, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association*, 84, 452-459.
- Diggle, P.J., Liang, K.Y., and Zeger, S.L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.
- Duncan, G.J., and Kalton, G. (1987). Issues of design and analysis of surveys across time. *International Statistical Review*, 55, 97-117.
- Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, 73, 43-56.
- Goldstein, H. (1995). *Multilevel Statistical Models* (2nd edition). New York: Halstead.
- Goldstein, H., Healy, M.J.R., and Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, 13, 1643-1655.
- Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. New York: Cambridge University Press.
- Herriot, R.A., and Kasprzyk, D. (1984). The survey of income and program participation. *Proceedings of the Social Statistics Section*, American Statistical Association, 107-116.
- Jones, R.H., and Ackerson, L.M. (1990). Serial correlation in unequally spaced longitudinal data. *Biometrika*, 77, 721-731.
- Jones, R.H., and Boadi-Boating, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics*, 47, 161-175.
- Jones, R.H., and Vecchia, A.V. (1993). Fitting continuous ARMA models to unequally spaced spatial data. *Journal of the American Statistical Association*, 88, 947-954.
- Jones, R.H. (1993). *Longitudinal Data with Serial Correlation. A State-space Approach*. New York: Chapman and Hall.
- Lawless, J.F. (1999). Event History Analysis and Longitudinal Surveys. Paper presented at the Conference on Analysis of Survey Data, Southampton, United Kingdom.
- Nathan, G. (1999). A Review of Sample Attrition and Representativeness in Three Longitudinal Surveys. GSS Methodology Series No. 13. London Office of National Statistics (ONS), United Kingdom.
- Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review*, 61, 317-337.
- Pfeffermann, D. (1996). The use of sampling weights for survey data analysis. *Statistical Methods in Medical Research*, 5, 239-261.
- Pfeffermann, D., Skinner, C.J., Goldstein, H., Holmes, D.J., and Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models (with discussion). *Journal of the Royal Statistical Society, Series B*, 60, 23-40.
- Pfeffermann, D., and Nathan, G. (forthcoming). Imputation for wave nonresponse: existing methods and a times series approach. To appear in: *Survey Nonresponse*, (Eds, R.M. Groves, D. Dillman, J.L. Eltinge, and R.J.A. Little). New York: John Wiley & Sons, Inc.
- Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 175-187.
- Sallas, W.H., and Harville, D. A. (1981). Best linear recursive estimation for mixed linear models. *Journal of the American Statistical Association*, 76, 860-869.
- Skinner, C.J., Holt, D., and Smith, T.M.F. (Eds.) (1989). *Analysis of Complex Surveys*. Chichester: Wiley.
- Skinner, C.J., and Holmes, D. (1999). Random Effects Models for Longitudinal Survey Data. Paper presented at the Conference on Analysis of Survey Data, Southampton, United Kingdom.
- Survey Research Center (1984). User Guide to the Panel Study of Income Dynamics. Ann Arbor, Michigan: Inter-university Consortium for Political and Social Research.
- Webber, M. (1994). The survey of labor and income dynamics: Lessons learned in testing. *Proceedings of the Annual Research Conference, US Bureau of the Census*, 85-99.
- Zimmerman, D.L., and Nunez-Anton, V. (1997). Structured antedependence models for longitudinal data. In: *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions*, (Eds. T.G. Gregoire et al.). Lecture Notes in Statistics, 22. New York: Springer Verlag, 62-76.