Quality & Quantity (2005) 39: 155–174
DOI 10.1007/s11135-004-1673-x
© Springer 2005

"Panelizing" Repeated Cross Sections

Female Labor Force Participation in The Netherlands and West Germany

BEN PELZER[1], ROB EISINGA[1] and PHILIP HANS FRANSES[2]
[1]*Department of Social Science Research Methods, University of Nijmegen, PO Box 9108, 6500 HE Nijmegen, The Netherlands. E-mail: b.pelzer@maw.kun.nl, E-mail: r.eisinga@maw.kun.nl.;* [2]*Econometric Institute, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. E-mail: franses@few.eur.nl*

Abstract. This paper considers the implementation of a non-stationary, heterogeneous Markov model for the analysis of binary dependent variables in a time series of repeated cross-sectional (RCS) surveys. The model offers the opportunity to estimate entry and exit transition probabilities and to examine the effects of time-constant and time-varying covariates on the hazards. We show how maximum likelihood estimates of the parameters can be obtained by Fisher's method-of-scoring and how to estimate both fixed and time-varying covariate effects. The model is exemplified with an analysis of the labor force participation decision of Dutch and West German women using ISSP (and other) data from 10 annual Dutch surveys conducted between 1987 and 1996 and 7 annual West German surveys conducted between 1988 and 1994. Some open problems concerning the application of the model are discussed.

Key words: cross-national surveys, repeated cross-sectional surveys, Markov model, pseudo panel data, The Netherlands, (West-) Germany.

1. Introduction

In the past few decades there has been a considerable expansion in the availability of repeated cross-sectional (RCS) surveys. Some important examples include the General Social Survey, the European Value Survey, and the International Social Survey Program (ISSP). This accumulation not only provides researchers with a growing opportunity to analyze over-time change but also raises questions about new analytic methodology for exploiting the properties of RCS data for longitudinal study.

Repeated cross-sectional data contain information on different cross-sectional units (typically individuals) independently drawn from the same population at multiple points in time and aim to provide a representative cross section of the population at each sample point. A limitation of this type of data for longitudinal research is that the sample units are not retained from one time period to the next. RCS data are therefore, in the context of dynamic modeling, generally regarded as inferior to genuine panel data, that is, repeated observations on the same individual units across occasions. An important advantage to using a matched panel file is that it provides a measure of gross individual change for each sample unit and that it enables us to use each unit as its own control. Panel data, however, may also be inferior to the available cross sections in terms of sample size, representativeness, and time period covered. The size of a panel is commonly reduced over time by the process of selective attrition, which may create serious biases in the analysis. Especially in the case of long-term panel surveys the panel may become unrepresentative as time proceeds. Moreover, logistical constraints often preclude tracking individual units through long periods of time, so that analyzing rolling cross-sectional data for the assessment of long-run change is the best one can do.

In this paper we discuss, for the case of binary dependent variables, a dynamic model originally considered by Moffitt (1990, 1993) that permits the identification and estimation of entry and exit transition rates from a time series of RCS samples. The model also offers the opportunity to examine the effects of covariates on the hazards. In doing so, we extend the framework put forth by Moffitt on two points: (i) a procedure is derived to obtain maximum likelihood (ML) estimates of the parameters and their dispersion, and (ii) the time-constant coefficient model is expanded to also incorporate time-varying coefficients. The proposed model is likely to be useful to researchers seeking to explain over-time change at the micro level in the absence of microlevel data. It should equally be of interest to researchers whose concern resides with the explanation of macrolevel trends as it reveals to them the microlevel contours underlying such trends.

The paper is organized as follows. Section 2 presents the model, discusses the ML estimation of the parameters and gives additional extensions and refinements. We then provide an example application [1] using a time series of cross-sectional data on female labor force participation taken from the Dutch and German omnibus surveys that incorporated the ISSP modules, i.e., the Dutch CULTURAL CHANGES surveys by the SCP and the German ALLBUS surveys by ZUMA and ZA. The paper concludes with some open problems requiring further study.

2. Dynamic Model for RCS Data

The problem of analyzing repeated cross-sectional data has attracted increasing attention in econometrics and other disciplines in the past several years. One class of models considered is the linear fixed effects model (Deaton, 1985; Nijman and Verbeek, 1990; Verbeek, 1996; Verbeek and

Nijman, 1992; 1993; Baltagi, 1995; Collado, 1997). In this approach individual observations are grouped into cohorts based on a time-invariant characteristic (typically date of birth) which results in a so-called pseudo panel with cohort aggregates. The studies are concerned with the conditions under which we can validly ignore the cohort nature of the averaged data and treat the pseudopanel of cohorts as if it were a panel of individuals. Moffitt (1993) has generalized this approach by considering models with a more dynamic structure and binary dependent variables. In his method actual grouping of the data into cohorts need not be done and the variation in the micro-data is utilized as part of the analytic procedure. This section discusses and elaborates his method. It is assumed in the sequel that the responses are observed at equally spaced discrete time intervals $t = 1, 2, \ldots$ and that the samples at periods t_j and t_k are independent if $j \neq k$. The symbol *it* is commonly used to indicate repeated observations on the same sample element *i*. As there can be no misunderstanding, this paper also uses the symbol *it* to index individuals in RCS samples.

2.1. FIRST-ORDER MARKOV MODEL

Suppose, for the moment, that we have a multinomial distribution with probabilities

$$
\begin{array}{ccc}
 & & y_{it} & \\
 & & \begin{array}{cc} 0 & 1 \end{array} & \\
y_{it-1} & \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} & \begin{array}{c} p_{0+} \\ p_{1+} \end{array} \\
 & & \begin{array}{cc} p_{+0} & p_{+1} \end{array} & 1
\end{array}
$$

Obviously, this distribution is only observed with panel data and not with a series of cross-sectional samples. If we define the cell probabilities so that they sum to unity across rows and set, $\mu_{it} = p_{01}/p_{0+},\ 1 - \mu_{it} = p_{00}/p_{0+},\ \lambda_{it} = p_{10}/p_{1+},$ and $1 - \lambda_{it} = p_{11}/p_{1+}$, then the matrix becomes

$$
\begin{array}{ccc}
 & & y_{it} \\
 & & \begin{array}{cc} 0 & 1 \end{array} \\
y_{it-1} & \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1 - \mu_{it} & \mu_{it} \\ \lambda_{it} & 1 - \lambda_{it} \end{pmatrix}.
\end{array}
$$

This expression is a two-state first-order Markov matrix of transition rates that records the probabilities of making each of the possible transitions from one time period to the next; e.g., μ_{it} represents the probability that the unit satisfying $y_i = 0$ at time $t-1$ subsequently satisfies $y_i = 0$ at time t. Note that the Markov process assumes that the underlying process of change can be described in terms of one-step transitions, i.e., the probability of occupying a state at time t depends only on the state occupied at time $t-1$. This first-order assumption implies that the dependency between successive transitions can be eliminated by conditioning on the previous state. Operationally this can be achieved, as we will show, by including the previous state in the model as a covariate predicting y_{it}. Also note that, if we let

$$p_{it} = P(Y_{it} = 1), \quad \mu_{it} = P(Y_{it} = 1 | Y_{it-1} = 0), \quad \lambda_{it} = P(Y_{it} = 0 | Y_{it-1} = 1),$$

then we have

$$E(Y_{it}) = p_{it} = \mu_{it}(1 - p_{it-1}) + (1 - \lambda_{it})p_{it-1} = \mu_{it} + \eta_{it}\, p_{it-1}, \tag{1}$$

where $\eta_{it} = 1 - \lambda_{it} - \mu_{it}$ As noted by Moffitt (1990), the accounting identity in Eq. (1) is the critical equation for estimating dynamic models with repeated cross-sectional samples as it relates the marginal probabilities p_{it} at t and p_{it-1} at $t-1$ to the probabilities of inflow (μ_{it}) and outflow (λ_{it}) from each of the two states. Obviously, the difficulty with using cross-sectional surveys is that the state-to-state transitions over time for each sample unit are not observed, but rather one observes at each of a number of times a distinct cross section of units and their current states. And it is immediately obvious that the hazards in (1) are not identified given only the marginal probabilities.[2] This implies that identification of the unobserved transitions over time in RCS data is only possible with the imposition of certain restrictions over i and/or t.

A popular restriction is to assume that the transition probabilities are the same during the period of time under consideration and that the individuals are in a steady state. Then the Markov process is said to have time-stationary and unit-homogeneous transition probabilities, hence $\mu_{it} = \mu$ and $\lambda_{it} = \lambda$ for all i and t. Using $\eta = 1 - \lambda - \mu$, it is easy to show that the long-run outcome of the t sets of successive transitions is $p_t = (\mu/(\mu + \lambda))(1 - \eta^t) + \eta^t p_{i0}$, which collapses to $p_t = \mu/(\mu + \lambda)$ as t goes to infinity.[3] This limiting result gives the long-run probability of being in a state. That is, for a time point sufficiently far in the future the probability is $\mu/(\mu + \lambda)$ that the state is '1'. Note that this probability does not depend on the initial probability p_{i0}. Hence there is a tendency as time passes for the probability of being in a state to be independent of the initial condition. Moreover, as Moffitt (1993) has argued, the initial probability refers to the value of the state prior to the beginning of the Markov process, for example the state of being unemployed at the beginning of an unemployment spell, rather than to the first observed outcome (which is p_{i1}). It is therefore assumed in many applications to finite-horizon situations that $p_{i0} = 0$ (see, e.g., Bishop et al., 1975). This time-invariant steady state model is the standard approach to the problem of estimating transition rates from aggregate frequency data in the statistical literature (see, e.g., Lee et al., 1970; Firth, 1982; Kalbfleish and Lawless, 1984; 1985; Lawless and McLeish, 1984; Li and Kwok, 1990; Hawkins, et al., 1996). The formulation has been applied in several economic studies, for example, by Topel (1983) in his study on employment duration and by McCall (1971) in his Markovian analysis of earnings mobility. Similar uses occur in the social science literature on intra-generational job mobility processes where it has come to be know as the "mover-stayer" model (see, e.g., Goodman, 1961; Bartholomew, 1996).

Because the assumption of stationarity and homogeneity is generally not plausible and frequently violated in applications (see, e.g., McFarland, 1970), it is desirable to relax this restriction. If we define the model as in Eq. (1) and let $p_{i0} = 0$ (or $t \to \infty$), it may be verified that p_{it} has the representation

$$p_{it} = \mu_{it} + \sum_{\tau=1}^{t-1} \mu_{i\tau}\Big(\prod_{s=\tau+1}^{t} \eta_{is}\Big), \tag{2}$$

where $\eta_{is} = 1 - \lambda_{is} - \mu_{is}$.[4] This reduced form equation for p_{it} accounts for time-dependence and heterogeneity in a flexible manner and it will therefore be maintained in the ensuing method.

To estimate the model in (2) with RCS data, Moffitt (1990; 1993) uses an instrumental variable estimation procedure. While repeated cross-sections lack direct information on the individual transition probabilities, they often do provide a set of time-invariant or time-varying covariates X_{it} that affect the hazards. The history of these covariates $(X_{it}, X_{it-1}, \ldots, X_{i1})$ can be employed to generate backward predictions for the transition probabilities $(\mu_{it}, \mu_{it-1}, \ldots, \mu_{i1}$ and $\lambda_{it}, \lambda_{it-1}, \ldots, \lambda_{i2})$ and thus for the marginal probabilities $(p_{it}, p_{it-1}, \ldots, p_{i1})$. Hence the basic idea is to model the current and past μ_{it}'s and λ it's in a regression setting as functions of current and backcasted values of time-invariant and time-varying covariates X_{it}. Parameter estimates of the covariates are thereupon obtained by substituting the hazard functions into Eq. (2). Of course, this estimation procedure can only be applied if an instrument for y_{it-1} can be constructed, that is, if one has available a vector of time-invariant or time-varying variables X_{it} which affect the transition probabilities. Moreover, the model can be validly estimated provided we assume that measured explanatory variables capture the differences between individuals that affect the hazards.

A common specification for the hazard functions uses a separate binary logistic regression for $P(Y_{it} = 1 | Y_{it-1} = y_{it})$, $y_{it} = 0, 1$. That is, we assume that

$$\operatorname{logit} P(Y_{it} = 1 | Y_{it-1} = 0) = \operatorname{logit}(\mu_{it}) = X_{it}'\beta, \text{ and}$$

$$\operatorname{logit} P(Y_{it} = 1 | Y_{it-1} = 1) = \operatorname{logit}(1 - \lambda_{it}) = X_{it}'\beta^{*},$$

where the parameters β and β^* may differ. Hence the model assumes that the effects of the covariates will differ depending on the previous response. A condensed form for the same general model is

$$\text{logit } P(Y_{it} = 1 | Y_{it-1} = y_{it-1}) = X_{it}^{'}\beta + y_{it-1}X_{it}^{'}\alpha, \tag{3}$$

where $\alpha = \beta^* - \beta$. This equation expresses the two regressions as a single dynamic model that includes as predictors both the previous response y_{it-1} (given that the intercept vector is included in X_{it}) and the interaction of y_{it-1} and the covariates X_{it}. Note that the transition matrix varies across both individuals and time periods because the hazards depend on the current and backcasted values of the covariates. Theoretical uses of Eq. (3) for panel data occur in Amemiya (1985), Diggle et al. (1994), and Hamerle and Ronning (1995). Boskin and Nold (1975) offer an application of a heterogeneous but stationary model with exogenous variables to the case of turnover in welfare based on panel data. See Toikka (1976) for an application of a three-state Markov model with exogenous variables to labor market choices (employed, unemployed and searching for a job, and withdrawal from employment) in which the transitions are estimated using frequency data disaggregated by sex.

According to Eq. (3) the transition rates are $\mu_{it} = F(X_{it}^{'}\beta)$ and $\lambda_{it} = 1 - F[X_{it}^{'}(\alpha + \beta)]$, where F is the logistic function. Maximum likelihood estimates of α and β can be obtained by maximization of the log likelihood function

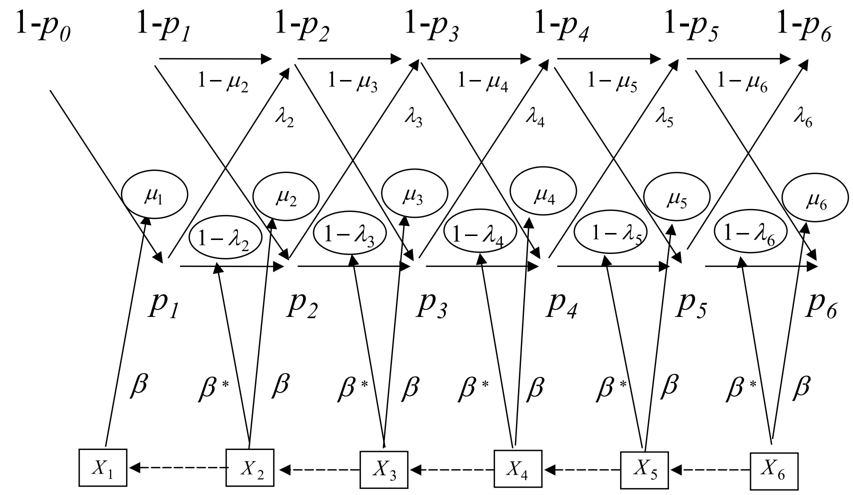

Figure 1. Graphical illustration of Markov model for RCS data.

$$LL = \sum_{t=1}^{T} \sum_{i=1}^{n_t} [y_{it} \log(p_{it}) + (1 - y_{it}) \log(1 - p_{it})] \tag{4}$$

with respect to the parameters, with p_{it} defined by Eq. (2). As indicated by Moffitt (1993), obtaining p_{it} by means of the reduced form equation is equivalent to 'integrating out' over all possible transition histories for each individual i at time t to derive an expression for the observed marginal probabilities. To see this, a graphical presentation of the model is given in Figure 1, omitting the subscript i for clarity.

The marginal probability p_{it} depends on the set of all possible transition histories for each individual i up to time t. That is, p_{it} is a polynomial in the transition rates μ_{it} and λ_{it}. The unobserved transition probabilities themselves are modeled as functions of current and backcasted values of time-invariant and time-varying covariates X_{it}. Hence an important feature of the model is that the transition probabilities and the marginal probabilities are estimated as a function of all available cross sections rather than simply the observations from the current time period. Thus estimates of the distribution at the beginning of the Markov chain, for example, are not determined solely by the sample obtained for the first time period but by all the samples.

2.2. MAXIMUM LIKELIHOOD ESTIMATION

Maximum likelihood fitting of the model in Eq. (2) requires the derivatives of the likelihood function (4) with respect to the parameters. The gradients of such models are frequently, as in Moffitt (1993), calculated by means of numerical differentiation, but there is no need to perform the maximization of the likelihood numerically if expressions are available for the derivatives. A major advantage of using analytical gradients is that they considerably speed up estimation. The gradients generate large and computationally cheap likelihood increases especially during the first iteration steps and thus considerable savings in computer time. Another advantage is that an asymptotic estimate of the dispersion matrix for the estimators is obtained from (the expectation of) the second-order derivatives of the likelihood surface. For ease of exposition, subscript i is omitted in the expressions of the derivatives and Eq. (2) is re-written as

$$p_t = \sum_{\tau=1}^{t} \mu_\tau \left(\prod_{s=\tau}^{t} \eta_s\right) \eta_\tau^{-1}, \tag{5}$$

where $\mu_\tau = (1 + e^{-1 \cdot (\beta x_\tau)})^{-1}$, $\eta_s = 1 - \lambda_s - \mu_s$, and $\lambda_s = (1 + e^{((\alpha+\beta)x_s)})^{-1}$. The first order partial derivatives of p_t in Eq. (5) with respect to the parameters β and α are

$$\frac{\partial p_t}{\partial \beta} = \sum_{\tau=1}^{t} \frac{\partial \mu_\tau}{\partial \beta} (\prod_{s=\tau}^{t} \eta_s) \eta_\tau^{-1} + \sum_{\tau=1}^{t-1} \sum_{s=\tau+1}^{t} \mu_\tau \frac{\partial \eta_s}{\partial \beta} (\prod_{\gamma=\tau+1}^{t} \eta_\gamma) \eta_s^{-1} \quad \text{and}$$

$$\frac{\partial p_t}{\partial \alpha} = \sum_{\tau=1}^{t-1} \sum_{s=\tau+1}^{t} \mu_\tau \frac{\partial \eta_s}{\partial \alpha} (\prod_{\gamma=\tau+1}^{t} \eta_\gamma) \eta_s^{-1}, \tag{6}$$

respectively, where $\partial \mu_\tau / \partial \beta = x_\tau(1 - \mu_\tau)\mu_\tau$, $\partial \eta_s / \partial \beta = x_s(1 - \lambda_s)\lambda_s - x_s(1 - \mu_s)\mu_s$, and $\partial \eta_s / \partial \alpha = x_s(1 - \lambda_s)\lambda_s$. Using these expressions we can calculate the derivatives of the log likelihood function with respect to the parameters.[5] The ML estimates are the values of the parameters for which the efficient scores (Rao, 1973) are zero. To obtain a solution to the equations resulting from setting $\partial LL/\partial \beta = \partial LL/\partial \alpha = 0$, we use a modified Newton method [6] called Fisher's method-of-scoring which provides an iterative search procedure for the computation of $\widehat{\beta}$ consisting of the iterations: $\widehat{\beta}^{(i+1)} = \widehat{\beta}^{(i)} + \varepsilon[\widehat{\mathbf{I}}(\widehat{\beta}^{(i)})]^{-1}(\partial LL(\hat{\beta}^{(i)}) \big/ \partial \beta)$ (see, e.g., Amemiya, 1981). The parameter ε denotes an appropriate step length which scales the parameter increments and $\hat{\mathbf{I}}(\hat{\beta}^{(i)})$ is an estimate of the Fisher information matrix $\mathbf{I}(\beta) = -\mathrm{E}[\partial^2 LL(\beta)/\partial \beta_j \partial \beta_k]$ evaluated at $\beta = \widehat{\beta}^{(i)}$, where $\partial^2 LL(\beta)/\partial \beta_j \partial \beta_k$ is the Hessian matrix. As a by-product of this iterative scheme, the method-of-scoring produces an estimate of the asymptotic variance–covariance matrix of the model parameters, being the inverse of the information matrix $\mathbf{I}^{-1}(\beta)$ evaluated at the values of the maximum likelihood estimates.

2.3. MODEL EXTENSION AND REFINEMENT

A drawback to the Markov model presented by Moffitt (1990; 1993) is that it assumes that the covariate effects are fixed over time, implying that the covariates are expected to have much the same impact over the period of time during which the observations were obtained.[7] This restriction cannot be expected to remain valid over long time periods and potentially biases the estimated effects, particularly those of time-varying variables and the baseline hazards. A question arises, however, as to what alternative model to consider if we drop the assumption of time-constant parameters. Even for moderate numbers of time periods, modifying continually the values of the parameters so as to allow the model to adapt itself to "local" conditions produces problems of over-parameterization. Due to the large number of parameters involved, this will often lead to the nonexistence of unique ML estimates. We try to avoid such problems by a parsimonious parameterization suitable for practical applications and introduce time variation into the model by allowing the regression coefficient to become polynomials in time using the expression $\beta_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \cdots + \gamma_d t^d$, where d is a positive integer specifying the degree of the polynomial. This parametric specification is particularly useful in situations where we have some prior expectations about how the covariate effects vary over time and if the effects evolve slowly. Further, the relative ease with which the likelihood function may be maximized adds to the usefulness of polynomials as practical tools for time dependence in the use of covariates. Of course, in practice it will be desirable to have models with low degree polynomials that combine parsimony of parameterization with fidelity to data.

A further way in which we accommodate the model is that whereas Moffitt defined the first observed outcome of the process $P(Y_{i1} = 1)$ to equal the transition probability μ_{i1}, we take $P(Y_{i1} = 1)$ to equal the state probability p_{i1}. That is, we assume that the Y_{i1}'s are random variables with a probability distribution $P(Y_{i1} = 1) = F(X_{it}^{'}\delta)$, where δ is a set of parameters to be estimated and F is the logistic function. The δ-parameters for the first observed outcome at t=1 are estimated simultaneously with the entry and exit parameters of interest at $t = 2, \ldots, T$. Recall that the probability vector at the beginning of the Markov chain is estimated as a function of all of the cross-sectional data, rather than simply the observations at t=1.

Finally, we also relax the assumption that the cross sections at each time t are of the same sample size. To ensure a potentially equal contribution of the cross-sectional samples to the likelihood, we use the weighted log likelihood function

$$LL^* = \sum_{t=1}^{T} \sum_{i=1}^{n_t} w_i [y_{it} \log(p_{it}) + (1 - y_{it}) \log(1 - p_{it})],$$

where $w_i = (\sum_{t=1}^{T} n_t)/Tn_t$, n_t is the number of observations of cross section t and T is the number of cross sections.

3. Application

Our empirical application employs ISSP data for married and unmarried cohabiting women aged 20–64 drawn from 10 annual Dutch (NL) surveys conducted in the period 1987–1996 and 7 annual West German (WG) surveys conducted in the "Alte Bundesländer" in the period 1988–1994. Because the ISSP surveys failed to provide some relevant covariates, additional information was taken from the omnibus surveys that incorporated the ISSP modules. The Dutch ISSP data were part of the omnibus survey CULTURAL CHANGES conducted by the Social and Cultural Planning Office (SCP). Because the SCP failed to conduct a survey in 1990, the cross sections were supplemented by data from the survey SOCIAL AND CULTURAL DEVELOPMENTS IN THE NETHERLANDS 1990 (SOCON) by the University of Nijmegen (Eisinga et al., 1992). The West German data were

Table I. Marginal fraction of female employment, $n = 6{,}411$ (NL) and 4,150 (WG)

	The Netherlands		West Germany	
Year	n_t	$y = 1$	n_t	$y = 1$
1987	586	0.276		
1988	582	0.325	869	0.420
1989	611	0.358	468	0.391
1990	839	0.455	792	0.509
1991	584	0.430	413	0.508
1992	637	0.425	690	0.525
1993	578	0.483	283	0.502
1994	609	0.435	635	0.572
1995	659	0.490		
1996	726	0.493		

taken from the ISSP surveys of 1989 and 1993 and the ALLBUS omnibus surveys of 1988, 1990–1992, and 1994 by ZUMA and ZA that incorporated the ISSP modules.

The labor market status y_{it} is defined to equal 1 if the women participates in the labor force (i.e., one of more hours of paid work per week) and 0 otherwise. Table I gives the number of respondents and the marginal distribution of participation over time in The Netherlands and West Germany. The table shows that over the period considered the female participation rate in the Netherlands almost doubled from about 28% in 1987 to around 49% in 1996. While the rates for West Germany are generally higher, the increase over time is smaller.

As time-varying covariates, the analysis employs (linear, quadratic and cubic terms in) age, number of children at three different age categories (<5, 5–17, ≥ 18 years of age), and the annual nationwide unemployment rate (%). The covariates completed education and religious upbringing (NL) or religion (WG) are taken to be fixed over time. Next to these variables the analysis also includes three initial conditions variables that capture the first entry into the process at age 20, the interaction of first entry with education and the interaction with the aggregate unemployment rate.[8] It is of interest to note that the individual observations were back-casted until the minimum age of 20, at which the first entry into the participation process is taken to have occurred. For observations whose back-casted value of age in a particular cross section was less than 20, the entry and exit rates for that time period were fixed to zero. Table II presents the parameter estimates for a time-constant–coefficient model specifying women's transition into and out

Table II. Time-constant Markov estimates of women's transition into and out of employment; $n = 6{,}411$ (NL) and 4,150 (WG) [a]

	Netherlands		West Germany	
	$\beta(\mu_t)^b$	$-\beta^*(\lambda_t)$	$\beta(\mu_t)$	$-\beta^*(\lambda_t)$
Fixed covariates [c]				
Intercept	−2.788*	−7.107	−3.688*	−18.861*
	(1.104)	(4.223)	(1.108)	(5.942)
Education completed	0.220*	−0.719*	0.344*	−0.072
	(0.066)	(0.105)	(0.081)	(0.130)
Religious upbringing (NL)	−0.148	−0.008	0.074	1.131*
/ religion (WG)	(0.125)	(0.178)	(0.200)	(0.399)
Varying covariates				
Age	0.179*	0.710*	0.177*	1.267*
	(0.051)	(0.335)	(0.045)	(0.451)
$\text{Age}^2 \div 100$	−0.281*	−1.936*	−0.283*	−3.300*
	(0.061)	(0.882)	(0.054)	(1.127)
$\text{Age}^3 \div 10{,}000$		1.804*		2.850*
		(0.760)		(0.919)
Number of children:				
< 5 years old	−0.770*	0.575*	−0.629*	3.847*
	(0.116)	(0.143)	(0.112)	(0.482)
5-17 years old	−0.460*	−0.066	−0.332*	0.585*
	(0.076)	(0.117)	(0.074)	(0.141)
≥18 years old	−0.083	−0.054	0.266*	0.504*
	(0.129)	(0.208)	(0.105)	(0.176)
Unemployment rate	−0.076	−0.160	0.078	−0.018
	(0.097)	(0.133)	(0.076)	(0.090)
Age20	5.223		3.045	
	(3.347)		(3.810)	
Age20 × education	−0.523		0.474	
	(0.681)		(0.893)	
Age20 × unemployment	−0.618		−0.433	
rate	(0.417)		(0.463)	
Log likelihood (LL^*)		−3706.729		−2416.144

* Significant at 5% level.
[a] Asymptotic estimates of standard errors in parentheses.
[b] The β-parameters represent the effect on entry (i.e., μ_t), the β^*-parameters the effect on $(1 - \lambda_t)$ and thus $-\beta^*$ the effects on exit (i.e., λ_t).
[c] Range of covariates: education completed (low–high): 1–4 (NL)/1–3 (WG); religious upbringing (NL) and religion (WG): 0 (no), 1 (yes); Age (back-cast) in years: 20–64; number of children (back-cast) <5: 0–4; Number of children (back-cast) 5–17: 0–7 (NL)/0–6 (WG); Number of children (back-cast) ≥18: 0–5; national unemployment rate (back-cast) in each year in percentages; Age20: 1 if age (back-cast) = 20, 0 if not.

of employment.[9] The model defines the first outcome to equal the transition probability μ_{i1}, as in Moffitt (1990; 1993), and not the state probability p_{i1}.

The first and third column in Table II present the effect of the variables on the transition from non-employment to employment in The Netherlands and West Germany, respectively. As can be seen, the parameters in both countries are well determined. Whereas education is significant in encouraging entry into the labor force, young children in the household (especially preschool children) negatively affect the entry decision. We also find that age has a substantial curvilinear effect on entry implying that the entry rates increase until a certain age after which they decline. The initial conditions variables indicate that higher unemployment rates and, in the Netherlands, higher education decrease the probability of entry at age 20. According to the standard errors, however, these variables have little impact on the hazards. The same goes for religious upbringing (NL) and religion (WG) and the aggregate unemployment rate.

The second and fourth column in Table II give the effect of the variables on the transition into non-employment. We find that in The Netherlands the exit rates are negatively affected by education and positively by the number of preschool children in the household. In West Germany the exit rates are unaffected by education, but positively affected by religion and the number of children of different ages. Particularly strong is the positive effect on exit of the number of preschool children in German households. The coefficients of the age terms imply that in both countries the incentives to end a job increase with age, but that the increase is not linear. The exit rates initially increase with age, temporarily decrease and thereupon increase again. In both countries the effect of the aggregate unemployment rate on the transition into non-employment is insignificant.

There are several arguments to anticipate that some of the covariate effects vary over time. First, the presence of young children the household may have become less of an impediment to women's employment in The Netherlands and West Germany. The extension of statutory maternity leave, the growing access to child care arrangements and the availability of non-parental supervision on schools, may all have eroded the effect of young children on the entry and exit decisions of mothers. Second, over the time period considered, the increase in women's schooling has contributed directly and indirectly, through wages, to an increase in women's labor supply. The educational expansion increased their opportunities in the labor market and gave way to an increasing attachment to paid work. The growth in real earning opportunities altered women's work decision in that it increased the costs of staying at home with an infant and thereby pulled women into the labor force. These changes are likely to have led to a strengthening of the effect of education on entry and exit. Third, religious secularization may have weakened the norms against women's participation in the labor force and we may thus expect to find a decreasing effect of religious upbringing and religion on entry and exit.

To examine these expectations, the baseline hazards and the effects of the covariates mentioned were allowed to vary over time. The effects of the age terms and the unemployment rate were held constant. We also separated the first observed outcome of the process from the subsequent ones and considered it to equal the state probability p_{i1} rather than the transition probability μ_{i1}. After some testing with several specifications, we decided to model all time-varying parameters in The Netherlands by a second-degree polynomial. Because of the smaller number of West Germany cross sections, the effects of the parameters on the entry rates were modeled by a second-degree polynomial, but their effects on the exit rates were designed by a first-degree polynomial. Further, the effect of religion on exit in West Germany turned out to be more or less constant over time and this parameter was therefore held time-constant.

According to the Akaike information criteria in Table III, that adjust the log likelihood for the number of estimated parameters, in both countries the time-varying–coefficient model slightly better describes the data than the time-constant–coefficient model. This indicates that pooling the estimates may be a misspecification, although we have not tested this hypothesis formally. The time-paths of the estimated parameters are displayed in Figure 2. It should be noted that the parameters at t=1 (i.e., 1987 in NL and 1988 in WG) represent the effect on the state probability and not the effect on entry.

Not surprisingly the parameter estimates change substantially if we allow for time variation. For The Netherlands, most of the time-paths traced out by the Markov coefficients are broadly consistent with the expectations: the declining effects of young children (under 18) on both entry and exit indicate positive reactions of mothers of preschool children to improvements in child care arrangements. Further, the growing positive effect of education on entry

Table III. Goodness of fit statistics; n = 6,411 (NL) and 4,150 (WG)

	Netherlands	West Germany
Time-constant–coefficient model		
Log likelihood (*LL**)	−3706.729	−2416.144
number of parameters	22	22
Akaike information criterion	1.163	1.175
Time-vaying–coefficient model		
Log likelihood (*LL**)	−3669.815	−2382.010
number of parameters	56	48
Akaike information criterion	1.162	1.171

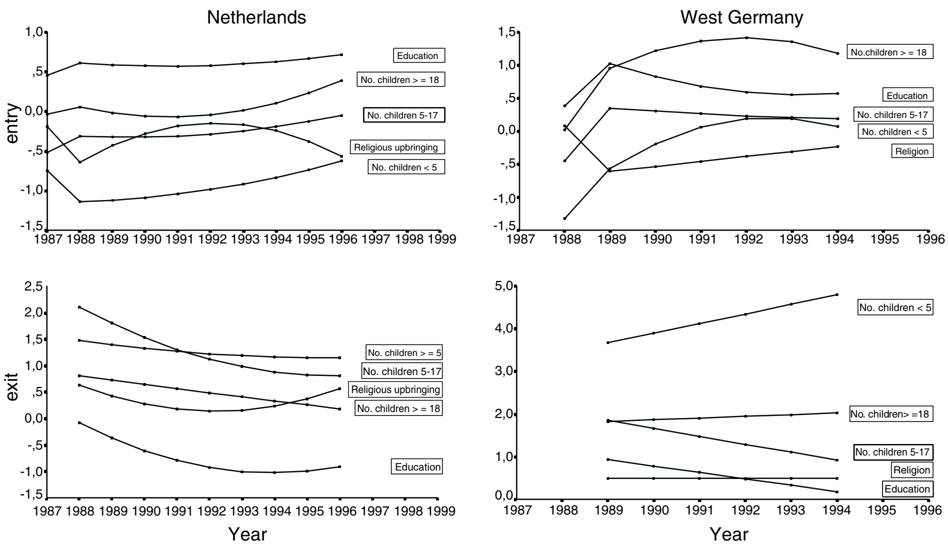

Figure 2. Estimated time-varying effects on entry (top) and exit (bottom).

and its growing negative effect on exit reveal women's increasing occupational aspirations. For West Germany, we see that the positive effects of education on both entry and exit have declined over time. Whereas the

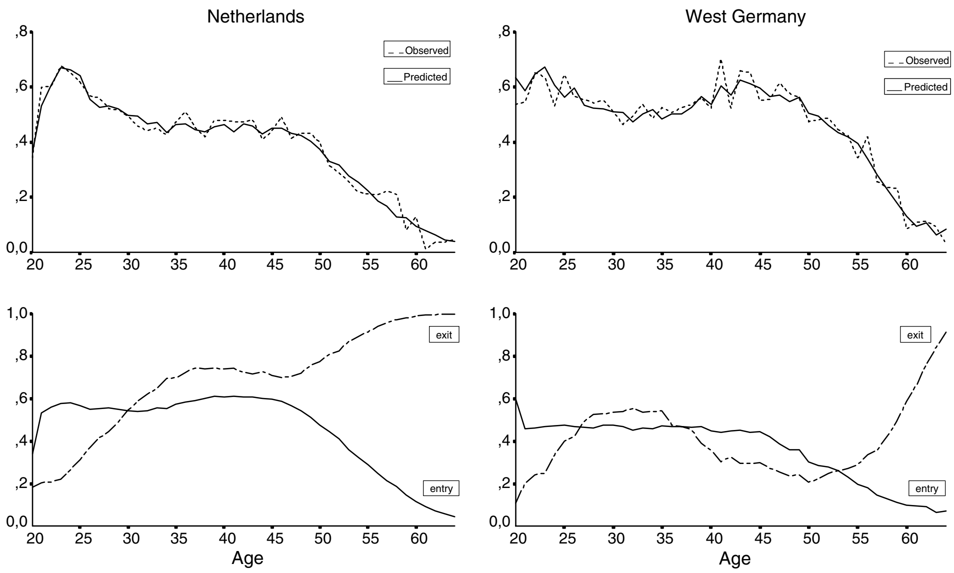

Figure 3. Observed and predicted employment probabilities (top) and entry and exit transition probabilities (bottom) by age.

negative effect of preschool children on entry has declined, the strong positive effect of preschool children on exit has increased over time. Hence most of the effects in West Germany are not consistent with the expectation of an increasing effect of education and a decreasing effect of the presence of young children.

To illustrate the model's ability in predicting life-cycle employment and non-employment patterns, Figure 3 (top) presents the observed and predicted marginal employment probabilities by age.[10] Figure 3 shows that in both countries the predicted probabilities are very similar to the observed. In The Netherlands, the participation rates increase substantially until the age of 24 but they are depressed (by the presence of preschool children) from age 25 to 34. The rates remain almost unchanged during age 35–44 and they are forced down again (by occupational pension) after the age of 45. For WG, we see that the high employment rates at age 24 decline until the age of 32, then increase until the age of 43 after which they fall again rapidly. Hence the most important difference between the countries is the substantial increase in participation in West Germany during the ages 32–43. This may be the result of either higher (re-) entry rates after childbearing or lower exit rates during childbearing and childrearing in West Germany.

To examine this issue, the bottom part of Figure 3 shows the life cycle profile of entry into and exit from the labor force, obtained from $t = 2, \ldots, T$.[11] As can be seen, the entry rates in the Netherlands decline slightly after age 23 (due to the impact of childrearing), increase slightly after the age of 32 (return to work) and then fall substantially past the age of 45. With respect to the entry rates the two countries are relatively similar, albeit that the German rates are lower. The countries differ substantially, however, with respect to the life cycle profile of exit. The exit rates in The Netherlands accelerate rapidly after age 25 (the presence of young children), remain high and relatively flat during age 36–46, and then increase again after age 46. In West Germany, on the other hand, the exit rates increase until the age of 27, remain flat during age 28–34, substantially decrease after age 35 and then increase again after the age of 50. Hence the most important difference between the two countries seems to be the strong decline in exit rates in West Germany during the ages 35–50. These life cycle profiles clearly visualize the employment interruption during childbearing and childrearing and the effect of occupational pension. It should be noted, however, that the rates are averages and that they thus confound within-cohort rates with across-cohorts rates. A more comprehensive analysis of these transitions could be conducted by verifying the results in panel data where sequences are known. Such an analysis, however, is beyond the scope of the present study and must be left for future research.

4. Conclusion

The overall conclusion that we draw from this example is that the proposed model can be a useful tool in applied work. It is not a panacea, nor does it supersede genuine panel designs, but it puts a series of one-shot surveys into perspective and it can certainly provide more refined results and interpretations than those available from a single cross-sectional study. Micro-data panel sets, without any question, offer the potential for the construction of more flexible and richer statistical models of transition dynamics than do those based upon cross-sectional information. However, while there has been a substantial increase of data archives holding vast collections of repeated cross-sectional data, panel data represent the exception of these collection efforts, rather than the rule. Moreover, a disadvantage to using pure panel surveys is the limited number of time points at which persons are usually re-interviewed. Hence the small number of time points in panel surveys has to be balanced against the lack of direct information on the transitions in long-run RCS data. The ideal situation would be to have complete histories of individual moves among states over a long time span. This life history information can be collected in both panel and RCS surveys through a retrospective interview.

Some problems we encountered in trying to model unobserved transitions over time using RCS data deserve to be mentioned. The application of the method presented here requires knowing the history of the explanatory variables for the individuals in the samples. We often have characteristics for which the history is unknown however. These characteristics may be relevant explanatory variables, but in many applications the analysis would omit them. Nevertheless, it is our believe that relatively rich dynamic models can be developed with a time series of RCS data. Many individual variables can be back-casted with considerable accuracy and many aggregate indicators are also measurable in the past.[12]

A somewhat related problem, common to all duration analyses, is that the model specification assumes that individual heterogeneity is due to the observed variables. It is likely, however, that unobserved and possibly unobservable variables including the initial conditions are also a source of population heterogeneity. The pre-sample history is lost by imposing an arbitrary survey window on the behavioral process, thus left-censoring the process and omitting events of interests associated with, or arising from, the periods prior to the first survey. The potential effect of this uncontrolled heterogeneity can bias the estimated effects of the explanatory variables included in the model. It is unknown, however, how serious the consequences of misspecification are if we have sufficiently flexible models for baseline hazards and time-varying covariates. The latter are often interpreted as caused by heterogeneity (Fahrmeier and Knorr-Held, 1997). Hence further investigation is needed on how much of the evidence in censored, for example by examining the application of mixture models which allow for residual heterogeneity. These models include an additional, individual-specific random error term (or nuisance parameter) in the linear predictor of the logistic function of the hazards to account for omitted variables (or extra-binary variance).

Another subject for future study is the extension to both higher-order and multi-state models. In practice the dependent variable may depend not on just the most recent observation but on other previous observations of the process as well. Although no essential new theory is involved in such an extension, a higher-order chain may have too many parameters in the model unless there are some structural constraints imposed on the hazards. An initial, computationally tractable way to improve over the example application presented here is to consider a first-order model that distinguishes exit into non-employment from exit into early retirement, where the latter is modeled as an absorbing state (Andersen, 1980: 304), implying that once entered it is never left.

Finally, our approach to imposing restrictions on the time-varying–coefficient model is through low degree polynomial functions. In some applications this parametric bases may not provide enough flexibility and local adaptiveness. It would therefore seem important to study the minimal requirements needed for a varying-coefficient model to yield uniquely identified parameter estimates. We can prove that under relatively mild conditions there always exists exactly one solution for the parameters, but we can only verify this for relatively simple Markov models, for example, those with constant terms only. Unfortunately, no complete set of identification rules has yet been found guaranteeing unique solutions in more complex models with continuous regressors. It is worthwhile to pursue this thorny problem further.

Acknowledgements

The Dutch ISSP data were collected by the Social and Cultural Planning Office (SCP) and were made available by courtesy of the Steinmetz Archive in Amsterdam. The Dutch data were supplemented by data from the 1990 SOCON survey by the University of Nijmegen. The West German data were taken from the ISSP and ALLBUS surveys by the Zentrum für Umfragen, Methoden und Analysen (ZUMA) and the Zentralarchiv für Empirische Sozialforschung (ZA). The stand-alone program *CrossMark* to do the computations reported here is available from the authors.

Notes

1. See Felteau et al. (1997) for an application to the marriage and fertility decisions of Canadian women using data from the Survey of Consumer Finances of Statistics Canada.
2. More generally, a higher-order Markov chain of order l on m states has $m^l(m-1)$ independent transition probabilities. Given m possible states, there are only $m-1$ unique state probabilities. Because $m^l(m-1) > m-1$ for $m > 1$ the transitions are not identified (see Tuman and Hannan, 1984: 297).
3. Let $p_{i1} = \mu + \eta p_{i0}$, $p_{i2} = \mu + \eta p_{i1} = \mu + \eta(\mu + \eta p_{i0}) = \mu(1+\eta) + \eta^2 p_{i0}$. Hence $p_{it} = \mu(1+\eta+\cdots+\eta^{t-1}) + \eta^t p_{i0} = \mu(1+\sum_{\tau=1}^{t-1}\eta^{t-\tau}) + \eta^t p_{i0}, = (\mu/(\mu+\lambda))(1-\eta^t) + \eta^t p_{i0}$. As $t \to \infty$, η^t tends to zero, thus $p_{it} = \mu/(\mu+\lambda)$. Obviously, this equation holds for $\eta \neq 1$, and, if $\eta = 1$, $\mu = \lambda = 0$.
4. Let $p_{i1} = \mu_{i1} + \eta_{i1}p_{i0}$, $p_{i2} = \mu_{i2} + \eta_{i2}p_{i1} = \mu_{i2} + \eta_{i2}(\mu_{i1} + \eta_{i1}p_{i0}) = \mu_{i2} + \mu_{i1}\eta_{i2} + p_{i0}\eta_{i1}\eta_{i2}$. Hence $p_{it} = \mu_{it} + (\mu_{it-1}\eta_{it} + \mu_{it-2}\eta_{it-1}\eta_{it} + \cdots + \mu_{i1}\eta_{i2}\cdots\eta_{it}) + p_{i0}\eta_{i1}\cdots\eta_{it} = \mu_{it} + \sum_{\tau=1}^{t-1}\mu_{i\tau}(\prod_{s=\tau+1}^{t}\eta_{is}) + p_{i0}\prod_{\tau=1}^{t}\eta_{it}$. As $t \to \infty$, $\prod_{\tau=1}^{t}\eta_{it}$ tends to zero, thus $p_{it} = \mu_{it} + \sum_{\tau=1}^{t-1}\mu_{i\tau}(\prod_{s=\tau+1}^{t}\eta_{is})$.
5. The partial derivative of (the contribution LL_i of observation i to) the log likelihood function LL with respect to p_t is $\partial LL_i/\partial p_t = (y-p_t)/p_t(1-p_t)$ and the partial derivative of LL with respect to the parameters can be obtained by the chain rule, for example, $\partial LL/\partial\beta = \partial LL/\partial p_t \cdot \partial p_t/\partial\beta$.
6. The modification consists in substituting the Hessian matrix by its estimated expectation. If an iterative procedure of the Newton-type is used, involving analytical derivatives, there is a choice between using either actual second derivatives or expected second derivatives, i.e., the Fisher information (or expected Hessian). According to Cox and Hinkley (1974: 308) and Greene (1993: 347–348) there is evidence that the latter is to be preferred because it performs better in practice.
7. It may be of interest to note that while this restriction is not necessary with true panel data, in practice most panel studies nevertheless impose the restriction of time-constant coefficients in the model specification (see Bell and Ritchie, 1997).
8. The potentially important initial conditions variable Age20 × children was not included in the analysis as the number of mothers aged 20 was insufficient to allow reliable estimation.
9. The time-invariant Markov model with constant terms only produced $\beta(\mu_t)$ coefficients of −1.099 and −0.484 and $-\beta^*(\lambda_t)$ coefficients of −0.841 and −0.583 in The Netherlands and WG, respectively. This implies constant annual transition rates of μ=0.252 and λ=0.301 in The Netherlands and μ=0.381 and λ=0.358 in West Germany.
10. The mean $\bar{p}_m$ for age category m was obtained as $\bar{p}_m = n_m^{-1}\sum_{i=1}^{n_m} p_{it}$, where n_m is the number of observations in age category m and p_{it} the predicted probability of observation i at the current time period t (i.e., when y_{it} was observed).
11. The means $\bar{\mu}_m$ and $\bar{\lambda}_m$ for age category m were obtained as a weighted average of the transitions up to t with weights defined by $w_k = (\sum_{t=k}^{T} n_t)^{-1}(T-1)^{-1}\sum_{j=2}^{T}\sum_{t=j}^{T} n_t$.
12. Obviously, it also depends on the time span of the repeated cross sections. If the cross sections concern a number of consecutive week-surveys, for example, many variables (e.g., income) can reasonably be treated as time-constant.

References

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Econometric Literature* 19: 1483–1536.

Amemiya, T. (1985). *Advanced Econometrics*. Oxford: Basil Blackwell.

Andersen, E. B. (1980). *Discrete Statistical Models with Social Science Applications*. Amsterdam: North-Holland.

Baltagi, B. H. (1995). *Econometric Analysis of Panel Data*. Chichester: Wiley.

Bartholomew, D. J. (1996). *The Statistical Approach to Social Measurement*. San Diego: Academic Press.

Bell, D. & Ritchie, F. (1997). *Time-varying parameters in panel models*. unpublished manuscript, Department of Economics, University of Stirling.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge MA: MIT Press.

Boskin, M. J. & Nold, F. C. (1975). A Markov model of turnover in aid to families with dependent children. *Journal of Human Resources* 10: 476–481.

Collado, M. D. (1997). Estimating dynamic models from time series of independent cross-sections. *Journal of Econometrics* 82: 37–62.

Cox, D. R. & Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman & Hall.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics* 30: 109–126.

Diggle, P. J., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Eisinga, R., Felling, A., Peters, J., Scheepers, P., & Schreuder, O. (1992). *Religion in Dutch Society 90. Documentation of a National Survey on Religious and Secular Attitudes in 1990*. Amsterdam: Steinmetz Archive.

Fahrmeir, L. & Knorr-Held, L. (1997). Dynamic discrete-time duration models: Estimation via Markov chain Monte Carlo. In: A. E. Raftery (ed.), *Sociological Methodology 1997,* San Francisco: Jossey-Bass, pp. 417–452.

Felteau, C., Lefebvre, P., Merrigan, Ph. & Brouillette, L. (1997). Conjugalité et fécondité des femmes Canadiennes: un modèle dynamique estimé à l'aide d'une série de coupes transversales. unpublished manuscript, CREFÉ, Université de Québec à Montréal.

Firth, D. (1982). Estimation of voter transition matrices from election data. M.Sc. Thesis, Department of Mathematics, Imperial College.

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association* 56: 841–868.

Greene, W. H. (1993). *Econometric Analysis* (2nd ed.). New York: MacMillan.

Hamerle, A. (1994). Panel-modelle für qualitative daten. *Allgemeines Statistisches Archiv* 78: 1–19.

Hamerle, A. & Ronning, G. (1995). Panel analysis for qualitative variables. In: G. Arminger, C. Clogg & M. E. Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum Press, pp. 401–451.

Hawkins, D. L., Han, C. P., & Eisenfeld, J. (1996). Estimating transition probabilities from aggregate samples augmented by haphazard recaptures. *Biometrics* 52: 625–638.

Kalbfleish, J. D. & Lawless, J. F. (1984). Least squares estimation of transition probabilities from aggregate data. *Canadian Journal of Statistics* 12: 169–182.

Kalbfleish, J. D. & Lawless, J. F. (1985). The analysis of panel data under a Markovian assumption. *Journal of the American Statistical Association* 80: 863–871.

Lawless, J. F. & McLeish, D. L. (1984). The information in aggregate data from Markov chains. *Biometrika* 71: 419–430.

Lee, T. C., Judge, G. G. & Zellner, A. (1970). *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*. Amsterdam: North-Holland.

Li, W. K. & Kwok, M. C. O. (1990). Some results on the estimation of a higher order Markov chain. *Communications in Statistics. Part B. Simulation and Computation* 19: 363–380.

McCall, J. J. (1971). A Markovian model of income dynamics. *Journal of the American Statistical Association* 66: 439–447.

McFarland, D. D. (1970). Intra-generational social mobility as a Markov process: Including a time-Stationary Markovian model that explains observed declines in mobility rates over time. *American Sociological Review* 35: 463–476.

Moffitt, R. (1990). The effect of the U.S. welfare system on marital status. *Journal of Public Economics* 41: 101–124.

Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics* 59: 99–123.

Nijman, Th. E. & Verbeek, M. (1990). Estimation of time-dependent parameters in linear models using cross-sections, panels, or both. *Journal of Econometrics* 46: 333–346.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: Wiley.

Verbeek, M. (1996). Pseudo panel data. In: L. Mátyás & P. Sevestre (eds.), *The Econometrics of Panel Data (2nd revised edn)*. Dordrecht: Kluwer Academic Publishers, pp. 280–292.

Verbeek, M. & Nijman, Th. (1992). Can cohort data be treated as genuine panel data? *Empirical Economics* 17: 9–23.

Verbeek, M. & Nijman, Th. (1993). Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections. *Journal of Econometrics* 59: 125–136.

Toikka, R. S. (1976). A Markovian model of labor market decision by workers. *American Economic Review* 66: 821–834.

Topel, R. H. (1983). On layoffs and unemployment insurance. *American Economic Review* 73: 541–559.

Tuma, N. B. & Hannan, M. T. (1984). *Social Dynamics. Models and Methods*. Orlando, FL: Academic Press.

Ben Pelzer is a Research Associate at the Department of Social Science Research Methods, University of Nijmegen.

Rob Eisinga is Professor of Quantitative Research Methods for Comparative Survey Research in the Department of Social Science Research Methods, University of Nijmegen.

Philip Hans Franses is Professor of Applied Econometrics at the Econometric Institute and Professor of Marketing Research in the Department of Marketing and Organization, Erasmus University Rotterdam.