# Recovering Transitions

# from

# Repeated Cross Sections

een wetenschappelijke proeve op het gebied van de Sociale Wetenschappen

## Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op vrijdag 24 februari 2006
des namiddags om 1.30 uur precies

door

Bernard Joseph Pelzer

geboren op 22 juli 1951 te Heerlen

*Promotor:*

Prof. dr. R. Eisinga

*Manuscript commissie:*

Prof. dr. A. Felling

Prof. dr. J. Hagenaars (Universiteit van Tilburg)

Prof. dr. ir. F. Willekens (Rijksuniversiteit Groningen)

ISBN-10: 9090201815
ISBN-13: 9789090201818

Printed by PrintPartners Ipskamp, Enschede
Cover design by Claudia Oomens
Layout by Albert Bakker

# Voorwoord

De precieze datum waarop deze studie begon is mij ontschoten. Op zekere dag kwam Rob Eisinga mijn werkkamer binnen met een klein vraagje. Het antwoord is iets uitgebreider geworden dan toen was te voorzien. Dat komt op de eerste plaats door mijn promotor, diezelfde Rob Eisinga. Zijn enthousiasme was zo aanstekelijk dat er geen ontkomen aan was: ik moest samen met hem dit mooie probleem te lijf te gaan! Zeer veel dank ben ik hem verschuldigd voor het continue verleggen van de grenzen en het aanreiken van nieuwe ideeën.

Veel anderen hebben bijgedragen aan het tot stand komen van dit proefschrift en daarvoor wil ik hun graag bedanken. Bert Felling, Jacques Hagenaars en Frans Willekens bedank ik voor hun bereidheid zitting te nemen in de manuscriptcommissie. Veel dank gaat uit naar twee leermeesters. Naar Jan Lammers, voor zijn uiterst duidelijke uitleg van talrijke onderwerpen uit de statistiek en voor de onvergetelijke discussies die we zo vaak voerden. Mijn geklungel aanschouwend heeft Jan mij vaak laten zien hoe het echt moest! Theo van der Weegen dank ik voor zijn rol als leermeester van het eerste uur en voor zijn antwoorden op statistische en wiskundige vragen waarvoor zijn deur immer open stond. Manfred te Grotenhuis bedank ik voor zijn kritische meedenken; in menig gesprek wist hij me met beide benen op de grond te houden. Bert Felling wil ik, ten tweeden male, bedanken voor de stimulerende kracht die hij altijd op mij heeft uitgeoefend. De medewerkers van de RTOG bedank ik voor hun deskundige adviezen op het gebied van wiskunde, statistiek en software ontwikkeling; vooral bij Frans Gremmen, Nol Bendermacher, Jan van Leeuwe en Pieter van Groenestijn sta ik in het krijt. Een speciaal dankwoord richt ik tot Harrie Hendriks voor zijn grondige wiskundige adviezen en tot Steven Teerenstra en Marcel Coenders voor het becommentariëren van het manuscript. Babette Pouwels bedank ik voor haar commentaar bij hoofdstuk 2 en 3. Claudia Oomens, Albert Bakker en

Gerti Maturbongs ben ik zeer erkentelijk voor hun creativiteit en precisie. Mijn collega's van de sectie Methoden bedank ik niet allen afzonderlijk, maar des te meer allen tezamen, voor de fijne werksfeer.

Mijn ouders bedank ik voor de mij geboden mogelijkheden, in een tijd waarin dat niet vanzelfsprekend was; mijn moeder bedank ik ook voor de rustige gezellige avonden in drukke tijden. Dick en Riek bedank ik voor het vaak 'er zijn', zodat ik mijn eigen gang kon gaan, verlost van huiselijke plichten. Greet dank ik voor haar aandacht, open (voor)deur en de koffie tussendoor. Levi, Sieme en Vince, de afgelopen maanden zag ik jullie te weinig, maar, werkend op Vince's kamer, hoorde ik jullie des te meer, trap op en af rennend, roepend, lachend, en soms scheldend. Door jullie aangenaam lawaai wist ik steeds dat er wel degelijk nog leven is na het werken! Mijn laatste dankwoord, Inge, is voor jou, voor je luisterend oor dat er altijd is voor mij en de jongens, voor je gezelligheid en voor je liefde.

Nijmegen, 27 december 2005
Ben Pelzer

# Contents

# 1 Introduction

This book is about a particular discrete-time Markov model for categorical data. Like many other Markov models for categorical data, the model presented can be used to estimate the probability that an individual (case, observational unit) will make a transition between the states of a categorical variable of interest during some period of time. In contrast to many other models, it enables the estimation of such probabilities using data from a number of cross sections, independently sampled at regular time intervals. Therefore, we refer to the model as the 'repeated cross sections (RCS) Markov model'. The model can employ time-constant and time-varying predictor variables to yield transition probabilities that vary over time and individuals.

Before going into more details, we will first present a few basic concepts related to statistical Markov models for categorical data and introduce some notation. Next, we will briefly discuss a selection of Markov models that were developed during the last fifty years, say. Finally, the RCS model is introduced and an empirical example is shown.

## 1.1 Basic concepts

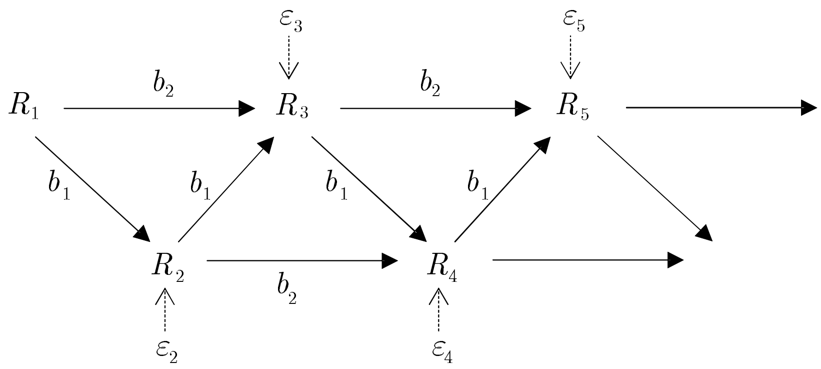
Markov models are about Markov processes. Roughly speaking, in such processes the outcome of the variable of interest at a given point in time is directly related to the outcome(s) of the same variable at a limited number of earlier time points. A textbook example of a Markov process is weather change. Rainy days often occur in numbers and, consequently, the amount of rain that falls on a given day is related to the amount that fell the day before. It is also related, though less strongly, to the amount of rain that fell two days ago. Going back in time more than 2 days, say, often results in a zero relation. Although from a meteorological viewpoint presumably not satisfactory, a Markov model to predict the amount of rainfall at a given day could be

$$R_t = b_0 + b_1 R_{t-1} + b_2 R_{t-2} + \varepsilon_t \qquad (1)$$

with $R_t$ denoting the rainfall (e.g., in millimetres) at day $t$, $R_{t-1}$ and $R_{t-2}$ the rainfall for the preceding two days and $e_t$ being the error associated with the model. In this model equation a limited 'history' of the dependent variable itself is used to predict its current value. The history here contains two time points in the past and hence the model is called a 'second order' Markov model. The terms 'autoregressive Markov model', 'autoregressive model' or simply 'AR-model' are also used for models in which the dependent variable is regressed on itself as it was observed at one or more earlier points in time. In Figure 1.1 the above equation is presented in the form of diagram that only considers the first five days of a possibly long series of daily observations that could have been made.

The solid arrows along with the regression effects $b_1$ and $b_2$ of Equation (1) represent the influence of rainfall of the previous two days. The dotted arrows refer to the contribution of the error term in (1) to each day's observed rainfall, from day 2 onwards. The second order property of the Markov process is represented by the omission of arrows between observations separated three or more days. There is for example no arrow going directly from $R_2$ to $R_5$ since, if controlled for $R_3$ and $R_4$, the effect of $R_2$ on $R_5$ is considered to be zero. The second order property does not imply that there is no effect of $R_2$ on $R_5$, but only that there is no *direct* effect.

Figure 1.1 Diagram of a second order Markov model



There are indirect effects, however, as the diagram shows, for example the effect from $R_2$ on $R_3$, from $R_3$ on $R_4$, and from $R_4$ on $R_5$. Thus, $R_2$ effects $R_5$ indirectly via $R_3$ and $R_4$. There are two more indirect paths, $R_2$, $R_3$, $R_5$ and $R_2$, $R_4$, $R_5$ by which the influence of $R_2$ on $R_5$ is passed on.

The dependent variable used in the above example is ratio-scaled and, apart from round-off error, can have an infinite number of possible outcomes or 'states'. If the variable of interest is categorical and hence has a finite number of states, the Markov process is typically referred to as a 'Markov chain'. The next example is more similar to the repeated cross sections Markov model, because it is concerned with a first order Markov chain, i.e., it deals with a categorical variable and uses a history of only 1 time point. Suppose there is an election taking place with two candidates, A and B say, running for president. Of the people who favoured candidate A one week prior to the election, 80% still favours A at election day, while 20% has switched to candidate B. Of those favouring B one week before the election, 90% still favours B at election day and 10% has switched to A. These percentages constitute the cell entries of Table 1.1.

From the high percentages on the diagonal of Table 1.1 it can be concluded that, at election day, most people for some reason stick to the preference they had one week earlier, whether it was A or B. Thus, for an arbitrary chosen individual, the outcome of the variable 'favoured candidate' at election day is strongly related to the outcome of that variable one week earlier. Let $\Upsilon_{i,t}$ and $\Upsilon_{i,t-1}$ be variables indicating whether A or B is favoured by person $i$ at election day $t$ and one week before, respectively, both taking value 0 if A and 1 if B is favoured. The following logistic

Table 1.1 Favoured candidate at election day and one week before

| | | election day | | |
|---|---|---|---|---|
| | | A | B | |
| one week before | A | 80% | 20% | 100% |
| | B | 10% | 90% | 100% |

Markov model equation may then be used for the probability that person $i$ favours B at day $t$:

$$\mathrm{P}(\Upsilon_{i,t}{=}1) \quad = \quad e^{b_0+b_1\Upsilon_{i,t-1}} \,/(1+\, e^{b_0+b_1\Upsilon_{i,t-1}}). \qquad (2)$$

Based on Equation (2) the observed value $\Upsilon_{i,t}$ can be expressed as $\Upsilon_{i,t} = P(\Upsilon_{i,t}{=}1) + \varepsilon_{i,t}$ with $\varepsilon_{i,t}$ denoting the error part of $\Upsilon_{i,t}$ not accounted for by $P(\Upsilon_{i,t}{=}1)$. In (2) only one past observation, $\Upsilon_{i,t\text{-}1}$, of the dependent variable itself is used as predictor variable, as opposed to two in the rainfall example. Hence, the model is first-order Markov. In general, such a model can be visualized by a diagram as the one in Figure 1.2, where the first five out of potentially much longer series of observations are shown. $\Upsilon_{i,4}$ could be the observation made at election day, then $\Upsilon_{i,3}$ would be the observation of the week before. Also, $\Upsilon_{i,2}$, $\Upsilon_{i,1}$ and $\Upsilon_{i,5}$ would then denote observations made two and three weeks before election day and one week after, respectively. The solid arrows indicate the direct influences of candidate preference one week before and $b_1$ the corresponding regression effect denoted in Equation (2). The dotted arrows represent the error term contributions to the $\Upsilon_{i,t}$ values. Note that, as in Figure 1.1, each observation $\Upsilon_{i,t}$ has an effect on all later observations, be it directly or indirectly.

Figure 1.2 Diagram of a first order Markov model

$$\begin{tikzcd}
\Upsilon_{i,1} \arrow[r, "b_1"] & \Upsilon_{i,2} \arrow[r, "b_1"] & \Upsilon_{i,3} \arrow[r, "b_1"] & \Upsilon_{i,4} \arrow[r, "b_1"] & \Upsilon_{i,5} \arrow[r] & {} \\
 & \varepsilon_{i,2} \arrow[u, dotted] & \varepsilon_{i,3} \arrow[u, dotted] & \varepsilon_{i,4} \arrow[u, dotted] & \varepsilon_{i,5} \arrow[u, dotted] &
\end{tikzcd}$$

The result of using predictor $\Upsilon_{i,t\text{-}1}$ in (2) is that two different outcomes can occur for the probability to favour candidate B at election day, one outcome for each candidate favoured one week earlier. Thus, the model is able to deal with the two different percentages for favouring B at election day, 20% and 90%, given in the example. In terms of Equation (2), the two possible outcomes of the probability to favour A at election day are given by

$$\text{P}(\Upsilon_{i,t}=1 \mid \Upsilon_{i,t-1}=0) = e^{b_0} \;/\; (1+ e^{b_0}) \qquad \text{and}$$
$$\text{P}(\Upsilon_{i,t}=1 \mid \Upsilon_{i,t-1}=1) = e^{b_0+b_1} \;/\; (1+ e^{b_0+b_1}).$$

Denoting these probabilities as conditional probabilities, conditional on the value of $\Upsilon_{i,t-1}$, shows that they are actually 'transition probabilities'. The first one refers to the transition from state A to B and the second one to the transition from state B to B. The true population values of these two probabilities as well as those for favouring A at election day given $\Upsilon_{i,t-1}$, would undoubtedly be considered valuable information by both candidates. In addition, Equation (2) can incorporate other predictor variables such as gender, race, age, highest completed education etc., an example of which is given below. By doing so, the model may reveal significant characteristics of voters who are willing to switch candidate or to stay with their earlier choice. This possibility to explore which type of individuals are inclined to make particular transitions is one of the reasons that Markov models have frequently been used, not only in voting transition research but in many other research areas as well.

The election example above concerned two states, A and B, of a categorical variable of interest $\Upsilon$ that was observed at two different time points. More generally, Markov models for categorical data can be used to study transitions between any finite number of states of a categorical variable that is observed at any finite number of time points: cases move from one state of $\Upsilon$ at a given time point 1, say, to another state or to the same state of $\Upsilon$ at time point 2, and again to another or the same state at time point 3, etc. If $\Upsilon$ has $M$ states then there are $M^2$ different transition probabilities when going from any previous time point $t-1$ to time point $t$. These transitions can be summarized in the so-called 'transition matrix' given in Table 1.2.

Table 1.2 Transition matrix for case $i$ at time point $t$

| | | | | $\Upsilon_{i,t}$ | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | . | . | . | $M$ | |
| $\Upsilon_{i,t-1}$ | 1 | $p_{i,t}(1,1)$ | $p_{i,t}(1,2)$ | . | . | . | $p_{i,t}(1,M)$ | 1 |
| | 2 | $p_{i,t}(2,1)$ | $p_{i,t}(2,2)$ | . | . | . | $p_{i,t}(2,M)$ | 1 |
| | . | . | . | . | . | . | . | 1 |
| | . | . | . | . | . | . | . | 1 |
| | . | . | . | . | . | . | . | 1 |
| | $M$ | $p_{i,t}(M,1)$ | $p_{i,t}(M,2)$ | . | . | . | $p_{i,t}(M,M)$ | 1 |

In this matrix $p_{i,t}(j,k)$ refers to the transition probability $P(\Upsilon_{i,t} = k \mid \Upsilon_{i,t-1} = j)$, that is the probability for case $i$ to be in state $k$ of $\Upsilon$ at time point $t$ given that the previous state at $t-1$ was $j$. Within a single row $j$ the probabilities sum to 1 since they cover all possible transitions a case can make from state $j$ at time point $t-1$ to all $M$ states at time point $t$.

With $T$ consecutive time points a case realizes $T-1$ transitions during the time period under study and, consequently, there are also $T-1$ transition matrices for each individual case. With the same $N$ individual cases observed at each time point there would be a total of $M^2(T-1)N$ transition probabilities. It is not surprising then that most Markov models assume some kind of structure to exist in what can become a very large amount of possibly different transition probabilities. An example of such a hypothesized structure is the assumption that the transition matrix is constant over time for each individual case. Such a model is said to be 'time-invariant' as opposed to a 'time-varying' model in which individual transition matrices are assumed to change over time. Another hypothesized structure is to assume is that for each time point $t$ the transition matrix is the same for all cases. Such a model is called 'individual homogeneous' as opposed to 'individual heterogeneous'.

If we let $\Upsilon$ be a binary variable with states 0 and 1, the transition matrix collapses to a 2x2 matrix with the probabilities in the left column the complement of those in the right column. If $\Upsilon$ is observed at only three consecutive time points then there are two separate transition matrices as represented in Table 1.3.

Table 1.3 Transition matrices for case $i$ for binary $\Upsilon$ and three time points

a)

| | | $\Upsilon_{i,2}$ | |
|---|---|---|---|
| | | 0 | 1 |
| $\Upsilon_{i,1}$ | 0 | $1-\mu_{i,2}$ | $\mu_{i,2}$ |
| | 1 | $\lambda_{i,2}$ | $1-\lambda_{i,2}$ |

b)

| | | $\Upsilon_{i,3}$ | |
|---|---|---|---|
| | | 0 | 1 |
| $\Upsilon_{i,2}$ | 0 | $1-\mu_{i,3}$ | $\mu_{i,3}$ |
| | 1 | $\lambda_{i,3}$ | $1-\lambda_{i,3}$ |

Table 1.3a contains the probabilities for all four transitions a case can possibly make from $t=1$ to $t=2$ and Table 1.3b those for the transitions from $t=2$ to $t=3$. For a binary variable $\Upsilon$ we prefer to use different symbols for particular transitions. We use $\mu_{i,t}$ to denote the 'entry-probability' or the probability to have entered state 1 at $t$, given state 0 at $t-1$. Also we use $\lambda_{i,t}$ to denote the 'exit-probability' or the probability to have exited state 1 at $t$ (or entered state 0 at $t$), given state 1 at $t-1$. In symbols: $\mu_{i,t} = P(\Upsilon_{i,t} = 1 \mid \Upsilon_{i,t-1} = 0)$ and $\lambda_{i,t} = P(\Upsilon_{i,t} = 0 \mid \Upsilon_{i,t-1} = 1)$. The complements of the diagonal probabilities $1-\mu_{i,t}$ and $1-\lambda_{i,t}$, refer to the probability to stay in state 0 or in state 1 of $\Upsilon$, respectively, i.e., $1-\mu_{i,t} = P(\Upsilon_{i,t} = 0 \mid \Upsilon_{i,t-1} = 0)$ and $1-\lambda_{i,t} = P(\Upsilon_{i,t} = 1 \mid \Upsilon_{i,t-1} = 1)$.

This concludes the introduction of concepts and notation. We will now discuss a selection of statistical Markov models for categorical data that have been developed in the past. Most of these models can deal with multi-state $\Upsilon$ variables. However, to explain a particular model we will often use a two state $\Upsilon$ and hence the symbols $\mu$ and $\lambda$.

## 1.2 A classification of Markov models

Different types of Markov models for studying longitudinal categorical data have been developed in the past. Especially since about 1950, many treatises of such models have appeared in the literature. One of several possible ways to classify these models is to focus on the type of $\Upsilon$ data observed in the applications. We will mention in some detail three main streams of models corresponding to three types of $\Upsilon$ data. These types of data can be summarized by cross-classifying the number of occasions at which each unit is observed by the aggregation level of the units. The result is shown in the table below.

| | | Aggregation level | |
|---|---|---|---|
| | | individual | aggregate |
| Number of occasions each unit is observed | >1 | 1<br>individual<br>panel data<br>(section 1.2.1) | 2<br>repeated aggregated<br>proportions<br>(section 1.2.2) |
| | 1 | 3<br>repeated<br>cross sections<br>(section 1.2.3) | |

As to the number of occasions, only the distinction between 'one' and 'more than one' occasion is relevant for the present purpose. With respect to the aggregation level, 'aggregate' refers to situations in which only data on aggregates or groups (e.g., school classes, voting districts, cities) of individual cases are available with the individual data that constitutes the aggregates being unobserved. Nevertheless, in the Markov models for aggregate data discussed below, the transitions pertain to the individuals that constitute the aggregates and not to the aggregates themselves. The data types numbered 1, 2 an 3 in the above table will be discussed in some detail below along with associated Markov models and relevant literature.

### 1.2.1 Models for individual panel data

In many applications, the values of $\Upsilon$ for all individual cases in a sample are observed at all, $T$ say, evenly spaced time points under study. This kind of data is known as 'individual panel data' or 'repeated measures'. The rainfall and the election example in the previous section were based on such data. For binary $\Upsilon$ with values 0 and 1 and $T = 3$, individual panel data can be presented as a set of 8 (= 2x2x2) sequences of values $\Upsilon_{i,1}$ , $\Upsilon_{i,2}$ and $\Upsilon_{i,3}$ followed by the corresponding number of cases for each sequence. An example using fictitious data is given in Table 1.4.

Table 1.4 Fictitious individual panel data for binary $\Upsilon$ at three time points

| $\Upsilon_{i,1}$ | $\Upsilon_{i,2}$ | $\Upsilon_{i,3}$ | frequencies |
|---|---|---|---|
| 0 | 0 | 0 | 5 |
| 0 | 0 | 1 | 2 |
| 0 | 1 | 0 | 1 |
| 0 | 1 | 1 | 2 |
| 1 | 0 | 0 | 3 |
| 1 | 0 | 1 | 1 |
| 1 | 1 | 0 | 3 |
| 1 | 1 | 1 | 8 |
| | | | N=25 |

For individual panel data, such as presented in Table 1.4, contingency tables of $\Upsilon_{t-1}$ by $\Upsilon_t$ for all time points $t$ can be constructed (dropping subscript $i$ for convenience and $\Upsilon_t$ denoting variable $\Upsilon$ observed at time point $t$). The cells of these tables contain the observed number of cases making all possible transitions. For the data of Table 1.4, with $T = 3$, there are two such tables, presented in Table 1.5 a/b. Anderson (1954) showed how observed transition frequencies, as those in Table 1.5a/b, can be used to obtain maximum likelihood (ML) estimates of the unknown underlying transition probabilities. Anderson and Goodman (1957) formally proved that, for individual panel data following a first-order Markov chain, the transition frequencies are sufficient statistics for the observed $\Upsilon$-sequences such as those presented in Table 1.4. Also, they discussed the asymptotic distributions of the ML estimates of the transition probabilities. Furthermore, they developed several statistical tests, for example tests related to the Markov process being time-varying or time-invariant and to the order of the Markov process.

Table 1.5 Transition frequencies for fictitious individual panel data

(a)

| | | $\Upsilon_2$ | |
|---|---|---|---|
| | | 0 | 1 |
| $\Upsilon_1$ | 0 | 7 | 3 |
| | 1 | 4 | 11 |

(b)

| | | $\Upsilon_3$ | |
|---|---|---|---|
| | | 0 | 1 |
| $\Upsilon_2$ | 0 | 8 | 3 |
| | 1 | 4 | 10 |

The Anderson/Goodman model is rather basic, since it cannot easily deal with a heterogeneous population and hence with different transition probabilities for different strata of the population. At the end of his 1954 article Anderson spends a few words on heterogeneity and suggests his Markov model to be applied to each stratum separately. He concludes with saying that the 'best' stratification variable may not always be observable. The idea to model different transition probabilities for unobservable strata was also applied by Blumen, Kogan and McCarthy (1955) in their 'stayer-mover' model. In this model, two different Markov chains are assumed to be 'mixed' within a population, one chain for 'stayers', who stay in the same state of $\Upsilon$ at each time point and another chain for 'movers', who may switch from one state to another. The stayers and movers can be considered as two unobserved strata, with each individual case having an unknown to be estimated probability of belonging to each stratum.

The idea to deal with heterogeneity in a regression like manner, by exploiting predictor variables that enable different outcomes of the transition probabilities for cases having different predictor values, was introduced some years later. In the election Equation (2), we could for example add Age as a predictor to obtain different transition probabilities for people of different ages, leading to:

$$\mathrm{P}(\Upsilon_{i,t}=1) \;=\; \frac{e^{b_0+b_1\Upsilon_{i,t-1}+b_2\mathrm{Age}_{i,t}+b_3\Upsilon_{i,t-1}*\mathrm{Age}_{i,t}}}{1+e^{b_0+b_1\Upsilon_{i,t-1}+b_2\mathrm{Age}_{i,t}+b_3\Upsilon_{i,t-1}*\mathrm{Age}_{i,t}}}\,. \tag{3}$$

In (3) the interaction product $\Upsilon_{i,t-1} * \mathrm{Age}_{i,t}$ enables a separate age-effect for each value of $\Upsilon_{i,t-1}$, i.e., for the entry and exit probability. Spilerman (1972, p278) elaborates on the advantages of such a regression approach as opposed to the earlier Markov models which required that 'all persons must transfer according to an identical transition array'. He employs a linear regression model to explain and 'project' (forecast) transitions of persons between geographic regions, with individual (e.g. occupation, race, age) as well as macro (city size) predictor variables. Amemiya (1985) speaks of 'exogenous variables' and Diggle, Liang and Zeger (1994) of 'covariates' to be used in Markov models as a way to deal with a population consisting of heterogeneous individuals. Cook and Ng (1997) and Stott (1997) further extend the regression approach by adding an (unobserved) random normal deviate to the logits of the entry and exit probabilities to account for even more, i.e. unobserved, heterogeneity, in addition to the heterogeneity modelled by the (observed) predictors.

Another powerful extension of the individual panel data Markov model has been proposed by Wiggins (1973), and is known as the latent Markov model. This model is meant to deal with possible measurement error in the observed values of $\Upsilon$. Instead of transitions between the states of the observed $\Upsilon$, the model pertains to transitions between the states of the 'true' but latent variable $Z$ that determines the outcome of the observed value of $\Upsilon$. Also, two or more $\Upsilon$ variables can be used as so-called 'indicator variables' of the true unobserved $Z$, as in linear structural equation models. Building on the work of Hagenaars (1990), Vermunt, Langeheine and Böckenholt (1999) show how a latent Markov model can be constructed that incorporates covariates affecting the latent variables.

### 1.2.2 Models for repeated aggregated proportions

Individual panel data or, to use the words of Lee, Judge and Zellner (1968, p1163) 'time-ordered data which reflects the movements of the micro units' are not always available. Instead, in many research problems the data pertain to higher level or macro units, such as geographical areas, voting districts, companies, schools etc., observed at a number of consecutive time points. More specifically, the data consist of aggregated proportions or frequency counts of individual cases in each state of a categorical variable $\Upsilon$ for each of a number macro units at two or more, evenly spaced, moments in time. As an example, suppose one is interested in voting transitions for an election taking place every four years. Assume that there are only two, and always the same two, political parties, 0 and 1 say, for voters to choose from at each separate election. For all voting districts of a particular city, the proportions of people voting for the two parties are known for three consecutive elections. For any given voting district, these proportions can be presented as in Table 1.6, where they form the margins of the subtables. Note that in Table 1.6a the column margins .4 and .6 are equal to the row margins of Table 1.6b, since both refer to $\Upsilon_2$, i.e., the results of the second election. Presenting the data this way, as was also done in Table 1.5 for individual panel data, points to the fact that the transition proportions are unknown, as symbolized by the question marks in the interior cells. Only the

Table 1.6 Repeated aggregated proportions for a fictitious voting district at three consecutive elections

(a)

|  |  | $\Upsilon_2$ |  |  |
|---|---|---|---|---|
|  |  | 0 | 1 |  |
| $\Upsilon_1$ | 0 | ? | ? | .8 |
|  | 1 | ? | ? | .2 |
|  |  | .4 | .6 | N=100 |

(b)

|  |  | $\Upsilon_3$ |  |  |
|---|---|---|---|---|
|  |  | 0 | 1 |  |
| $\Upsilon_2$ | 0 | ? | ? | .4 |
|  | 1 | ? | ? | .6 |
|  |  | .3 | .7 | N=100 |

'marginal' outcome of each election is observed but the individual voting-sequences, like those in Table 1.4, are not. Hence, the transition frequencies/proportions of individuals making any particular transition cannot be inferred from the unobserved sequences. For individual panel data, precisely these transition proportions provide the information necessary for estimating the unknown transition probabilities. Thus, compared to individual panel data, in models for repeated aggregated proportions a lesser amount of data is observed and as a result the estimation of the entry and exit probabilities is less straightforward. The lack of data in each separate macro unit is compensated, however, by obtaining aggregated proportions for (as) many macro units (as possible), all of which are thought to be governed by transition rules or probabilities that are either the same for all or in some way related to each other.

When using repeated aggregated proportions specific attention should be given the following. Consider the individual cases that make up a particular macro unit at each time point. In the most ideal situation the macro unit consists of the same individuals at all time points involved. For voting districts in the above example this would mean that there is no inflow (coming of voting age, immigration) or outflow (dead, emigration) of voters during the eight year period covering the three elections. Obviously, such ideal situation will seldom occur in practice: due to in- and outflow the macro units do not consist of the same (number of) individuals at each point in time. In the most extreme situation all individuals in a macro unit will have been replaced by others at the next point in time. The models for repeated aggregated proportions can nevertheless be applied in situations where in- and outflow occurs, provided that two assumptions are satisfied. These are discussed below.

The individuals making up a given macro unit at time point $t-1$ can be split up in a stay-group and an outflow-group: those who belong to the stay-group are still in the macro unit at the next time point $t$, while those in the outflow-group are not. The first assumption to be satisfied concerns the outflow-group. At time point $t-1$, the $\Upsilon$ proportions observed for the entire macro unit must also apply to the outflow-group. To put it differently, the stay-group and the outflow-group at time point $t-1$ must be homogeneous in their $\Upsilon$ proportions, which for both groups have to be equal to the observed proportions of the entire macro unit at $t-1$. Furthermore, at the next time point $t$, the individuals who then make up the macro unit, consist in part of an inflow-group: these are the individuals who were not yet in that same unit at the previous time point $t-1$. The second assumption concerns this inflow-group. The inflow-group of a given macro unit at time point $t$ had (or would have had if, e.g., they would have been of voting age) the same $\Upsilon$ proportions at time point $t-1$ as the proportions actually observed for that macro unit at time point $t-1$. If both assumptions are met, the models for repeated aggregated proportions can be applied even if the macro units involved are subject to inflow and outflow of individual cases.

From the many publications on Markov models dealing with repeated aggregated proportions during the last 50 years, we selected a few that, in our view, can be considered important for the development of this branch of models. Again, Goodman (1953, 1959) appears as a pioneer. Inferring micro characteristics (here: transition probabilities which are thought to hold for each individual in each macro unit) from macro data (here: observed proportions) is an example of what is often called 'ecological inference'. Goodman's approach is known as 'ecological regression'. From Table 1.6a the following equation can be derived between the left column margin and the row margins, involving the transition probabilities $1-\mu_2$ and $\lambda_2$ that were introduced before: $.4 = .8(1-\mu_2) + .2\lambda_2$. If the observed column marginal .4 does not depend totally on the values of $1-\mu_2$ and $\lambda_2$ but also contains some error, then the equation becomes $.4 = .8(1-\mu_2) + .2\lambda_2 + e_i$, with $e_i$ expressing the error part for voting district $i$, say. A similar expression can be given for the other districts, with different marginal values and errors but with the same parameters $1-\mu_2$ and $\lambda_2$ that, together with the error, 'generate' the value of the left column margin given the row margins. Thus, $1-\mu_2$ and $\lambda_2$ can be thought of as the unknown parameters in the linear regression of the left marginal column proportion on the two marginal row proportions of Table 1.6a. The same holds for $1-\mu_3$ and $\lambda_3$ and the corresponding margins of Table 1.6b. Goodman specifies conditions under which this ecological regression method can be meaningfully applied to estimate the unknown transition probabilities. However, the method suffers from a number of shortcomings, that are very carefully pointed out by King (1997).

Lee, Judge and Zellner (1968) describe a maximum likelihood (ML) procedure to estimate the transition probabilities. They argue that for a given macro unit the observed totals (i.e., numbers of individual cases) in all categories of $\Upsilon$ at time point $t$, given those at time point $t-1$, can be regarded as if they arise from a multinomial distribution. In terms of Table 1.6a this would for example imply that, given the row proportions, the distribution of the left column total is binomial$(n,p)$ with $n=100$ and $p=.8(1-\mu_2)+.2\lambda_2$. So, for each voting district in the city under study the $\Upsilon_2$ frequencies follow a binomial distribution with the success probability depending on the observed $\Upsilon_1$ proportions and the unknown $\mu_2$ and $\lambda_2$. Given the observed proportions of $\Upsilon_1$ and $\Upsilon_2$ of all voting districts in the city, ML estimates of $\mu_2$ and $\lambda_2$ (and $\mu_3$ and $\lambda_3$) can be obtained following the procedure described by the authors. Building on the supposed binomial (or multinomial) distribution of the category totals of $\Upsilon_t$, given those of $\Upsilon_{t-1}$, they also develop a Bayesian approach in which they use a multivariate beta distribution as a prior for the transition probabilities.

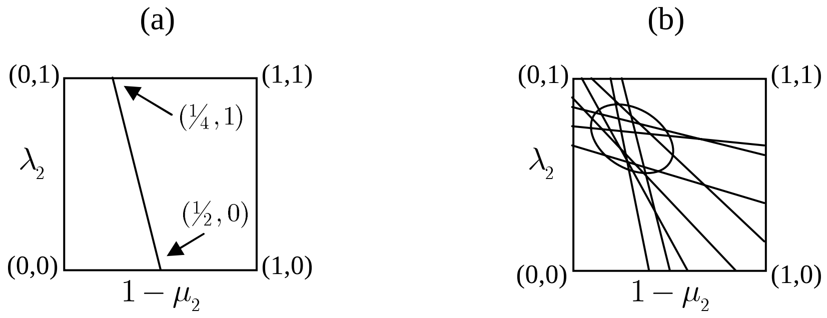
Hawkes (1969) shows that the distribution of the observed totals in all categories of $\Upsilon_t$, given those of $\Upsilon_{t-1}$, is not a simple multinomial but a weighted sum of multinomials. Translated to Table 1.6a, the distribution of the left column total, given the two row totals, is binomial$(80,1-\mu_2)$ + binomial$(20,\lambda_2)$ being the sum of the unobserved upper and lower left cell frequencies which are thought to be binomial$(80,1-\mu_2)$ and binomial$(20,\lambda_2)$ respectively. Hawkes (1969) also derives the means and (co)variances of the joint distribution of all $\Upsilon_t$ category totals, and approximates this distribution by a multivariate normal which forms the base of his ML approach. Also, he proposes two other procedures, one being a slightly adjusted form of Goodman's ecological regression, accounting for different 'district' sizes. Hawkes's third model considers $\mu_2$ and $\lambda_2$ for a particular district to be randomly drawn from some joint distribution. The expectation and covariance matrix of this distribution are the statistics to be estimated.

An important contribution has also been made by MacRae (1977). She extends the notion of the exact distribution of the observed totals in all categories of $\Upsilon_t$ given those of $\Upsilon_{t-1}$. The distribution is described in detail and has come to be known as the 'convolution of multinomials'. A special instance of this distribution, the 'convolution of two binomials' is discussed by McCullagh and Nelder (1989) in a ML setting. Furthermore, MacRae (1977) shows how to account for heterogeneity by using exogenous macro level variables in the ML approach; these variables are allowed to affect the transition probabilities using a logistic link function. Brown and Payne (1986) use the same convolution of multinomials but they compound it with a Dirichlet distribution to account for unobserved heterogeneity, thereby giving rise to what they have called the 'aggregate compound multinomial'.

We would like to mention two more recent contributions to the estimation of transition probabilities for repeated aggregated proportions. The first one uses the maximum entropy criterion which is closely related to the information-theory developed by Shannon (1948) and Jaynes (1957). Willekens (1982) shows how, in the analysis of multiway contingency tables, the maximization of entropy can be used to obtain expected cell frequencies if only marginal totals are known. Golan, Judge and Miller (1996) describe entropy more generally as a criterion that is especially appropriate to solve 'ill-posed' problems that arise in situations where data are 'limited, partial, aggregated and incomplete'. The related criterion of cross-entropy can be used when prior information about transition probabilities is available. Karantininis (2002) applies this approach to model size-changes over time of pork farms in Denmark during the period 1984-1998. Also, the effects of exogenous macro-level variables on transition probabilities can be assessed. Karantininis (2002) investigates effects of pig feed prices, pork prices and interest rate on the transition probabilities.

The second more recently developed approach is King's (1997) so-called 'ecological inference' (EI) method. For each 'district' the true values of $1-\mu_2$ and $\lambda_2$ range from 0 to 1. Hence, they can be represented by a point located somewhere in the unit square formed by a horizontal axis representing the value of $1-\mu_2$ and a vertical axis representing the value of $\lambda_2$, as shown in Figure 1.3a. For the district of Table 1.6, the relation $.4 = .8(1-\mu_2) + .2\lambda_2$ (between the left column and the two row propor-

Figure 1.3 Tomography lines in the unit square



tions of Table 1.6a) implies that for this district the point in the unit square must be located on the line given by equation $\lambda_2 = 2 - 4(1-\mu_2)$. The intersection of this line and the unit square is drawn in Figure 1.3a between the points $(\frac{1}{4},1)$ and $(\frac{1}{2},0)$. King calls this line 'tomography line' for a given district. In Figure 1.3b the tomography lines of all 8, say, districts of a city are shown, representing all possible true values of $1-\mu_2$ and $\lambda_2$ for each district. There's a high concentration of lines in the upper left corner of the square, especially in the area bounded by the ellipse. Following King, this would indicate that, for each district, a low value of $1-\mu_2$ and a high value of $\lambda_2$ is more likely than, say, the reverse. To determine which values of $1-\mu_2$ and $\lambda_2$ are most likely, King assumes the district values of $1-\mu_2$ and $\lambda_2$ to be drawn from a truncated bivariate normal distribution (TBN). This distribution can be imagined as a Gaussian hat rising above the unit square, being higher for more likely combinations of $1-\mu_2$ and $\lambda_2$ and lower for less likely ones. The distribution is truncated since it has to fit in the unit square. To estimate the exact form and place of the TBN above the square the locations of all tomography lines in the unit square are used. The ellipse in Figure 1.3b actually represents a contour line of the TBN after form and place were estimated for the given eight tomography lines. Each line segment located within the ellipse area defines the most likely values of $1-\mu_2$ and $\lambda_2$ for the corresponding district. What is more, imagine a plane perpendicular to the unit square passing through the tomography line for a given district. Also, imagine the intersection of this plane with the TBN, like a very thin piece of a pie. Projecting this intersection on the axes of $1-\mu_2$ and $\lambda_2$ yields the so-called posterior distribution of both $1-\mu_2$ and $\lambda_2$ for the district in question, the modes of which correspond to the most likely values. Thus, King's EI procedure not only provides information about the entry and exit rates averaged over all districts, but also about the two rates in each separate district.

A more detailed discussion of other specific models for repeated aggregated proportions is given by King, Rosen and Tanner (2004). More important, the authors present an overview of the latest developments in this field. Also, they show many applications to real data and evaluations of models using artificial data.

### 1.2.3 Models for repeated cross sections

Repeated cross sections (RCS) data consist of a number of cross sections independently sampled at consecutive time points. The separate cross sections are usually made up of different individual cases for each of which the value of $\Upsilon$ is observed. As opposed to repeated aggregated proportions, RCS data refer to data from individual cases, as do individual panel data. In contrast with the other data types, in RCS data there is typically only one observation available for each (individual) unit in each cross section. If the population is small in relation to the sample size, it may happen that the same individuals are sampled in two or more cross sections. These individuals, however, cannot be traced and consequently, a Markov model cannot borrow strength in any way of such 'hidden individual panel data' being part of the repeated cross sections.

Suppose there are three cross sections, A, B and C say, observed at time point 1, 2 and 3 respectively. The $\Upsilon$ data observed can then be represented as in Table 1.7 where they constitute the margins of the three subtables labelled A, B and C to refer to the corresponding cross section.

Table 1.7 Repeated cross sections at three time points

| (A) | $\Upsilon_1$ |  |
|---|---|---|
|  | 0 | 1 |
|  | 80 | 20 |

| (B) |  | $\Upsilon_2$ |  |  |
|---|---|---|---|---|
|  |  | 0 | 1 |  |
| $\Upsilon_1$ | 0 | ? | ? | ? |
|  | 1 | ? | ? | ? |
|  |  | 48 | 72 |  |

| (C) |  | $\Upsilon_3$ |  |  |
|---|---|---|---|---|
|  |  | 0 | 1 |  |
| $\Upsilon_2$ | 0 | ? | ? | ? |
|  | 1 | ? | ? | ? |
|  |  | 60 | 140 |  |

These margins represent the numbers of individual cases in each $\Upsilon$ category at each point in time, that is, in each cross section.

As for Table 1.6 with aggregated proportions, presenting RCS data by way of the empty cross tables of $\Upsilon_{t-1}$ by $\Upsilon_t$, as in subtables 1.7B and 1.7C, reveals many unknown quantities in RCS data as compared to individual panel data. In addition to these cross tables, the columns of which contain the $\Upsilon$ totals of cross sections B and C, an extra frequency table 1.7A is needed to show the totals of the first cross section A. Note that, with RCS data, not only are the transition frequencies unknown for each particular transition, but also unknown are the $\Upsilon$ totals at time point $t-1$ of the cross section that is actually observed at time point $t$. Thus, for cross section B only the totals 48 and 72 at $t=2$ are observed and nothing is known about the $\Upsilon$ values of the 120 individuals of cross section B at time point 1. Therefore, from the point of view of studying transitions, of the three data types considered here, RCS data have the highest level of missing data, hence higher than repeated aggregated proportions for which at least both margins of the cross tables of $\Upsilon_{t-1}$ by $\Upsilon_t$ are observed.

Although many models and techniques have been developed specifically for analyzing RCS data[1], this is not true for Markov chain models. Given the large number of unknown quantities in RCS data, this is not surprising. In contrast with this scarcity of models however is the abundance of RCS data (with the same variables observed in every cross section) in many different research areas. Clearly, the availability of categorical Markov models would create new perspectives for analyzing these data, with attention also focused on the dynamics (i.e., the transitions) behind the observable states of individuals over time. The model presented in the chapters to follow illustrates the great but rather unexplored potential of RCS data for studying individual transitions over time.

The basic version of the RCS Markov model for categorical data was proposed by Robert Moffitt (1990) who shows how data of even a single cross section can be exploited to model transitions over time, by using the ages of the individuals in the cross section to represent the time axis along which transitions take place. Moffitt applies this 'one cross section' version of the model to investigate how a person's marital status is influenced by

[1] See e.g. special issue "Analysis of repeated cross-sectional data" of Statistica Neerlandica, Journal of the Netherlands Society for Statistics and Operations Research (2001, volume 55, nr. 2).

the amount of benefit received from the U.S welfare system. Since, in general, only female-headed households with no able-bodied male present are eligible, negative effects were expected of benefit rate on the probability to transit from 'unmarried' to 'married' and positive effects for transition in the reverse direction. Moffitt shows that the model indeed is able to display such effects, using data from a single cross section of the U.S. Current Population Survey of 1985.

Later (Moffitt 1993) essentially the same model is presented, but now for data of *repeated* cross sections. Instead of respondent's age, the time axis now refers to the time points the cross sections were observed. The model is applied to investigate female labor supply using 21 annual cross sections of the U.S. Current Population Survey over the period 1968-1988. Transitions between the states 'employed' and 'unemployed' of women (white, married, aged 20-59) are explained by a number micro and macro predictor variables. Furthermore, it is interesting to note that in the same paper, Moffitt puts the model in a broader perspective of different types of models to be applied to RCS data, one of which is an (autoregressive) Markov model for interval level ϒ data.

We have extended Moffitt's model for categorical data in a number of ways, each of which is described in the following chapters and illustrated with an application to real data. Also, we developed standalone computer software for applying the model, the user manual of which is included in Appendix 1. We shall now briefly discuss the model and offer an example application.

## 1.3 Repeated cross sections Markov model for categorical data

In the presentation of the model that follows now, we first show a most simple model. This one uses data from three cross sections and has no predictor variables. The basic model equation is explained and a straightforward estimation procedure of the unknown transition probabilities is shown. The next model presented is about dealing with more then three

Table 1.8 Unknown probabilities for three time points

| $\Upsilon_1$ | |
|---|---|
| 0 | 1 |
| $1-p_1$ | $p_1$ |

| | | $\Upsilon_2$ | | |
|---|---|---|---|---|
| | | 0 | 1 | |
| $\Upsilon_1$ | 0 | $1-\mu$ | $\mu$ | $1-p_1$ |
| | 1 | $\lambda$ | $1-\lambda$ | $p_1$ |
| | | $1-p_2$ | $p_2$ | |

| | | $\Upsilon_3$ | | |
|---|---|---|---|---|
| | | 0 | 1 | |
| $\Upsilon_2$ | 0 | $1-\mu$ | $\mu$ | $1-p_2$ |
| | 1 | $\lambda$ | $1-\lambda$ | $p_2$ |
| | | $1-p_3$ | $p_3$ | |

cross sections and, again, using no predictor variables. At that place we introduce the maximum likelihood estimation procedure. Finally, we discuss an application using real data of five cross sections and show how predictor variables can be incorporated into the model.

### 1.3.1 Three cross sections, no predictor variables

Suppose there are three cross sections A, B and C, say, observed at time point 1, 2 and 3, respectively. Also assume that the unknown true values of the entry ($\mu_{i,t}$) and exit ($\lambda_{i,t}$) probabilities are time-invariant and homogeneous. Hence, we omit the subscripts to indicate time point and individual and simply use $\mu$ and $\lambda$. Also, we use $p_1$, $p_2$ and $p_3$ to denote the 'state 1' probabilities at time point 1, 2 and 3 respectively. The probabilities can be presented in (cross) tables, one for each point in time, as in Table 1.8.

In the title, nor in the subtabels of Table 1.8, reference is made to the observed cross sections A, B and C. This is because the probabilities $p$, $\mu$ and $\lambda$ do not pertain to a particular cross section but, instead, to the time-points that are denoted by the subscript of $\Upsilon$ above the subtables. For example, $p_2$ denotes the probability for an individual to be in state 1 of $\Upsilon$ at timepoint $t$=2, no matter what particular cross section that individual belongs to. Also, $\mu$ and $\lambda$ denote the entry and exit probabilities for $t$=2 and $t = 3$ for all individuals of all three cross sections.

From the middle subtable of Table 1.8 one can deduce that for the probability $p_2$ to be in state 1 at $t = 2$ the equation $p_2 = p_1(1-\lambda)+(1-p_1)\mu$ must hold. Likewise, from the right subtable of Table 1.8 it can be derived that $p_3 = p_2(1-\lambda)+(1-p_2)\mu$. In general, for $t \geq 2$, it holds that:

$$p_t = p_{t-1}(1-\lambda) + (1-p_{t-1})\mu \,. \tag{4}$$

Equation (4) is the kernel of the RCS Markov model. For individual heterogeneous and time variant models, subscripts for individuals and time points could be added to $p$, $\mu$ and $\lambda$. According to Equation (4) the probability $p_t$ for an individual to be in state 1 at a given time point $t$ is the sum of $p_{t-1}(1-\lambda)$, i.e., the probability of being in state 1 at the previous time point and staying there until the next, and of $(1-p_{t-1})\mu$, the probability of not being in state 1 at the previous time point but switching to it at the next. Hence the two 'previous' probabilities $p_{t-1}$ and $1-p_{t-1}$ and the two transition probabilities $\lambda$ and $\mu$ together determine the 'next' probability $p_t$ to be in state 1. Obviously, the fact that the previous state of $\Upsilon$ is not observed in RCS data implies that in the model equation, both previously possible $\Upsilon$ states must be taken into account as well as both transitions, that can lead to state 1 at the next time point. It is useful to compare Equation (4) with Equation (5) below, which could be applied if individual panel data would be available:

$$p_t = \Upsilon_{i,t-1}(1-\lambda) + (1-\Upsilon_{i,t-1})\mu \,. \tag{5}$$

In individual panel data, instead of $p_{t-1}$ in (4), the observed value of $\Upsilon_{i,t-1}$ can be used in the equation and consequently (5) results in either $p_t = 1-\lambda$ or $p_t = \mu$, depending on the observed value of $\Upsilon_{i,t-1}$ being 1 or 0. Thus, applying (5) to individual panel data boils down to dividing the individuals into two groups, depending on the value of $\Upsilon_{i,t-1}$; each group has its own probability $1-\lambda$ or $\mu$ to be in state 1 at time point $t$ and this state 1 probability is actually the transition probability one is interested in. (Equation (2) is an example of applying (5) to individual panel data using a logistic expression for $p_t$.) In contrast to individual panel data, in RCS data the value of $\Upsilon_{i,t-1}$ is not observed and therefore dividing the individuals on the basis of such value is impossible. Consequently, following Equation (4), there is only one probability $p_t$ which applies to all individuals, while there are two transition probabilities $1-\lambda$ and $\mu$ which also apply to all individuals. One could argue that this is the price to pay for having cross-sectional data only, which enable the use of (4) as a model equation but not (5). However, as will be shown below and in the chapters to follow, the more macro and/or micro level covariate information about the individuals in the cross sections is available, the less of a problem this limitation of RCS data becomes.

Note that in (4) each 'next' probability $p_t$ is thought to directly depend only on the immediately preceding probability $p_{t-1}$ and not on $p_{t-2}$, $p_{t-3}$ etc. Hence, (4) is a first order Markov model equation. Following (4), for $p_t$ one may write $p_t = p_{t-1}(1-\lambda-\mu)+\mu$ and thus for $p_2$ and $p_3$ it holds that

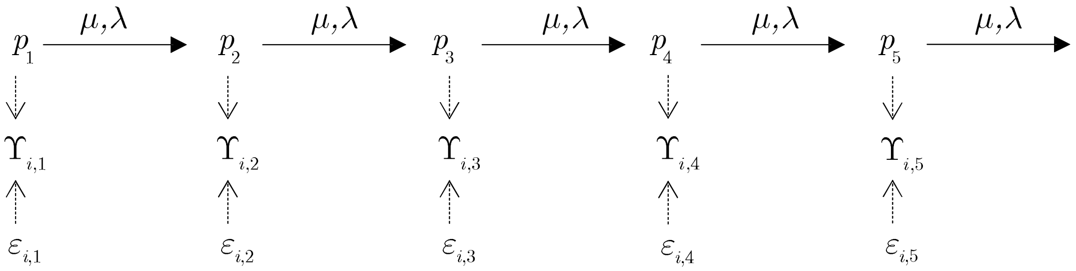
$$p_2 = p_1(1-\lambda-\mu)+\mu\,, \tag{6}$$
$$p_3 = p_1(1-\lambda-\mu)^2+\mu(1-\lambda-\mu)+\mu\,. \tag{7}$$

Equation (7) results if, in $p_3 = p_2(1-\lambda-\mu)+\mu$, one substitutes for $p_2$ the expression $p_1(1-\lambda-\mu)+\mu$ given in (6). From (6) and (7) it follows that, given the values of $p_1$, $\mu$ and $\lambda$, those of $p_2$ and $p_3$ automatically result. That is, the probability to be in state 1 on the first time point and the two transition probabilities together determine the probability to be in state 1 at the two subsequent time points. There are only two subsequent state 1 probabilities here, $p_2$ and $p_3$, but the same would hold for $p_4$, $p_5$, etc. In general, the following expression for $p_t$ applies, for $t \geq 2$:

$$p_t = p_1(1-\lambda-\mu)^{t-1}+\mu\sum_{\tau=0}^{t-2}(1-\lambda-\mu)^{\tau}\,. \tag{8}$$

Equation (8) shows that each subsequent $p_t$ can be reduced to $p_1$, $\mu$ and $\lambda$. These last three probabilities are therefore the quantities of interest to be estimated. The reduction of each subsequent $p_t$ to $p_1$, $\mu$ and $\lambda$ is visualized in Figure 1.4. The figure shows a diagram of the RCS Markov model for the first five time points of a potentially longer series of observations. The solid arrows represent the influence of the immediately preceding state 1 probability on the next, with $\mu$ and $\lambda$ as parameters determining the strength of this influence. Note e.g. that, if $1-\lambda=\mu$, from Equation (4) if follows that $p_t=\mu$ for all $t \geq 2$ so that $p_t$ only depends on $\mu$ and not on $p_{t-1}$, which would imply that $\Upsilon_{i,t}$ and $\Upsilon_{i,t-1}$ are independent. Also, if $\mu=\lambda=0$, then from (4) it follows that $p_t=p_{t-1}$ for all $t \geq 2$, implying that the influence of $p_{t-1}$ on $p_t$ is at its maximum or that $\Upsilon_{i,t-1}$ completely determines $\Upsilon_{i,t}$. The dotted arrows are related to the fact that $\Upsilon_{i,t}$ can be expressed as $\Upsilon_{i,t}=p_t+\varepsilon_{i,t}$, i.e., both $p_t$and $\varepsilon_{i,t}$ determine the value of

Figure 1.4 Diagram of the first order RCS Markov model



$\Upsilon_{i,t}$. In the diagram $p_5$ is completely determined by $p_4$ (i.e., there is no other arrow pointing to $p_5$), $p_4$ completely by $p_3$ etc. all the way back to $p_1$ and hence $p_5$ can be expressed completely in terms of $p_1$, $\mu$ and $\lambda$.

Given data for only three cross sections, a fairly simple procedure for estimating $p_1$, $\mu$ and $\lambda$ is to employ the sample moments $\bar{\Upsilon}_1$, $\bar{\Upsilon}_2$ and $\bar{\Upsilon}_3$ as estimates of the true or population moments $p_1$, $p_2$ and $p_3$, respectively, where $\bar{\Upsilon}_t$ denotes the mean $\Upsilon$-value for the individuals of the cross section observed at $t$, which is equal to the observed state 1 proportion for that cross section. Doing so, one obtains the following equations:

$$\bar{\Upsilon}_1 = \hat{p}_1, \tag{8}$$
$$\bar{\Upsilon}_2 = \hat{p}_1(1 - \hat{\lambda} - \hat{\mu}) + \hat{\mu}, \tag{9}$$
$$\bar{\Upsilon}_3 = \hat{p}_1(1 - \hat{\lambda} - \hat{\mu})^2 + \hat{\mu}(1 - \hat{\lambda} - \hat{\mu}) + \hat{\mu}. \tag{10}$$

with (9) and (10) following from (6) and (7). Solving these equations for the unknowns $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ yields the estimates

$$\hat{p}_1 = \bar{\Upsilon}_1, \tag{11}$$
$$\hat{\mu} = \bar{\Upsilon}_2 - (\bar{\Upsilon}_3 - \bar{\Upsilon}_2)(\bar{\Upsilon}_2 - \bar{\Upsilon}_1)^{-1}\ \bar{\Upsilon}_1, \tag{12}$$
$$\hat{\lambda} = 1 - \bar{\Upsilon}_2 - (\bar{\Upsilon}_3 - \bar{\Upsilon}_2)(\bar{\Upsilon}_2 - \bar{\Upsilon}_1)^{-1}\ (1 - \bar{\Upsilon}_1). \tag{13}$$

For the state 1 proportions $\bar{\Upsilon}_1 = .2$, $\bar{\Upsilon}_2 = .6$ and $\bar{\Upsilon}_3 = .7$ of the data in Table 1.7, applying (12) and (13) results in the estimates $\hat{\mu} = .55$ for the entry and $\hat{\lambda} = .20$ for the exit probability.

This simple example demonstrates that, as Moffitt (1990) has noted, it is actually possible to estimate a dynamic model using cross-sectional data, given that certain assumptions are met. A crucial assumption made was the time invariance of $\mu$ and $\lambda$ which, together with the assumption of individual homogeneity, resulted in only two unknown transition probabilities to be estimated, instead of the four $\hat{\mu}_2$, $\hat{\mu}_3$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ if only individual homogeneity would have been assumed, but time varying entry and exit probabilities. However, with the only unknowns being $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ the equations (8), (9) and (10) can be solved. The time invariance assumption would theoretically speaking not be needed for a Markov model applied to individual panel data. However, in applications of Markov models, assumptions regarding time invariance of the transition probabilities are very often made, for example to achieve parsimonious models. Furthermore, as will be shown below, for the RCS Markov model the time invariance assumption can be relaxed if covariates are incorporated into the model.

We previously noted that the marginal proportions at $t-1$ are unknown for the cross section observed at $t$. For the data of Table 1.7 the marginal frequencies for ($\Upsilon = 0$, $\Upsilon = 1$) are (80,20), (48,72) and (60,140) for cross section A, B and C, respectively. For the 120 cases of cross section B one may for example guess that, at $t = 1$, 80% of them were in state 0 and 20% in state 1. In other words, one may assume that, at $t = 1$, the $\Upsilon$ proportions of cross section B are equal to the $\Upsilon$ proportions observed at $t = 1$ for cross section A. In the RCS Markov model a similar assumption is made, be it not with respect to the unknown proportions of $\Upsilon$ at $t = 1$ of cross section B, but with respect to the underlying probability $p_1$ to have been in state 1 at time point 1. That is, the cases of cross section B are assumed to have had the same probability to be in state 1 at $t = 1$ as the cases of cross section A had at the moment they were actually observed. The same assumption is made for the value of $p_1$ of the cases of cross section C; also, for these cases the value of $p_2$ is assumed to be equal to the value of $p_2$ for the cases of cross section B.

It is relevant to note that evaluating (12) and (13) does not always produce estimates $\hat{\mu}$ and $\hat{\lambda}$ that lie in the feasible range (0,1). If the true data generating process does not behave as was assumed, i.e., first order Markov, individually homogeneous and time invariant, one could find values for $\hat{\mu}$ and/or $\hat{\lambda}$ outside the range (0,1). A few examples in which this is the case are worth to consider in some detail.

Suppose that for the means of the three cross sections it holds that $(\overline{\Upsilon}_3 - \overline{\Upsilon}_2)(\overline{\Upsilon}_2 - \overline{\Upsilon}_1)^{-1} = 1$, i.e., $\overline{\Upsilon}_3 - \overline{\Upsilon}_2 = \overline{\Upsilon}_2 - \overline{\Upsilon}_1$. This would imply that the proportion individuals in state 1 increases (or decreases) over time with a constant value, as e.g. with $\overline{\Upsilon}_1 = .6$, $\overline{\Upsilon}_2 = .7$ and $\overline{\Upsilon}_3 = .8$. Such a linear proportional growth of individuals in state 1 cannot be the result of a first order, time invariant, individual homogenous Markov process. This can be explained as follows. Suppose the values of $p_3$, $p_2$ and $p_1$ are such that $p_3 - p_2 = p_2 - p_1 = c$ with $c$ being some constant for which $-.5 \leq c \leq .5$ holds. Since both first order equations $p_2 = p_1(1-\lambda) + (1-p_1)\mu$ and $p_3 = p_2(1-\lambda) + (1-p_2)\mu$ should hold, subtracting the first from the second implies that $p_3 - p_2 = (p_2 - p_1)(1-\lambda) + (p_1 - p_2)\mu$ should hold and thus also that $c = c(1-\lambda) - c\mu$ and $\mu = -\lambda$ should hold, which of course could only be the case if either $\mu < 0$ or $\lambda < 0$.

As another example, suppose that $\overline{\Upsilon}_1 = \overline{\Upsilon}_2 \neq \overline{\Upsilon}_3$ or in terms of the population means $p_1 = p_2 \neq p_3$. The fact that $\overline{\Upsilon}_2 - \overline{\Upsilon}_1 = 0$ would cause a division by zero in (12) and (13) and would thus result in no solution for $\hat{\mu}$ and $\hat{\lambda}$. Again, the supposed sequence of means $p_1 = p_2 \neq p_3$ could never arise from the assumed Markov process, since if so, the equations $p_1 = p_2 = p_1(1-\lambda) + (1-p_1)\mu$ and $p_3 = p_2(1-\lambda) + (1-p_2)\mu = p_1(1-\lambda) + (1-p_1)\mu$ should both be valid and hence it should be so that $p_1 = p_3$ which is contradictory to $p_1 \neq p_3$.

As a last example, suppose $\overline{\Upsilon}_1 = \overline{\Upsilon}_2 = \overline{\Upsilon}_3$ or in words: the proportional distribution of $\Upsilon$ is in equilibrium. Again, division by zero due to $\overline{\Upsilon}_2 - \overline{\Upsilon}_1 = 0$ causes (12) and (13) to not yield a solution for $\hat{\mu}$ and $\hat{\lambda}$, which should be interpreted here as 'no unique solution'. This can be understood as follows. Suppose $p_1 = p_2 = p_3 = .2$. Knowing the true value of .2 does not provide enough information to conclude anything about the true values of $\mu$ and $\lambda$. As long as the first order equations $p_t = p_{t-1}(1-\lambda) + \left(1-p_{t-1}\right)\mu$ hold, which for both $t = 2$ and $t = 3$ result in $.2 = .2(1-\lambda) + .8\mu$ or $\lambda = 4\mu$, any pair of values for $\mu$ and $\lambda$ is equally acceptable for which $\mu \leq .25$ and $0 \leq \lambda \leq 1$ (the constraint on the value of $\mu$ follows from $.2 = .2(1-\lambda) + .8\mu$ implying that both $.2(1-\lambda) \leq .2$ and $.8\mu \leq .2$ must be true, resulting in $\lambda \leq 1$ and $\mu \leq .25$). Taking e.g. the values $\mu = .25$, $\lambda = 1$ and $p_1 = .2$ results in the equilibrium $p_2 = p_3 = ... = .2$, as would the values $\mu = .1$ and $\lambda = .4$ do or the values $\mu = 0$, $\lambda = 0$. In this last situation there would be no switching from 0 to 1 or from 1 to 0, meaning that each individual simply remains in the same state over time which is the most obvious way to achieve an equilibrium in $p_t$. The conclusion that the time invariant individual homogenous model does not produce unique estimates of the transition probabilities in case of an equilibrium in $p_t$ is not to be considered a serious drawback in the context of social research: the behaviour of many individuals over a longer time period will hardly ever be ruled by such a simple Markov process.

### 1.3.2 More than three cross sections, no predictor variables, maximum likelihood

We now turn to a situation in which there are data of four cross sections available, the extension to five or more being straightforward. Suppose that, in addition to the data presented in Table 1.7, there is a fourth cross section consisting of 100 individuals, say, 75 of which are in state 1. As before, the transition probabilities are considered time invariant and individual homogeneous. Then one can write for $p_4$:

$$p_4 = p_1(1-\lambda-\mu)^3 + \mu(1-\lambda-\mu)^2 + \mu(1-\lambda-\mu) + \mu \, . \tag{14}$$

There are three equations now, (6), (7) and (14), that express each $p_t$, with $t \geq 2$, as a function of $p_1$, $\mu$ and $\lambda$. To obtain estimates for $p_1$, $\mu$ and $\lambda$, one could use the sample means $\overline{\Upsilon}_1$, $\overline{\Upsilon}_2$, $\overline{\Upsilon}_3$ and $\overline{\Upsilon}_4$ as estimates of $p_1$, $p_2$, $p_3$ and $p_4$. This would again yield (8), (9) and (10) for $\hat{p}_1$, $\hat{p}_2$ and $\hat{p}_3$, while for $\hat{p}_4$ the equation

$$\overline{\Upsilon}_4 = \hat{p}_1(1-\hat{\lambda}-\hat{\mu})^3 + \hat{\mu}(1-\hat{\lambda}-\hat{\mu})^2 + \hat{\mu}(1-\hat{\lambda}-\hat{\mu}) + \hat{\mu} \tag{15}$$

arises which is the sample equivalent of (14). In total then, there would be four equations in only three unknowns $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$, and hence a unique and perfectly fitting solution for $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ generally does not exist. It would exist only if for the estimates $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$, as could be obtained from only evaluating (11) through (13), Equation (15) would happen to hold. Usually, this will not be the case due to sampling error in the cross sections, causing the sample moments $\overline{\Upsilon}_t$ to deviate from the corresponding probabilities $p_t$. To account for this error in $\overline{\Upsilon}_t$, the expression for each of the four $\overline{\Upsilon}_t$ is extended with an error term $\hat{e}_t$:

$$\overline{\Upsilon}_1 = \hat{p}_1 + \hat{e}_1 , \tag{16}$$
$$\overline{\Upsilon}_2 = \hat{p}_1(1-\hat{\lambda}-\hat{\mu}) + \hat{\mu} + \hat{e}_2 , \tag{17}$$

$$\bar{\Upsilon}_3 = \hat{p}_1(1-\hat{\lambda}-\hat{\mu})^2 + \hat{\mu}(1-\hat{\lambda}-\hat{\mu}) + \hat{\mu} + \hat{e}_3, \tag{18}$$
$$\bar{\Upsilon}_4 = \hat{p}_1(1-\hat{\lambda}-\hat{\mu})^3 + \hat{\mu}(1-\hat{\lambda}-\hat{\mu})^2 + \hat{\mu}(1-\hat{\lambda}-\hat{\mu}) + \hat{\mu} + \hat{e}_4. \tag{19}$$

In the case of three cross sections, only the first three of the above equations apply. As shown before, values of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ can then be found that exactly 'reproduce' $\bar{\Upsilon}_1$, $\bar{\Upsilon}_2$ and $\bar{\Upsilon}_3$ , i.e., with $\hat{e}_1 = \hat{e}_2 = \hat{e}_3 = 0$. Therefore, in the discussion of the estimators $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ for three cross sections, paying attention to the error involved in $\bar{\Upsilon}_t$ was not necessary. However, with more than three cross sections, values of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ for which all error terms are exactly zero will, in general, not exist. Therefore, some criterion is needed to determine which values of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ are 'best'. To this aim the maximum likelihood (ML) criterion can be deployed. Following (16) through (19) each $\bar{\Upsilon}_t$ can be written as $\hat{p}_t + \hat{e}_t$, with $\hat{p}_t$ being the estimate of the true state 1 probability $p_t$. Assuming that the number of individuals in state 1 in the cross section at $t$ follows a binomial distribution with expectation $\hat{p}_t$ and assuming independence of the binomials for the four cross sections, the likelihood $\ell$ of a given triplet $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ is given by:

$$\ell = \binom{100}{80}\hat{p}_1^{80}(1-\hat{p}_1)^{20} * \binom{120}{72}\hat{p}_2^{72}(1-\hat{p}_2)^{48} *$$
$$\binom{200}{140}\hat{p}_3^{140}(1-\hat{p}_3)^{60} * \binom{100}{75}\hat{p}_4^{75}(1-\hat{p}_4)^{25} .$$

In words, the value of $\ell$ represents the probability of obtaining exactly 80, 72, 140, and 75 'successes' given 100, 120, 200 and 100 'draws' with success probabilities $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$ and $\hat{p}_4$, respectively. Clearly, $\ell$ is a function of the four $\hat{p}_t$ and thus a function of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$. Choosing values of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ such that the resulting value of $\ell$ is at its maximum means that those values are considered 'best' for which the probability is highest that the observed state 1 marginals of the four cross sections would occur. By applying this criterion, standard ML theory can be applied which, among others things, offers the possibility to perform hypothesis tests on the true values of $p_1$, $\mu$ and $\lambda$. Examples of such tests will be given below. For now, we only present the estimated values of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ that maximize the likelihood. These are: $\hat{p}_1 = .2006$, $\hat{\mu} = .54$ and $\hat{\lambda} = .17$. Based on these estimates the resulting values for $\hat{p}_2$, $\hat{p}_3$ and $\hat{p}_4$ are $.5937$, $.7076$ and $.7409$, respectively. It can be concluded that the four estimates $\hat{p}_t$ are very close to the corresponding observed sample proportions .20, .60. ,70, and .75, respectively. Following ML theory, one can calculate the deviance of the current model from the saturated model (see e.g. Collett, 1991). Like Pearson's $X^2$, the deviance is a goodness of fit measure which, for the current example, follows a $\chi^2$ distribution with 5 (=number of cross sections) minus 3 (=number of parameters estimated, i.e., $\hat{p}_t$, $\hat{\mu}$ and $\hat{\lambda}$) makes 2 df, provided that the current model holds. Both Pearson's $X^2$ and the deviance are .12 here, this value being indicative of the good fit of the model predicted proportions to the proportions actually observed.

Applying the ML principle means that some method is needed to search the values of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ for which the likelihood function $\ell$ obtains its maximum. These values can in general not be evaluated analytically but, instead, must be determined by some iterative search procedure. During the search process the likelihood of candidate values of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ is evaluated. To prevent these values from falling outside the (0,1) range, a function can be applied that links each of the three probabilities to a parameter that is actually estimated. We used the logit link function implying that, instead of directly estimating values of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$, the logits of these probabilities are estimated. The logit link function is also used to incorporate predictor variables in the model as will be shown below.

We finally note that the ML method can also be applied if there are only three cross sections. The ML estimates of $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ are then equal to those obtained by applying (11), (12) and (13), provided these are in the (0,1) range. Using the ML approach has the advantage that not only point estimates but also standard errors of (the logits of) $\hat{p}_1$, $\hat{\mu}$ and $\hat{\lambda}$ are easily obtained.

### 1.3.3 Including predictor variables and an empirical application

The potential application of the RCS Markov model is enhanced if the cross sections include data related to the unknown values of $p_1$, $\mu_t$ and $\lambda_t$. By employing these data as predictor variables for $p_1$, $\mu_t$ and $\lambda_t$, the model becomes more refined in the sense that individuals with different predictor values may have different initial and transition probabilities. Furthermore, if for a given individual the value for a predictor variable changes during the time period under study, the transition probabilities for that individual change over time. As a consequence, by using predictors variables, the resulting Markov model can be individual heterogeneous and time varying.

To link predictor values to the initial, entry and exit probabilities we use the logit link function. To let $p_1$, $\mu_t$ and $\lambda_t$ depend on the predictors sex and age, say, the following expressions would be used for the logits of the three probabilities:

$$\text{logit}\,(p_{sa,1}) = b_0 + b_1 \text{ sex} + b_2 \text{ age}_1, \tag{20}$$
$$\text{logit}\,(\mu_{sa,t}) = b_0^{\mu} + b_1^{\mu} \text{ sex} + b_2^{\mu} \text{ age}_t, \tag{21}$$
$$\text{logit}\,(\lambda_{sa,t}) = b_0^{\lambda} + b_1^{\lambda} \text{ sex} + b_2^{\lambda} \text{ age}_t \tag{22}$$

where the subscript '$sa$,' denotes the dependence on sex and age. Given these logits, not only the corresponding probabilities $p_{sa,1}$, $\mu_{sa,t}$ and $\lambda_{sa,t}$ can easily be derived, but also all other state 1 probabilities $p_{sa,t}$ for $t > 1$ by using the Markov identity

$$p_{sa,t} = p_{sa,t-1}(1 - \lambda_{sa,t}) + (1 - p_{sa,t-1})\ \mu_{sa,t}\,. \tag{23}$$

In the logit expressions (20), (21) and (22) the variable age is subscripted with a time index, either $1$ or $t$, since its value changes over time for each individual. Note that in the Markov identity given in Equation (23) the value of $p_{sa,t-1}$ actually depends on the respondents age at $t-1$. (It would therefore be more precise to use $p_{sa_{t-1},t-1}$ instead of $p_{sa,t-1}$. However, the latter is preferred since the link between time point and corresponding age value is obvious.) In the logits, the variable sex has no time index, since its value does not change over time. For the respondents of the cross section observed at $t = 3$, say, the value of $\text{age}_1$ can of course be derived from the observed age at $t = 3$. To use Moffitt's words: one can 'backcast' the age variable in time, in order to derive previous age values. Obviously, such 'backcasting' in time of predictor variables is easy for time-constant variables like a person's sex and race. For certain time-varying variables like income, marital status, attitudes etc., backcasting is often not possible and this limits the possibility to include these variables as predictors. Other time-varying variables like age and macro variables, like unemployment rate or GNP, are easy to backcast and, hence, to employ as predictors in the model. Also, for certain variables it may be valid to backcast them for only 1 or 2 time points, say, if they can be assumed to remain stable for such short period of time.

In (20), (21) and (22) there are nine $b$ parameters, three for each probability. Note that $b_1^\mu$ denotes the effect of a respondents sex on *each* of the $T-1$ entry probabilities $\mu_t$, where $T$ is the total number of time points or cross sections available. Hence the effect of sex on the entry probability is assumed not to vary over time. The same holds for the effect of age and the effects of sex and age on the $T-1$ exit probabilities $\lambda_t$. These rather restrictive assumptions can be loosened in a number of ways to which we shall return in the chapters to follow.

If sex takes the values 1 or 2 and age ranges from 18 to 70 years, then, for the model equations (20), (21) and (22), the likelihood is given by

$$\ell = \prod_{t=1}^{T}\prod_{s=1}^{2}\prod_{a=18}^{70} \binom{n_{tsa}}{n_{1,tsa}} p_{sa,t}^{n_{1,tsa}} (1-p_{sa,t})^{n_{tsa}-n_{1,tsa}}$$

where $n_{tsa}$ denotes the number of individuals in the cross section of time point $t$ with sex $s$ and age $a$, of which $n_{1,tsa}$ are in state 1 of $\Upsilon$. Each $p_{sa,t}$ in the function $\ell$ depends on all previous $\mu_{sa,t}$, $\lambda_{sa,t}$ and/or $p_{sa,t}$ which in turn depend on the nine parameters. The maximum likelihood estimates of the parameters are the ones that maximize the value of $\ell$. In this example the value of $\ell$ would depend on 2 (70-18+1)=106 independent draws from as many binomial distributions for each of the $T$ time points, given that each sex/age combination would occur in each cross section. The predictor variables thus break down the cross sections into groups of individuals, each of whom has the same initial and transition probabilities. In the extreme case, each group contains only a single individual, in which case the likelihood turns into

$$\ell = \prod_{t=1}^{T}\prod_{i=1}^{n_t} \left[\Upsilon_{i,t} p_{i,t} + (1-\Upsilon_{i,t})(1-p_{i,t})\right]$$

with $n_t$ denoting the number individuals in the cross section observed at time point $t$. Every individual case may have its own unique set of values for the initial and transition probabilities, different from those of all other individuals. This explains why the RCS model was earlier classified as a model appropriate for individual (as opposed to aggregate) data. However, the parameters that determine the individual specific probabilities are common to many or even all individuals.

We will now illustrate the model using actual data. In doing so, we will also discuss some additional model details. The data we will use are taken from the 5 'Social and Cultural Developments in the Netherlands' (SOCON) surveys, conducted in 1979, 1985, 1990, 1995 and 2000. In each survey, a sample of the Dutch population aged 18 to 70 is interviewed on a number of social-cultural issues. One point of interest is the degree to which respondents maintain a cultural conservative attitude. This attitude is measured with a set of items, one of which is: 'Do you think it should be possible for a woman to have an abortion without further preface, if she wants to?' with response categories 'yes', 'no' and 'no opinion'. We restrict the analysis to the responses 'yes' and 'no' (97% for the five surveys together) and exclude respondents with 'no opinion'.

Table 1.9 gives proportions 'yes' (i.e., pro-abortion) and the totals for 3*41=123 different combinations of age and education. The rightmost column of Table 1.9 reveals a large increase between 1995 and 2000 of the proportion of people who feel that women should have the possibility of an abortion without preface.

The column labelled 'age' contains the history of (maximally five) age values for the period starting in 1979 up until the year that the respondents were interviewed. A dash '-' denotes that these respondents did not reach the age of 18 yet at the time that the survey took place. We explained above that in the RCS model the initial probability $p_1$ is used as the starting point for the Markov process under study. For the current example $p_1$ refers to the year 1979. However, the respondents in the 2000 cross section with age profile '---00' were 2 to 6 years of age in 1979. Consequently, for these respondents the initial probability $p_1$ would represent the probability to have a certain opinion about abortion at childhood age. This unwanted situation is typically encountered if the data cover a relatively long time period. A first but rather crude way to get rid of the problem is to select only those respondents of each cross section who were old enough to have (had)

Table 1.9 Pro-abortion attitude by survey year, age and education in the SOCON surveys conducted in 1979, 1985, 1990, 1995 and 2000

| survey year | age | low educ | middle educ | high educ | per survey year |
|---|---|---|---|---|---|
| 1979 | 0* | .31 (103)** | .27 ( 73) | .46 ( 99) | .36 (978) |
| 1979 | 1 | .35 (253) | .37 (153) | .48 (122) | |
| 1979 | 2 | .24 (115) | .41 ( 37) | .35 ( 23) | |
| 1985 | -0 | .33 ( 81) | .36 (118) | .41 (142) | .39(2885) |
| 1985 | 00 | .35 (108) | .34 (171) | .52 (188) | |
| 1985 | 01 | .32 (151) | .42 (178) | .51 (137) | |
| 1985 | 11 | .30 (396) | .43 (402) | .55 (264) | |
| 1985 | 12 | .28 (122) | .42 ( 88) | .41 ( 41) | |
| 1985 | 22 | .22 (135) | .25 ( 87) | .30 ( 76) | |
| 1990 | --0 | .39 ( 59) | .44 ( 87) | .43 ( 83) | .44(2250) |
| 1990 | -00 | .41 (101) | .40 (118) | .52 (126) | |
| 1990 | 000 | .38 ( 21) | .33 ( 36) | .49 ( 51) | |
| 1990 | 001 | .37 ( 76) | .34 ( 90) | .52 (123) | |
| 1990 | 011 | .43 (116) | .53 (100) | .62 (130) | |
| 1990 | 111 | .42 (206) | .44 (175) | .57 (160) | |
| 1990 | 112 | .18 ( 67) | .50 ( 62) | .51 ( 43) | |
| 1990 | 122 | .29 ( 78) | .33 ( 33) | .42 ( 36) | |
| 1990 | 222 | .26 ( 35) | .43 ( 23) | .60 ( 15) | |
| 1995 | ---0 | .50 ( 30) | .50 ( 46) | .56 ( 61) | .41(1933) |
| 1995 | --00 | .48 ( 21) | .46 ( 70) | .48 ( 89) | |
| 1995 | -000 | .32 ( 31) | .40 ( 42) | .55 ( 62) | |
| 1995 | -001 | .21 ( 24) | .17 ( 42) | .48 ( 81) | |
| 1995 | 0001 | .27 ( 22) | .34 ( 35) | .37 ( 54) | |
| 1995 | 0011 | .35 ( 52) | .40 ( 73) | .46 (111) | |
| 1995 | 0111 | .34 ( 73) | .52 ( 97) | .49 (106) | |
| 1995 | 1111 | .35 (101) | .41 ( 99) | .49 (114) | |
| 1995 | 1112 | .26 ( 57) | .51 ( 45) | .43 ( 46) | |
| 1995 | 1122 | .21 ( 56) | .36 ( 28) | .34 ( 50) | |
| 1995 | 1222 | .28 ( 60) | .31 ( 26) | .45 ( 29) | |
| 2000 | ---00 | .50 ( 6) | .52 ( 29) | .81 ( 21) | .62(902) |
| 2000 | --000 | .33 ( 3) | .53 ( 17) | .73 ( 33) | |
| 2000 | --001 | .63 ( 8) | .50 ( 12) | .70 ( 23) | |
| 2000 | -0001 | .54 ( 13) | .64 ( 22) | .70 ( 33) | |
| 2000 | -0011 | .64 ( 11) | .57 ( 30) | .69 ( 32) | |
| 2000 | 00011 | .50 ( 12) | .65 ( 17) | .74 ( 19) | |
| 2000 | 00111 | .58 ( 26) | .63 ( 49) | .73 ( 55) | |
| 2000 | 01111 | .57 ( 28) | .67 ( 61) | .74 ( 53) | |
| 2000 | 11111 | .48 ( 21) | .57 ( 35) | .69 ( 35) | |
| 2000 | 11112 | .40 ( 25) | .61 ( 23) | .73 ( 30) | |
| 2000 | 11122 | .45 ( 29) | .58 ( 12) | .54 ( 24) | |
| 2000 | 11222 | .45 ( 29) | .54 ( 13) | .62 ( 13) | |

Total N=8948

* Ages for the five periods, '- ' = less than 18, 0=18-30, 1=31-55, 2=56-70 years old; e.g. '--001' means: less than 18 years in 1979 and 1985, 18 through 30 years in 1990 and 1995, and 31 through 55 years of age in 2000.

** Proportion with pro-abortion attitude in the subclass (subclass total)

Table 1.10 Birth cohort, initial year and initial probability for the SOCON cross sections conducted in 1979, 1985, 1990, 1995 and 2000

| birth cohort | initial year | initial probability | N |
|---|---|---|---|
| 1909-1961 | 1979 | $p_1$ | 7141 |
| 1962-1967 | 1985 | $p_2$ | 1109 |
| 1968-1972 | 1990 | $p_3$ | 505 |
| 1973-1977 | 1995 | $p_4$ | 193 |
| 1978-1982 | 2000 | $p_5$ | 63 |

a valid response in the year of the first cross section. For the SOCON data one could, for example, select respondents who were at least 18 years of age in 1979. For the cross section of 2000 this would mean that only respondents of at least 39 years of age would be retained, which is obviously not desirable. A second solution to this problem is to let the initial probability of a respondent not necessarily relate to the year of the first survey but, instead, to the earliest survey year in which the respondent is at least 18 years old, say. For the respondents of the 2000 cross section with age profile '---00' the initial probability then refers to the year 1995. These respondents make their entrance into the model in 1995 and participate in only a single transition, the one from 1995 to 2000. Consequently, the probabilities $p_1$, $p_2$ and $p_3$ are irrelevant for them, $p_4$ and $p_5$ being the only state probabilities that matter. Furthermore, for these cases $p_4$ serves as the initial probability and hence is modelled by a logistic link function, similar to the one for $p_1$ in (20). For other respondents $p_4$ is modelled by the usual Markov accounting identity $p_4 = p_3(1-\lambda_4)+(1-p_3)\mu_4$. In general, in this second way of treating respondents who are too young in the year(s) of the first cross section(s), it is the respondents birth cohort that determines which of all $T$ probabilities $p_t$ serves as the initial probability. For the SOCON data, the birth cohort, initial year and initial probability are shown in Table 1.10.

The rightmost column in Table 1.10 shows that for a considerably large number of respondents (1870) the initial year comes after 1979, which justifies the second method of treating respondents who were too young in the year(s) of the first survey(s). Note that there are 63 respondents in the youngest birth cohort 1978-1982 who cannot have participated in any transition. Because our main interest here are the transitions, we excluded these 63 respondents from this analysis, which is thus based on 8948 respondents.

Table 1.11 summarizes different models applied to the data on attitude towards abortion. For this dependent variable, we used the values 0 and 1 to denote the anti-abortion and pro-abortion attitude, respectively. As predictors we employed the variables age and education. As to the initial probability, we expect older and lower educated people to be more conservative and thus to be more opposed to abortion than younger and higher educated people. With respect to the transition probabilities, we expect older and lower educated people to have lower entry (switch to pro-abortion) and higher exit (switch to anti-abortion) probabilities than younger and higher educated people. Furthermore, an extra intercept parameter (named $i_{2000}$ in Table 1.11) was added to the logits of the entry and exit probabilities for the 1995-2000 transition. This was done since we expect a higher entry and/or lower exit probability for the 1995-2000 transition than for the previous three, because of the large proportional increase of 'yes' answers in the 2000 survey.

For each of the models presented in Table 1.11, we started with a full model including all predictor effects in each logit equation. We then searched for a more parsi-monious model using a stepwise backward procedure[2]. The restricted models thus achieved, are the ones presented in Table 1.11.

In model 1 both age and education are treated as interval level variables. For age, the values 0, 1, 2 were used for the categories 18-30, 31-55 and 56-70 years of age, respectively; for education, we used the values -1, 0, +1 to denote low, middle and high education, respectively. The estimates of all parameters selected by the backward procedure are in the expected directions, except for the effect of age on the probability to change from pro-abortion to anti-abortion: older people are apparently less inclined to

[2] From the predictor variables that were not significant at the 5% level in one of the logit equations, we selected the one that was least significant. That predictor was removed from the equation and the model re-estimated. This step was repeated until all the remaining predictors had significant effects. After each removal step, predictors removed earlier were reentered into the equation and tested for significance at the 5% level. The most significant one, if any, was incorporated anew in the equation. The process of removing and re-entering predictors was continued until no more predictors could be removed and none could be reentered.

Table 1.11 Parameter estimates and goodness of fit

| | | initial probability | entry probability (anti → pro) | exit probability (pro → anti) |
|---|---|---|---|---|
| **Model 1** | | | | |
| Pearson $X^2$=146.7 | age | -.197 * | -.710 * | -.649 * |
| df =112, sig=.02 | education | .200 * | | -.853 * |
| LL=-337.31 | $i_{2000}$ | | 1.433 * | |
| | intercept | -.344 * | .165 | .834 |
| **Model 2** | | | | |
| Pearson $X^2$ =111.1 | $ic_{1909\text{-}1961}$ | -.536 * | | |
| df=110, sig=.45 | $ic_{1962\text{-}1967}$ | .013 | | |
| LL=-321.41 | $ic_{1968\text{-}1972}$ | .236 | | |
| | $ic_{1973\text{-}1977}$ | .571 * | | |
| | age 31-55 | .108 | .201 * | |
| | age 56-70 | -.369 * | -.491 | |
| | education | .238 * | .507 * | |
| | $i_{2000}$ | | 1.468 * | |
| | intercept | | -1.274 * | -.820 * |
| **Model 3** | | | | |
| Pearson $X^2$ =255.5 | $ic_{1909\text{-}1961}$ | .089 | | |
| df=232, sig=.14 | $ic_{1962\text{-}1967}$ | -.062 | | |
| LL=-542.99 | $ic_{1968\text{-}1972}$ | .157 | | |
| | $ic_{1973\text{-}1977}$ | .406 * | | |
| | age 31-55 | .186 * | .249 * | |
| | age 56-70 | -.307 | -.405 * | |
| | education | .243 * | .529 * | |
| | pchurch | -.814 * | | .985 * |
| | $i_{2000}$ | | 1.396 * | |
| | intercept | | -1.268 * | -1.526 * |

* indicates significance at the 5% level

change their attitude. However, according to the Pearson's $X^2$ the fit of the model is not very good.

Model 2 includes two refinements in order to improve the fit. First, age was treated as a nominal predictor for the following reason. Table 1.9 reveals that in 9 of the 15 combinations of survey year and education, the highest proportion pro-abortion is found for respondents who are in the middle age category in the given survey year (value 1 for the rightmost digit of the age profile) and not, as we initially expected, for the youngest respondents. Also, if the abortion attitude is (logistically) regressed on age and education, with all cross sections taken together, it turns out that treating age as a nominal variable greatly improves the log- likelihood of this logistic model; this is not true for education. If age has a nonlinear effect on the logits of the marginal state 1 probabilities in this logistic model, this may also be the case for the transition probabilities. Therefore, in model 2 we employed two dummy variables for age, the youngest age category functioning as the reference category. The second refinement deals with the different age cohorts discussed earlier. After omitting the youngest cohort containing 63 respondents, there are four cohorts left. For these, the observed proportions pro-abortion are .42, .44, .48 and .55 respectively, younger cohorts being less opposed to abortion than older cohorts. To enable the model to also recover this trend in the initial probabilities of the cohorts (if such would be necessary after controlling for the other effects) we used a baseline intercept for the oldest cohort which is labelled $ic_{1909\text{-}1961}$; for the other three cohorts we used dummy variables, labelled $ic_{1962\text{-}1967}$, $ic_{1968\text{-}1972}$ and $ic_{1973\text{-}1978}$, the effects of which represent deviations from the baseline. During the search procedure for a parsimonious model, the three dummies were treated as a block, meaning that if only one was significant, none of them were fixed to zero. The same block treatment was applied to the two age dummies.

The results of model 2 confirm the expectations, albeit that many predictors located at the exit-transition part of the model are insignificant. Apparently, neither age nor education is (significantly) related to the attitude change from pro-abortion to anti-abortion. As to the significant effects, the cohort estimates follow a similar rising pattern over time as the observed cohort proportions do. Note that only the youngest of the four cohorts, born in period 1973-1977, differs significantly from the baseline value -.536 of the oldest cohort. The age effects are nonlinear, as we presumed, with the middle age category having the highest value for the initial as well as for the entry probability. The effects of education are also as expected, with higher education associated with a higher initial probability, i.e., a more permissive initial attitude towards abortion; also, higher education implies a higher probability to change from anti-abortion to pro-abortion. The effect of $i_{2000}$ on the entry probability has a positive sign, again as we expected it to be. Note that, as was intended, the fit of the model has improved dramatically compared to the fit of model 1.

Without presenting the data necessary to perform the analysis, we finally show the results of still another model in which an additional predictor is used. It concerns the predictor labelled 'pchurch' in Table 1.11 which refers to whether or not (value 1 and 0, respectively) the respondent's parents were member of a Christian church or religious community. If so, we expect the respondent to have a more conservative attitude and, consequently, a lower entry and a higher exit probability. The variable was treated to be constant over time. The results are shown under model 3 in Table 1.11. Focusing only on the two significant church effects, it can be seen that both are in the expected direction. Furthermore, parents' church membership appears to be the only predictor of the ones used here that has a significant effect on the probability to change attitude from pro-abortion to anti-abortion. As to the fit, it should be noted that Pearson's $X^2$ and the likelihood of model 3 cannot be compared directly with those of model 1 and 2 since the number of (aggregate) cases is not the same. For model 1 and 2 there are 123 cases (i.e., combinations of 41 age-profiles and 3 education levels) while for model 3 there are 245 cases. Nevertheless, based on the significance level of Pearson's $X^2$ it can be concluded that the fit for model 3 is lower than the one for model 2.

This concludes our introductory treatment of the RCS Markov model. The following chapters provide additional details along with a number of model extensions and applications in diverse research areas.

Chapter 2 starts with a brief comparison of RCS data and individual panel data as to the opportunities for investigating individual-level change. The structure of the RCS Markov model is explained and the exact initial time point at which a Markov process starts is contrasted with the initial time point as we defined it above, i.e., the time point of the first cross section. A concise expression is derived for the probability $p_{it}$ of case $i$ to be in state 1 at time point $t$, which forms the basis for the log-likelihood formula. The procedure of maximum likelihood estimation is explained and the derivatives of $p_{it}$ with respect to the regression parameters in the logit link functions are given. The Fisher scoring algorithm, that is used to find the maximum of the log-likelihood function, is briefly discussed along with the estimated covariance matrix of the parameter estimates. Furthermore, for these parameters it is explained how polynomials in time may be used to overcome the assumption of regression parameters being constant for the total time period covered by all cross sections. The chapter offers an application of the model to cross-sectional data on female labor force participation in the Netherlands and West Germany for the 1987-1996 period.

Chapter 3 resembles Chapter 2 in that it starts with a description of the model, the maximum likelihood estimation procedure and a few model extensions. Also, the application is similar to the one in the previous chapter, using the same dependent and independent variables and the data covering roughly the same time period. However, instead of cross sections, in Chapter 3 we use individual panel data, the waves of which are treated as independent cross sections. In doing so, we are able to compare the transition frequencies for the panel waves with transitions frequencies predicted by the RCS model.

Chapter 4 contains two model extensions. The first deals with the possibility to use predictor variables that are not backcastable. In the basic model as proposed by Moffitt, such (time-varying) predictors, like income, cannot be used, which may seriously limit the potential application of the model. The second extension deals with modelling unobserved heterogeneity. Due, for example, to influences from the (unmodelled) period prior to the first observed cross section, or due to the fact that important predictor variables are not available in the data, the distributions of the $\Upsilon_{i,t}$ may seriously deviate from the modelled binomial distributions and consequently, the likelihood function may be incorrect. We propose a method to take account of such unmodelled sources of variation. Furthermore, the RCS model is compared with ecological inference approaches for the estimation of transition probabilities using cross-sectional data. Such approaches are frequently used in political science studies. The application in this chapter again uses individual panel data. The dependent variable is vote intention for the U.S. presidential elections of 1976. Both model extensions just mentioned are applied. Panel transition frequencies and RCS model predictions are compared. Also, the results of a individual panel model are compared with those of the RCS model.

In Chapter 5 the properties of the ML estimators are investigated for the example data used. These are again panel data, containing 13 annual waves of Dutch households for the 1986-1998 period. The dependent variable concerns the presence (or absence) of a PC in the household. Three different procedures are employed to gain insight into the properties of the ML estimators of the particular model that we used. These include a

MCMC procedure, parametric bootstrapping and cross-sectional sub sampling, the last procedure implying that from each panel wave a sub sample is taken in such a way that each sub sample contains data from different respondents.

Chapter 6 deals with the problem that, due to the many unobserved quantities in RCS data, finding the values for the model parameters that maximize the likelihood may not be easy. In practice, the likelihood function may have multiple modes with only slight differences in the likelihood associated with each mode. We use a Bayesian approach to enquire into the likelihood function's shape for a simple simulated data example. In addition, it is shown how, in a Bayesian analysis, a small individual panel data set can be used to provide prior information about the RCS model parameters. This information is thereupon used to come to well defined posterior distributions of the RCS model parameters that could not have been obtained without the 'panel' priors.

Chapter 7 gives a summary of the preceding chapters and an overview of our future activities to further develop the RCS Markov model's potential.

Appendix 1 contains a user manual of the stand-alone computer program *Crossmark* that we developed for estimating the RCS Markov model. Apart from ML estimation using Fisher's scoring method, the program can perform Metropolis sampling and parametric bootstrapping. Most model extensions discussed are included in the program.

## References

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

Anderson, T.W. 1954. Probability Models for Analyzing Time Changes in Attitudes. In *Mathematical thinking in the social sciences,* ed. P.F. Lazarsfeld. New York: The Free Press, pp. 17-66.

Anderson, T.W. and L. A. Goodman. 1957. "Statistical Inference about Markov Chains," *The Annals of Mathematical Statistics,* 28: 89-110.

Blumen, I., M. Kogan and P.J. Mac Carthy. 1955. *The Industrial Mobility of Labour as a Probability Process.* Ithaca, NY: Cornell University Press.

Brown, P.J. and C.D. Payne. 1986. "Aggregate Data, Ecological Regression an Voting Transitions," *Journal of the American Statistical Association*, 81: 452-460.

Cook, R. J. and E.T.M. Ng. 1997. "A Logistic-bivariate Normal Model for Overdispersed Two-State Markov Processes," *Biometrics*, 53: 358-364.

Collett, D. 1991. *Modelling Binary Data.* London: Chapman and Hall.

Diggle, P.J., K. Liang and S.L. Zeger. 1994. *Analysis of Longitudinal Data.* Oxford: Clarendon Press.

Golan, A., G. Judge, and D. Miller, *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* Wiley.

Goodman, L.A. 1953. "Ecological Regressions and Behavior of Individuals," *American Sociological Review*, 18: 663-664.

Goodman, L.A. 1959. "Some Alternatives to Ecological Correlation," *The American Journal of Sociology*, 64: 610-625.

Hagenaars, J. A. 1990. *Categorical Longitudinal Data: Log-linear Panel, Trend and Cohort Analysis*. Sage.

Hawkes, A.G. 1969. "An Approach to the Analysis of Electoral Swing," *Journal of the Royal Statistical Association. Series A (General)*, 132: 68-79.

King, G. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

Janes E. T. 1957. "Information theory and statistical mechanics," *Physics Review*, 106: 620-630.

Karantininis, K. 2002. "Information-based Estimators for the Non-stationary Transition Probability Matrix: an Application to the Danish Pork Industry," *Journal of Econometrics*, 107: 275-290.

King, G., O. Rosen, and M.A. Tanner. 2004. *Ecological Inference New Methodological Strategies.* Cambridge: Cambridge University Press.

Lee, T.C., G.G. Judge, and A. Zellner. 1970. "Maximum Likelihood and Bayesian Estimation of Transition Probabilities," *Journal of the American Statistical Association*, 63: 1162-1179.

MacRae, E.C. 1977. "Estimation of Time-varying Markov Processes with Aggregate Data," *Econometrica*, 45: 183-198.

McCullagh, P. and J.A. Nelder. 1989. *Generalized Linear Models*. London: Chapman and Hall.

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.

Shannon, C.E. 1948. "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 27: 379-423.

Spilerman, S. 1972. "The Analysis of Mobility Processes by the Introduction of Independent Variables into a Markov Chain," *American Sociological Review*, 37: 277-294.

Stott, D. 1997. *SABRE 3.1: Software for the Analysis of Binary Recurrent Events.* http://www.cas.lancs.ac.uk/software/sabre3.1/sabre.html

Vermunt, J.K., R. Langeheine, and U. Böckenholt. 1999. "Discrete-time Discrete-state Latent Markov Models with Time-constant and Time-varying Covariates," *Journal of Educational and Behavioral Statistics*, 24: 179-207.

Wiggins, L.M. 1973. *Panel Analysis*. Amsterdam: Elsevier.

Willekens, F. 1982. Multidimensional Population Analysis with Incomplete Data. In *Multidimensional Mathematical Demography*, ed.K.C. Land and A. Rogers. New York: Academic Press, pp. 43-112.

# 2 "Panelizing" Repeated Cross Sections

Female Labor Force Participation in
the Netherlands and West-Germany

This chapter[a] 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.

[a] This chapter has been published as Pelzer, B.J., R. Eisinga, and P.H. Franses. 2005. "Panelizing" Repeated Cross Sections. *Quality & Quantity,* 39: 155-174.

## 2.1 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.

## 2.2 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.

### 2.2.1 First-order Markov model

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

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

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

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

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., $\mu_{it}$ represents the probability that the unit satisfying $y_i = 0$ at time $t - 1$ subsequently satisfies $y_i = 1$ 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

$$\begin{aligned} p_{it} &= P(Y_{it} = 1), \\ \mu_{it} &= P(Y_{it} = 1 \mid Y_{it-1} = 0), \\ \lambda_{it} &= P(Y_{it} = 0 \mid Y_{it-1} = 1), \end{aligned}$$

then we have

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

where $\eta_{it} = 1 - \lambda_{it} - \mu_{it}$. As noted by Moffitt (1990), the accounting identity in Equation (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 $(\mu_{it})$ and outflow $(\lambda_{it})$ 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, Fienberg, and Holland 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, Judge, and Zellner 1970; Firth 1982; Kalbfleish and Lawless 1984, 1985; Lawless and McLeish 1984; Li and Kwok 1990; Hawkins, Han and Eisenfeld 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 Equation (1) and let $p_{i0}=0$ (or $t\rightarrow\infty$), it may be verified that $p_{it}$ has the representation

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

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 $(X_{it},X_{it-1},\dots,X_{i1})$

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 $\mu_{it}$'s and $\lambda_{it}$'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 Equation (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 $P(Y_{it} = 1 \mid Y_{it-1} = y_{it}),\ \ y_{it} = 0{,}1$. That is, we assume that

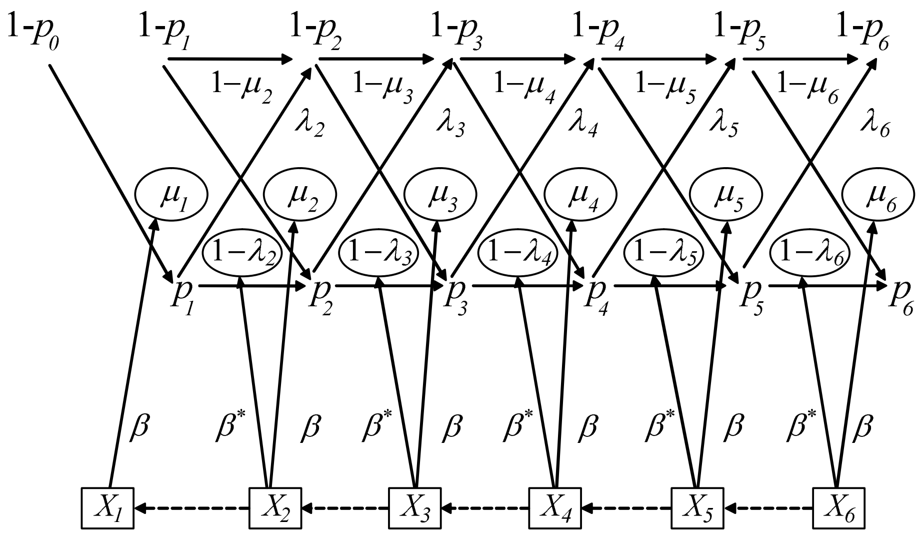
$$\begin{aligned} \text{logit } P(Y_{it} = 1 \mid Y_{it-1} = 0) &= \text{logit}\,(\mu_{it}) &&= X_{it}^{'}\beta, \text{ and} \\ \text{logit } P(Y_{it} = 1 \mid Y_{it-1} = 1) &= \text{logit}\,(1 - \lambda_{it}) &&= X_{it}^{'}\beta^{*}, \end{aligned}$$

where the parameters $\beta$ and $\beta^{*}$ 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

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

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 Equation (3) for panel data occur in Amemiya (1985), Diggle, Liang and Zeger (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

Figure 2.1 Graphical illustration of Markov model for RCS data



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

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

with respect to the parameters, with $p_{it}$ defined by Equation (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 2.1, 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 $\mu_{it}$ and $\lambda_{it}$. 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.

### 2.2.2 Maximum likelihood estimation

Maximum likelihood fitting of the model in Equation (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 Equation (2) is re-written as

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

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 Equation (5) with respect to the parameters $\beta$ and $\alpha$ are

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

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

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 $\varepsilon$ 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.

### 2.2.3 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 $\mu_{i1}$, 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 $\delta$ is a set of parameters to be estimated and $F$ is the logistic function. The $\delta$-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

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

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

Table 2.1 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 | .276 | | |
| 1988 | 582 | .325 | 869 | .420 |
| 1989 | 611 | .358 | 468 | .391 |
| 1990 | 839 | .455 | 792 | .509 |
| 1991 | 584 | .430 | 413 | .508 |
| 1992 | 637 | .425 | 690 | .525 |
| 1993 | 578 | .483 | 283 | .502 |
| 1994 | 609 | .435 | 635 | .573 |
| 1995 | 659 | .490 | | |
| 1996 | 726 | .493 | | |

## 2.3 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 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 woman participates in the labor force (i.e., one or more hours of paid work per week) and 0 otherwise. Table 2.1 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, $\geq$ 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 2.2 presents the parameter estimates for a time-constant-coefficient model specifying women's transition into and out of employment.[9] The model defines the first outcome to equal the transition probability $\mu_{i1}$, as in Moffitt (1990, 1993), and not the state probability $p_{i1}$.

The first and third column in Table 2.2 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 2.2 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 in 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 $\mu_{i1}$. 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

Table 2.2 Time-constant Markov estimates of women's transition into and out of employment; n=6,411 (NL) and 4,150 (WG) [a]

| | The 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 | .220* | -.719* | .344* | -.072 |
| | *(.066)* | *(.105)* | *(.081)* | *(.130)* |
| Religious upbringing (NL) | -.148 | -.008 | .074 | 1.131* |
| / religion (WG) | *(.125)* | *(.178)* | *(.200)* | *(.399)* |
| | | | | |
| *Varying covariates* | | | | |
| Age | .179* | .710* | .177* | 1.267* |
| | *(.051)* | *(.335)* | *(.045)* | *(.451)* |
| $\text{Age}^2 \div 100$ | -.281* | -1.936* | -.283* | -3.300* |
| | *(.061)* | *(.882)* | *(.054)* | *(1.127)* |
| $\text{Age}^3 \div 10{,}000$ | | 1.804* | | 2.850* |
| | | *(.760)* | | *(.919)* |
| Number of children: | | | | |
| < 5 years old | -.770* | .575* | -.629* | 3.847* |
| | *(.116)* | *(.143)* | *(.112)* | *(.482)* |
| 5-17 years old | -.460* | -.066 | -.332* | .585* |
| | *(.076)* | *(.117)* | *(.074)* | *(.141)* |
| $\geq$18 years old | -.083 | -.054 | .266* | .504* |
| | *(.129)* | *(.208)* | *(.105)* | *(.176)* |
| Unemployment rate | -.076 | -.160 | .078 | -.018 |
| | *(.097)* | *(.133)* | *(.076)* | *(.090)* |
| Age20 | 5.223 | | 3.045 | |
| | *(3.347)* | | *(3.810)* | |
| Age20 × education | -.523 | | .474 | |
| | *(.681)* | | *(.893)* | |
| Age20 × unemployment rate | -.618 | | -.433 | |
| | *(.417)* | | *(.463)* | |
| Log likelihood ($LL^*$) | -3706.729 | | -2416.144 | |

* Significant at 5% level.

[a] Asymptotic estimates of standard errors in parentheses.

[b] The $\beta$-parameters represent the effect on entry (i.e., $\mu_t$), the $\beta^*$-parameters the effect on $(1-\lambda_t)$ and thus $-\beta^*$ the effects on exit (i.e., $\lambda_t$).

[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 (backcast) in years: 20-64; Number of children (backcast) <5: 0-4; Number of children (backcast) 5-17: 0-7 (NL) / 0-6 (WG); Number of children (backcast) $\geq$18: 0-5; National unemployment rate (backcast) in each year in percentages; Age20: 1 if age (backcast) = 20, 0 if not.

Table 2.3 Goodness of fit statistics; n=6,411 (NL) and 4,150 (WG)

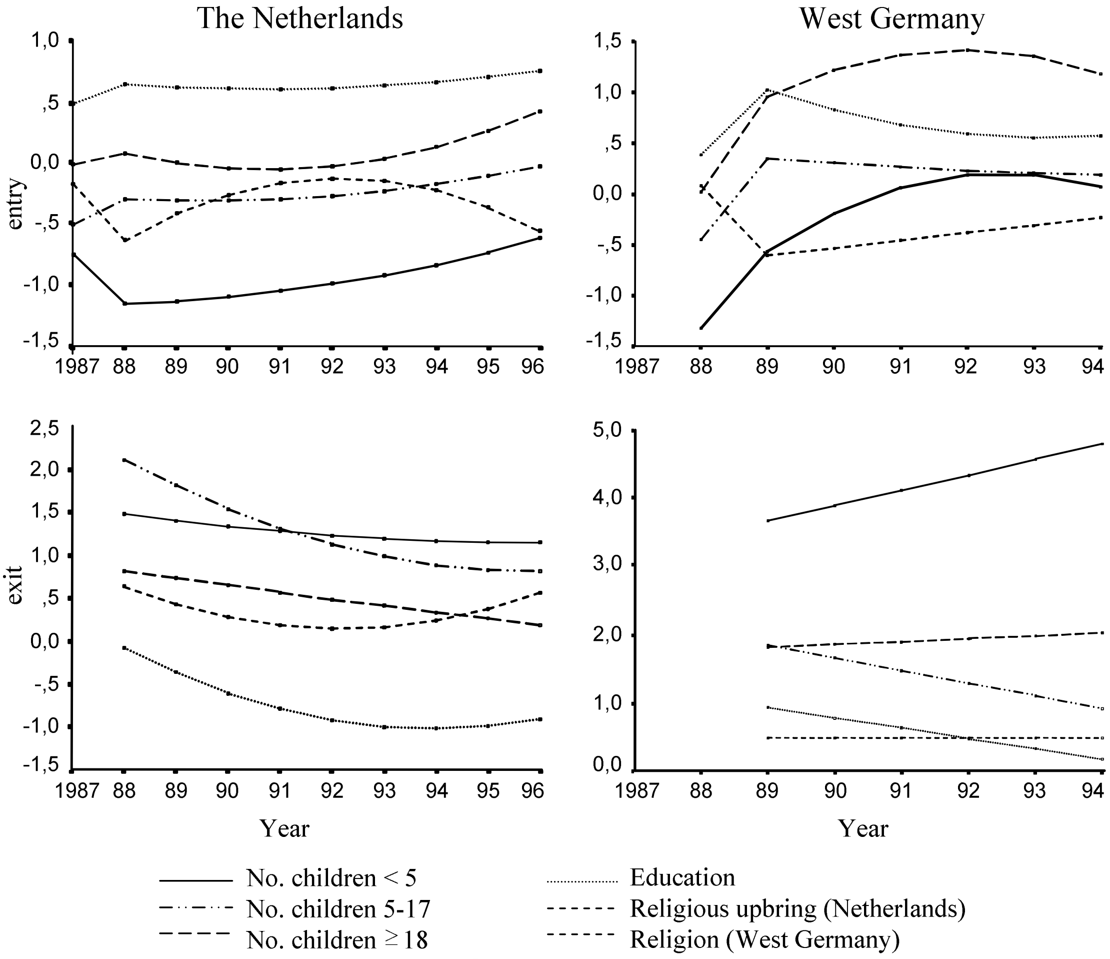
| | The Netherlands | West Germany |
|---|---:|---:|
| *Time-constant-coefficient model* | | |
| Log likelihood ( $LL^*$ ) | -3706.729 | -2416.144 |
| number of parameters | 22 | 22 |
| Akaike information criterion | 1.163 | 1.175 |
| *Time-vaying-coefficient model* | | |
| Log likelihood ( $LL^*$ )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

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 2.3, 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 2.2. 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 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 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

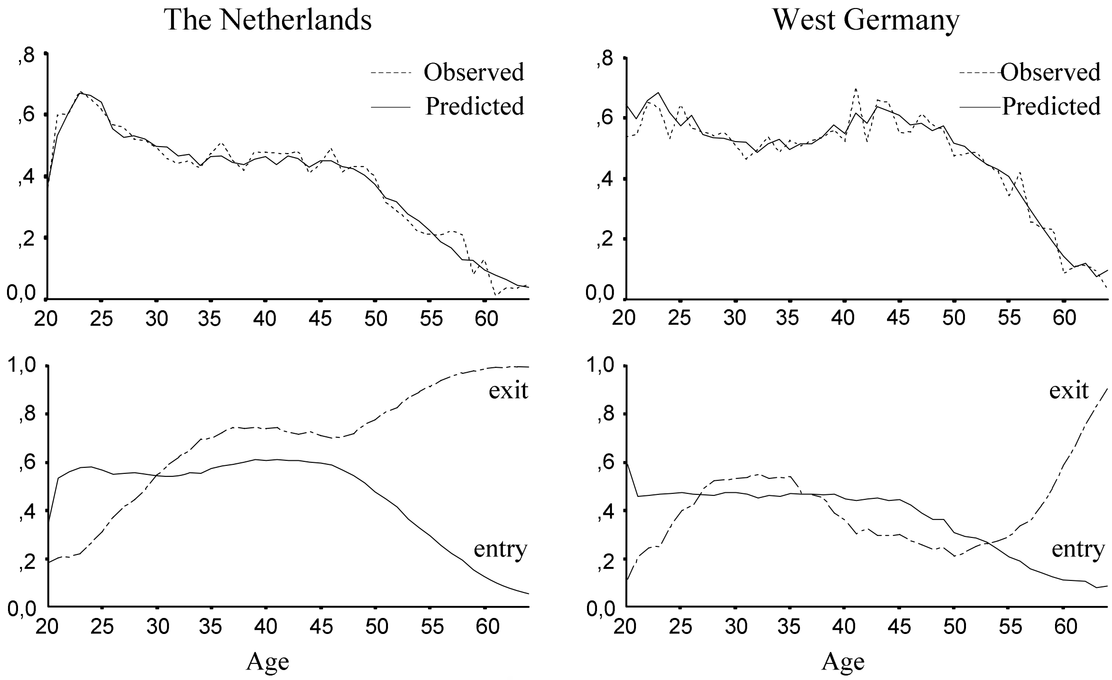
Figure 2.2 Estimated time-varying effects on entry (top) and exit (bottom)



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 2.3 (top) presents the observed and predicted marginal employment probabilities by age.[10] The figure 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 West Germany, 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

Figure 2.3 Observed and predicted employment probabilities (top) and entry and exit transition probabilities (bottom) by age



(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 2.3 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.

## 2.4 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-cast 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 $p_{it} = \mu(1+\eta+\cdots+\eta^{t-1}) + \eta^t p_{i0} = \mu(1+\Sigma_{\tau=1}^{t-1}\eta^{t-\tau}) + \eta^t p_{i0}$ $= (\mu/(\mu+\lambda))(1-\eta^t) + \eta^t p_{i0}$. As $t \to \infty$, $\eta^t$ 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 $p_{it} = \mu_{it} + (\mu_{it-1}\eta_{it} + \mu_{it-2}\eta_{it-1}\eta_{it} + \cdots + \mu_{i1}\eta_{i2}\cdots\eta_{it}) + p_{i0}\eta_{i1}\cdots\eta_{it} = \mu_{it} + \Sigma_{\tau=1}^{t-1}\mu_{i\tau}(\Pi_{s=\tau+1}^{t}\eta_{is}) + p_{i0}\Pi_{\tau=1}^{t}\eta_{it}$. As $t \rightarrow \infty$, $\Pi_{\tau=1}^{t}\eta_{it}$ tends to zero, thus $p_{it} = \mu_{it} + \Sigma_{\tau=1}^{t-1}\mu_{i\tau}(\Pi_{s=\tau+1}^{t}\eta_{is})$.
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 -.484 and -$\beta^*(\lambda_t)$ coefficients of -.841 and -.583 in the Netherlands and West Germany, respectively. This implies constant annual transition rates of $\mu = .252$ and $\lambda = .301$ in the Netherlands and $\mu = .381$ and $\lambda = .358$ in West Germany.
10. The mean $\overline{p}_m$ for age category $m$ was obtained as $\overline{p}_m = n_m^{-1}\Sigma_{i=1}^{n_m} p_{it}$, 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 $\overline{\mu}_m$ and $\overline{\lambda}_m$ for age category $m$ were obtained as a weighted average of the transitions up to $t$ with weights defined by $w_k = (\Sigma_{t=k}^{T} n_t)^{-1}(T-1)^{-1}\Sigma_{j=2}^{T}\Sigma_{t=j}^{T} n_t$.
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 and 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.

# 3 Estimating Transition Probabilities from a Time Series of Independent Cross Sections

This chapter[a] considers the implementation of a nonstationary, heterogeneous Markov model for the analysis of a binary dependent variable in a time series of independent cross sections. The model, previously considered by MOFFITT (1993), 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 ML 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 women using data from the Socio-economic Panel[b] (SEP) study conducted in the Netherlands between 1986 and 1995. We treat the panel data as independent cross sections and compare the employment status sequences predicted by the model with the observed sequences in the panel. Some open problems concerning the application of the model are also discussed.

[a] This chapter has been published as Pelzer, B., R. Eisinga, and P.H. Franses. 2001. Estimating Transition Probabilities from a Time Series of Independent Cross Sections. *Statistica Neerlandica*, 55: 249-262. Our thanks to Marno Verbeek for his helpful comments on a previous draft.

[b] The data for *Socio-economic Panel (SEP)* utilized in this paper were collected by STATISTICS NETHERLANDS (NCBS) and were made available by the SCIENTIFIC STATISTICAL AGENCY (WSA) of the NETHERLANDS ORGANIZATION FOR SCIENTIFIC RESEARCH (NWO).

## 3.1 Introduction

The increasing availability of repeated cross-sectional (RCS) surveys not only provides researchers with a growing opportunity to analyze over-time change but also raises questions about new methodology for exploiting these data for longitudinal study. RCS data contain information on different cross-sectional units (typically individuals) independently drawn from a 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 units across occasions. Obviously, 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 repeated 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 we can do.

This paper discusses, for the case of a binary dependent variable, a dynamic model previously treated briefly by MOFFITT (1990 1993) that permits the 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. It is therefore likely to be useful to researchers seeking to explain over-time change at the micro level in the absence of microlevel data. The paper is organized as follows. Section 2 discusses the model, parameter estimation and some refinements of the model. Section 3 provides an example application using panel data on female labor force participation taken from the Socio-economic Panel (SEP) study conducted in the Netherlands between 1986 and 1995. We treat the panel data as independent cross sections and compare the predictions of the Markov model for RCS data with the observations in the panel. Section 4 concludes.

## 3.2 Dynamic model for RCS data

The problem of analyzing repeated cross-sectional data has attracted increasing attention in econometrics and other disciplines in the last several years. One class of models considered is the linear fixed effect model (BALTAGI 1995, COLLADO 1997, DEATON 1985, GIRMA 2000 2001, NIJMAN and VERBEEK 1990, VERBEEK 1996, VERBEEK and NIJMAN 1992 1993, VERBEEK and VELLA 2000). 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 pseudo panel 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 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 $j \neq k$. Other discussions of the model include FELTEAU *et al.* (1997) and MEBANE and WAND (1997).

### 3.2.1 First-order Markov model

Suppose we have the following two-state first-order Markov matrix of transition rates in which the cell probabilities sum to unity across rows:

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

This expression records the probabilities of making each of the possible transitions from one time period to the next; e.g., $\mu_{it}$ represents the probability that the unit satisfying $y_i = 0$ at time $t-1$ subsequently satisfies $y_i = 1$ at time $t$. Recall that the first-order 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 assumption implies that the dependency between successive transitions can be eliminated by conditioning on the previous state. Operationally this can be achieved by including the previous state in the model as a covariate predicting $y_{it}$. Also note that, if we let

$$p_{it} = P(Y_{it} = 1),\ \mu_{it} = P(Y_{it} = 1 \mid Y_{it-1} = 0), \text{ and } \lambda_{it} = P(Y_{it} = 0 \mid Y_{it-1} = 1)$$

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 $\eta_{it} = 1 - \lambda_{it} - \mu_{it}$. The accounting identity in (1) is the elemental equation for estimating dynamic models with repeated cross-sectional samples as it relates the marginal probabilities $p_{it}$ and $p_{it-1}$ to the probabilities of inflow $(\mu_{it})$ and outflow $(\lambda_{it})$ 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. 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 both time-stationary and unit-homogeneous, 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 $p_{it}$ based on $t$ sets of successive transitions is $p_{it} = (\mu/(\mu + \lambda))(1 - \eta^t) + \eta^t p_{i0}$, which collapses to $p_{it} = \mu/(\mu + \lambda)$ as $t$ goes to infinity. The limiting result for $p_{it}$ gives the long-run probability of being in a state, i.e., for a time point sufficiently far in the future the probability that the state is 1 is $\mu/(\mu+\lambda)$. 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 noted by MOFFITT (1993), 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, FIENBERG, and HOLLAND 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., FIRTH 1982, HAWKINS, HAN and EISENFELD 1996, KALBFLEISH and LAWLESS 1984 1985, LAWLESS and MCLEISH 1984, LEE, JUDGE, and ZELLNER 1970, LI and KWOK 1990). 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 known as the 'mover-stayer' model (see, e.g., BARTHOLOMEW 1996, GOODMAN 1961).

Because the assumption of stationarity and homogeneity is generally not plausible and frequently violated in applications, it is desirable to relax this restriction. If we define the model as in Equation (1) and let $p_{i0} = 0$ (or $t \to \infty$), it is easy to verify that $p_{it}$ has the representation

$$p_{it} = \mu_{it} + \sum_{\tau=1}^{t-1} \left[ \mu_{i\tau} \prod_{s=\tau+1}^{t} \eta_{is} \right], \tag{2}$$

where $\eta_{is} = 1 - \lambda_{is} - \mu_{is}$. 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 the following estimation procedure. While repeated cross sections lack direct information on the individual transitions, they often do provide a set of time-invariant or time-varying covariates $X_{it}$ that affect the hazards. If so, the history of these covariates ($X_{it}, X_{it-1}, \dots, X_{i1}$) can be employed to generate backward predictions for the transition probabilities ($\mu_{it}, \mu_{it-1}, \dots, \mu_{i1}$ and $\lambda_{it}, \lambda_{it-1}, \dots, \lambda_{i2}$) and thus for the marginal probabilities ($p_{it}, p_{it-1}, \dots, p_{i1}$). Hence the basic idea is to model the current and past $\mu_{it}$'s and $\lambda_{it}$'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 Equation (2).

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

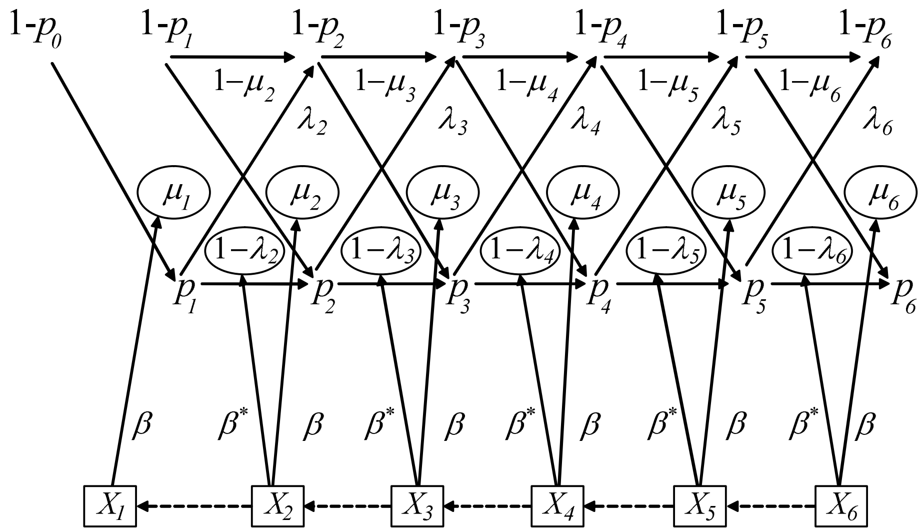
$$\begin{aligned} \text{logit } P(Y_{it} = 1 \mid Y_{it-1} = 0) &= \text{logit}(\mu_{it}) &&= X_{it}\beta, \text{ and} \\ \text{logit } P(Y_{it} = 1 \mid Y_{it-1} = 1) &= \text{logit}(1 - \lambda_{it}) &&= X_{it}\beta^*, \end{aligned}$$

where $\beta$ and $\beta^*$ are two potentially different sets of parameters. 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

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

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 (3) for panel data occur in AMEMIYA (1985), DIGGLE, LIANG and ZEGER (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

Figure 3.1 Graphical illustration of Markov model for RCS data



searching for a job, and withdrawal from employment) in which the transitions are estimated using frequency data disaggregated by sex.

According to Equation (3) the transition rates are $\mu_{it} = F(X_{it}\beta)$ and $\lambda_{it} = 1 - F\left[X_{it}(\alpha + \beta)\right]$, where $F$ - in this article - is the logistic function. Maximum likelihood estimates of $\alpha$ and $\beta$ can be obtained by maximization of the log likelihood function

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

with respect to the parameters, with $p_{it}$ defined by (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 marginal probability $p_{it}$. A graphical presentation of the model illustrating this is given in Figure 3.1, 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 $\mu_{it}$ and $\lambda_{it}$. 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 the available cross sections rather than simply the observations from the current time period. Thus estimates of the transitions 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.

### 3.2.2 ML estimation

Maximum likelihood fitting of the model in Equation (2) requires the derivatives of the likelihood function (4) with respect to the parameters. For ease of exposition, subscript $i$ is omitted in the expressions of the derivatives and Equation (2) is re-written as

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

where $\mu_\tau = (1+e^{-x_\tau\beta})^{-1}$, $\eta_s = 1-\lambda_s-\mu_s$, $\lambda_s = (1+e^{x_s(\alpha+\beta)})^{-1}$, and $x_\tau$ and $x_s$ the current and backcasted values of the covariates at $t=\tau$ and $t=s$, respectively. The first order partial derivatives of $p_t$ in Equation (5) with respect to the parameters $\beta$ and $\alpha$ are

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

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

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. 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

Fisher's method-of-scoring which provides an iterative search procedure for the estimation of $\beta$ and $\alpha$. Let $\theta$ be the vertical concatenation of the column vectors $\beta$ and $\alpha$, then the iteration scheme is $\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + \varepsilon[\widehat{\mathbf{I}}(\hat{\theta}^{(i)})]^{-1}(\partial LL(\hat{\theta}^{(i)}) \big/ \partial\theta)$ (see, e.g., AMEMIYA 1981). The parameter $\varepsilon$ denotes an appropriate step length that scales the parameter increments and $\hat{\mathbf{I}}(\hat{\theta}^{(i)})$ is an estimate of the Fisher information matrix $\mathbf{I}(\theta) = -\mathrm{E}[\partial^2 LL(\theta) / \partial\theta_j\partial\theta_k]$ evaluated at $\hat{\theta}^{(i)}$, where $\partial^2 LL(\theta) / \partial\theta_j\partial\theta_k$ 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 $\mathbf{I}^{-1}(\theta)$ evaluated at the values of the maximum likelihood estimates.

### 3.2.3 Some model extensions

A potential drawback to the model presented by MOFFITT (1990 1993) is that it assumes that the effects of the covariates are fixed over time, implying that they are expected to have much the same impact over the period of time during which the observations were obtained. This restriction may not be valid for long time periods and potentially biases the estimated effects. An alternative model that could be considered is to allow the regression coefficient to become polynomials in $t$ 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. Obviously, in practice it will be desirable to have models with low degree polynomials that avoid problems of overparametrization (i.e., nonexistence of unique ML estimates) and that combine parsimony of parametrization with fidelity to data. Another way in which we may accommodate the model is that whereas Moffitt defined the first observed outcome of the process $P(Y_{i1} = 1)$ to equal the transition probability $\mu_{i1}$, 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 $\delta$ is a set of parameters to be estimated and $F$ is the logistic function. The $\delta$-parameters for the first observed outcome at $t = 1$ are estimated simultaneously with the entry and exit parameters of interest at $t = 2, \ldots, T$. Moreover, recall that the probability vector at the beginning of the Markov chain is estimated as a function of all cross-sectional data, rather than simply the observations at

$t = 1$. Finally, we may also relax the implicit 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

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

where $w_t = \bar{n} / n_t$, with $\bar{n} = \Sigma_{t=1}^{T} n_t / T$, $n_t$ is the number of observations of cross section $t$ and $T$ is the number of cross sections.

## 3.3 Application

Our empirical application employs panel data on female labour force participation of Dutch women aged 20-64 drawn from the Socio-economic Panel (SEP) study conducted by STATISTICS NETHERLANDS in the period 1986-1995. The panel data were treated as if they were a temporal sequence of cross sections of unrelated women (i.e., no estimate of cov $(y_t, y_{t-1})$ is available in the data used to estimate the Markov model). These data were used because they allow us to verify the results of the Markov model. The labor market status $y_{it}$ is defined to equal 1 if the woman participates in the labor force at time $t$ and 0 otherwise. Table 3.1 gives the number of observations (including panel inflow and outflow), the marginal distribution of participation over time, and the observed annual entry and exit transitions rates in the panel. The table shows that over the period considered the female participation rate in the panel increased from about 40% in 1986 to around 56% in 1995. It also shows that both the panel entry and exit transition rates are relatively low. The analysis uses only covariates that are generally available in repeated cross-sectional surveys. As time-varying covariates, the analysis employs age in four different age categories (20-34, 35-44, 45-54, 55-64 years of age), the number of children at three different age categories (< 5, 5-17, $\geq$ 18 years of age), and the annual nationwide unemployment rate (in %). The covariate completed education is taken to be fixed over time. Next to these

Table 3.1 Marginal fraction of women's employment and observed annual entry and exit transition rates

| year | $n_t$ | inflow<br>(age 20) | outflow<br>(age 64) | $\bar{y}_t$ | $\bar{y}_t \mid y_{t-1} = 0$ | $1-\bar{y}_t \mid y_{t-1} = 1$ |
|---|---|---|---|---|---|---|
| 1986 | 2,302 | 52 | 21 | .400 | | |
| 87 | 2,299 | 18 | 33 | .406 | .076 | .109 |
| 88 | 2,306 | 39 | 28 | .425 | .097 | .106 |
| 89 | 2,308 | 30 | 28 | .432 | .087 | .109 |
| 90 | 2,316 | 36 | 36 | .448 | .107 | .113 |
| 91 | 2,288 | 8 | 47 | .476 | .127 | .105 |
| 92 | 2,241 | 0 | 41 | .515 | .128 | .074 |
| 93 | 2,200 | 0 | 39 | .525 | .097 | .086 |
| 94 | 2,161 | 0 | 48 | .526 | .077 | .082 |
| 95 | 2,113 | 0 | 36 | .557 | .121 | .066 |

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 its interaction with the aggregate unemployment rate. The potentially important interaction of first entry with number of children was not included, as the number of mothers aged 20 was insufficient to allow reliable estimation. It is of interest to note that the individual observations were backcasted until the minimum age of 20, at which the first entry into the participation process is taken to have occurred. If for an observation the backcasted value of age in a particular cross section was less than 20, the entry and exit rates at that time period were fixed to zero.

First a simple time-stationary Markov model with constant terms only was applied to the data using the software program ***CrossMark*** (which is available upon request). This model produced a $\beta(\mu_{t>1})$ of -.222 and a $-\beta^*(\lambda_{t>1})$ of -.078. These estimates imply constant transition rates of $\mu = .445$ and $\lambda = .480$; hence implausibly high values that amply exceed those reported in Table 3.1. The model was thereupon extended to a nonstationary, heterogeneous Markov model by including the covariates reported above. The results are shown in Table 3.2.

The parameters in the first column show the effect of the variables on the employment state probability $p_{i1}$ at $t = 1$, estimated for all

Table 3.2 Markov repeated cross section estimates for women's transition into and out of employment, $n$ =22,534

| | $\delta(p_{t=1})$ [a] | $\beta(\mu_{t>1})$ | -$\beta^{*}(\lambda_{t>1})$ |
|---|---|---|---|
| Intercept | -.027 *(.099)* | -.684 *(.468)* | -1.877* *(.670)* |
| Education | .322* *(.031)* | .347* *(.043)* | -.570* *(.067)* |
| Age [b]: | | | |
| 35-44 years old | -.199* *(.079)* | -1.287* *(.127)* | -2.190* *(.287)* |
| 45-54 years old | -1.198* *(.095)* | -1.592* *(.203)* | -.311 *(.309)* |
| 55-64 years old | -2.187* *(.115)* | -3.139* *(.439)* | 1.290* *(.240)* |
| Number of children: | | | |
| < 5 years old | -1.543* *(.094)* | -.214* *(.089)* | 2.066* *(.151)* |
| 5-17 years old | -.438* *(.036)* | -.017 *(.052)* | .220* *(.107)* |
| ≥18 years old | -.176* *(.054)* | .091 *(.105)* | -.253 *(.179)* |
| Unemployment rate | | -.225* *(.067)* | .052 *(.093)* |
| Age20 [b] | | .853 *(1.599)* | |
| Age20 × education | | .306 *(.209)* | |
| Age20 × unemployment rate | | .283 *(.191)* | |
| Log likelihood ($LL$*) | | | -12760.67 |

* Significant at 5% level (based on the estimated information matrix).

[a] Estimates of standard errors in parentheses. The $\beta$ -parameters represent the effect on $\mu_t$, the $\beta^*$ -parameters the effect on $(1-\lambda_t)$, and thus -$\beta^*$ the effect on $\lambda_t$.

[b] Reference category Age=20-34 years; Age20: 1 if age = 20, 0 if not.

observations in the model. As can be seen, the parameters are well determined, with employment positively affected by education and negatively by age and the number of children (particularly preschool children) in the household. The second column in Table 3.2 presents the effect of the variables on the transition from non-employment to employment. Whereas education is significant in encouraging entry into the labor force, young children in the household and the aggregate unemployment rate negatively affect the entry decision. We also find that age has a negative effect on entry implying that the entry rates decline with age. The initial conditions variables indicate that higher unemployment rates and higher education increase the probability of entry at age 20. According to the standard errors, however, these variables have little impact on the hazards. The third column gives the effect of the variables on the transition into non-employment. We find that the exit rates are negatively affected by education and positively by the number of school and preschool children in the household. The coefficients of the age terms imply that the incentives to end a job initially decrease with age but they are forced up again (presumably by occupational pension) after the age of 54. The effect of the aggregate unemployment rate on the transition into non-employment is insignificant.

Because there are substantive arguments to anticipate that the effect of some of the covariates (intercept, number of young children, education) may vary over time, several tests with different time-varying-coefficient models were applied to the data. These models, however, describe the data only slightly better (in terms of goodness-of-fit) than the time-constant-coefficient model and their results are therefore not reported here. We instead concentrate on an examination of the fit of the estimated model presented in Table 3.2 in terms of predictions. There are several ways to do so. One is to compare the actual sample frequency of all possible labor force participation sequences from 1986 to 1995 with the estimated expected frequency of each sequence. The latter were computed as follows. With *T* sample periods, we have $\Sigma_{t=1}^{T} 2^t$ different sequences (which in the present application equals 2,046) ranging in length from 1 (e.g., '0') to *T* (e.g., '0101010101') . We define the probability of a sequence of length *t* for each observation $i$ of cross section *t* as

$$\tilde{p}_i(\tilde{y}_1, ..., \tilde{y}_t) = P(Y_{i1} = \tilde{y}_1 \cap \ldots \cap Y_{it} = \tilde{y}_t),$$

where $\tilde{y}_1, \ldots, \tilde{y}_t = 0{,}1$. Hence

$$\tilde{p}_i(\tilde{y}_1) = P(Y_{i1} = \tilde{y}_1) = \tilde{y}_1 p_{i1} + (1 - \tilde{y}_1)(1 - p_{i1}),$$

where $p_{i1}$ is $P(Y_{i1} = 1)$. For $t > 1$, we have

$$\tilde{p}_i(\tilde{y}_1, ..., \tilde{y}_t) = \tilde{p}_i(\tilde{y}_1)\, \Pi_{\tau=2}^{t}(p_{00} + p_{01} + p_{10} + p_{11}),$$

where $p_{00} = (1 - \tilde{y}_{\tau-1})(1 - \tilde{y}_\tau)(1 - \mu_{i\tau})$, $p_{01} = (1 - \tilde{y}_{\tau-1})\tilde{y}_\tau \mu_{i\tau}$, $p_{10} = \tilde{y}_{\tau-1}$ $(1 - \tilde{y}_\tau)\lambda_{i\tau}$, and $p_{11} = \tilde{y}_{\tau-1}\tilde{y}_\tau(1 - \lambda_{i\tau})$. The mean value of $\tilde{p}_i(\tilde{y}_1, ..., \tilde{y}_t)$ for all observations of cross section *t* was obtained as $\tilde{p}(\tilde{y}_1, ..., \tilde{y}_t) =$ $\Sigma_{i=1}^{n_t}$ $\tilde{p}_i(\tilde{y}_1, ..., \tilde{y}_t) / n_t$. The estimated expected absolute frequency $\tilde{f}(\tilde{y}_1, ..., \tilde{y}_t)$ of each participation sequence was thereupon computed by evaluating $\tilde{f}(\tilde{y}_1, ..., \tilde{y}_t) = \tilde{p}(\tilde{y}_1, ..., \tilde{y}_t)\ n_t$.

Table 3.3 Relative frequencies of estimated expected $(\tilde{y}_{t-1}, \tilde{y}_t)$ transitions at sample period $T$ and estimated expected minus observed proportions, $n$=22,534

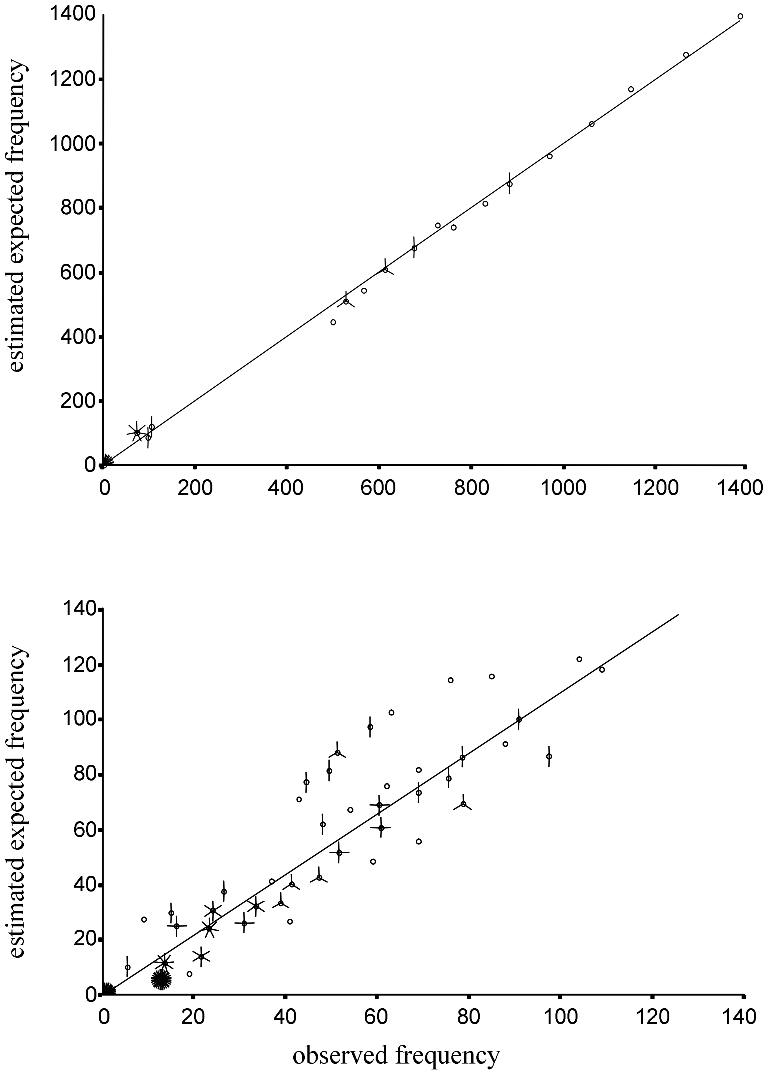
| $T$ | $n_t$ | estimated expected | | | | expected - observed | | | | $\chi^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | (00) | (01) | (11) | (10) | (00) | (01) | (11) | (10) | |
| 2 | 2,299 | .556 | .053 | .354 | .037 | .006 | .008 | -.007 | -.007 | 6.03 |
| 3 | 2,306 | .540 | .056 | .365 | .039 | .009 | -.001 | -.003 | -.005 | 1.78 |
| 4 | 2,308 | .521 | .060 | .381 | .038 | .000 | .010 | -.002 | -.009 | 8.57 |
| 5 | 2,316 | .495 | .067 | .400 | .038 | -.008 | .007 | .012 | -.011 | 10.49 |
| 6 | 2,288 | .469 | .059 | .432 | .040 | -.007 | -.010 | .025 | -.008 | 10.96 |
| 7 | 2,241 | .448 | .053 | .460 | .039 | .000 | -.013 | .011 | .003 | 8.41 |
| 8 | 2,200 | .441 | .038 | .482 | .039 | .011 | -.008 | .003 | -.006 | 6.12 |
| 9 | 2,161 | .441 | .031 | .491 | .037 | .011 | -.005 | .001 | -.007 | 5.47 |
| 10 | 2,113 | .435 | .033 | .500 | .032 | .028 | -.023 | -.001 | -.003 | 36.68 |

An initial examination is to compare the expected with the observed first-order transitions over the time period of our data. Table 3.3 shows the relative frequencies of the estimated expected $(\tilde{y}_{t-1}, \tilde{y}_t)$ transitions and the differences between the expected and the observed relative frequencies. As can be seen, the predicted frequencies are concentrated in the continuous work (11) and continuous nonwork (00) categories. Further, while for some time periods the discrepancies between the predicted and the observed proportions are significant at the .05 level, most differences are very small. This implies that both the mover and the stayer frequencies are predicted fairly well.

A further examination of the fit of the model reported here is to compare the estimated expected and the actually observed absolute frequencies of all 2,046 employment status sequences. Because it is unfeasible to tabulate all frequencies, they are graphically displayed in Figure 3.2 together with the OLS regression lines.

The top part of the figure displays the predicted and the actual frequencies of all possible employment profiles, but highlights the relatively small number of sequences with high frequencies. These sequences concern the continuous participation and continuous nonparticipation categories. The bottom part of the figure zooms in on the employment sequences with relatively low frequencies in the 0-140 range. Visual inspection suggests close agreement between the estimated expected frequencies predicted by the RCS Markov model and the observed

Figure 3.2 Estimated expected versus observed frequencies of 2,046 employment status sequences and OLS regression lines



frequencies of the spells in the panel. The (unreported) longitudinal profiles indicate that most women remain employed or non-employed throughout the observation interval and that proportionally few women move into and out of the labor force frequently.

## 3.4 Conclusion

The overall conclusion that we draw from this example is that the proposed model can be a useful tool in applied work. It obviously does not supersede genuine panel designs, but it definitely puts a series of one-shot surveys into perspective and it provides more refined results than would be available from a single cross-sectional study. Microdata panel sets 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. RCS data are cheaper to collect and they do not suffer from problems of non-random attrition which plague panel data. Moreover, a disadvantage to using pure panel surveys is the limited number of units followed and the limited number of time points at which these units are usually re-interviewed. These limitations have to be balanced against the lack of direct information on the transitions in long-run RCS data.

Some problems we encountered in trying to model unobserved transitions over time using RCS data deserve to be mentioned. The application of the model 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 backcasted with considerable accuracy and many aggregate indicators are also measurable in the past. Moreover, our experiments have shown that it is also possible to specify a model with two different sets of parameters for both $\mu$ and $\lambda$ , i.e., one for the past transition rates and a separate one for the transition at the current time period. This offers the opportunity to also include relevant non-backcastable covariates in the (current part of the) Markov model.

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 initial conditions are also a source of population heterogeneity. The presample 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. Hence further investigation is needed on how much of the evidence is censored.

# References

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

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

Baltagi, B. H. (1995), *Econometric analysis of panel data*, Wiley, Chicester.

Bartholomew, D. J. (1996), *The statistical approach to social measurement*, Academic Press, San Diego.

Bishop, Y. M. M., S. E. Fienberg and P. W. Holland (1975), *Discrete multivariate analysis: theory and practice*, MIT Press, Cambridge MA.

Boskin, M. J. and F. C. Nold (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.

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

Diggle, P. J., K. Y. Liang and S. L. Zeger (1994), *Analysis of longitudinal data*, Clarendon Press: Oxford.

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

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

Girma, S. (2000), A quasi-differencing approach to dynamic modelling from a time series of independent cross-sections, *Journal of Econometrics* **98**, 365-383.

Girma, S. (2001), A note on dynamic modelling from short and heterogeneous pseudo panels, *Statistica Neerlandica* **55**, 238-247.

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

Hamerle, A. and G. Ronning (1995), Panel analysis for qualitative variables, in: G. Arminger, C. Clogg and M. E. Sobel (eds.), *Handbook of statistical modeling for the social and behavioral sciences*, Plenum Press, New York, 401-451.

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

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

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

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

Lee, T.C., G.G. Judge and A. Zellner (1970), *Estimating the parameters of the Markov probability model from aggregate time series data*, North-Holland, Amsterdam.

Li, W.K. and M. C. O. Kwok (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.

Mebane, W.R. and J. Wand (1997), *Markov chain models for rolling cross-section data: how campaign events and political awareness affect vote intentions and partisanship in the United States and Canada*. Paper presented at the 1997 Annual Meeting of the Midwest Political Science Association, Chicago Il.

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. and M. Verbeek (1990), Estimation of time-dependent parameters in linear models using cross-sections, panels, or both, *Journal of Econometrics* **46**, 333-446.

Rao, C. R. (1973), *Linear statistical inference and its applications*, Wiley, New York.

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.

Verbeek, M. (1996), Pseudo panel data, in: L. Mátyás and P. Sevestre (eds.). *The econometrics of panel* data (2nd revised ed.), Kluwer Academic Publishers, Dordrecht, 280-292.

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

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

Verbeek, M. and F. Vella (2000), *Estimating dynamic models from repeated cross sections*. Paper presented at the International Conference on the Analysis of Repeated Cross-sectional Surveys, June 15-16, 2000, Nijmegen, Netherlands.

# 4 Inferring Transition Probabilities from Repeated Cross Sections

This chapter[1] discusses a nonstationary, heterogeneous Markov model designed to estimate entry and exit transition probabilities at the micro level from a time series of independent cross-sectional samples with a binary outcome variable. The model has its origins in the work of Moffitt and shares features with standard statistical methods for ecological inference. We outline the methodological framework proposed by Moffitt and present several extensions of the model to increase its potential application in a wider array of research contexts. We also discuss the relationship with previous lines of related research in political science. The example illustration uses survey data on American presidential vote intentions from a five-wave panel study conducted by Patterson in 1976. We treat the panel data as independent cross sections and compare the estimates of the Markov model with both dynamic panel parameter estimates and the actual observations in the panel. The results suggest that the proposed model provides a useful framework for the analysis of transitions in repeated cross sections. Open problems requiring further study are discussed.

[1] This chapter has been published as Pelzer, B.J., R. Eisinga, and P.H. Franses. 2002. Inferring Transition Probabilities from Repeated Cross Sections. *Political Analysis,* 10: 113-133. We would like to thank sincerely the anonymous reviewers for their most helpful comments on an early version of the chapter. Their remarks and guidance led to many improvements in the presentation of the manuscript. Some elements of the model considered here were published previously by the authors (2001). The data utilized in this chapter were made available by the Inter-university Consortium for Political and Social Research (ICPSR). The data for *Presidential Campaign Impact on Voters: 1976 Panel, Erie, Pennsylvania, and Los Angeles* were originally collected by Thomas E. Patterson. Neither the collector of the original data nor the Consortium bears any responsibility for the analysis or interpretation presented in this paper. The data and the stand-alone program *CrossMark* (including documentation), needed to replicate the ML estimations reported here, are available at the *Political Analysis* Web site.

## 4.1 Introduction

Surveys that trace the same units across occasions provide the most powerful sorts of data for dynamic analysis of political phenomena. However, repeated observations are often unavailable, and many panel data sets that do exist are of limited time coverage. This shortcoming combined with potential drawbacks such as nonrandom attrition and conditioning restrict the use of panel data for the analysis of long-term political change.

In the absence of suitable panel data, repeated cross-sectional (RCS) surveys carried out with a regular periodicity may provide a viable alternative. These data do not suffer from problems of selective attrition that often plague panel data. Moreover, there exists an abundance of high-quality RCS data and many repeated cross-sectional surveys are available for relatively long time periods, some of which continue to accumulate. Given the importance of dynamics in political studies and the lack of panel data on many important issues, it would be of great advantage if RCS data could somehow be used for the estimation of longitudinal models with a dynamic structure. The objective of this paper is to explore those possibilities. Specifically, our purpose here is to present a nonstationary, heterogeneous Markov model for the analysis of a binary dependent variable in a time series of independent crosssectional samples. The model has its origins in the work of Moffitt (1990, 1993) and shares features with standard statistical methods for ecological or cross-level inference as outlined, for example, by Achen and Shively (1995) and King (1997). It offers the opportunity to estimate individual-level entry and exit transition rates and to examine the effects of timeconstant and time-varying covariates on the transitions. Previous discussions of (aspects of) the model include those by Felteau et al. (1997), Mebane and Wand (1997) and Pelzer et al. (2001).

The following section first presents the basic Markov model for RCS data as proposed by Moffitt and subsequently discusses several extensions of his approach and its relationship with related research in political science. Section 4.3 provides an example application using panel data on American presidential vote intentions from a five-wave survey conducted by Patterson (1980) in 1976. We treat these data as independent cross sections and compare (i) the parameter estimates obtained from the Markov model for RCS data with the estimates obtained from a dynamic panel model and (ii) the transitions predicted by the model with the actual transitions in the panel. We do not aim to present a very detailed analysis of the electoral data here. The subject matter itself is not the ultimate object and we also ignore the potential biases due to panel mortality. Our interest here is to calibrate a rather unfamiliar statistical technique on a reasonably well-understood set of data to increase the understanding of the model rather than offer an immediate analysis of voter preferences and a detailed subject-matter interpretation. Most of our substantive results correspond to well-accepted political science findings. Yet more crucial to our topic is that the validation results suggest that the model can provide a useful tool for inferring individual-level transition probability estimates in the absence of transition data. We conclude with a discussion of open problems requiring further study[2].

## 4.2 Estimating transition probabilities with RCS data

### 4.2.1 Basic model

Consider a two-state Markov matrix of transition rates in which the cell probabilities sum to unity across rows. For this 2×2 table, we define the following three terms, where $y_{it}$ denotes the value of the binary random variable $y$ for unit $i$ at time $t$: $p_{it} = P(y_{it} = 1)$, $\mu_{it} = P(y_{it} = 1 \mid y_{it-1} = 0)$, and $\lambda_{it} = P(y_{it} = 0 \mid y_{it-1} = 1)$. These marginal and conditional probabilities, respectively, give rise to the well-known flow equation

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

where $\eta_{it} = 1 - \lambda_{it} - \mu_{it}$. This accounting identity — also used in Goodman's ecological regression (Goodman 1953; King 1997) — is the

[2] It is assumed in this paper that the responses are observed at evenly spaced discrete time intervals $t$ =1, 2, . . . , and that the samples at periods $t_j$ and $t_k$ are independent if $j \neq k$. The subscript $it$ is commonly used to indicate repeated observations on the same sample element $i$. However, to simplify notation, this paper uses the subscript $it$ to index nonpanel individuals in RCS samples.

elemental equation for estimating dynamic models with repeated cross sections as it relates the marginal probabilities $p_i$ at $t$ and $t-1$ to the entry ($\mu_{it}$) and exit ($\lambda_{it}$) transition probabilities. Clearly, a dynamic analysis of repeated cross sections is difficult because the surveys are "incomplete" in the sense that they do not assess directly the state-to-state transitions over time for each individual unit. That is, there is no information on the temporal covariances $(y_{it}, y_{it-1})$ available in the data, and this information gap implies that some identifying constraints over $i$ and/or $t$ must be imposed to estimate the unobserved transitions uniquely.

Different types of restrictions may be called upon (see Moffitt 1990). A rather restrictive approach frequently applied in the statistical literature is to assume *a priori* that the transition probabilities are time-invariant and unit-homogeneous, hence $\mu_{it} = \mu$ and $\lambda_{it} = \lambda$ for all $i$ and $t$. It is easy to show that in this case the long-run steady-state outcome of $p_{it}$ is $p_{it} = \mu/(\mu + \lambda)$.[3] Some early references relating to models of this type include those that estimate transition rates from aggregate frequency data (e.g., Lee et al. 1970; Lawless and McLeish 1984; Kalbfleish and Lawless 1984, 1985). The formulation has also been used in applied economic studies (McCall 1971; Topel 1983), in the famous mover–stayer model of intragenerational job mobility (Goodman 1961; Bartholomew 1996), and in electoral studies on voter transitions (e.g., Firth 1982). The assumption, however, that individual differences in transitions are not present in the population lacks plausibility in many empirical applications. Many populations studied are heterogeneous in the sense that they comprise variation in transitions between units within periods and within units over time. Consequently, as noted by Hawkins and Han (2000), studies that assume a time-invariant Markov model with a homogeneous transition probability matrix have typically found their estimates to be highly inefficient.

Moffitt (1993) proposed a model that relaxes the assumption of a time-invariant and unit-homogeneous population. If we define the model as

[3] Let $p_{i1} = \mu + \eta p_{i0}$, $p_{i2} = \mu + \eta p_{i1} = \mu + \eta(\mu + \eta p_{i0}) = \mu(1+\eta) + \eta^2 p_{i0}$, where $\eta = 1 - \lambda - \mu$. Hence $p_{it} = \mu(1 + \eta + \cdots + \eta^{t-1}) + \eta^t p_{i0} = \mu\,(1 + \Sigma_{\tau=1}^{t-1} \eta^{t-\tau}) + \eta^t p_{i0}$ $= (\mu/(\mu+\lambda))(1-\eta^t) + \eta^t p_{i0}$. As $t \to \infty$, the polynomial $\eta^t$ tends to 0, thus $p_{it} = \mu/(\mu+\lambda)$. Obviously, this equation holds for $-1 < \eta < 1$, as there is no steady-state outcome if $|\eta| = 1$ (see also Bishop et al., 1975: 261-262, and Ross, 1993: 152-153).

in Equation (1), it is straightforward to show that the reduced form for $p_{it}$ is

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

where $\eta_{is} = 1 - \lambda_{is} - \mu_{is}$, assuming $p_{i0} = 0$ or $t \to \infty$.[4] By explicitly allowing for time dependence and unit heterogeneity, this dynamic version of Equation (1) is better suited to yield a more informative model, as it imposes no *a priori* homogeneous structure on the transitions.

The framework Moffitt (1993) uses to estimate Equation (2) is based on the following observation. While RCS data lack direct information on transitions in opinions, preferences, choices, and other individual characteristics, they often do provide a set of time-invariant and time-varying covariates $\mathbf{x}_{it}$ that affect the hazards (i.e., the entry and exit transition probabilities). If so, the history of these covariates (i.e., $\mathbf{x}_{it}, \mathbf{x}_{it-1}, \ldots, \mathbf{x}_{i1}$) 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 key here is to model the current and past $\mu_{it}$ and $\lambda_{it}$ in a regression setting as functions of current and backcasted values of time-invariant and time-varying covariates $\mathbf{x}_{it}$. The parameter estimates of the covariates are obtained by substituting the hazards into Equation (2). The hazards themselves are specified as $\mu_{it} = F(\mathbf{x}_{it}\beta)$ and $\lambda_{it} = 1 - F(\mathbf{x}_{it}\beta^*)$, where $F$ — in the current paper — is the logistic link function [Moffitt (1993) uses the probit]. Hence, it is assumed that

$$\text{logit}\,(\mu_{it}) = \mathbf{x}_{it}\beta \quad \text{and} \quad \text{logit}\,(1 - \lambda_{it}) = \mathbf{x}_{it}\beta^*, \tag{3}$$

where $\beta$ and $\beta^*$ are two potentially different sets of parameters associated with two potentially different sets of covariates $\mathbf{x}_{it}$. This regression setup offers the opportunity to estimate transition probabilities that vary across individuals and — if the model includes time-varying covariates — time

[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 $p_{it} = \mu_{it} + (\mu_{it-1}\eta_{it} + \mu_{it-2}\eta_{it-1}\eta_{it} + \cdots + \mu_{i1}\eta_{i2}\cdots\eta_{it}) + p_{i0}\eta_{i1}\cdots\eta_{it} = \mu_{it} + \Sigma_{\tau=1}^{t-1} \mu_{i\tau} (\Pi_{s=\tau+1}^{t} \eta_{is}) + p_{i0} \Pi_{\tau=1}^{t} \eta_{it}$. As $t \to \infty$, $\Pi_{\tau=1}^{t} \eta_{it}$ tends to 0, thus $p_{it} = \mu_{it} + \Sigma_{\tau=1}^{t-1} \mu_{it} (\Pi_{s=\tau+1}^{t} \eta_{is})$. Obviously, we get the same form for $p_{it}$ if we let $p_{i0} = 0$.

periods. Note that it is assumed that the regression coefficients are fixed over time. This is the fundamental restriction Moffitt (1993) imposes to secure the identifiability of the parameters. There is, however, no need to invoke the assumption of time-constant parameters if we have a sufficient number of cross sections. We will return to this point momentarily. Maximum likelihood (ML) estimates of $\beta$ and $\beta^*$ can be obtained by maximization of the log-likelihood function

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

with respect to the parameters, where $T$ is the number of cross sections and $n_t$ the number of units of cross section $t$.[5] As Moffitt (1993) notes, obtaining $p_{it}$ by means of Equation (2) is equivalent to "integrating out" over all possible transition histories for each individual $i$ at time $t$ to derive an expression for the marginal probability estimates. To convey this idea, compare the contribution to the likelihood of the $i$th case at time $t$ in panel data with the likelihood contribution of the same case in RCS data. For a first-order transition model of binary recurrent events the contribution can be written as

$$L_{it} = \mu_{it}^{y_{it}(1-y_{it-1})}(1-\lambda_{it})^{y_{it}y_{it-1}}(1-\mu_{it})^{(1-y_{it})(1-y_{it-1})}\lambda_{it}^{(1-y_{it})y_{it-1}} \tag{5}$$

(e.g., Stott 1997). Hence, conditional on $y_{it}$ and $y_{it-1}$, the likelihood contribution in binary panel data simplifies to a single transition probability estimate. In the Markov model for RCS data proposed by Moffitt (1993), however, the contribution of the $i$th case is given by

$$L_{it} = \left[\mu_{it}(1-p_{it-1}) + (1-\lambda_{it})p_{it-1}\right]^{y_{it}}\left[(1-\mu_{it})(1-p_{it-1}) + \lambda_{it}p_{it-1}\right]^{1-y_{it}}. \tag{6}$$

In this formulation the likelihood contribution is not a single hazard but, rather, a weighted sum of two transition probabilities. Note that in the Markov model for RCS data the transition probabilities are estimated as a

[5] If the samples of the repeated cross-sectional surveys have an unequal number of observations, it may be desirable ensure a potentially equal contribution of the cross-sectional units to the likelihood by using the weighted log likelihood function $LL^* = \Sigma_{t=1}^{T}\Sigma_{i=1}^{n_t} m_t\ell\ell_{it}$, where $m_t = \bar{n}/n_t$, with $\bar{n} = \Sigma_{t=1}^{T}n_t/T$.

function of all of the available cross-sectional samples rather than simply the observations from the current time period (Mebane and Wand 1997). This full information strategy expresses the notion that in RCS data different individuals are observed over time, but individuals sharing the same covariate values are considered to be exchangeable in the sense that their transition histories are assumed to be identical. Also, note from the comparison that some efficiency is likely to be lost if we use RCS data instead of a comparable panel data set with the same sample size. But too much should not be made of mentioning differences in the efficiency of estimators since repeated cross-sectional surveys typically have a larger effective sample size than pure panels (see Heckman and Robb 1985; Moffitt 1990).

## 4.2.2 Modifications and extensions of the model

### *4.2.2.1 Infinite time horizon and initial condition*

The Markov model presented in Equation (2) assumes that either $p_{i0} = 0$ or $t \to \infty$. The latter does not imply that the model is appropriate only in an infinite-horizon setting. Successful application of the model, as our example shows, does not even require data from a large number of time points. In fact, given good instrumental variables, two cross-sectional samples would be sufficient. Also, inferences in the model are not conditional on the observed units and we do not want to make inferences to some notational or hypothetical population. The model is used to make probability statements about a well-defined sample (or target) population from which the purposive repeated samples were selected. The infinite-horizon notation does imply, however, that there is a tendency as time passes for the probability of being in a state to become independent of the initial condition at $t = 0$. For this reason the initial condition is often regarded as a matter of minor importance in Markov modeling and in many applications involving finite-horizon situations it is assumed that $p_{i0} = 0$ (Bishop et al. 1975). It may be objected that this assumption is not very realistic for social and political phenomena, which are often characterized by features such as inertia and state dependence. It is clear, however, that when the number of time points grows large, the weight of the initial observations in the likelihood becomes negligible and it is appropriate to ignore this issue.

As noted by Moffitt (1993), the initial probability (i.e., $p_{i0}$) refers to the value of the state prior to the start of the Markov process (for example, the state of being below voting age at the beginning of a vote/nonvote sequence) rather than to the first observed outcome (which is $p_{i1}$). If the initial states are known and fixed, they can be included in the model as additional explanatory variables. For example, initial condition variables can be used to capture the first entry into the vote/nonvote process at voting age 18 and, if appropriate, to capture the interaction of first entry with other characteristics such as race and education (see Moffitt 1993). To do so, one backcasts the individual observations until the minimum age of 18, at which the first entry into the process is assumed to have occurred, and estimates $p_i$ for the individuals aged 18 (which is not necessarily $p_{i1}$). If for an individual the backcasted value of age in a particular cross section is 18 or less, the entry and exit transition probabilities at that time period are fixed to 0. So if it is appropriate to assume that an individual is at the start of a new process, the initial state can be incorporated into the model. But for most individuals in the samples we do not have access to the process from the beginning. The first observed outcome for these individuals cannot be assumed fixed as it is determined by the process generating the sample observations. Getting around this problem is difficult, but it might be solved, at least in part, as follows. Moffitt (1993) assumes that $p_{i0} = 0$ and defines $P(y_{i1} = 1)$ to equal the transition probability $\mu_{i1}$. In many applications this assumption is untenable and it seems more plausible simply to take $P(y_{i1} = 1)$ to equal the state probability $p_{i1}$. Thus for all of the cross-sectional samples the model starts with $p_{i1}$ instead of $\mu_{i1}$, invoked by the assumption that $p_{i0} = 0$. That is, one assumes that the $y_{i1}$'s are random variables with a probability distribution $P(y_{i1} = 1) = F(\mathbf{x}_{it}\delta)$, where $\delta$ is a set of parameters to be estimated and $F$ is the logistic link function. The $\delta$ parameters for the first observed outcomes at $t = 1$ are estimated simultaneously with the entry and exit parameters of interest at $t = 2,...,T$. Note, again, that the probability vector at the beginning of the observed Markov chain, $p_{i1}$, is estimated as a function of all cross-sectional data, rather than simply the observations at $t = 1$.

#### *4.2.2.2 ML estimation*

Maximum likelihood estimation requires the (analytic or numerical) derivatives of the loglikelihood function with respect to the parameters. If we suppress the subscript $i$ for the moment to avoid cumbersome notation, the first-order partial derivatives of $\ell\ell$ with respect to the parameters $\beta$ and $\beta^*$ are

$$
\begin{aligned}
\frac{\partial \ell\ell}{\partial \beta} &= \frac{\partial \ell\ell}{\partial p_t} \cdot \frac{\partial p_t}{\partial \beta} = \frac{y_t - p_t}{p_t(1-p_t)} \cdot \left( \frac{\partial p_{t-1}}{\partial \beta} \eta_t + \frac{\partial \mu_t}{\partial \beta}(1 - p_{t-1}) \right), \\
\frac{\partial \ell\ell}{\partial \beta^*} &= \frac{\partial \ell\ell}{\partial p_t} \cdot \frac{\partial p_t}{\partial \beta^*} = \frac{y_t - p_t}{p_t(1-p_t)} \cdot \left( \frac{\partial p_{t-1}}{\partial \beta^*} \eta_t - \frac{\partial \lambda_t}{\partial \beta^*} p_{t-1} \right),
\end{aligned} \tag{7}
$$

where $\partial \mu_t / \partial \beta = x_t \mu_t (1 - \mu_t)$ and $\partial \lambda_t / \partial \beta^* = -x_t \lambda_t (1 - \lambda_t)$. Fisher's method-of-scoring (Amemiya 1981) may be used to obtain both the ML parameter estimates and an estimate of the asymptotic variance-covariance matrix of the model parameters. Further details about the method-of-scoring procedure, including the analytic derivatives of $p_t$ with respect to the parameters, are provided by Pelzer et al. (2001).

#### *4.2.2.3 Nonbackcastable covariates*

The estimation strategy proposed by Moffitt (1993) involves searching the cross-sectional data files for variables taking known values in the past. Clearly, time-invariant characteristics such as sex, race, cohort, and completed education are candidates, and time-specific aggregates measurable in the past may also enter the model. But variables such as age are usable too, as are age-related variables such as the number of children at different ages, since knowledge of the current age implies knowledge of age in any past year. However, in many application settings we have time-dependent covariates that the basic model would omit because the past histories are unknown. To incorporate these "nonbackcastable" variables, we may adopt a model with two different sets of parameters for both $\mu_{it}$ and $\lambda_{it}$, i.e., one for the current transition probability estimates and a separate one for the preceding estimates. Define $\mathbf{v}_{it}$ as a vector of nonbackcastable variables and $\zeta$ as the associated parameter vector. One can then write

$$\operatorname{logit}(\mu_{it}) = \begin{cases} \mathbf{x}_{it}\beta^{**} + \mathbf{v}_{it}\zeta & \text{for } t \\ \mathbf{x}_{it}\beta & \text{for } t-1,\ldots,1. \end{cases} \tag{8}$$

A similar model may be specified for $\lambda_{it}$. These specifications offer the opportunity to express the current transition probability estimates as a logistic function of both the backcastable and nonbackcastable variables. The expression also affords a test — useful for efficiency gains — of the hypothesis $\beta^{**} = \beta$. Whether variables can be backcasted with reasonable accuracy obviously also depends on the time span of the repeated cross-sectional data. If, for example, the samples concern a limited number of consecutive week surveys, even nonbackcastable variables such as income may reasonably be treated as time-constant. Also, the model can easily be adjusted so that backcasting is performed for a limited number of time periods. Restricted backcasting may be preferred if only the immediate history is known or if covariates can safely be assumed to be constant only for a particular number of time points.

### *4.2.2.4 Time-varying covariate effects*

Another drawback of the basic model is that it assumes that the parameters of the covariates are fixed over the time period during which the repeated cross-sectional samples were obtained. As indicated above, this is the critical identifying restriction Moffitt imposed to estimate the parameters. However, the assumption of time-constant coefficients cannot be expected to remain valid for long periods of time and thus potentially biases the estimated effects. Relevant changes in the population and events that intervene in consecutive cross sections induce variation in the population parameters. There are at least two approaches to deal with time dependence. One is to use a fully parametric approach, not pursued in this paper, and to allow the regression coefficients to become a specific function of time using, for example, the polynomial function $\beta_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \cdots + \gamma_d t^d$, where the positive integer $d$ specifies the degree of the polynomial. Alternatively, one may use a partially parametric approach, as in this paper, divide the time axis into discrete time periods, and assume that the parameters are constant within but vary across time periods. An advantage of the fully parametric approach is that it often requires that fewer additional parameters be estimated, but in some applications it may not provide enough flexibility and local adaptiveness. It will also be necessary in the fully parametric approach to have models with low-degree polynomials to avoid nonexistence of unique ML estimates. The partially parametric approach is particularly useful when little is known about the form of the time dependence. Obviously, in this approach too, continually modifying the values of the parameters so as to allow the model to adapt itself to local conditions produces problems of overparameterization.

### *4.2.2.5 Unobserved heterogeneity*

The framework discussed by Moffitt (1993) assumes that the differences in transitions within the population depend only on variation in the observed variables used as covariates in the model. However, the assumption that the model includes all relevant variables is rarely even approximately true in social and political science practice. Therefore, another useful extension of the basic model is to include an additional, individual-specific random error term, $\varepsilon_i$, in the linear predictor of the transition probabilities to account for omitted variables, at least insofar as the omitted variables are time-invariant for each individual. In this so-called logistic-normal mixture model we have logit $(\mu_{it}^*) = \mathbf{x}_{it}\beta + \gamma_0\varepsilon_i$ and logit $(1-\lambda_{it}^*) = \mathbf{x}_{it}\beta^* + \gamma_1\varepsilon_i$, where $\gamma_0$ and $\gamma_1$ are the coefficients of the random variable $\varepsilon_i$ having zero mean and unit variance (Collett, 1991). Hence $\mu_{it}^*$ and $(1-\lambda_{it}^*)$ have a logistic-normal distribution, e.g., $\text{logit}(\mu_{it}^*) \sim N(\mathbf{x}_{it}\beta, \gamma_0^2)$. This model has the marginal log likelihood

$$LL = \sum_{t=1}^{T}\sum_{i=1}^{n_t} \int_{-\infty}^{\infty} [y_{it}\log(p_{it}^*) + (1-y_{it})\log(1-p_{it}^*)] \cdot f(\varepsilon_i)\, d\varepsilon_i\,, \tag{9}$$

where $p_{it}^* = \mu_{it}^*(1-p_{it-1}^*) + (1-\lambda_{it}^*)p_{it-1}^*$ and $f(\varepsilon)$ the probability density function of the standard normal random variable $\varepsilon_i$. To integrate this likelihood with respect to the distribution of $\varepsilon_i$, we approximate the integral by the Gauss-Hermite formula for numerical integration, i.e.,

$\int_{-\infty}^{\infty} f(z)e^{-z^2}dz \approx \Sigma_{j=1}^{q} w_j f(z_j)$, where $z_j$ are the nodes of the quadrature formula and $w_j$ the associated weights. The integrated log likelihood then becomes

$$LL = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \sum_{j=1}^{q} \pi^{-\frac{1}{2}} \; w_j[y_{it} \log(p_{it}^{j*}) + (1-y_{it})\log(1-p_{it}^{j*})], \qquad (10)$$

where $p_{it}^{j*} = \mu_{it}^{j*} + \Sigma_{\tau=1}^{t-1} \mu_{i\tau}^{j*} \Pi_{s=\tau+1}^{t}(1-\lambda_{is}^{j*} - \mu_{is}^{j*})$, $\text{logit}(\mu_{it}^{j*}) = \mathbf{x}_{it}\beta + \gamma_0 z_j \sqrt{2}$, $\text{logit}(1-\lambda_{it}^{j*}) = \mathbf{x}_{it}\beta^* + \gamma_1 z_j \sqrt{2}$, $w_j$ are the fixed quadrature probabilities, and $z_j$ are the nodes at the mass points *j* of the *q* quadrature. Their values are tabulated in standard tables for specified numbers of quadrature points (e.g., Stroud and Secrest 1966). Our application below uses a 20-point Gaussian quadrature and $\pi^{-0.5}w_j$ and $z_j\sqrt{2}$ as fixed probabilities and mass points, respectively. Note that the model employs a single random error term, $\varepsilon_i$, for both $\mu_{it}^{*}$ and $\lambda_{it}^{*}$. Additional insight into the nature of heterogeneity could be provided by more general models that fit two independent Gaussian random variables or, preferably, a bivariate normal random effect (see Cook and Ng 1997). Also, the model assumes that the unmeasured variables for each individual are constant over time. For example, among the unmeasured (or not accurately measured) factors determining voter preferences, characteristics such as personality traits, political knowledge, and features of the local political system are likely to differ considerably among voters and to remain reasonably stable over time. Nevertheless, controlling for heterogeneity caused by unobserved time-invariant variables may be insufficient in empirical applications. Further, although relatively little is known about individual-specific heterogeneity in Bernoulli models of the kind considered here, our limited Monte Carlo experiments indicate that a large quantity of individual observations is needed to estimate the random effects accurately (see also Heckman 1981).

Our limited experience also supports the notion that ignoring heterogeneity in the current model is unlikely radically to change parameter estimates, but it may lead to underestimation of the standard errors and thus to misleading tests (Morgan 1992, p. 287). Traditional likelihood-ratio testing should not be used to test for the significance of the ancillary variance parameter $\gamma$ because the difference in deviance for a model including the random effect and a (nested) model excluding the random effect (i.e., $-2\Delta LL$) cannot be assumed to have a $\chi^2$ distribution (Collett 1991). The hypothesis tested here is that $\gamma = 0$. Since variances are by definition nonzero, positive quantities, the alternative is one-sided and the distribution of the likelihood-ratio test statistic under the null hypothesis is generally not known. For this situation Snijders and Bosker (1999, pp. 90–91) suggested determining the tail probability of $-2\Delta LL$ for the $\chi^2$ distribution with df equal to the number of additional parameters, and then to halve this tail value to obtain the $p$ value for testing the significance of the random effect. Finally, as noted by Moffitt (1993), uncontrolled heterogeneity in the transitions generates serial correlation in the model and thereby affects the form of the reduced-form expression (2). Hence, the presence of such time-dependent structure complicates matters considerably as $p_{t-1}$ influences $p_t$ in a nonlinear way.

### 4.2.3 Related lines of research

#### *4.2.3.1 Shrinking logical bounds*

The partition Equation (1) implies the familiar restriction, customarily attributed to Duncan and Davis (1953), that $\mu_{it} = p_{it}/(1-p_{it-1}) - p_{it-1}/(1-p_{it-1})\kappa_{it}$, where $\kappa_{it} = 1-\lambda_{it}$. This identity is used by King (1997) in his ecological inference method to construct a so-called tomography plot. The axes of this plot represent the parameters $\kappa_{it}$ and $\mu_{it}$, and the linear constraint on each individual $i$ inherent in Equation (1) is represented by a tomography line with intercept $p_{it}/(1-p_{it-1})$ and slope $-p_{it-1}/(1-p_{it-1})$ that goes through the point $(\kappa_{it}, \mu_{it})$. The lines have a limited range of angles (i.e., all have a negative slope) and they all intersect the 45° line of $\mu_{it} = \kappa_{it}$ at $(p_{it}, p_{it})$. Since the estimated probabilities are guaranteed to lie in the (0, 1) range, we have $\mu_{it} \in (L\mu_{it}, U\mu_{it})$ and $\kappa_{it} \in (L\kappa_{it}, U\kappa_{it})$, where the lower $(L)$ and upper $(U)$ bounds of these intervals are defined by the min and max operators

$$L\mu_{it} = \max\left(0, \frac{p_{it} - p_{it-1}}{1 - p_{it-1}}\right) \le \mu_{it} \le \min\left(\frac{p_{it}}{1-p_{it-1}}, 1\right) = U\mu_{it} \qquad \text{(11a)}$$

and

$$L\kappa_{it} = \max\left(0, \frac{p_{it} - (1 - p_{it-1})}{p_{it-1}}\right) \le \kappa_{it} \le \min\left(\frac{p_{it}}{p_{it-1}}, 1\right) = U\kappa_{it} \qquad \text{(11b)}$$

(see King 1997). Hence the estimated values of $\mu_{it}$ and $\kappa_{it}$ are constrained to lie on that part of the tomography line that intersects the feasible region defined by the logical boundary points. Since the limits are related (e.g., $L\mu_{it} = (p_{it}/1 - p_{it-1}) - (p_{it-1}/1 - p_{it-1})U\kappa_{it}$), the tomography line corresponds to the main diagonal of the rectangular region defined by the lower and upper bounds. Also, because the estimates produced are restricted to lie on the diagonal, they satisfy $\mu_{it} = a_{it} - b_{it}\kappa_{it}$, where $a_{it} = (U\mu_{it}U\kappa_{it} - L\mu_{it}L\kappa_{it})\,(U\kappa_{it} - L\kappa_{it})^{-1}$ and $b_{it} = (U\mu_{it} - L\mu_{it})(U\kappa_{it} - L\kappa_{it})^{-1}$ (see Chambers and Steel 2001).

The estimation procedure considered here implicitly takes into account the bounds and thereby restricts the range of feasible estimates of $\mu_{it}$ and $\kappa_{it}$. This is accomplished simply by constraining the individual probabilities to lie within the admissible range (0, 1). Clearly, explicit assumptions about the relative magnitude of $\mu_{it}$ and $\kappa_{it}$ would allow one to narrow the bounds beyond the logical limits. For example, in studies of U.S. interparty electoral transition it may be assumed, in the spirit of Shively (1991), that the probability that a Democrat at $t-1$ repeats a vote for that party at $t$ is greater than the probability that a non-Democrat at $t-1$ shifts to the Democrats at $t$. This assumption translates into the restriction that $\kappa_{it} > \mu_{it}$ (i.e., $\eta_{it} > 0$). Such a restriction is difficult to justify in general, however, and we would not expect it to be the case for every single voter. Because there is also no algebraic requirement in Equation (1) that $\eta_{it} > 0$, we would not recommend using this assumption universally. Finally, note that if the entry and 1-exit transitions are equal to each other (i.e., $\mu_{it} = \kappa_{it}$), identity (1) reduces to $p_{it} = \mu_{it}$.

#### *4.2.3.2 Ecological panel inference and two-stage auxiliary instrumental variables*

The framework considered here is related to both the ecological panel inference (EPI) method of Penubarti and Schuessler (1998) and the two-stage auxiliary instrumental variables (2SAIV) approach of Franklin

(1989). The EPI method and the one presented here are the same in that both intend to derive micro-level conclusions from repeated cross sections, but they are methodologically quite different in their strategy. The former uses a cross-sectional data set to construct a limited number of demographic profiles, which amounts to grouping the individual data according to the values of the observed covariates and aggregating within the groupings (i.e., summing counts and totals to obtain proportions). If one has available two consecutive cross sections, this aggregate information can be used to obtain the margins of the 2×2 transition table for each profile that — using King's (1997) method of ecological inference — allows one to track changes in the dependent variable of interest. As Penubarti and Schuessler (1998) note, the number of possible combinations of values of the covariates should not be too large relative to the sample size to obtain reasonably reliable aggregates. Hence the method has a problem with sparse data, where sparse means that for every pattern of covariate values we have only a small number of observations. Also note that inferences in EPI are at the level of profiles (based on individuals sharing the same values of the observed covariates) rather than at the level of individuals. The method allows one to trace demographic profiles over time rather than individuals as their profiles might change. In the instrumental variable method presented here actual grouping of the cross-sectional data in observed covariate patterns need not be done. In fact, in the extreme case each individual observation may have its own pattern of covariates. Hence what is special for the current model is that the variation and information in the individual data is fully exploited. Further, while it might be possible to extend the EPI approach to more complex situations involving multiple surveys, the method is likely to face difficulties if the number of cross sections and the number of time-varying covariates become large and if we have important nonbackcastable covariates. Our procedure is also closely related to the intriguing framework presented by Franklin (1989), who proposed a two-stage auxiliary instrumental variables (2SAIV) method of estimating across (panel and other) data sets. It differs, however, in at least three ways. First, while the two-stage instrumental variables method uses auxiliary data to generate predicted values for a right-hand-side variable in the equation of interest in a main data set, the current model is full information in the sense that all subsequent data sets are used in the ML estimation. Second, 2SAIV estimators assume that the (auxiliary and main) data sets derive from the same underlying population. In the current model important events and relevant population changes can in principle be included in the model as additional covariates. Of course, if these events and changes are not in any way related to the variables included, there is no reason to adjust the model. Third, the 2SAIV method as presented by Franklin (1989) assumes that the relationships between the auxiliary measures and the measures of interest are time-invariant. Given a sufficient number of cross sections, the procedure presented here offers the opportunity to verify and, if needed, to relax the assumption of time invariant relationships.

### 4.2.4 *Quantities of interest and potential applications*

The model presented above may be used for different purposes. One is to understand the individual-level relation between covariate effects and transitions in a binary response variable, under Markov assumptions. Another potential goal is to estimate transition probabilities when individual sequence information is not available. The empirical application below illustrates how the model can be used to provide information on individual electoral transitions and the role of voting-related covariates when exact voting sequences are unknown. While our illustration example uses bimonthly data, the model is typically designed to estimate transition probabilities from repeated cross sections covering long-term periods. An example is the analysis of labor force participation decisions of Dutch women over the 1986–1995 period by Pelzer et al. (2001)[6]. Probably the most obvious application in political science is the examination of voter transitions. However, all kinds of political science research problems concerning transitions and involving a binary outcome could benefit from the proposed model, provided that one has available good instruments to predict the unobserved transitions. It may also be noted that not only is the model suitable for examining transitions over historical or calendar time, but also it can be used to study changes in developmental time over age,

[6] 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 consisting of 15 repeated cross sections of the years 1975 to 1993.

i.e., to study life cycle history issues (see Moffitt 1990). Our program *CrossMark* may be used to do the computations.[7]

## 4.3 Application

### 4.3.1 Data

The empirical illustration employs election-year panel data on U.S. presidential vote intention drawn from the campaign study conducted by Patterson (1980) in Erie, PA, and Los Angeles, CA, in 1976. These five-wave bimonthly panel data were also used by Sigelman (1991) in his panel ecological inference study. As indicated above, the purpose of this example is to illustrate the model rather than to provide a definitive analysis of the data. The panel data were treated as if they were a temporal sequence of cross sections of the electorate. That is, no information on the cov($y_t, y_{t-1}$) is available in the data file used for the Markov analysis. The application uses panel data because they provide a check of the ability of the Markov approach to recover known party-switching transitions. Some caution is warranted in interpreting the results, however, as the individual transition probability estimates are based on observations that are not independent. The binary outcome variable $y_{it}$ is defined to equal 1 if the voter $i$ prefers the Democratic party or candidate (i.e., Carter) at time period *t* and 0 otherwise [i.e., Republican party or candidate (Ford) and others].

Table 4.1 provides some summary descriptive statistics. It gives the number of observations including panel inflow and outflow, the marginal distribution of $y_{it}$ over time, and the observed entry and exit transition rates in the panel. The table shows that, despite substantial bimonthly turnover,

[7] The program *CrossMark* is free software and can be freely used and distributed. The main characteristic of the program is the implementation of the Fisher-scoring estimation algorithm. The software is programmed in Delphi but distributed as a compiled version running independently from Delphi or any software on theWindows platform. *CrossMark* does all of the computations reported here including ML estimation, weighting, fixing probabilities, random effect parameter estimation, and (by tricking the program) dynamic panel analysis. The software is available at the *Political Analysis*Web site. Those interested in SPSS Matrix or Gauss versions of the program (with fewer options) should contact the authors.

Table 4.1 Marginal fraction of Democratic vote intention and observed entry and exit transition rates

| year.month | $n_t$ | inflow | outflow | $\overline{y}_t$ | $\overline{y}_t \vert y_{t-1} = 0$ | 1-$\overline{y}_t \vert y_{t-1} = 1$ |
|---|---|---|---|---|---|---|
| 1976.02 | 856 | | | 0.384 | | |
| 04 | 790 | 142 | 208 | 0.460 | 0.248 | 0.178 |
| 06 | 792 | 153 | 151 | 0.471 | 0.170 | 0.176 |
| 08 | 727 | 90 | 155 | 0.465 | 0.203 | 0.229 |
| 10 | 691 | 80 | 116 | 0.457 | 0.140 | 0.138 |

with values ranging from 0.138 to 0.248, almost half of the respondents continue to prefer the Democratic presidential candidate over time. It is important to note that across the five waves of data a substantial number of sample members attrites from the panel. Because some nonrespondents from one wave are recruited back into the sample at subsequent waves, both monotone and nonmonotone participation patterns occur. The current model is special in that it includes all respondents, i.e., both nonattritors and attritors.

The survey also provides information on sociodemographic characteristics and attitudes toward the presidential candidates. The analysis presented here uses only variables that would generally be available in repeated cross-sectional surveys. As backcastable variables, the analysis employs vote choice at the preceding election (i.e., whether the respondent voted for either Nixon or Ford in 1972), race, education, age, and sex. All of these covariates are assumed to be fixed over the survey's duration. In addition to these time-constant variables, the analysis also includes several nonbackcastable covariates. These include (i) whether the respondent identifies him/herself as Democrat or not, (ii) responses to the statements "It doesn't make much difference whether a Republican or a Democrat is elected President" and "All in all, Gerald Ford has done a good job as President," (iii) measures of (un)favorable feelings toward the candidates Ford and Carter, and (iv) opinions about their specific qualities [i.e., very (un)trustworthy, excellent/poor leader, and great deal of/almost no ability]. The responses to the two statements and the candidate images were all registered on 7-point Likert-type scales, running from "strongly disagree" to "strongly agree" and "unfavorable" to "favorable."

Table 4.2 Markov repeated cross section parameter estimates of backcastable variables only for transitions into and out of Democratic vote intention

| | $\delta(p_{t=1})$ | $\beta(\mu_t)$ | $-\beta^*(\lambda_t)$ |
|---|---|---|---|
| Voted Nixon in 1972 | -1.14 (.03) | -1.36 (.04) | |
| Voted McGovern in 1972 | 1.30 (.03) | 1.58 (.11) | -0.56 (.28) |
| Black | 0.96 (.07) | | -2.29 (.38) |
| Education | -0.29 (.00) | -0.23 (.01) | |
| Age | -0.01 (.00) | -0.08 (.00) | -0.10 (.02) |
| Female | | | 0.73 (.21) |
| Constant | 0.82 (.09) | 3.47 (.45) | 2.67 (.65) |
| | | | |
| Number of observations | 3856 | | |
| Log likelihood | -2142.48 | | |

*Note*. Standard errors in parentheses. The $\beta$ parameters represent the effect on $\mu_t$, $\beta^*$ the effect on $(1-\lambda_t)$, and thus $-\beta^*$ the effect on $\lambda_t$.

## 4.3.2 Model estimation

First, a time-stationary Markov model with constant terms only was applied to the data. This model produced the parameters $\beta(\mu_{t>1})$=−0.238 and $\beta^*(\lambda_{t>1})$=0.034 and a corresponding maximum log-likelihood value of $LL = -2643.56$. These estimates imply constant transition rates of $\mu$=0.44 and $\lambda$=0.51, hence implausibly high values that amply exceed the observed rates as reported in Table 4.1. The model was then extended to a nonstationary, heterogeneous Markov model by including the backcastable covariates reported above. The results are shown in Table 4.2. The parameters in the second column show the effects of the backcastable variables on the probability of a Democratic vote at $t=1$ (i.e., $p_{i1}$) estimated for all cases. As can be seen, the parameters are well determined, with a Democratic preference positively affected by being black and a vote for McGovern in 1972 and negatively by education and a vote for Nixon at the prior election. The third column in Table 4.2 presents the effects of the variables on the transitions from non-Democratic (i.e., Republican and others) to Democratic. Whereas a previous vote for McGovern is significant in encouraging entry into a Democratic preference, the entry decisions are negatively affected by education, age, and a 1972 vote for Nixon. The last column gives the effects on the transitions into non-Democratic. We

Table 4.3 Markov repeated cross section estimates for backcastable and nonbackcastable variables*

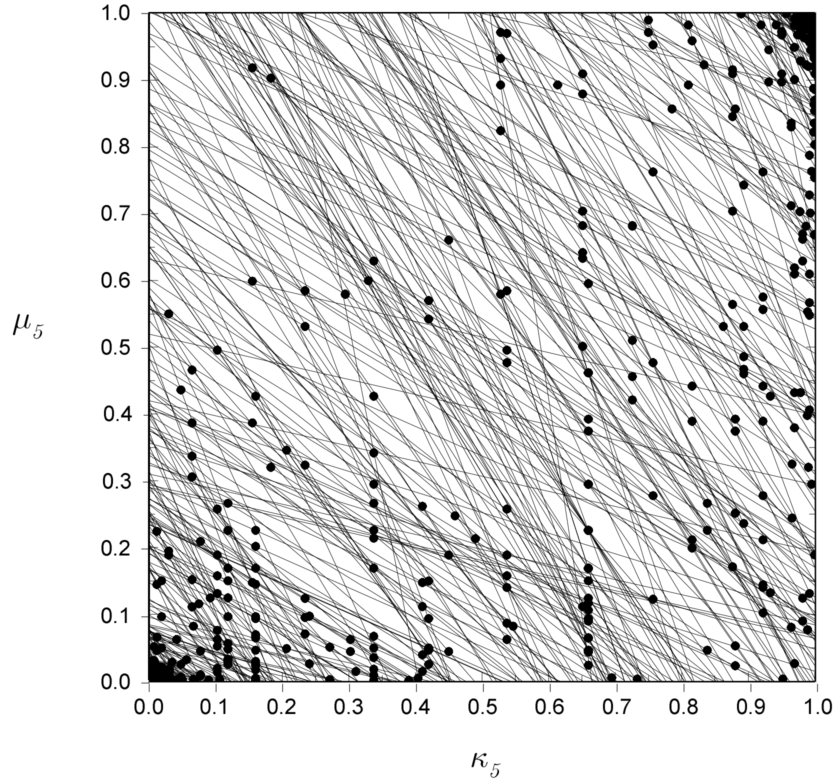
| | $\delta(p_{t=1})$ | $\beta(\mu_t)$ | *time* | $-\beta^*(\lambda_t)$ | *time* |
|---|---|---|---|---|---|
| *Backcastable variables* | | | | | |
| Voted Nixon in 1972 | -0.93 (0.23) | -0.61 (0.35) | 2,4 | 1.47(0.71) | 2,4 |
| Voted McGovern in 1972 | 0.57 (0.21) | 0.96 (0.32) | 2 | | |
| Black | | 1.39 (0.59) | 2 | | |
| Education | | | | 0.71 (0.19) | 2,3,4 |
| Constant | -1.37 (0.18) | -1.09 (0.33) | 2,3,4,5 | -4.58 (1.06) | 2,3,4,5 |
| *Nonbackcastable variables* | | | | | |
| Self-identification | 1.87 (0.19) | 2.59 (0.54) | 2,3 | -3.15 (0.79) | 3 |
| as Democrat | | 1.58 (0.70) | 5 | -2.85 (0.79) | 4 |
| Indifferent toward Democratic or Republican president | -0.19 (0.05) | | | 0.43 (0.12) | 2,3,4 |
| Ford | | | | | |
| good job as president | | -0.39 (0.18) | 4,5 | 0.63 (0.16) | 2,3,4 |
| favorable feelings | -0.28 (0.05) | -0.31 (0.10) | 2 | 0.97 (0.21) | 5 |
| | | -1.34 (0.38) | 4 | | |
| trustworthiness | | -1.12 (0.34) | 5 | 1.40 (0.40) | 4 |
| leadership | | -0.39 (0.13) | 3 | | |
| ability | | | | 1.35 (0.33) | 2,5 |
| Carter | | | | | |
| favorable feelings | | 0.38 (0.11) | 2,3 | -0.69 (0.18) | 3,4 |
| | | 1.23 (0.31) | 4,5 | -1.81 (0.35) | 5 |
| trustworthiness | | 1.36 (0.16) | 4,5 | | |
| leadership | | | | -0.75 (0.31) | 4 |
| ability | | | | -1.24 (0.53) | 2 |
| Constant | | -1.28 (0.56) | 2,3 | 3.19 (0.73) | 3 |
| | | -1.86 (0.82) | 4 | 2.80 (0.84) | 4 |
| | | -1.69 (0.88) | 5 | | |
| $\gamma$ | | 0.71 (0.86) | 2,3,4,5 | 0.12 (1.90) | 2,3,4,5 |
| Number of observations | 3856 | | | | |
| Log likelihood | -1431.04 | | | | |

* Standard errors in parentheses. The columns labeled *time* indicate the discrete time periods pertaining to the parameters.

find that the exit rates are negatively affected by a vote for McGovern in 1972, being black, and age and positively by sex (female).

Table 4.3 reports the regression estimates of a transition model that has all of the variables (including those with unknown history) along with the random effects to account for potential overdispersion. Wald and likelihood-ratio tests revealed no significant difference between the effects of the backcastable variables on the current transitions and their effects on the past transitions. The table therefore presents a single parameter for the backcastable covariates. Further, because there are reasons to believe that the effects of the nonbackcastable covariates may vary over the period leading up to the election, several tests with different time-varying coefficient models of varying degrees of simplicity were applied to the data. The model shown in Table 4.3 best describes the data in terms of goodness of fit. The likelihood-ratio statistic may also be computed to assess the statistical significance of the improvement in fit that results from including the nonbackcastable variables and the random effects. But it is clear from the log-likelihood values reported in Tables 4.2 and 4.3 that the enlarged model provides a much better fit. The second column in Table 4.3 again shows the estimated effects on the state probability $p_{i1}$. Whereas the effects of a 1972 vote for McGovern and identification with the Democrats turn out to be positive, the effects of a vote for Nixon, favorable feelings toward Ford, and indifference toward the future president's leaning are negative. The third and fifth columns provide the effects on the entry and exit rates, respectively, with respect to a Democratic vote. The columns labeled "$Time$" indicate the time periods pertaining to the (time-varying) parameters. For example, favorable feelings toward Carter have an effect on $\mu_t$ of 0.38 at time=2, 3 and an effect of 1.23 at time=4, 5. Most of the parameters are again well determined and consistent with those commonly reported in the literature. In short, a positive attitude toward the Republican (Democratic) candidate Ford (Carter) decreases (increases) the entry rates and increases (decreases) the exit rates. The stronger respondents think of themselves as being Democrat, the higher (lower) their entry (exit) transition rates. The two random effect parameters, $\gamma$, are insignificant. The difference in deviance between the model in question and the model that omits the random effects is $-2\Delta LL = 0.262$, which is obviously not significant even if we were to halve the $p$ value. For the analyses reported

Figure 4.1 Tomography lines (691) for current entry and 1-exit transitions at sample period $t = 5$



below the parameters were therefore estimated anew with the ancillary parameters $\gamma$ restricted to 0.

The tomography lines for one time period are singled out for discussion purposes. Figure 4.1 shows for all $i$ at $t = 5$ the lines $\mu_{i5} = (p_{i5}/1 - p_{i4}) - (p_{i4}/1 - p_{i4})\kappa_{i5}$, where $\kappa_{i5} = 1 - \lambda_{i5}$. The 691 lines all have a negative slope, and they all intersect the 45° line of $\mu_{i5} = \kappa_{i5}$ at $(p_{i5}, p_{i5})$. The permissible range of the parameters for an individual can be obtained by projecting the line onto the horizontal (for $\kappa_{i5}$) and vertical (for $\mu_{i5}$) axes. Note that while most of the point estimates are below the 45° line, for a substantial number of cases $\mu_{i5}$ exceeds $\kappa_{i5}$. In fact, almost 25% of the observations fail to conform to the restriction that $\kappa_{it} > \mu_{it}$. Hence, incorporating the external assumption that party loyalty rates exceed entry rates would most likely lead to incorrect conclusions. Visual inspection of Figure 4.1 also suggests a strong relationship between $\mu_{i5}$ and $\kappa_{i5}$, with low(high) entry rates corresponding with high (low) exit rates. Also note that most of the predictions tend to approach the basically ideal situation of either extremely high or extremely low transition probability estimates. The estimates themselves clearly exhibit a bimodal distribution. Had the instrumental variables been weaker, the two modes would be less well separated or even unimodal.

### 4.3.3 Model validation

It may be of interest to report how the parameter estimates compare to the estimates we would get using a standard dynamic panel estimator. This comparison indicates how much is lost by modeling the panel data as an RCS data set. Most closely related to the RCS transition model is a first-order Markov model for panel data as discussed, for example, by Amemiya (1985), Diggle et al. (1994), and Hamerle and Ronning (1995). Their model uses a separate logistic regression for $P(y_{it} = 1 \mid y_{it-1} = 0,1)$ and can be written logit $P(y_{it} = 1 \mid y_{it-1} = 0,1) = \boldsymbol{x}_{it}\beta + y_{it-1}\boldsymbol{x}_{it}\alpha$, where $\alpha = \beta^{*} - \beta$. This equation thus expresses two regressions as a single dynamic logistic model that includes as predictors both the previous response $y_{it-1}$ and the interaction of $y_{it-1}$ and the covariates $\boldsymbol{x}_{it}$. Because $y_{t-1}$ is missing for some respondents, the estimates of the two models reported in Table 4.4 were obtained from an analysis of the respondents with a valid score on both $y_t$ and $y_{t-1}$. As can be seen, the parameter estimates of the two models are rather similar, except for the constant terms. The signs are all identical and there are no gross discrepancies in magnitude. Also note that, again except for intercepts, the ratio of the parameter estimates to the standard errors is very much alike for the two models, implying that they lead to similar test statistics. Hence the RCS estimators compare rather favorably with the dynamic panel estimators in the sense that a panel analysis of the data would not markedly alter the substantive results.

To understand how well the RCS Markov model reproduces the actual observations in the panel, we may examine its efficacy in various ways. One is to assess the fit of the model in terms of prediction errors, using the mean squared error (MSE). The error measures are given in Table 4.5. The MSE tends to zero if $\mu_{it}(\lambda_{it})$ tends to approach 0 or 1, and the lower the error rate, the better the model predicts. Table 4.5 indicates that the MSEs are remarkably low and that over time they gradually lean to the

Table 4.4 Markov repeated cross section (RCS) and Markov panel parameter estimates*

| | $\beta(\mu_t)$ | | | $-\beta^*(\lambda_t)$ | | |
|---|---|---|---|---|---|---|
| | *RCS* | *panel* | *time* | *RCS* | *panel* | *time* |
| Voted Nixon in 1972 | -0.43(0.40) | -0.67(0.26) | 2,4 | 1.83(0.73) | 0.67 (0.29) | 2,4 |
| Voted McGovern in 1972 | 0.58(0.46) | 0.29(0.35) | 2 | | | |
| Black | 0.77(0.72) | 0.26(0.75) | 2 | | | |
| Education | | | | 0.46(0.15) | 0.04 (0.08) | 2,3,4 |
| Self-identification | 2.94(0.43) | 1.75(0.21) | 2,3 | -2.43(0.71) | -1.78 (0.37) | 3 |
| as Democrat | 1.09(0.59) | 1.05(0.50) | 5 | -2.42(0.75) | -0.75 (0.39) | 4 |
| Indifferent toward Democratic or Republican president | | | | 0.38(0.10) | 0.29 (0.05) | 2,3,4 |
| Ford | | | | | | |
| good job as president | -0.41(0.16) | -0.54(0.13) | 4,5 | 0.43(0.13) | 0.20 (0.07) | 2,3,4 |
| favorable | -0.26(0.10) | -0.20(0.08) | 2 | 2.91(0.88) | 1.17 (0.23) | 5 |
| feelings | -1.32 (0.29) | -1.06(0.22) | 4 | | | |
| trustworthiness | -0.99(0.26) | -0.87(0.21) | 5 | 1.27(0.36) | 0.41 (0.14) | 4 |
| leadership | -0.38(0.12) | -0.31(0.10) | 3 | | | |
| ability | | | | 1.33(0.34) | 0.24 (0.11) | 2,5 |
| Carter | | | | | | |
| favorable | 0.26(0.09) | 0.46(0.08) | 2,3 | -0.75(0.18) | -0.43 (0.09) | 3,4 |
| feelings | 1.02(0.21) | 1.12(0.18) | 4,5 | -3.10(0.86) | -1.02 (0.20) | 5 |
| trustworthiness | 1.13(0.26) | 0.67(0.18) | 4,5 | | | |
| leadership | | | | -0.71(0.33) | -0.48 (0.16) | 4 |
| ability | | | | -1.41(0.33) | -0.46 (0.12) | 2 |
| Constant | -2.45(0.64) | -3.04(0.58) | 2,3,4,5 | -5.92(1.58) | -2.40 (0.68) | 2,3,4,5 |
| | 0.35(0.78) | -0.06(0.67) | 3 | 5.07(1.59) | 2.36 (0.87) | 3 |
| | -1.56(1.63) | 0.18(1.24) | 4 | 2.64(2.31) | 2.58 (1.19) | 4 |
| | -2.77(1.79) | -0.59(1.34) | 5 | | | |
| Number of observations | | 2572 | | | | |
| Log likelihood RCS** | | -885.97 | | | | |
| Log likelihood panel | | -798.66 | | | | |

* Standard errors in parentheses. The columns labeled *time* indicate the discrete time periods pertaining to the parameters.

** The log likelihood of the RCS model is obtaining after excluding the contribution of the 856 observations at $t = 1$ of -381.44.

Table 4.5 Mean squared errors

| | *t* | | | |
|---|---|---|---|---|
| | *2* | *3* | *4* | *5* |
| $\mu: \quad n_t^{-1}\sum_{i=1}^{n_t}(y_{it}^{*} - \mu_{it})^2$ | 0.146 | 0.123 | 0.068 | 0.049 |
| $\lambda: \quad n_t^{-1}\sum_{i=1}^{n_t}(1 - y_{it}^{**} - \lambda_{it})^2$ | 0.155 | 0.121 | 0.126 | 0.069 |

*Note.* $y_{it}^{*} = (y_{it} | y_{it-1} = 0)$, and $y_{it}^{**} = (y_{it} | y_{it-1} = 1)$.

ideal situation of perfect separation between the $y_{it} = 0$ and the $y_{it} = 1$ groups. Also note that the summary measures suggest that the model does somewhat better in terms of predicting entry than it does in predicting exit. Another way to examine the performance of the model is to compare the actual sample frequency of all possible bimonthly (0,1) voting sequences with the estimated expected frequency of each sequence. [8]

Before discussing the findings it is important to note that while the model predicts the current probabilities at time point *t* (i.e., $p_{it}$, $\mu_{it}$ and $\lambda_{it}$) very well, it does not in general reproduce the past probabilities at $t-1$, $t-2$, etc., equally well. The reason is that the past probabilities are predicted by the backcastable variables only, and they are not very good predictors. This obviously hampers the estimation of the expected frequencies.We therefore decided to “backcast” the nonbackcastable variables a single time period (by assuming them to be constant for two consecutive time periods $t-1$ and $t$) and subsequently computed the expected frequencies. Table 4.6 compares the estimated expected and the actually observed absolute frequencies of all 62 (0,1) voting sequences. The longitudinal voting profiles indicate that both the observed and the predicted frequencies are concentrated in the continuous Democratic and

[8] The estimated expected frequencies were computed as follows. With $T$ sample periods, we have $\Sigma_{t=1}^{T} 2^t$ different (0,1) sequences (which in the present application equals 62) ranging in length from 1 (e.g., ‘0’) to $T$ (e.g., ‘11111’) . We define the probability of a sequence of length $t$ for observation $i$ of cross section $t$ as $\tilde{p}_i(\tilde{y}_1,\ldots,\tilde{y}_t) = P(y_{i1} = \tilde{y}_1 \cap \ldots \cap y_{it} = \tilde{y}_t)$, where $\tilde{y}_1,\ldots,\tilde{y}_t = 0{,}1$. Hence $\tilde{p}_i(\tilde{y}_1) = P(y_{i1} = \tilde{y}_1) = \tilde{y}_1 p_{i1} + (1 - \tilde{y}_1)(1 - p_{i1})$, where $p_{i1}$ is $P(y_{i1} = 1)$. For $t > 1$, we have $\tilde{p}_i(\tilde{y}_1,\ldots,\tilde{y}_t) = \tilde{p}_i(\tilde{y}_1)\Pi_{\tau=2}^{t}(p_{00} + p_{01} + p_{10} + p_{11})$, where $p_{00} = (1 - \tilde{y}_{\tau-1})(1 - \tilde{y}_\tau)(1 - \mu_{i\tau})$, $p_{01} = (1 - \tilde{y}_{\tau-1})\tilde{y}_\tau \mu_{i\tau}$, $p_{10} = \tilde{y}_{\tau-1}(1 - \tilde{y}_\tau)\lambda_{i\tau}$, and $p_{11} = \tilde{y}_{\tau-1}\tilde{y}_\tau(1 - \lambda_{i\tau})$. The estimated expected absolute frequency $\tilde{f}(\tilde{y}_1,\ldots,\tilde{y}_t)$ of each participation sequence was obtained by evaluating $\tilde{f}(\tilde{y}_1,\ldots,\tilde{y}_t) = \Sigma_{i=1}^{n_t} \tilde{p}_i(\tilde{y}_1,\ldots,\tilde{y}_t)$.

Table 4.6 Frequencies of observed (*Obs*) and estimated expected (Exp) (non-)Democratic vote intention sequences

| Sequence[*] | *Obs* | Exp | Δ | Sequence[*] | *Obs* | Exp | Δ |
|---|---|---|---|---|---|---|---|
| 0 | *527* | 524 | -3 | 00001 | *7* | 9 | 2 |
| 1 | *329* | 332 | 3 | 00010 | *9* | 3 | -6 |
| 00 | *309* | 296 | -13 | 00011 | *14* | 13 | -1 |
| 01 | *102* | 104 | 2 | 00100 | *9* | 13 | 4 |
| 10 | *46* | 50 | 4 | 00101 | *2* | 2 | 0 |
| 11 | *213* | 219 | 6 | 00110 | *2* | 3 | 1 |
| 000 | *223* | 207 | -16 | 00111 | *11* | 8 | -3 |
| 001 | *37* | 40 | 3 | 01000 | *8* | 10 | 2 |
| 010 | *26* | 20 | -6 | 01001 | *5* | 1 | -4 |
| 011 | *66* | 69 | 3 | 01010 | *3* | 0 | -3 |
| 100 | *25* | 26 | 1 | 01011 | *4* | 2 | -2 |
| 101 | *13* | 14 | 1 | 01100 | *10* | 7 | -3 |
| 110 | *20* | 20 | 0 | 01101 | *4* | 3 | -1 |
| 111 | *160* | 174 | 14 | 01110 | *4* | 5 | 1 |
| 0000 | *160* | 157 | -3 | 01111 | *33* | 29 | -4 |
| 0001 | *30* | 24 | -6 | 10000 | *9* | 18 | 9 |
| 0010 | *12* | 18 | 6 | 10001 | *3* | 2 | -1 |
| 0011 | *14* | 14 | 0 | 10010 | *1* | 1 | 0 |
| 0100 | *13* | 13 | 0 | 10011 | *4* | 1 | -3 |
| 0101 | *10* | 4 | -6 | 10100 | *3* | 5 | 2 |
| 0110 | *15* | 12 | -3 | 10101 | *2* | 1 | -1 |
| 0111 | *43* | 40 | -3 | 10110 | *0* | 2 | 2 |
| 1000 | *12* | 19 | 7 | 10111 | *4* | 2 | -2 |
| 1001 | *5* | 2 | -3 | 11000 | *9* | 6 | -3 |
| 1010 | *5* | 6 | 1 | 11001 | *0* | 1 | 1 |
| 1011 | *4* | 5 | 1 | 11010 | *1* | 0 | -1 |
| 1100 | *12* | 11 | -1 | 11011 | *3* | 3 | 0 |
| 1101 | *4* | 6 | 2 | 11100 | *9* | 7 | -2 |
| 1110 | *23* | 13 | -10 | 11101 | *11* | 3 | -8 |
| 1111 | *114* | 132 | 18 | 11110 | *9* | 12 | 3 |
| 00000 | *140* | 138 | -2 | 11111 | *91* | 114 | 23 |

[*] A binary digit represents a spell occurring over the sample periods $t$, where 1 refers to Democrat and 0 to non-Democrat. The first spell starts at $t = 1$ and the sequences end at the observation period $t$. The frequencies were obtained only for respondents with a valid score on $y_1$ through $y_t$ in the panel.

the continuous non-Democratic vote categories. Hence most voters remain loyal to their initial preference and proportionally few change their vote intention frequently. What is encouraging is the ability of the model to recover sequence membership, even in the presence of recurrent vote switching. Table 4.6 indicates quite clearly that for most sequences the estimated expected frequency predicted by the RCS transition model matches the observed frequency in the panel data well. The only notable exceptions are the highly populated consecutive Democratic vote categories (i.e., the arrays of 1s). However, even for these sequences the model performance is quite good. Hence these findings illustrate that, in this application at least, the model is well able to recover the actual transitions in the panel.

## 4.4 Conclusion

The benefits of repeated cross sections for longitudinal analysis of social and political phenomena have long been understated. Moreover, they are generally regarded as inferior to panel data. It is often thought, for example, that it is inherently impossible to estimate micro-level dynamic models with independent cross sections. As Moffitt (1990, 1993) and others (e.g., Heckman and Robb 1985) have shown, however, this is not correct. Obviously, the estimation of dynamic models with cross-sectional samples is hampered by the lack of information about lagged variables, but these data can nevertheless sometimes be used to identify longitudinal estimators. One important advantage to using panel data is that they provide a measure of gross individual change for each sample unit. However, panel data are often not available and they may also be inferior to the available repeated cross sections in terms of sample size, time period covered, and representativeness.

There has been a considerable expansion in the availability of repeated cross-sectional surveys in the past few decades. 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. The Markov model for cross-level inference presented here can help us estimate binary transitions when it is either impossible or impractical to collect panel information on these events. Our example application shows that the model captures voters with very different entry and exit transition probabilities.

More important, it yields parameters that are fairly consistent with those of a dynamic panel model and it produces transition frequency estimates that are remarkably consistent with the actual observations in the panel. The results thus demonstrate that the proposed model can be used to identify transition probabilities accurately solely on the basis of repeated cross sections and hence to coax panel conclusions out of nonpanel data.

Obviously, generalizing from one particular example is hazardous and there are certainly caveats in applying the model. The prerequisite for adequate application is to have good instruments for the unobserved transitions. In the example reported above the covariates predict the transitions very well but the poor predictions of the past probabilities may serve as a cautionary tale. Uncritical application of the method with weak instrumental variables has the very real danger of leading to incorrect inferences. Hence cautious application and careful data analysis seem warranted.

This warning also implies that the model is not ready for prime-time application. The most prominent subject for future work concerns an examination of the importance of the quality of the instrumental variables by Monte Carlo simulation study. In addition, although the current model promises to be useful in different settings, there are some extensions that we are currently exploring that may further enhance its applicability. One is to use multistate models. Although no essential new theory is involved in such an extension, these models may have too many parameters unless there are some structural constraints imposed on the transitions. A computationally tractable way is to consider three-state models with one absorbing or death state implying that once this state is entered it is never left (Andersen 1980, p. 304). Further, our approach to imposing restrictions on time-varying parameters is to use a fully or partially parametric strategy. In some applications these parametric bases may not provide enough flexibility. It would therefore seem important to study the minimal requirements needed for a varying-coefficient model to yield uniquely identified parameters. We can prove that under relatively mild conditions there always exists exactly one solution for the parameters, but we can verify this only for relatively simple Markov models with constant terms only. Unfortunately, no complete set of identification rules has yet been found guaranteeing unique solutions in more complex models with continuous covariates. It is worthwhile to pursue this thorny problem further.

Another next step is to use Bayesian methods, similar to King et al. (1999) and Rosen et al. (2001), next to ML estimation. A limitation of ML is that it is basically a large-sample inferential approach.With small or moderate-sized data sets, the log likelihood may have a nonnormal shape and asymptotic theory may not work well. It is unknown, however, how large the sample should be for the standard errors based on the information matrix of the current model to yield reliable inferences. One approach to study this small sample problem is to analyze the data by Markov chain Monte Carlo (MCMC) methods. An initial study of this problem is reported by Pelzer and Eisinga (2002).

Finally, it has frequently been argued that King's ecological inference solution can fruitfully be adapted to repeated cross sections (e.g., King et al. 1999; Davies Withers 2001). Despite the steady development in ecological analysis toward more sophisticated statistical modeling, little has been done to date on developing models that draw panel inference from nonpanel data [Franklin (1989), Sigelman (1991), and Penubarti and Schuessler (1998) are notable exceptions]. It is our belief that the approach presented here, when properly enhanced, has the potential to make a significant contribution to political (and other) inquiry.

# References

Achen, Christopher H., and W. Phillips Shively. 1995. *Cross-Level Inference*. Chicago: University of Chicago Press.

Amemiya, Takeshi. 1981. "Qualitative Response Models: A Survey." *Journal of Econometric Literature* 19:1483–1536.

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Oxford: Basil Blackwell.

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

Bartholomew, David J. 1996. *The Statistical Approach to Social Measurement*. San Diego, CA: Academic Press.

Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul W. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Chambers, R. L., and D. G. Steel. 2001. “Simple Methods for Ecological Inference in 2×2 Tables.” *Journal of the Royal Statistical Society. Series A* 164:175–192.

Collett, D. 1991. *Modelling Binary Data*. London: Chapman and Hall.

Cook, Richard J., and Edmund T. M. Ng. 1997. “A Logistic-Bivariate Normal Model for Overdispersed Two-State Markov Processes.” *Biometrics* 53:358–364.

Davies Withers, Suzanne. 2001. “Quantitative Methods: Advancement in Ecological Inference.” *Progress in Human Geography* 25:87–96.

Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Duncan, Otis Dudley, and Beverly Davis. 1953. “An Alternative to Ecological Correlation.” *American Sociological Review* 18:665–666.

Felteau, Claude, Pierre Lefebvre, Philip Merrigan, and Liliane Brouillette. 1997. *Conjugalité et Fécondité des Femmes Canadiennes: Un Modèle Dynamique Estimé à l’aide d’une Série de Coupes Transversales*. Montreal: CREFÉ, Université de Québec à Montréal.

Firth, David. 1982. *Estimation of Voter Transition Matrices from Election Data*.M.Sc. thesis. London: Department of Mathematics, Imperial College London.

Franklin, Charles H. 1989. “Estimation across Data Sets: Two-Stage Auxiliary Instrumental Variables Estimation (2SAIV).” *Political Analysis* 1:1–23.

Goodman, Leo A. 1953. “Ecological Regressions and the Behavior of Individuals.” *American Sociological Review* 18:663–666.

Goodman, Leo A. 1961. “Statistical Methods for the Mover-Stayer Model.” *Journal of the American Statistical Association* 56:841–868.

Hamerle, Alfred, and Gerd Ronning. 1995. Panel Analysis for Qualitative Variables. In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, eds. Gerhard Arminger, Clifford Clogg, and Michael E. Sobel. New York: Plenum Press, pp. 401–451.

Hawkins, D. L., and C. P. Han. 2000. “Estimating Transition Probabilities from Aggregate Samples Plus Partial Transition Data.” *Biometrics* 56:848–854.

Heckman, James J. 1981. Statistical Models for Discrete Panel Data. In *Structural Analysis of Discrete Data with Econometric Applications*, eds. Charles F. Manski and Daniel McFadden. Cambridge MA: MIT Press, pp. 114–178.

Heckman, James J., and Richard Robb, Jr. 1985. “Alternative Methods for Evaluating the Impact of Interventions: An Overview.” *Journal of Econometrics* 30:239–267. Kalbfleish, J. D., and J. F. Lawless. 1984. “Least Squares Estimation of Transition Probabilities from Aggregate Data.” *Canadian Journal of Statistics* 12:169–182.

Kalbfleish, J. D., and J. F. Lawless. 1985. “The Analysis of Panel Data under a Markovian Assumption.” *Journal of the American Statistical Association* 80:863–871.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Cambridge: Cambridge University Press.

King, Gary, Ori Rosen, and Martin Tanner. 1999. “Binomial-Beta Hierarchical Models for Ecological Inference.” *Sociological Methods and Research* 28:61–90.

Lawless, J. F., and D. L. McLeish. 1984. “The Information in Aggregate Data from Markov Chains.” *Biometrika* 71:419–430.

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

McCall, John J. 1971. “AMarkovian Model of Income Dynamics.” *Journal of the American Statistical Association* 66:439–447.

Mebane, Walter R., and Jonathan Wand. 1997. “*Markov Chain Models for Rolling Cross-Section Data: How Campaign Events and Political Awareness Affect Vote Intentions and Partisanship in the United States and Canada*.” Paper presented at the 1997 Annual Meeting of the Midwest Political Science Association, Chicago, IL.

Moffitt, Robert. 1990. “The Effect of the U.S. Welfare System on Marital Status.” *Journal of Public Economics* 41:101–124.

Moffitt, Robert. 1993. “Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections.” *Journal of Econometrics* 59:99–123.

Morgan, B. J. T. 1992. *Analysis of Quantal Response Data*. London: Chapman and Hall.

Patterson, Thomas E. 1980. *The Mass Media Election: How Americans Choose Their President*. New York: Praeger.

Pelzer, Ben, and Rob Eisinga. 2002. “Bayesian Estimation of Transition Probabilities from Repeated Cross Sections.” *Statistica Neerlandica* 56:23–33.

Pelzer, Ben, Rob Eisinga, and Philip H. Franses. 2001. “Estimating Transition Probabilities from a Time Series of Repeated Cross Sections.” *Statistica Neerlandica* 55:248–261.

Penubarti, Mohan, and Alexander A. Schuessler. 1998. *Inferring Micro- from Macrolevel Change: Ecological Panel Inference in Surveys*. Los Angeles: University of California.

Rosen, Ori, Wenxin Jiang, Gary King, and Martin Tanner. 2001. “Bayesian and Frequentist Inference for Ecological Inference: The $R \times C$ Case.” *Statistica Neerlandica* 55:133–155.

Ross, Sheldon M. 1993. *Introduction to Probability Models* (5th ed.). San Diego, CA: Academic Press.

Shively, W. Phillips. 1991. “A General Extension of the Methods of Bounds, with Special Application to Studies of Electoral Transition.” *Historical Methods* 24:81–94.

Sigelman, Lee. 1991. “Turning Cross Sections into a Panel: A Simple Procedure for Ecological Inference.” *Social Science Research* 20:150–170.

Snijders, Tom, and Roel Bosker. 1999. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage.

Stott, David. 1997. *SABRE 3.0: Software for the Analysis of Binary Recurrent Events*. http://www.cas.lancs.ac.uk:80/software/ (Dec. 2001).

Stroud, A. H., and Don Secrest. 1966. *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice–Hall.

Topel, Robert H. 1983. “On Layoffs and Unemployment Insurance.” *American Economic Review* 73:541–559

# 5 Ecological Panel Inference from Repeated Cross Sections

This chapter[1] presents a Markov chain model for the estimation of individual-level binary transitions from a time series of independent repeated cross-sectional (RCS) samples. Although RCS samples lack direct information on individual turnover, it is demonstrated here that it is possible with these data to draw meaningful conclusions on individual state-to-state transitions. We discuss estimation and inference using maximum likelihood, parametric bootstrap, and Markov chain Monte Carlo approaches. The model is illustrated by an application to the rise in ownership of computers in Dutch households since 1986, using a 13-wave annual panel data set[2]. These data encompass more information than we need to estimate the model, and this additional information allows us to assess the validity of the parameter estimates. We examine the determinants of the transitions from have-not to have (and back again) using well-known socioeconomic and demographic covariates of the digital divide. Parametric bootstrap and Bayesian simulation are used to evaluate the accuracy and the precision of the ML estimates, and the results are also compared with those of a first-order dynamic panel model. To mimic genuine repeated cross-sectional data, we additionally analyze samples of independent observations randomly drawn from the panel. Software implementing the model is available.

[1] This chapter has been published as Pelzer, B., R. Eisinga, and P.H. Franses. 2004. Ecological Panel Inference from Repeated Cross Sections. In *Ecological Inference. New Methodological Strategies*, eds, G. King, O. Rosen, and M.A. Tanner. Cambridge MA: Cambridge University Press, pp. 188-205.

[2] The data for the Socio-Economic Panel used in this paper were collected by Statistics Netherlands and were made available by the Scientific Statistical Agency of the Netherlands Organization for Scientific Research. Our program *CrossMark* implements all the simulations and estimations reported here. It is programmed in Delphi but distributed as a standalone program running under Windows. The program (including documentation) is free software and available from the first author (b.pelzer@maw.ru.nl).

## 5.1 Introduction

It has sometimes been argued that King's ecological inference model can be adapted and fruitfully applied to independent repeated cross-sectional (RCS) samples (see, e.g., Penubarti and Schuessler 1998, King, Rosen and Tanner 1999). To date, however, surprisingly little research has been devoted to the development of cross-level inference models that draw panel conclusions from nonpanel data.[3] Moreover, the existing approaches to ecological panel inference are implicitly or explicitly grouping methods, which suffer from small-sample-size restrictions. The individual observations are typically grouped into a limited number of observed covariate patterns, based on time-invariant characteristics (e.g., sex, race). For each covariate pattern, the margins of a transition table are obtained by aggregating within the groupings, and this aggregate information is subsequently used to track changes in the dependent variable of interest. Obviously, such grouping methods are likely to face difficulties (such as sparse-data problems) if the number of covariates and/or the number of repeated cross sections become large.

In this chapter we consider a transition inference model for RCS data with a more dynamic and more flexible structure. In the model proposed here, the micro observations need not be divided into (fixed) groups to obtain sample aggregates. In fact, the variation in the individual covariates is utilized as part of the estimation procedure. The model therefore takes full advantage of the individual survey data and provides full information on the effects of covariates entering the model.

There are several reasons for investigating dynamic models for RCS data. One is the lack of genuine panel data. Panel designs are, rightfully, highly regarded for the opportunity they offer to measure transitions of state or value from repeated observations on the same sample units. For many research issues, however, adequate panel data are rather hard to come by or simply unavailable. Another major reason is that panel data are

[3] Studies that are related to this topic include Franklin (1989), Moffitt (1990 1993), Sigelman (1991), Mebane and Wand (1997), Penubarti and Schuessler (1998). The model presented by Quinn (2004) is also of relevance. The framework discussed here has, in its basic form, been proposed by Moffitt (1990 1993). Pelzer, Eisinga, and Franses (2002) discuss the (dis)similarities between this model and the ecological panel inference (EPI) method of Penubarti and Schuessler (1998) and the two-stage auxiliary instrumental variables (2SAIV) approach of Franklin (1989).

potentially subject to nonsampling biases. An important such bias is sample attrition that results from the progressive loss of (often selective groups of) respondents willing to participate in the data collection. While nonresponse is also a limitation for cross-sectional surveys, it is a more serious problem for panel data because nonresponse often accumulates over time. A related limitation is that it is often difficult to ensure that changes in the target population are reflected in the panel. While panels are typically designed to be representative of the population at the beginning of the study, the panel ages over time, and few panels are, in addition to providing longitudinal data, also designed to permanently provide fully representative information of the population by continuous renewal of the sample.

A large number of cross-sectional surveys conducted by public and private organizations are repeated at regular time intervals. These repeated cross-sectional surveys do not suffer from panel mortality and reflect changes in the universe that cannot be taken into account by a panel study. Estimating individual transitions from such data has an air of performing an impossible task, of obtaining information from nowhere. Indeed, it is often argued that panel data are absolutely needed to study individual-level change (e.g., Kish 1987, p. 167). While individual change is obviously only *visible* in panel data, we will show that this argument is not correct and that data from successive, separately drawn samples can be used to validly estimate transitions using a model that is no more magical than the use of 'plug-in' estimates and bridging assumptions in other areas of statistical modeling.

The outline of this chapter is as follows. Section 5.2 presents a Markov transition model for repeated cross sections designed to deal specifically with binary responses. The model has its origins in the work of Moffitt (1990 1993). We briefly review its main features and discuss maximum likelihood (ML), parametric bootstrap and Markov chain Monte Carlo (MCMC) approaches to estimation and inference. Section 5.3 considers an application of the model to the rise in computer penetration rates in Dutch households from 1986 to 1998, using annual panel data from the Socio-Economic Panel (SEP) survey of Statistics Netherlands. We examine the determinants of the transitions from 'have-not' to 'have' (and back again) using well-known socioeconomic and demographic covariates of the digital divide. Parametric bootstrap and Bayesian simulation are used to evaluate the accuracy and the precision of the RCS Markov ML estimates, and the results are also compared with those of a first-order dynamic panel model. To mimic genuine RCS data, we additionally analyze samples of independent observations randomly drawn from the panel. The summary in Section 5.4 concludes the chapter.

## 5.2 Estimating transitions from RCS data

### 5.2.1 Binary transition model

Obviously, the estimation of dynamic models with repeated cross-sectional data is hampered by the lack of information about lagged variables. Let $y_{it}$ denote the observed response for the binary random variable $y$ of unit $i$ at time period $t$. The crucial characteristic of RCS data is that $y_{it}$ is observed, but $y_{it-1}$ is not. Consequently, no estimate of the serial covariance of successive $y_{it}$ is available in RCS data. This does not imply that dynamic models cannot be estimated with repeated cross sections. However, it does imply that estimation of the unobserved transitions is possible only by putting certain constraints on the transitions for unit $i$ and/or time period $t$.

Consider a 2×2 transition table in which the internal cell values sum to unity across rows. If we define $p_{it} = P(y_{it=1})$, $\mu_{it} = P(y_{it} = 1 \mid y_{it-1} = 0)$, and $\lambda_{it} = P(y_{it} = 0 \mid y_{it-1} = 1)$ then we have the well-known accounting equation

$$E(y_{it}) = p_{it} = \mu_{it}(1 - p_{it-1}) + (1 - \lambda_{it})p_{it-1}. \tag{1}$$

This identity is recognized as the equivalent of Equation 0.4 presented in King, Rosen and Tanner (2004). It is the critical equation that needs to be solved in estimating dynamic models with repeated cross sections, as it relates the marginal probabilities ($p_{it}$ and $p_{it-1}$) to the entry ($\mu_{it}$) and exit ($\lambda_{it}$) transition probabilities. A more concise form for the same equation is $p_{it} = \mu_{it} + \eta_{it}p_{it-1}$, so that $\eta_{it} = 1 - \lambda_{it} - \mu_{it}$. It is also sometimes convenient to define $\kappa_{it} = 1 - \lambda_{it} = P(y_{it} = 1 \mid y_{it-1} = 1)$. If we recursively substitute for $p_{it}$ in Equation (1) and derive its reduced form in terms of past $\mu_{it}$ and $\lambda_{it}$, then we get

$$p_{it} = \mu_{it} + \sum_{\tau=1}^{t-1}\left[\mu_{i\tau}\prod_{s=\tau+1}^{t}\eta_{is}\right] + p_{i0}\prod_{\tau=1}^{t}\eta_{it}\,. \tag{2}$$

This is the model equation that will be used in this chapter. It is obviously not uniquely solvable with RCS data without identifying constraints. Several types of restrictions may be used in this context.

One is to impose some direct restraint on the patterns of the unobserved $\mu_{it}$ and $\lambda_{it}$. For example, the parameters in Equation (2) are clearly identifiable with RCS data if we take the transition probabilities to be homogeneous with respect to both units $i$ and time periods $t$. With the assumption that $\mu_{it} = \mu$ and $\lambda_{it} = \lambda$ for all $i$ and $t$, the long-run value of $p_{it}$ in Equation (2) reduces to $p_{it} = \mu/(\mu + \lambda)$ (see, e.g., Ross 1993, pp. 152-153). Models with this type of homogeneity have been studied extensively in the statistical literature, and they have been applied in various economic, social, and political science studies (see Pelzer, Eisinga and Franses 2002, for additional references).

The model proposed here uses a different type of restriction. This restriction may be imposed if the cross-sectional data include covariates $\mathbf{x}_{it}$ that are measurable in the past (by 'backcasting'), and if the current and lagged $\mathbf{x}_{it}$ affect $\mu_{it}$ and $\lambda_{it}$. In that case, the covariates $\mathbf{x}_{it}, \mathbf{x}_{it-1}, \ldots, \mathbf{x}_{i1}$ can be employed to obtain current and backward predictions of the entry $(\mu_{it}, \mu_{it-1}, \ldots, \mu_{i1})$ and exit $(\lambda_{it}, \lambda_{it-1}, \ldots, \lambda_{i2})$ transition probabilities, by specifying

$$\mu_{it} = F(\mathbf{x}_{it}\beta) \qquad \text{and} \qquad \lambda_{it} = 1 - F(\mathbf{x}_{it}\beta^{*})\,. \tag{3}$$

In these equations $\beta$ and $\beta^{*}$ are two different sets of $k$-dimensional parameters associated with two potentially different sets of (time-invariant or time-varying) $k$-dimensional covariates $\mathbf{x}_{it}$, and $F$ is the — in this paper logistic — link function. Estimates of the model parameters are obtained by substituting Equation (3) into (2).

The critical identifying restriction used here is that the regression parameters are taken to be constant over time, but this constancy assumption may easily be relaxed if we have a sufficient number of repeated cross sections. We may use a semiparametric approach that assumes the parameters to be constant within but different across discrete time periods, or we can model the parameters as a function of time using polynomials or splines. For example, in our empirical illustration below, we introduce time variation into the model by allowing the baseline entry rates (i.e., the constant parameter) to become a first-degree polynomial in time. This is accomplished simply by including the variable time in the model. It is important to note that the underlying Markov chain is not assumed to be homogeneous in the model proposed here, implying that the entry and exit transition probabilities may vary across both units $i$ and time periods $t$. Also note that to obtain $p_{it}$, we actually integrate (sum) over all possible unobserved state-to-state transition paths for each individual unit $i$, starting at $t = 1$ and ending at the cross-sectional observation period $t$. This implies that the probabilities are estimated as a function of all the available cross-sectional samples, rather than simply the observations from the current time period.

Other, perhaps more implicit assumptions underlying the application of the model are that $p_{i0} = 0$, that all the covariates $\mathbf{x}_{it}$ included in the model should have known values in the past, and that the estimation of the entry and exit transitions depend exclusively on variations in the covariates observed. With respect to the first assumption, it should be noted that $p_{i1}$ is the first observed outcome and $p_{i0}$ the value of the state prior to the first outcome. It is generally difficult to incorporate the prior state into the model, and we could invoke the restriction that $p_{i1} = 0$, the consequence of which would be that $p_{i1} = \mu_{i1}$. However, because in many applications the latter assumption is untenable, we prefer to use a separate logistic function for the cross section at $t = 1$, i.e., $P(y_{i1} = 1) = F(\mathbf{x}_{it}\delta)$. The $\delta$-parameters are estimated simultaneously with the entry and exit parameters of interest at $t = 2,...,T$, and they are estimated as a function of all cross-sectional data, rather than simply the observations at $t = 1$.

If some of the covariates are 'nonbackcastable' (i.e., if their past history is unknown), the model may be modified by estimating two different sets of parameters for both $\mu_{it}$ and $\lambda_{it}$: one for the current transition probability estimates and a separate one for the preceding estimates. If we denote the time-dependent covariate with unknown past history by $\mathbf{v}_{it}$ and the associated parameter vector representing the effect on $\mu_{it}$ by $\zeta$, then we have $\text{logit}(\mu_{it}) = \mathbf{x}_{it}\beta^{**} + \mathbf{v}_{it}\zeta$ for cross section $t$, and $\text{logit}(\mu_{it}) = \mathbf{x}_{it}\beta$ for the cross sections $1,\ldots,t-1$. This specification allows one to express the current transition probability estimates as a logistic function of both backcastable and nonbackcastable variables. A similar model may be specified for $\lambda_{it}$. It should be noted here that in our application below we assume that $\beta^{**} = \beta$.

If the assumption that all relevant variables are included in the model is not a realistic one, it may be useful to include an individual-specific random error term $\varepsilon_i$ in the linear predictor of the transition probabilities to account for omitted variables, at least insofar as these variables are time-invariant for each individual. In this logistic-normal mixture model we have $\text{logit}(\mu_{it}) = \mathbf{x}_{it}\beta + \gamma_0\varepsilon_i$ and $\text{logit}(1-\lambda_{it}) = \mathbf{x}_{it}\beta^{*} + \gamma_1\varepsilon_i$, where $\gamma_0$ and $\gamma_1$ are coefficients of the random variable $\varepsilon_i$ having zero mean and unit variance. To estimate the parameters, the (marginal) likelihood of this model may be integrated with respect to the distribution of $\varepsilon_i$ using the Gauss-Hermite quadrature approximation. While likelihood inference about the parameters is possible, it is worth noting that accurate estimation of $\gamma_0$ and $\gamma_1$ from the data themselves is difficult, unless the number of observations is large. As unobserved heterogeneity is not examined in the empirical application below, we will not elaborate on this topic here. Pelzer, Eisinga, and Franses (2002) provide further details.

Finally, it may be useful to outline the commonalities and differences between the ecological analysis of aggregate data and the Markov model for repeated cross-sectional data proposed here. As noted by Sigelman (1991) and Penubarti and Schuessler (1998), drawing panel inferences at the micro-level from repeated cross sections constitutes an ecological inference problem. To demonstrate this point, consider the following partially observed transition table for a population in which there is an absence of both recruitment (immigration or birth) and losses (emigration or death):

| | $Y_t = 0$ | $Y_t = 1$ | |
|---|---|---|---|
| $Y_{t-1} = 0$ | | | $N_{t-1}^0$ |
| $Y_{t-1} = 1$ | | | $N_{t-1}^1$ |
| | $N_t^0$ | $N_t^1$ | $N$ |

In this closed population the marginal distributions are known and fixed, and the ecological inference problem arises because the aggregate measures of change are observed, but the interior cells are not. The two margins provide (at least some) information on the cells, and the accounting identity ensures that the Duncan and Davis (1953) bounds (also termed Fréchet bounds in the statistical literature) will obtain. If we have available a sufficiently large number of transition tables for consecutive time points, an ecological inference model such as presented by Quinn (2004) may be applied to the data.

The situation is somewhat different if the data are drawn from a time series of independent samples of the population of interest. In that case, the marginal values are estimates of the true population parameters and thus themselves subject to error (Cho 1998). And this implies that the bounds too will be known only up to sampling error. If the sample sizes are large, one may be willing to take the margins as fixed and error-free and use the samples to obtain the marginal proportions of the transition table, as presented in the left panel below.

| | $Y_t = 0$ | $Y_t = 1$ | |
|---|---|---|---|
| $Y_{t-1} = 0$ | | | $p_{t-1}^0$ |
| $Y_{t-1} = 1$ | | | $p_{t-1}^1$ |
| | $p_t^0$ | $p_t^1$ | 1 |

| | $Y_t = 0$ | $Y_t = 1$ | |
|---|---|---|---|
| $Y_{t-1} = 0$ | | | |
| $Y_{t-1} = 1$ | | | |
| | $p_{it}^0$ | $p_{it}^1$ | 1 |

If the data are limited to $y_{it}$, we could apply the inference model proposed here, using a Markov model with constant terms only. If we additionally observe covariates, we could also aggregate the micro data into covariate patterns, as in Penubarti and Schuessler (1998), to obtain the marginal distributions of the transition table for each pattern and thus ranges of feasible entries that are consistent with the margins. King's EI could then be used to exploit the information provided by the bounds (using covariate patterns as equivalents to precincts in the analysis of voting). The number of patterns obviously should not be too large relative to the sample size, to obtain reasonably reliable aggregates. Hence the method is likely to suffer from small-sample-size restrictions.

Also note that in using this grouping method, inferences are at the level of individuals sharing the same values of the observed covariates, that is, at the level of the covariate patterns, rather than at the level of individuals. This allows one to trace fixed groups over time rather than individuals, whose covariate values might change. Thus, the method is applicable only if we have a sufficient number of observations for every covariate value and if, in addition, the covariates are time-invariant (so that the sample population can be divided into groups with fixed membership). It faces difficulties in cases of time-varying or nonbackcastable covariates, and these difficulties increase if the number of repeated cross sections becomes large.

The empirical application discussed in Section 5.3 may be used to illustrate the issue. The covariates used in that example include education, age, number of household members, income, and time. The number of covariate patterns observed is 10,510, and the average number of observations per pattern is 2.5. Even if we were to categorize the variable age into three different age categories, as is done in the estimation procedure, the number of covariate patterns would still be large (1,053) and, accordingly, the number of observations per group low (about 25 on average). That is, the group sizes in this example are simply too small for us to ignore the presence of sampling error. And this implies that the data at hand cannot be used to fruitfully compare the performance of our model with the EI grouping method. That is a very interesting and important topic, but one left for future research with other data.

As indicated, what is special for the current model is that the information available in the repeated cross sections is fully exploited. In the model proposed here, there is no grouping of the data, and in the extreme case each individual unit may have its own covariate pattern. This means, as illustrated in the right panel above, that in our procedure only one of the margins ($y_{it}$) is available for inference, and the other one ($y_{it-1}$) is not. And this in turn implies that in our model the repeated cross sections themselves cannot provide any deterministic, informative restrictions on the entries. Consequently, the inference problem in the model proposed here is greater (in the sense of a larger number of unknowns) than in the applications where the margins are (assumed to be) known. The approach proposed here is to completely express the marginal probabilities $p_{it}$ in terms of $\mu_{it}$ and $\kappa_{it}$, recursively, so that estimating the latter automatically renders the former. Also, Equation (1) may be rearranged into $\mu_{it} = p_{it}/(1-p_{it-1})-$

$p_{it-1}/(1-p_{it-1})\kappa_{it}$, where $\kappa_{it}=1-\lambda_{it}$. This expression resembles the equation that King (1997) termed 'tomography line'. Since the estimated marginal probabilities $p_{it}$ and $p_{it-1}$ are guaranteed to lie in the (0,1) range, bounds are enforced on the maximum likelihood estimators of $\mu_{it}$ and $\kappa_{it}$. These upper and lower limits are not informative as in the Duncan and Davis (1953) methods of bounds, however, but rather logical limits implied by the model.

## 5.2.2 Estimation and simulation

### *5.2.2.1 Maximum likelihood estimation*

The method of maximum likelihood may be used to estimate the parameters in Equation (3) ─ plugged into (2) ─ along with their (co)variances. For a sample of $n$ statistically independent observations ─ where each observation is treated as a single draw from a Bernoulli distribution ─ with success probability $p_{it}$, the model (2) has the log likelihood function

$$\ell\ell=\sum_{t=1}^{T}\sum_{i=1}^{n_t}\ell\ell_{it}=\sum_{t=1}^{T}\sum_{i=1}^{n_t}\left[y_{it}\log(p_{it})+(1-y_{it})\log(1-p_{it})\right],$$

where $T$ is the number of cross sections and $n_t$ the number of units of the cross-sectional sample at time period $t$. Maximization of this function has to be performed iteratively and requires the derivatives of the log likelihood with respect to the (vector of) parameters, $\theta$, say. If we suppress subscript $i$ to ease notation, the first order derivatives with respect to $\theta$ are

$$\frac{\partial\ell\ell_t}{\partial\theta}=\frac{y_t-p_t}{p_t(1-p_t)}\cdot\frac{\partial p_t}{\partial\theta},$$

where

$$\frac{\partial p_t}{\partial\theta}=\frac{\partial\mu_t}{\partial\theta}+\frac{\partial p_{t-1}}{\partial\theta}\eta_t+p_{t-1}\frac{\partial\eta_t}{\partial\theta}.$$

If $\theta$ is used to estimate $\mu_t$, then $\partial \mu_t / \partial \theta = \mathbf{x}_t \mu_t (1 - \mu_t)$ and $\partial \eta_t / \partial \theta = -\partial \mu_t / \partial \theta$. If it is used for $\lambda_t$, then $\partial \mu_t / \partial \theta = 0$ and $\partial \eta_t / \partial \theta = \mathbf{x}_t \lambda_t (1 - \lambda_t)$. The values for $\partial p_t / \partial \theta$ can be obtained by recursive substitution, setting $p_0 = 0$ and $\partial p_0 / \partial \theta = 0$, and starting from $\partial p_1 / \partial \theta = \partial \mu_1 / \partial \theta = \mathbf{x}_1 \mu_1 (1 - \mu_1)$. The second derivatives are

$$\frac{\partial^2 \ell\ell_t}{\partial \theta\, \partial \theta'} = -\frac{(y_t - p_t)^2}{p_t^2 (1 - p_t)^2} \cdot \frac{\partial p_t}{\partial \theta} . \frac{\partial p_t}{\partial \theta'} + \frac{y_t - p_t}{p_t (1 - p_t)} . \frac{\partial^2 p_t}{\partial \theta\, \partial \theta'},$$

where

$$\frac{\partial^2 p_t}{\partial \theta\, \partial \theta'} = \frac{\partial^2 p_{t-1}}{\partial \theta\, \partial \theta'} . \eta_t + \frac{\partial p_{t-1}}{\partial \theta'} . \frac{\partial \eta_t}{\partial \theta} + \frac{\partial^2 \mu_t}{\partial \theta\, \partial \theta'} . (1 - p_{t-1}) - \frac{\partial \mu_t}{\partial \theta'} . \frac{\partial p_{t-1}}{\partial \theta},$$

with $\partial^2 \mu_t / \partial \theta\, \partial \theta' = \mathbf{x}_t' \mathbf{x}_t \mu_t\ (1 - \mu_t)\ (1 - 2\mu_t)$. Again, if we set $\partial^2 p_0 / \partial \theta\, \partial \theta' = \partial p_0 / \partial \theta = \partial p_0 / \partial \theta' = 0$, the values for $\partial^2 p_t / \partial \theta\, \partial \theta'$ can be obtained recursively, starting from $\partial^2 p_1 / \partial \theta\, \partial \theta' = \partial^2 \mu_1 / \partial \theta\, \partial \theta'$.

The parameter estimates may be obtained by Newton's method, which uses the Hessian matrix of the actual second derivatives. To speed up computation, we may avoid calculating the exact Hessian by approximating it instead by the expected second derivatives, and use Fisher's method of scoring. Here we will follow the latter approach. In addition to providing parameter estimates, the Fisher optimization algorithm produces as a by-product an estimate of the asymptotic variance-covariance matrix of the model parameters, given by the inverse of the estimated information matrix evaluated at the converged values of the estimates. Each element of the inverse of the information matrix is a minimum variance bound for the corresponding parameter, and the positive square roots of the diagonal elements of this matrix (i.e., the standard errors of the estimated coefficients) may be used for significance tests and to construct confidence intervals.

According to asymptotic theory, ML estimators become progressively more unbiased and more normally distributed, and achieve the minimum possible variance more closely, as the sample size increases (see, e.g., King 1989). However, these asymptotic assumptions may be violated in our complex Markov chain model. Moreover, the estimators in our model have essentially unknown properties for small to moderate sample sizes, and we cannot present any guidelines as to when a sample is sufficiently large for the asymptotic properties to be closely approximated. It is therefore important to investigate the behavior of the estimators of the parameters in Equation (2) by examining their finite-sampling distribution. The bootstrap and MCMC simulations provide useful tools in this situation.

### *5.2.2.2 Parametric bootstrap simulation*

The bootstrap uses Monte Carlo simulation to empirically approximate the probability distribution of the parameter estimates and other statistics, rather than relying on assumptions about its shape, that may only be asymptotically correct. The technique used here is the model-based parametric bootstrap (Davison and Hinkley 1997). For the parametric bootstrap, resamples are taken from the original data via a fitted parametric model to create replicate data sets, from which the variability of the quantities of interest can be assessed. In the repeated simulations, it is assumed that both the form of the deterministic component of the model and the nature of the stochastic component are known. Bootstrap samples are generated using the same fixed covariates as in the original sample and a set of predetermined values for the parameters, allowing only the stochastic component to change randomly from sample to sample. By this means, many bootstrap samples are generated, each of which provides a set of estimates of the parameters that may then be examined for their bias, variance, and other distributional properties and used for bootstrap confidence intervals and hypothesis testing. The parametric bootstrap re-sampling procedure is implemented here according to the following algorithm:

1. Estimate the unknown parameter $\theta$ according to the model (2), using the original sample $\{\mathrm{x}_{it}, y_{it}\}, i = 1,\ldots,n_t, t = 1,\ldots,T$, with the estimate denoted as $\hat{\theta}$, and obtain the fitted values $\hat{p}_{it}$ of the probability that the binary dependent variable $y_{it} = 1$.
2. For each $\mathrm{x}_{it}$ in the original sample $\{\mathrm{x}_{it}, y_{it}\}$, generate a value of the bootstrap dependent variable $y_{it}^{*}$ by random sampling from a Bernoulli distribution with success probability given by $\hat{p}_{it}$.
3. Use the bootstrap sample $\{\mathrm{x}_{it}, y_{it}^{*}\}$ to fit the parameter estimate $\theta^{*}$.

4. Repeat Steps 2 and 3 $R$ times, yielding the bootstrap replications denoted as $\hat{\theta}_1^*,\ldots,\hat{\theta}_R^*$. The empirical distribution of these replications is used to approximate the finite sample distribution of $\hat{\theta}$.

In this study we look at the density of the values of $\hat{\theta}^*$ under resampling of the fitted model to examine bias and variance and to see if it is multi-modal, skewed, or otherwise nonnormal. To obtain an accurate empirical approximation, we use $R$=5,000 replications of the original data set. While the bootstrap estimates of bias and variance under the fitted model are important in their own right, parametric resampling may also be useful in testing problems when standard approximations do not apply or when the accuracy of the approximation is suspect. The key to applying the bootstrap for hypothesis testing is to transform the data so that the null hypothesis is true in the bootstrap population. That is, we simulate data under the null hypothesis so that bootstrap resampling resembles sampling from a population for which the null hypothesis holds (Hall and Wilson 1991). The bootstrap hypothesis test compares the observed value in the original sample with the $R$ values $\hat{\theta}_1^*,\ldots,\hat{\theta}_R^*$, which are obtained from samples independently generated under the null model that satisfies $H_0$. The bootstrap $P$-value may then be obtained by $p^*(\hat{\theta}) = P(\hat{\theta}^* \geq \hat{\theta} \mid H_0) = R^{-1}\sum_{i=1}^{R} I(\theta^* \geq \hat{\theta})$ , where the indicator $I(.)$ equals one if the inequality is satisfied and zero if not (Davison and Hinkley 1997). We reject the null hypothesis if the selected significance level exceeds $p^*(\hat{\theta})$.

### *5.2.2.3 Markov chain Monte Carlo simulation*

Another powerful tool next to MLE and parametric bootstrap is Bayesian simulation, which is easily implemented using Markov chain Monte Carlo (MCMC) methods. Bayesian data analysis is not concerned with finding the parameter values for which the likelihood reaches the global maximum. It is primarily concerned with generating samples from the posterior distribution of the parameters given the data and a prior density, and this distribution may be asymmetric and/or multimodal. Other advantages of the Bayesian approach include the possible incorporation of any available prior information and the ability to make inferences on arbitrary functions of the parameters or predictions concerning specific individual units in the sample (see Pelzer and Eisinga 2002). A popular method for MCMC simulation is Metropolis sampling (Tanner 1996). The Metropolis sampler obtains a chain of draws from the posterior multivariate distribution $\pi(\theta \mid y)$ of the parameter $\theta$. In sampling from the unknown target distribution, the algorithm uses a known auxiliary density $A$— e.g., a (multivariate) uniform or normal distribution — to select candidate parameters $\theta^c$. The Metropolis algorithm proceeds as follows:

1. Choose a starting value for the parameter (e.g., the ML estimates).
2. Randomly draw parameter $\theta^c$ from $A$, a symmetric proposal distribution with mean equal to the previous draw $\theta$ and an arbitrary variance.
3. If $\pi(\theta^c \mid y) \geq \pi(\theta \mid y)$, add the candidate $\theta^c$ to the chain of draws. If $\pi(\theta^c \mid y) < \pi(\theta \mid y)$, calculate the ratio $r = \pi(\theta^c \mid y) / \pi(\theta \mid y)$, and add $\theta^c$ with probability $r$ to the chain of draws.
4. If candidate $\theta^c$ is not added to the accepted draws in Step 3, add $\theta$ so that two successive elements of the chain have the same parameter value $\theta$. Else proceed with the next step.
5. Repeat Steps 2-4 $K$ times, yielding a sample from the posterior distribution of $\theta$.

In the Markov chain sampling used here, we assumed a priori that we are ignorant of the values of the parameters (i.e., have a vague prior belief). This implies that $\pi(\theta \mid y)$ equals the likelihood of $\theta$. Once stationarity has been achieved, a value from a chain of draws from the Metropolis algorithm is supposed to have the same distribution as the target density. We ran the Metropolis algorithm $K$ =100,000 times, excluding an initial burn-in of 10,000 samples, and subsequently obtained the mean, standard deviation, and limits of the 95% credibility interval of $\theta$.

Table 5.1 Proportions of PC ownership in Dutch households over time, 2,028 cases

| Year | $\overline{y}_t$ | $\overline{y}_t \mid y_{t-1} = 0$ | $(1-\overline{y}_t) \mid y_{t-1} = 1$ |
|---|---|---|---|
| 1986 | .12 | | |
| 1987 | .15 | .05 | .10 |
| 1988 | .20 | .08 | .12 |
| 1989 | .24 | .08 | .13 |
| 1990 | .28 | .08 | .08 |
| 1991 | .31 | .09 | .09 |
| 1992 | .36 | .11 | .09 |
| 1993 | .38 | .10 | .13 |
| 1994 | .41 | .10 | .09 |
| 1995 | .44 | .13 | .11 |
| 1996 | .48 | .13 | .07 |
| 1997 | .51 | .14 | .09 |
| 1998 | .57 | .19 | .07 |

## 5.3 Application

### 5.3.1 PC penetration in Dutch households

The major concern of this section is how the RCS Markov model performs in practice. The empirical application is concerned with modeling the rise in computer penetration rates in Dutch households in the 1986─1998 period using data from the Socio-Economic Panel (SEP) collected by Statistics Netherlands. The reason for using this 13-wave annual household panel study is that it offers the opportunity to check the estimation results against the panel findings. However, it is important to note that in the RCS Markov analysis below the panel data are treated as if they were observations of a temporal sequence of 13 independent cross-sectional samples. That is, no use is made of information about lagged values of $y_{it}$.

The binary dependent variable $y_{it}$ is defined to equal one if the household owns a personal computer and zero if not. Table 5.1 reports the proportions of Dutch households with a PC in the 1986─1998 along with the observed entry and exit transition rates. As can be seen, there is a marked upward time trend in PC ownership, from 12% in 1986 to 57% in

1998. While the entry rates (i.e., $\overline{y}_t \mid y_{t-1} = 0$) also show an increase over time, the exit rates (i.e., $1 - \overline{y}_t \mid y_{t-1} = 1$) reveal erratic change.

It is clear from previous studies which structural determinants explain systematic variation in the presence of a PC in homes. The most important covariates – in the Netherlands as elsewhere – are educational attainment, age, the size of the household, and household income (see, e.g., OECD 2001). These variables are included in the SEP household study, but they would generally also be available in a repeated cross-sectional survey. The time-varying variable age of head of household (hereafter *age*) is categorized into three different age categories (18-34, 35-54, and 55+ years). The time-varying variable number of household members is constructed from cross-sectional information about the number and the ages of the children in the household and the presence of a spouse. It is assumed that a family with children has two adults. The variable highest completed education of head of household (hereafter *education*) is taken to be fixed over time. In addition to these backcastable variables, the analysis also includes the temporary, nonbackcastable covariate household income. The variable used here is the standardized (i.e., corrected for size and type of household) disposable household income, categorized into quintiles.

## 5.3.2 RCS Markov model

### *5.3.2.1 Maximum likelihood*

The first model fitted was a time-stationary Markov chain with constant terms only. This model produces the parameters $\beta(\mu_t) =$ -2.543 and $\beta^*(\lambda_t) =$ -3.310 and a log-likelihood value of $LL =$ -15,895.214. These estimates imply constant transition probabilities $\mu = .073$ and $\lambda = .035$, and hence predicted rates that underestimate the observed sample frequencies reported in Table 5.1. The model was subsequently modified to a nonstationary, heterogeneous Markov model by adding the covariates reported above. In analyzing the data with this model, it became apparent that the covariates have a substantial effect on the transition from have-not to have, but that they contribute little to the explanation of the reverse transition. We therefore decided to model the exit transitions using a constant term only. Further, it turned out that the inclusion of a linear time trend in the prediction of obtaining a computer appreciably improves the fit. We therefore included the variable time in the model. This inclusion implies, as indicated in Section 5.2.1, that we drop the assumption of a time-constant intercept and allow the baseline entry rates to increase linearly over time. The results are reported in the second column of Table 5.2.

The top part of the table gives the estimated effects on the marginal probabilities $p_{i1}$. The table indicates that both education and the number of household members positively affect the presence of a PC in homes. While there is no significant difference in PC ownership between the 18-34 year age group and others aged 35-54, ownership is significantly more widespread among the younger age group than among those aged 55 and over. The middle part of Table 5.2 presents the effects on the transition from have-not to have with respect to PC ownership. The results show that educational attainment of head of household, household size, household income, and time have a positive effect on obtaining a computer. This finding confirms the conclusion of cross-sectional studies that computer ownership has spread most rapidly among affluent, well-educated families with children (OECD 2001). The coefficients of the age terms again imply similar entry rates among younger and middle age groups. The older age group has considerably lower access rates. The parameter estimate of the constant term for $\lambda_{it}$ is shown in the bottom part of the table. An intercept of -2.292 implies a time-constant exit transition probability of $\lambda = .092$ (i.e., $\kappa = .908$), which perfectly matches the observed mean frequency of .092.

#### *5.3.2.2 Parametric bootstrap*

As indicated, the benefit of parametric simulation is that the bootstrap estimates give empirical evidence that likelihood theory can be trusted, while providing alternative methods for calculating measures of uncertainty if this theory is unreliable. To examine the sampling distribution of the parameter estimates, we generated $R = 5{,}000$ bootstrap samples according to the algorithm given in Section 5.2.2.2. Table 5.2 provides for each parameter the mean and the sample standard deviation of the bootstrap estimates. In some applications of likelihood methods the variability of

Table 5.2 ML, parametric bootstrap and MCMC estimates of RCS Markov model and ML estimates of first-order panel model, observations 26,364

| | | RCS Markov | | |
|---|---|---|---|---|
| | ML[a] | Bootstrap[b] | MCMC[b] | Panel ML[a] |
| $\delta(p_{t=1})$ | | | | |
| constant | -3.713 (.202) | -3.718 (.205) | -3.754 (.232) | -3.606 (.276) |
| | | [4.137 -3.318] | [-4.225 -3.327] | |
| education | .382 (.054) | .381 (.055) | .393 (.056) | .364 (.072) |
| | | [.271 .489] | [.288 .504] | |
| age 35-54 | -.058 (.119) | -.057 (.121) | -.037 (.120) | .092 (.170) |
| | | [-.294 .181] | [-.284 .197] | |
| age 55 and over | -.852 (.162) | -.859 (.165) | -.842 (.178) | -.782 (.252) |
| | | [-1.201 -.551] | [-1.207 -.513] | |
| no. of household members | .331 (.042) | .332 (.043) | .327 (.038) | .310 (.061) |
| | | [.248 .417] | [.249 .397] | |
| $\beta(\mu_{t=2,\ldots,13})$ | | | | |
| constant | -6.336 (.121) | -6.344 (.124) | -6.339 (.130) | -5.116 (.138) |
| | | [-6.586 -6.110] | [-6.605 -6.105] | |
| education | .368 (.023) | .369 (.023) | .365 (.026) | .245 (.029) |
| | | [.323 .413] | [.310 .414] | |
| age 35-54 | .137 (.049) | .137 (.050) | .129 (.049) | -.098 (.067) |
| | | [.042 .238] | [.037 .224] | |
| age 55 and over | -1.364 (.066) | -1.365 (.065) | -1.362 (.067) | -1.270 (.089) |
| | | [-1.494 -1.240] | [-1.499 -1.226] | |
| no. of household members | .421 (.018) | .422 (.018) | .425 (.020) | .375 (.023) |
| | | [.387 .457] | [.389 .470] | |
| income | .438 (.015) | .438 (.015) | .438 (.016) | .230 (.022) |
| | | [.408 .468] | [.403 .467] | |
| time | .218 (.009) | .218 (.009) | .219 (.010) | .171 (.008) |
| | | [.201 .236] | [.198 .240] | |
| $\beta^*(\lambda_{t=2,\ldots,13})$ | | | | |
| constant | -2.292 (.132) | -2.300 (.133) | -2.307 (.198) | -2.284 (.039) |
| | | [-2.576 -2.058] | [-2.779 -1.938] | |
| $\ell\ell$ | -12,895.106 | | | -7,766.304 |

[a] Standard errors in parentheses.

[b] The mean is reported as the point estimate, the standard deviation in parentheses, and the 95th percentile interval in brackets. The parametric bootstrap results are based on $R$=5,000 bootstrap samples from the original data, and the MCMC findings on $K$=100,000 Metropolis sampler posterior estimates.

Table 5.3 Parametric bootstrap estimates, based on $R = 5{,}000$ bootstrap samples

| | | | | | Excess | Jarque- |
|---|---|---|---|---|---|---|
| | Bias $\times 10^2$ | Bias $\div$ sd | rmse | Skewness | kurtosis | Bera |
| $\delta(p_{t=1})$ | | | | | | |
| constant | -.493 | -.024 | .205 | -.098* | .094 | 9.812* |
| education | -.089 | -.016 | .055 | -.008 | .061 | .796 |
| age 35-54 | .107 | .009 | .121 | .032 | -.026 | 1.008 |
| age 55 and over | -.729 | -.044 | .165 | -.179* | .104 | 28.954* |
| no. of household members | .128 | .030 | .043 | .028 | -.078 | 1.985 |
| $\beta(\mu_{t=2,\ldots,13})$ | | | | | | |
| constant | -.862 | -.070 | .124 | -.033 | -.012 | .931 |
| education | .066 | .029 | .023 | -.050 | -.037 | 2.405 |
| age 35-54 | .040 | .008 | .050 | .070 | -.067 | 5.225 |
| age 55 and over | -.059 | -.009 | .065 | -.052 | .000 | 2.285 |
| no. of household members | .084 | .047 | .018 | .010 | -.025 | .224 |
| income | .065 | .043 | .015 | -.032 | .044 | 1.260 |
| time | .022 | .025 | .009 | .008 | -.104 | 2.338 |
| $\beta^*(\lambda_{t=2,\ldots,13})$ | | | | | | |
| constant | -.789 | -.059 | .133 | -.293* | .296* | 89.691* |

*Note*. The bootstrap estimate of bias $(= \overline{\theta}_{\text{bootstrap}} - \theta_{\text{ML}})$ is multiplied by 100, and $\text{rmse} = (\text{sd}^2 + \text{bias}^2)^{.5}$. The standard errors of skewness and excess kurtosis are .035 and .069, respectively. The Jarque-Bera (1980) test statistic for normality has an asymptotic $\chi_2^2$ distribution; the 5% critical value is 5.991.

* significant at the .05 level.

likelihood quantities may be grossly over- or underestimated. As the table shows, however, the misestimation is small enough to be unimportant here. The bootstrap mean values are close to the ML estimates, and the sample standard deviations are similar to the likelihood-based standard errors. The bootstrap estimates of bias and other distributional properties are given in Table 5.3.

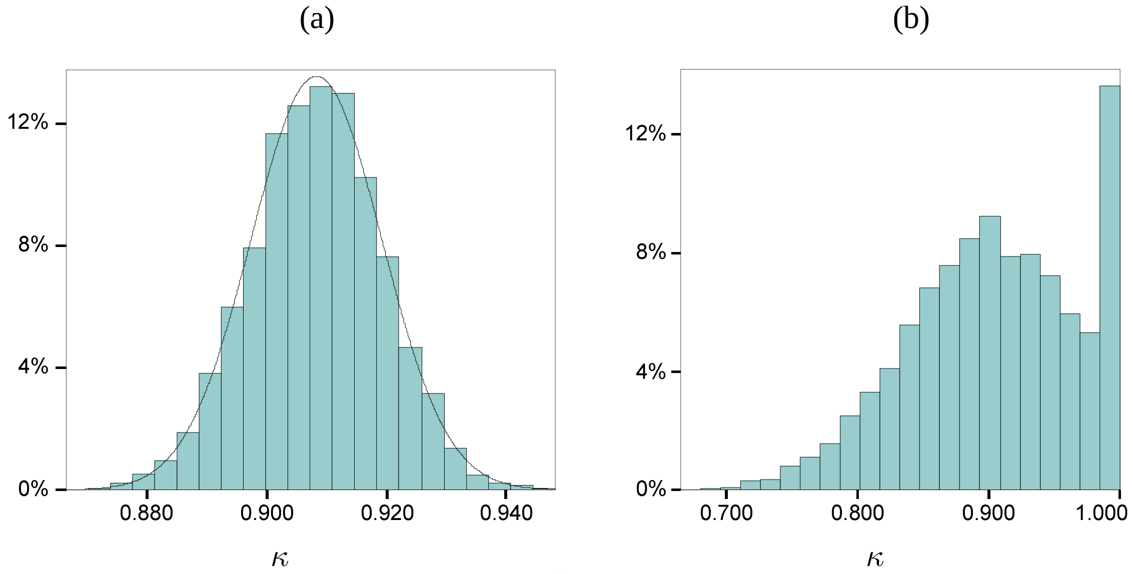
The ML estimates of the model parameters appear to be only slightly biased, the largest absolute bias being .0086. When the estimated bias is expressed as a percentage of the parameter estimate (not reported in Table 5.3), the largest differences between standard theory and the bootstrap results are found for the parameter $\delta(p_{i1})$ of the age 35-54 dummy, for which the percentage bias is 1.85%. All other parameters have percentage biases less than 1%. The parameters also tend to have a small bias compared to the magnitude of their standard deviation. A frequently applied rule of thumb is that a good estimator should be biased by less than 25% of its standard deviation (Efron and Tibshirani 1993). As can be seen in Table 5.3, the ratios of estimated bias to standard deviation are all much smaller than .25. Small values are also found for the root mean square error, which takes into account both standard deviation and bias. The bootstrap sample variance may be compared to the estimated ML variance using a chi-square test to examine whether the sample variance from the bootstrap is significantly larger than the variance from ML (Ratkowsky 1983). For none of the parameters is the bootstrap variance significantly in excess of the ML variance. The largest value was again found for the $\delta(p_{i1})$ parameter of the age 35-54 dummy. The statistic $\chi^2 = (N-1)(\hat{\sigma}^2_{bootstrap} / \hat{\sigma}^2_{ML})$ is distributed as chi-square with 4,999 degrees of freedom (df), a transform of which may be closely approximated by the standard normal distribution, yielding, for this dummy variable, $z = \sqrt{2\chi^2} - \sqrt{2\text{df} - 1} = 1.857$.

Table 5.3 also reports skewness, the excess kurtosis, and the Jarque-Bera (1980) statistic, which may be used to test whether the estimators are normally distributed. The null hypothesis of normality is only rejected for the constant and the age 55+ parameter of $\delta(p_{i1})$, and for the constant term parameter of $\beta^*(\lambda)$. The distribution of the latter is somewhat peaked, and all three estimates have an extended tail to the left. The normal approximation is least accurate for the $\beta^*(\lambda)$ constant. However, even for this estimate the deviation from normality is negligible. The same goes for the distribution of $\kappa$ $[= (1 + \exp(\beta^*(\lambda)))^{-1}]$, shown in Figure 5.1a. The histogram shows no visible departure of the $\kappa$ estimates from those expected for a normally distributed random variable.

### *5.3.2.3 Markov Chain Monte Carlo*

The Metropolis sampler posterior estimates for each parameter are reported in Table 5.2. The findings are based on $K = 100{,}000$ samples, excluding 10,000 samples for initial settling. Inspection of the posterior means reveals that there are no gross discrepancies in magnitude with the ML estimates. The MCMC standard deviations and the ML standard errors are also similar to one another. The same goes for the 95th percentile intervals of

Figure 5.1 Histogram of ML estimates of $\kappa$ (a) for 5,000 bootstrap samples from the original full data, with normal curve superimposed, and (b) for 5,000 cross-sectional samples of 2,028 observations, one observation per household.



the parametric bootstrap estimates and the Bayesian credibility intervals. Thus Bayesian and frequentist methods for obtaining estimates produce roughly similar results.

In sum, according to both parametric bootstrap and MCMC simulations, the maximum likelihood estimators in this application are almost unbiased, with a variance close to the minimum variance bound, and a distribution close to normal. This implies that the ML point estimates of the parameters are accurate and that the inverse of the Fisher information matrix may be used as a good estimate of the covariance matrix of the parameter estimates.

### 5.3.3 Dynamic panel model

It is compelling to compare the RCS Markov ML estimates with the corresponding parameter estimates of a dynamic panel model that allows for first-order dependence. Most directly related to the RCS Markov model is a panel model that specifies a separate logistic regression for $P(y_{it} = 1 \mid y_{it-1} = 0,1)$, and includes $y_{it-1}$ as an additional predictor. This model can conveniently be written in a single equation as logit $P(y_{it} = 1 \,|\, y_{it-1} = 0{,}1) = \mathbf{x}_{it}\beta + y_{it-1}\mathbf{x}_{it}\alpha$, where $\alpha = \beta^{*} - \beta$ (see Amemiya 1985, Diggle, Liang and Zeger 1994, Beck, Epstein, Jackman, and O'Halloran 2001).

The results of applying this logistic model to the binary panel data are shown in the right most columns of Table 5.2. A comparison of the RCS Markov and panel estimates indicates that most of the findings are insensitive to choice of model. The point estimates of all parameters, except perhaps the coefficients for age 35-54 and those for income, are rather similar, and the standard errors also correspond.

Note that the standard errors of the entry parameters are somewhat smaller for the RCS Markov model than for the panel data analysis. This may seem to be counterintuitive, as it would appear to show that more efficient estimates are produced when lagged $y_{it}$-values are unknown than when they are known. It should be noted, however, that the two models differ in the number of observations per parameter. The RCS Markov model uses 24,336 observations (excluding the observations at $t = 1$) to estimate seven $\beta(\mu_t)$ and one $\beta^{*}(\lambda)$ parameter, hence 3,042 observations per parameter. In the panel model we have 16,431 observations to estimate seven $\beta(\mu_t)$ — i.e., 2,347 observations per parameter — and 7,905 observations to estimate $\beta^{*}(\lambda)$. This explains, at least intuitively, the somewhat smaller (larger) standard errors of the entry (exit) parameters in the RCS Markov model. The differences are modest, however, and inferences about the parameters do not change appreciably with the choice of model. Moreover, the two models predict equal transition probabilities $\mu_{it}$ and $\lambda_{it}$ for all individual cases (not reported), and the accuracy of the two models as judged by a ROC curve analysis is almost identical (the area under the ROC curve for the $(y_t \mid y_{t-1} = 0)$ observations is .763 for the RCS Markov model and .768 for the panel model).

Only with respect to the likelihood is the RCS Markov model clearly inferior to the panel model. However, the two models differ in the computation of $p_{it}$ and thus also of the likelihood. In binary panel data, the marginal probability $p_{it}$ is either $\mu_{it}$ or $1 - \lambda_{it}$, conditional on $y_{it-1}$, and the likelihood contribution can be written as $\ell_{it} = \mu_{it}^{y_{it}(1-y_{it-1})}(1-\lambda_{it})^{y_{it}y_{it-1}}$ $(1-\mu_{it})^{(1-y_{it})(1-y_{it-1})}\lambda_{it}^{(1-y_{it})y_{it-1}}$. In the RCS Markov model, however, the marginal probability $p_{it}$ is always a weighted sum of two probabilities — $\mu_{it}$ and $\lambda_{it}$ — weighted by $p_{it}$, and the likelihood is given by $\ell_{it} =$

$[\mu_{it}(1-p_{it-1})+(1-\lambda_{it})p_{it-1}]^{y_{it}}\,[(1-\mu_{it})(1-p_{it-1})+\lambda_{it}p_{it-1}]^{1-y_{it}}$. This implies that even if panel and RCS data produce identical transition probabilities $\mu_{it}$ and $\lambda_{it}$, the two likelihood functions may differ because of $p_{it-1}$. The likelihood values are identical only if $p_{it-1}$ is equal to $y_{it-1}$; that is, if the lagged covariates perfectly predict the previous response.

### 5.3.4 Samples of independent observations

As indicated, in the RCS Markov model the panel data are treated as independent cross sections, implying that there is no information on autocov $(y_{it}, y_{it-1})$ available in the data file used for analysis. Nevertheless, the best way to make sure that the results are not artifacts is to analyze independent observations. To do so, we randomly draw (without replacement) samples of 2,028 different households from the (2,028×13=) 26,364 panel observations, where each sample consists of 13 separate sets — one for each time period — of 156 households. Hence each household is selected only once in the 'cross-sectional' sample. The total number of possible 'cross-sectional' samples in our application is approximately $10^{2,242}$ ($\approx \Pi_{s=0}^{12}$ $\{(2{,}028 - s \times 156)!\ /\ [156!\,(2{,}028 - 156 - s \times 156)!]\}$). We randomly drew 5,000 samples and analyzed each data set separately, using maximum likelihood estimation.

Table 5.4 reports the average values of the parameters across the samples along with the standard deviation divided by $\sqrt{13}$. A comparison of the Tables 5.2 and 5.4 suggests that for almost all parameters the mean values are close to the MLE obtained for the original full sample size. The only noticeable difference is in the constant term parameter of $\beta^{*}(\lambda)$. This mismatch can be explained by referring to the distribution for $\kappa$, shown in Figure 5.1b. For several 'extreme', small samples the true maximum of the likelihood function is attained when $\kappa$ takes the boundary value of $\kappa = 1$. This implies that the true MLE of $\beta^{*}(\lambda)$ is minus infinity and the Fisher optimization algorithm thus fails to converge.

Since the re-sample size is much smaller than the original sample size, it is not surprising that there is a large drop in efficiency relative to the estimates from the original full sample. However, dividing the standard deviations by $\sqrt{26{,}364 / 2{,}028} = \sqrt{13}$ scales them back to the standard errors of the parameters in the original sample. As can be seen, the standard

Table 5.4 Mean and standard deviation $(\div\sqrt{13})$ of the RCS Markov ML estimates for 5,000 samples of 2,028 observations, one for each household

| | $\delta(p_{t=1})$ | $\beta(\mu_{t=2,\dots,13})$ | $\beta^*(\lambda_{t=2,\dots,13})$[a] |
|---|---|---|---|
| constant | -3.845 (.199) | -6.426 (.120) | -2.389 (.260) |
| education | .403 (.046) | .366 (.027) | |
| age 35-54 | -.045 (.118) | .147 (.045) | |
| age 55 and over | -.785 (.160) | -1.423 (.063) | |
| no. of household members | .343 (.032) | .431 (.018) | |
| income | | .447 (.015) | |
| time | | .223 (.010) | |

*Note*. Each sample is drawn without replacement and consists of 13 sets - one for each time period - of size 156. The standard deviation, divided by $\sqrt{13}$, is reported in parentheses.

[a] Excluding 440 samples with $\beta^*(\lambda_t) \leq$ -8 (i.e., $\kappa$ > .9996).

deviations in Table 5.4 agree well with the ML standard errors reported in Table 5.2, the exception again being the constant parameter of $\beta^*(\lambda)$.

### 5.3.5 Parametric bootstrap test

Under parametric bootstrap, hypothesis testing is remarkably easy. We simply need to fit the hypothesized null model, generate bootstrap replications under the assumptions of this model, and calculate the measure we wish to test, both for the real data and for the $R$ sets of bootstrap data. If the value from the real data is among the 5% of the most extreme values in the combined set of $R+1$ values, the hypothesis is rejected at the .05 level of significance. For illustrative purposes, we selected a single sample from the 'cross sections' of size 2,028, with ML estimates close to those reported in Table 5.2. The estimated value for $\kappa$ in this sample was .916. Now consider testing the hypothesis $H_0 : \kappa \geq .999$ against the one-sided alternative $H_1 : \kappa < .999$ ($H_0 : \kappa = 1$ would be a theoretically implausible hypothesis to test for all cases). In $R$=4,999 bootstrap resamples from $H_0$, we found 51 values less then or equal to .916, so the $p^*$-value is 51/5,000 = .0102. This finding leads us to reject the null hypothesis for this particular sample.

## 5.4 Summary

Repeated cross-sectional surveys have become an important data source for research over the past decades. The accumulation of these surveys offers researchers from various disciplines a growing opportunity to analyze longitudinal change. Dynamic models for the analysis of repeated cross sections are, however, relatively rare, and one may even argue that there is an increasing lag between the availability of data and models to analyze them.

The results presented here illustrate the usefulness of exploiting repeated cross-sectional surveys to identify and to estimate 0-1 transition probabilities, which are generally thought to be nonestimable from RCS data. The bootstrap and MCMC findings for the PC ownership example suggest that the maximum likelihood RCS Markov model produces reliable estimates in large samples. It also turns out that, in our empirical application at least, the RCS Markov model performs almost as well as a first-order dynamic panel model. To rule out artificial results, samples of independent observations from the panel data were also analyzed, with similar results to those for the full sample.

This paper has made some necessary first steps in exploring a largely unknown area, and many relevant topics could not be covered here. For example, in some contexts (e.g., the empirical illustration discussed here) it is pretty clear from previous studies or theory which covariates are likely to be important and how they are related, at least qualitatively, to the dependent variable of interest. In other cases, especially in complex data from an unfamiliar field, covariate selection may be far from obvious. An important part of the analysis is then a preliminary analysis to search for a suitable model. This involves not just inspecting the adequacy of the initial model, but doing so in a way that will suggest an improvement of the model and bring to light possibly unsuspected features of the data.

A difficult problem in model specification is that it is not always possible from the data themselves to obtain a clear indication of how to improve the model (and how important it is to do so). It may also happen that different models fit the data roughly equally well and that any choice between them has to be made on grounds external to the data.

Further, it is obvious that estimating the 'nonestimable' is possible only by making assumptions. The validity of the assumptions, however, cannot be assessed from the data under study. Consequently, findings are always conditional on the appropriateness of the assumed model, which in a fundamental sense is not testable. An appropriate statistical framework then is to consider how sensitive the results are to model assumptions. An important subject for future work is therefore to develop sensitivity analysis tools (such as influence diagnostics) and to study the stability of the results under different model specifications and small modifications or perturbations of the data.

Topics to be studied by further Monte Carlo work are the distributional properties of the estimators in different model specifications and the sensitivity of inference procedures to varying sample sizes. In addition to parametric bootstrap, nonparametric resampling could be used to examine the robustness of specification. Nonparametric simulation requires generating artificial data without assuming that the original data have some particular parametric distribution. Finally, although the impetus behind developing the methodology presented here came from the intend to dynamically model RCS data, it would be of interest to apply the model to panel data with missing observations for $y_{t-1}$. The Markov chain model could then be used, in conjunction with a first-order panel model for observations with nonmissing $y_{t-1}$, to obtain model-based imputations for the missing data.

# References

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Oxford: Basil Blackwell.

Beck, Nathaniel, David Epstein, Simon Jackman, and Sharyn O'Halloran. 2001. *Alternative Models of Dynamics in Binary Time-Series—Cross-Section Models: The Example of State Failure.* Paper presented at the 2001 Annual Meeting of the Society for Political Methodology, Atlanta, GA.

Davison, A.C., and D.V. Hinkley. 1997. *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.

Diggle, Peter J., Kung-Yee Liang, and Scott L. Zeger. 1994. *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Duncan, Otis Dudley, and Beverly Davis. 1953. An Alternative to Ecological Correlation. *American Sociological Review* 18:665-666.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Franklin, Charles H. 1989. Estimation across Data Sets: Two-Stage Auxiliary Instrumental Variables Estimation (2SAIV). *Political Analysis* 1:1-23.

Hall, Peter, and Susan R. Wilson. 1991. Two Guidelines for Bootstrap Hypothesis Testing. *Biometrics* 47:757-762.

Jarque, Carlos M., and Anil K. Bera. 1980. Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. *Economics Letters* 6:255-259.

King, Gary. 1989. *Unifying Political Methodology. The Likelihood Theory of Statistical Inference.* Cambridge: Cambridge University Press.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.* Cambridge: Cambridge University Press.

King, Gary, Ori Rosen, and Martin Tanner. 1999. Binomial-beta Hierarchical Models for Ecological Inference. *Sociological Methods and Research* 28:61-90.

King, Gary, Ori Rosen, and Martin Tanner. 2004. *Ecological Inference. New Methodological Strategies.* Cambridge: Cambridge University Press.

Kish, Leslie. 1987. *Statistical Design for Research*. New York: Wiley.

Mebane, Walter R., and Jonathan Wand. 1997. *Markov Chain Models for Rolling Cross-Section Data: How Campaign Events and Political Awareness Affect Vote Intentions and Partisanship in the United States and Canada.* Paper presented at the 1997 Annual Meeting of the Midwest Political Science Association, Chicago Il.

Moffitt, Robert. 1990. The Effect of the U.S. Welfare System on Marital Status. *Journal of Public Economics* 41:101-124.

Moffitt, Robert. 1993. Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-sections. *Journal of Econometrics* 59:99-123.

OECD. 2001. *Understanding the Digital Divide*. Pdf file available at http://www.oecd.org/pdf/M00002000/M00002444.pdf (June 2003).

Pelzer, Ben and Rob Eisinga. 2002. Bayesian Estimation of Transition Probabilities from Repeated Cross Sections. *Statistica Neerlandica* 56:23-33.

Pelzer, Ben, Rob Eisinga, and Philip Hans Franses. 2002. Inferring Transition Probabilities from Repeated Cross Sections. *Political Analysis* 10:113-133.

Penubarti, Mohan, and Alexander A. Schuessler. 1998. *Inferring Micro- from Macrolevel Change: Ecological Panel Inference in Surveys.* Los Angeles CA: University of California.

Quinn, Kevin. 2004. Ecological Inference in the Presence of Temporal Dependence. In *Ecological Inference. New Methodological Strategies*, eds, G. King, O. Rosen, and M.A. Tanner. Cambridge MA: Cambridge University Press, pp. 207-232.

Ratkowsky, David A. 1983. *Nonlinear Regression Modeling: A Unified Practical Approach*. New York: Marcel Dekker.

Ross, Sheldon M. 1993. *Introduction to Probability Models (5th ed.)*. San Diego CA: Academic Press.

Sigelman, Lee. 1991. Turning Cross Sections into a Panel: A Simple Procedure for Ecological Inference. *Social Science Research* 20:150-170.

Tam Cho, Wendy K. 1998. Iff the Assumption Fits ...: A Comment on the King Ecological Inference Solution. *Political Analysis* 7: 143-163.

Tanner, Martin. 1996. *Tools for Statistical Inference*. New York: Springer.

# 6 Bayesian Estimation of Transition Probabilities from Repeated Cross Sections

This chapter[1] discusses some simple practical advantages of Markov chain Monte Carlo (MCMC) methods in estimating entry and exit transition probabilities from repeated independent surveys. Simulated data are used to illustrate the usefulness of MCMC methods when the likelihood function has multiple local maxima. Actual data on the evaluation of an HIV prevention-intervention program among drug users are used to demonstrate the advantage of using prior information to enhance parameter identification. The latter example also demonstrates an important strength of the MCMC approach, namely the ability to make inferences on arbitrary functions of model parameters.

[1] This chapter has been published as Pelzer, B. and R. Eisinga. 2002. "Bayesian Estimation of Transition Probabilities from Repeated Cross Sections," *Statistica Neerlandica,* 56: 23-33.

## 6.1 Introduction

Paap (2002) has shown that MCMC methods are not just a new set of techniques that exploit modern computing technology. Rather, they allow researchers to work with statistical models (and data) previously considered intractable. These include models with dynamics in latent variables, hierarchical, mixture, item-response and nonresponse models and combinations of these model types (see Congdon 2001). While the main advantage is estimation in complex models, Bayesian simulation has also some less sweeping but useful aspects. This short communication is concerned with the problem of estimating binary transition probabilities from independent repeated cross-sectional (RCS) data and aims to demonstrate some practical advantages of Bayesian statistics based on the following issues.

Any model that can be estimated by maximum likelihood can obviously also be estimated by Bayesian simulation. However, when the likelihood function is asymmetric or has multiple local maxima, evaluating the likelihood only around the global maximum, as in ordinary maximum likelihood estimation (MLE), may produce inaccurate information about the distributions of the parameters. Bayesian simulation has an advantage in these circumstances because it is not concerned with finding the parameter values for which the likelihood reaches the global maximum. It is primarily concerned with generating samples from the posterior distribution of the parameters given both the data and a prior density and this distribution may be asymmetric and multimodal. Simulated data will be used to illustrate this. Also, identification may be less of a problem in Bayesian analysis compared with classic approaches such as MLE. While unidentified parameters cannot be estimated in MLE, in the Bayesian approach it is possible to use an ‘informative’ prior that can provide identification. Our example below is concerned with a simple type of Bayesian data combination, in which the posterior determined from a small sized panel data set is used as the prior for a subsequent analysis of repeated cross-sectional data to yield a set of identified parameters. Finally, MCMC offers the opportunity to make inferences on arbitrary functions of model parameters. We will use this ability to derive samples from the posterior distribution of entry and exit transition probabilities in RCS data.

## 6.2 Estimating binary transitions from RCS data

We will first briefly present the model we use to estimate transition probabilities from repeated cross sections. Consider a two-state Markov matrix of transition rates in which the cell probabilities sum to unity across rows. For this $2 \times 2$ table, we define the following three terms, were $Y_{it}$ denotes the value of the binary random variable $Y$ for observation $i$ at time point $t$: $p_{it} = P(Y_{it} = 1)$, $\mu_{it} = P(Y_{it} = 1 \mid Y_{it-1} = 0)$, and $\lambda_{it} = P(Y_{it} = 0 \mid Y_{it-1} = 1)$. These probabilities give rise to the equation $p_{it} = \mu_{it}(1 - p_{it-1}) + (1 - \lambda_{it})p_{it-1} = \mu_{it} + \eta_{it}p_{it-1}$, where $\eta_{it} = 1 - \lambda_{it} - \mu_{it}$. If we let the initial probability $p_{i0} = 0$ (or $t \to \infty$), it is straightforward to show that the reduced form for $p_{it}$ is

$$p_{it} = \mu_{it} + \sum_{\tau=1}^{t-1}\left(\mu_{i\tau}\prod_{s=\tau+1}^{t}\eta_{is}\right).$$

To estimate this equation with repeated independent cross-sectional data, current and backcasted values of time-invariant and time-varying covariates $X_{it}$ (i.e., $X_{it}, X_{it-1}, \ldots, X_{i1}$) are employed to generate backward predictions of the transition probabilities ($\mu_{it}, \mu_{it-1}, \ldots, \mu_{i1}$ and $\lambda_{it}, \lambda_{it-1}, \ldots, \lambda_{i2}$) and thereby of the marginal probabilities ($p_{it}, p_{it-1}, \cdots, p_{i1}$). The transition probabilities themselves are specified as $\mu_{it} = F(X_{it}\beta)$ and $\lambda_{it} = 1 - F(X_{it}\beta^*)$, where $F$ is the logistic link function and $\beta$ and $\beta^*$ are two potentially different sets of parameters associated with two potentially different sets of covariates $X_{it}$. To incorporate “non-backcastable” variables (i.e., time-dependent covariates for which past histories are unknown) into the model, two different sets of parameters are estimated for both $\mu_{it}$ and $\lambda_{it}$: one for the current transition probability estimates and a separate one for the preceding estimates. If we define $Z_{it}$ as a vector of nonbackcastable variables and $\zeta$ as the associated parameter vector representing the effect on $\mu_{it}$, we can write $\text{logit}(\mu_{it}) = X_{it}\beta^{**} + Z_{it}\zeta$ for cross section $t$, and $\text{logit}(\mu_{it}) = X_{it}\beta$ for the cross sections $1, \ldots, t-1$. In our applications below we assume that $\beta^{**} = \beta$. Also, we define the first observed outcome of the process, $\text{P}(Y_{i1} = 1)$, to equal the state probability $p_{i1}$ (rather then the transition probability $\mu_{i1}$) and assume that the $Y_{i1}$'s are random variables with a probability distribution $\text{Prob}(Y_{i1} = 1) = F(X_{it}\delta)$, where $F$ is the logistic function and $\delta$ a set of parameters to be estimated. ML estimates of $\beta$, $\beta^*$ and $\delta$ can be obtained by maximizing the log likelihood $LL = \Sigma_{t=1}^{T}\Sigma_{i=1}^{n_t}\ell\ell_{it} = \Sigma_{t=1}^{T}\Sigma_{i=1}^{n_t}\ [y_{it}\log(p_{it}) + (1 - y_{it})\log(1 - p_{it})]$ with respect to the parameters, where $n_t$ is the number of observations of cross section $t$ and $T$ is the number of cross sections. Fisher's method-of-scoring may be used for maximum likelihood estimation. If we suppress the subscript $i$ and define $p_0 = 0$, the first order partial derivatives of $\ell\ell$ with respect to the parameters $\beta$ and $\beta^*$ are

$$\frac{\partial\,\ell\ell}{\partial\beta} = \frac{\partial\ell\ell}{\partial p_t}\cdot\frac{\partial p_t}{\partial\beta} = \frac{y_t - p_t}{p_t(1-p_t)}\cdot\left(\frac{\partial p_{t-1}}{\partial\beta}\eta_t + \frac{\partial\mu_t}{\partial\beta}(1 - p_{t-1})\right)$$

and

$$\frac{\partial\ell\ell}{\partial\beta^*} = \frac{\partial\ell\ell}{\partial p_t}\cdot\frac{\partial p_t}{\partial\beta^*} = \frac{y_t - p_t}{p_t(1-p_t)}\cdot\left(\frac{\partial p_{t-1}}{\partial\beta^*}\eta_t - \frac{\partial\lambda_t}{\partial\beta^*}p_{t-1}\right),$$

where $\partial\mu_t / \partial\beta = x_t\mu_t(1 - \mu_t)$ and $\partial\lambda_t / \partial\beta^* = -x_t\lambda_t(1 - \lambda_t)$. Further details about the model are provided by Moffitt (1993) and Pelzer, Eisinga and Franses (2001, 2002). In the examples below, the ML estimates were used as starting values of the Markov chain to reduce the period required for burning-in the sampler.

## 6.3 Multimodal likelihood function and Bayesian simulation

The likelihood function can have multiple local maxima with some distributions and models and assuring oneself that a local maximum is indeed the global maximum can be computationally difficult or intensive. Also, if the likelihood function is not well behaved around its maximum, standard errors produced by MLE can lead to unreliable inferences. Markov chain algorithms for sampling from the posterior offer a more complete picture of the uncertainty in the estimation of the unknown parameters. We will illustrate this with a simulated data set. For this simulation, we generated data for $T = 5$ cross sections with $n_t = 2,500$ observations each, using to the following equations and parameter values:

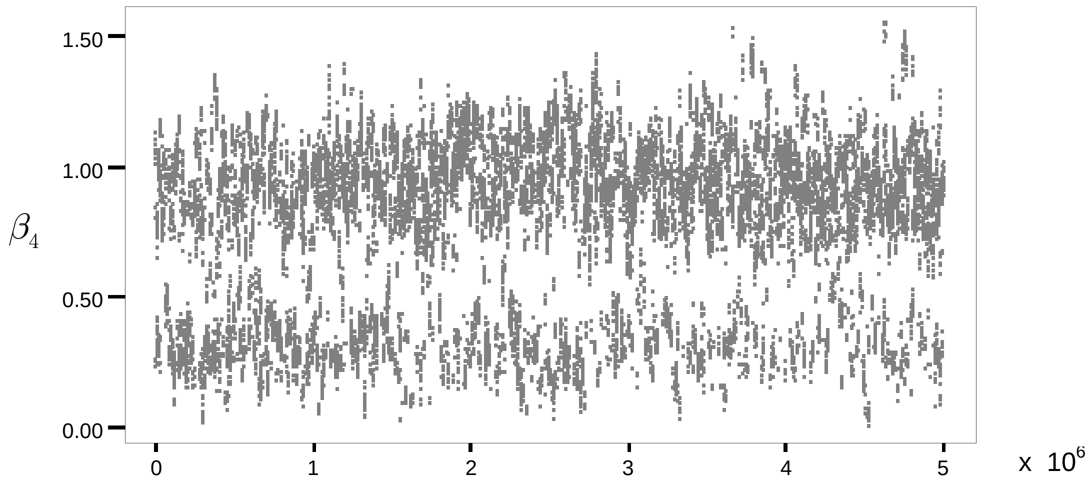
$$\begin{array}{lll} p_{i1} = \mu_{i1} & & \text{for } t = 1 \\ p_{it} = \mu_i(1 - p_{it-1}) + (1 - \lambda_i)p_{it-1} & & \text{for } t = 2,\ldots,5 \\ & & \\ \text{logit}(\mu_{i1}) = \beta_1 + \beta_2 X_i & \beta_1 = -.69 & \beta_2 = .25 \\ \text{logit}(\mu_i) = \beta_3 + \beta_4 X_i & \beta_3 = -1.09 & \beta_4 = .25 \\ \text{logit}(\lambda_i) = -\beta_1^* - \beta_2^* X_i & \beta_1^* = 0 & \beta_2^* = .75\,. \end{array}$$

The $X_i$ values were drawn from the standard normal distribution and subsequently rounded to the nearest integer. The values ranged from -4 to +4, with about 38% of the observations having zero values. Note that the $X_i$ values were fixed over time. Also note that the transition probabilities $\mu_i$ and $\lambda_i$ were taken to be time-constant. The $Y_{it}$ values were sampled from a Bernoulli $(p_{it})$ distribution, $t = 1,\ldots,5$.

The intercept values $\beta_1$, $\beta_3$ and $\beta_1^*$ were selected so that for observations with zero $X_i$ values the marginal probabilities equal $p_{i1} = p_{it} = \mu_i / (\mu_i + \lambda_i)$, which is .334. This steady state condition is reflected in the marginal distribution of the simulated $Y_{it}$, the proportions of $Y_{it} = 1$ being .34, .35, .36, .37, and .34 for the respective cross sections. In a steady state condition, different sets of parameter estimates for $\beta$ and $\beta^*$ may yield an (almost) identical maximum likelihood, especially if the covariate $X_i$ has weak effects. If $X_i$ has no effect at all, we may as well remove it from the model. However, for a model with intercept parameters only, infinitely many estimates satisfy $p_{i1} = p_{it} = \mu_i / (\mu_i + \lambda_i)$ and thus produce an identical maximum likelihood. The covariate $X_i$ reduces the infinitely many ML estimates to a single one, but there still may be many sets of point estimates that yield nearly similar maximum likelihood. MCMC techniques, which seek to characterize the posterior distribution of the regression parameters, can be usefully applied here.

The Metropolis algorithm is often used to generate samples from the posteriors (Tanner 1996). As is well known, this scheme is potentially inefficient when confronted with posteriors with multiple peaks, especially if they are well separated. Multimodal target distributions (especially if the starting values trap us near one of the modes) lead to a poorly mixing chain that stays in small regions of the parameter space for long periods of time.

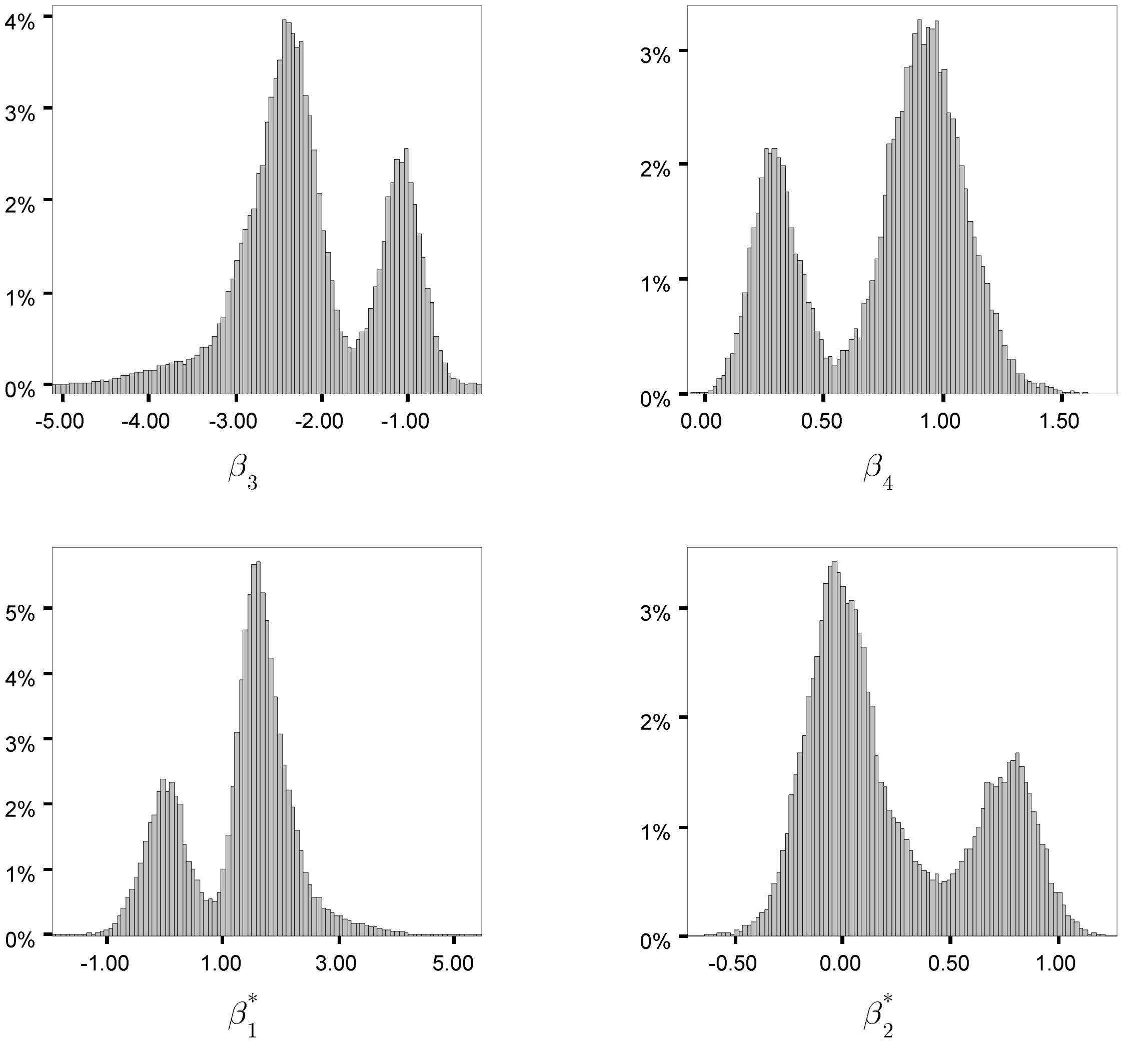
Figure 6.1 Trace of $\beta_4$ (for visual clarity, every 100th sample from the chain is displayed)



The result is that a very large number of random draws is needed to locate the modes. An algorithm that is more efficient in these circumstances is parallel tempering (Liu 2001). Parallel tempering uses a number of chains to traverse the full parameter space, each chain being updated $M$ times by the Metropolis algorithm. After $M$ updates, a swap of the states of two randomly chosen adjacent chains is proposed and this swap is accepted with a particular probability. This swapping mechanism enables parallel chains to explore the entire parameter space, jumping over "narrow" bridges with low likelihood, from one modal area to an other one. To promote a visit of all the modes, the Markov chains can be "heated" to different "temperatures": a "hot" chain is, when going through the $M$ Metropolis updates, more willing to accept parameter proposals with a low likelihood than a "cold" chain. Heating chains is especially effective when the modes are well separated. If the modes are close, heating the chains may be superfluous.

We started the analysis with ML estimation, using Fisher-scoring and the true parameter values as starting values. The resulting parameter estimates were $\hat{\beta}_1 = -.70$, $\hat{\beta}_2 = .27$, $\hat{\beta}_3 = -1.04$, $\hat{\beta}_4 = .28$, $\hat{\beta}_1^* = -.05$, $\hat{\beta}_2^* = .77$, and the corresponding log likelihood was –7729.72. These estimates are close to the true values used in simulating the data. We subsequently performed a number of MCMC analyses using different heating schemes. The trace plots of $\beta_3$, $\beta_4$, $\beta_1^*$ and $\beta_2^*$ all showed two

Figure 6.2 Posterior distributions of $\beta_3$, $\beta_4$, $\beta_1^*$ and $\beta_2^*$



sample bands, indicating that the posterior distributions are bimodal. However, the unheated chain appeared to mix poorly. The "hotter" the heated chains, the poorer the mixing of the unheated chain. We therefore decided to use multiple unheated chains in our simulation. The final analysis employed 20 unheated chains with uninformative priors. Five million samples were run, discarding 50,000 samples for initial settling. Figure 6.1 plots $\beta_4$ against sample iteration number (the other parameters are not displayed as their traces are very similar).

The figure displays two well separated sample bands. The upper mode is located near a value of approximately .90 and the lower mode is close to .25, i.e., the true value of $\beta_4$. The posterior probability distributions of the $\mu_i$ and $\lambda_i$ parameters are shown in Figure 6.2. As can be seen, the distributions are all bimodal. Also note that for all the parameters, the true values are located near the modes with the lowest density. To verify additionally that the likelihood has two different modes, ML estimation was performed using the modes with the highest density as starting values. The resulting parameter estimates were $\hat{\beta}_1 = -.69$, $\hat{\beta}_2 = .27$, $\hat{\beta}_3 = -2.30$, $\hat{\beta}_4 = .92$, $\hat{\beta}_1^* = 1.50$, $\hat{\beta}_2^* = -.02$. These estimates correspond with the high-density modes in Figure 6.2. The log likelihood obtained was -7728.47; hence slightly smaller then the log likelihood of the previous analysis.

These results indicate that models and data with multimodal posteriors may easily cause the unwary ML user to get misleading results. A properly implemented MCMC method will produce the entire parameter distribution and thus reveal asymmetric or multimodal posteriors. In addition, under ML estimation we would compute the mode of the likelihood function and use the local curvature to construct confidence intervals. Consider how odd it would be to use this procedure here. Since standard confidence intervals step on to some fixed distance from the mean and assume a normal parameter density, they completely ignore potentially multimodal or asymmetric features of the distribution. An advantage of Bayesian simulation is that it aims to recover the posterior density without the assumption of normality.

## 6.4 Bayesian data combination

Likelihood-based estimation can be troublesome when the parameters are barely identified or unidentified. In practice, however, additional knowledge may exist about the parameters. This information can, when incorporated in a Bayesian analysis as an informative prior, help to produce uniquely defined estimates. In the example below previously estimated model parameters computed from a different data set are combined with new observations to yield an updated set of identified parameters.

Table 6.1 is based on data presented by Hawkins and Han (2000) taken from an evaluation study of a HIV prevention-intervention program

Table 6.1 Repeated cross section and partial-transition data*

Repeated cross section data (n=1,337)

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *u* = | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | Area = | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | Sex = | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| | Talk = | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |
| time(*t*) | Likely = | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 1 | | 11 | 28 | 12 | 18 | 5 | 29 | 6 | 15 | 7 | 46 | 9 | 29 | 5 | 18 | 4 | 13 |
| 2 | | 14 | 28 | 15 | 32 | 7 | 29 | 11 | 29 | 8 | 54 | 12 | 23 | 6 | 40 | 6 | 19 |
| 3 | | 7 | 31 | 3 | 20 | 2 | 34 | 4 | 9 | 5 | 40 | 6 | 35 | 1 | 31 | 1 | 12 |
| 4 | | 10 | 38 | 6 | 23 | 7 | 35 | 5 | 14 | 2 | 33 | 6 | 24 | 0 | 32 | 7 | 22 |
| 5 | | 9 | 36 | 7 | 22 | 2 | 36 | 4 | 11 | 2 | 34 | 4 | 16 | 3 | 36 | 0 | 22 |

Partial-transition data (n=215)

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| time | *u* = | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 |
| (*t*, *t+1)* | $u^*$ = | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1, 2 | | 1 | 0 | 1 | 1 | 1 | 2 | 0 | 0 | 2 | 1 | 0 | 0 | 0 | 11 | 0 | 0 |
| 2, 3 | | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 10 | 0 | 2 |
| 3, 4 | | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 4 | 0 | 4 |
| 4, 5 | | 0 | 1 | 0 | 0 | 1 | 5 | 2 | 3 | 0 | 1 | 0 | 0 | 0 | 3 | 0 | 1 |
| | *u* = | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 |
| | $u^*$ = | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1, 2 | | 1 | 0 | 0 | 0 | 0 | 2 | 1 | 3 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 1 |
| 2, 3 | | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 1 | 1 | 4 |
| 3, 4 | | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 5 | 0 | 1 | 0 | 0 | 0 | 3 | 0 | 3 |
| 4, 5 | | 0 | 0 | 1 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 2 |
| | *u* = | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 | 5 | 5 | 5 | 5 | 6 | 6 | 6 | 6 |
| | $u^*$ = | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 |
| 1, 2 | | 1 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 6 | 1 | 1 |
| 2, 3 | | 0 | 1 | 0 | 0 | 1 | 6 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 9 | 0 | 0 |
| 3, 4 | | 0 | 1 | 0 | 0 | 0 | 4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 1 |
| 4, 5 | | 0 | 1 | 0 | 0 | 1 | 6 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 1 |
| | *u* = | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 8 | 7 | 7 | 7 | 7 | 8 | 8 | 8 | 8 |
| | $u^*$ = | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 |
| 1, 2 | | 0 | 0 | 1 | 1 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 2, 3 | | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 |
| 3, 4 | | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 4, 5 | | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |

*Source:* Reprinted with permission of the International Biometric Society from Hawkins and Han (2000).

* The index $u$ is used to present the partial-transition data economically.

among drug injectors attempting to modify high-risk behaviors such as sharing unbleached needles to inject drugs. The study consisted of repeated independent surveys conducted at five consecutive time-points in two geographical areas, i.e., an intervention area that underwent various intervention efforts and a comparison area that underwent no intervention. The variable of interest was knowledge of the risk of the transmission of HIV through sharing unclean needles, as measured by responses to the question "how likely is it that you will get AIDS if you share, but don't clean with bleach, drug needles?" The responses of the 1,337 drug users were one of two categories of LIKELY (0=not likely, 1=very likely). Explanatory variables include the time-constant covariates AREA (0=comparison, 1=intervention) and SEX (0=male, 1=female) and the time-varying covariate TALK (0=no, 1=yes). The latter variable records responses to the question "In the last 2 months, has anyone talked to you about AIDS, HIV, or cleaning needles with bleach?". In addition to the independent cross-sectional data, shown in the top part of Table 6.1, the study also collected partial-transition data (for pairs of consecutive waves) from a small sample of 215 drug users. The partial-transition data obtained by haphazard recaptures are shown in the bottom part of Table 6.1.

The repeated cross-sectional data alone can be used to estimate relatively simple transition models such as those with both time-constant intercepts and time-constant covariate effects. But models that drop this assumption are likely to produce problems of overparametrization. That is, it seems impossible to estimate more complex models without the partial-transition data. Several methods can be used to combine the two types of data. The procedure pursued here is based on the idea that the partial-transition data set provides useful auxiliary information about the behavior of the parameters in the repeated cross-sectional context. We therefore first analysed the partial-transition data separately using the Metropolis sampler with a non-informative prior for the regression parameters. The non-informative prior was approximated by a normal distribution with zero mean and variance $10^6$. The means and the variance-covariance matrix of the estimated model parameters were thereupon transferred into the analysis of the repeated cross-sectional data. That is, they were used to construct a multivariate normal prior. Without this prior the problem would be overparameterized and the parameters would be unidentifiable. The regression parameters at $t=1$ were assumed to follow independent normal

Table 6.2 Metropolis sampler posterior estimates*

| | $\delta(p_{t=1})$ | $t$ | $\beta(\mu_t)$ | $t$ | $\beta^*(1-\lambda_t)$ |
|---|---|---|---|---|---|
| Area | .504 (.287)<br>[-.052, 1.071] | 2-5 | .661 (.734)<br>[-.669, 2.195] | 2-5 | .667 (.337)<br>[.025, 1.350] |
| Sex | .186 (.302)<br>[-.390, .798] | 2-5 | 1.430 (.633)<br>[.206, 2.696] | 2-5 | -.082 (.333)<br>[-.777, .555] |
| Talk | .490 (.300)<br>[-.089, 1.085] | 2-5 | -.449 (.693)<br>[-1.929, .811] | 2-5 | 1.133 (.329)<br>[.505, 1.799] |
| Constant | .690 (.272)<br>[.148, 1.222] | 2 | -.686 (.794)<br>[-2.279, .846] | 2 | 1.046 (.425)<br>[.257, 1.919] |
| | | 3-5 | 1.123 (.682)<br>[-.097, 2.571] | 3 | 1.436 (.449)<br>[.621, 2.378] |
| | | | | 4 | .975 (.363)<br>[.299, 1.727] |
| | | | | 5 | 1.190 (.367)<br>[.499, 1.937] |

* The mean of the last 100,000 samples is reported as the point estimate. The standard deviation is reported in parenthesis and the limits of the 95% credibility interval in brackets

distributions with zero mean and variance $10^6$ (i.e., diffuse or non-informative priors). In the Markov chain sampling, we run the Metropolis algorithm 100,000 times excluding an initial burn-in of 5,000 samples. The posterior estimates are shown in Table 6.2.

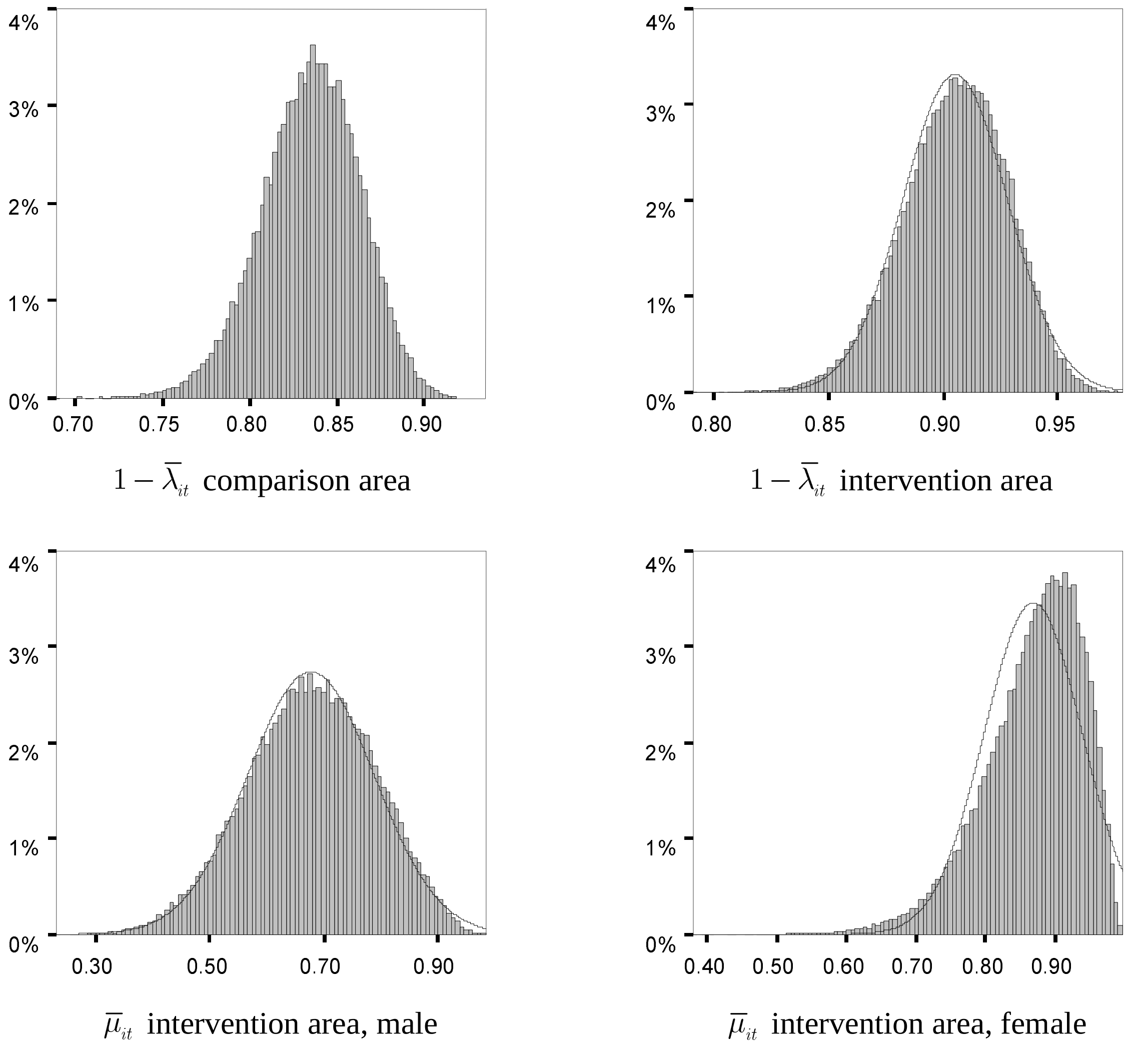
One notes from this table that the entry decisions are affected by SEX. That is, the transition probabilities from the “unlikely” to the “very likely” response are higher among females than they are among males. Both AREA and TALK affect the probability of staying in the “very likely” category (i.e., the $(1-\lambda_t)$ transition). Hence these probabilities are higher in the intervention area and among those reporting TALK=”yes”.

Of course, other methods could be used to analyse these data. One is to simply pool the two data sets and to analyse the combined data using either maximum likelihood or Bayesian analysis with uninformative priors. When the same model is specified under these approaches, estimates from the ML and Bayesian procedures are close (even with these relatively small samples) and they would converge asymptotically. However, sometimes external constraints prohibit explicit data pooling. In many instances previous observations will not be available and even if they were, estimating the whole data might be so time-consuming that shortcut procedures using only the new data, and the estimates of the old as priors, would be appealing. Also, in many research problems data acquisition and data evaluation proceed in stages. Bayesian updating - the transfer of previously estimated model parameters to a new context - can reduce the need for a large data collection in the next stage.

A final issue we would like to address is that MCMC can be employed to obtain inference reaching beyond point estimates and approximate standard errors. A particular strength of the Markov chain Monte Carlo approach is the ability to make inferences on arbitrary functions of model parameters. Moreover, anything we wish to know about this function can be discovered up to any degree of accuracy via random sampling from the density distribution. We may, for example, obtain a sample from the posterior distribution of the mean entry and exit transition probabilities.

The top part of Figure 6.3 shows the mean posterior $(1-\lambda_t)$ transition probabilities for the two study areas. The distributions illustrate the beneficial effect of the intervention program on $(1-\lambda_t)$. The two densities presented in the bottom part of Figure 6.3 display the mean $\mu_t$ for males and females in the experimental area. These figures indicate considerable gender differences and they also show that the distribution for females is asymmetric.

Figure 6.3 Posterior distributions of average transition probabilities in comparison and intervention area (normal curve superimposed)



$1-\overline{\lambda}_{it}$ comparison area

$1-\overline{\lambda}_{it}$ intervention area

$\overline{\mu}_{it}$ intervention area, male

$\overline{\mu}_{it}$ intervention area, female

## 6.5 Conclusion

We have presented two simple examples to illustrate the strengths of modern tools for Bayesian simulation. A straightforward advantage of the MCMC approach is that it provides estimates when traditional maximum likelihood struggles. Bayesian simulation recovers the posterior precisely, without any need to rely on assumptions about the shape of the likelihood function. This feature may help one to arrive at a deeper understanding of the problem of interest.

# References

Congdon, P. (2001), *Bayesian statistical modelling*, Wiley, Chichester.

Hawkins, D.L., and C.P. Han (2000), Estimating transition probabilities from aggregate samples plus partial transition data, *Biometrics* **56**, 848-854.

Liu, J.S. (2001). *Monte Carlo strategies in statistical computing*, Springer, New York.

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

Paap, R. (2002). What are the advantages of MCMC based inference in latent variable models? *Statistica Neerlandica* **56**, 2-22.

Pelzer, B., R. Eisinga and Ph.H. Franses (2001). Estimating transition probabilities from a time series of independent cross sections, *Statistica Neerlandica* **55**, 249-262.

Pelzer, B., R. Eisinga and Ph.H. Franses (2001). *Inferring transition probabilities from repeated cross sections: A cross-level inference approach to US presidential voting.* Econometric Institute Report 2001-21, Erasmus University Rotterdam.

Tanner, M. (1996), *Tools for statistical inference*, Springer, New York.

# 7 Summary and Preview

The foregoing chapters considered a finite state, discrete time, first order Markov model, particularly designed for the analysis of individual-level cross-sectional data observed at a number of consecutive time points. The basic model is due to Moffitt (1990, 1993), who showed that dynamic models do not necessarily need dynamic data which record individual transitions. Although in cross-sectional data, each individual is observed at only one point in time, the proposed Markov model is formulated in terms of individual transition probabilities for all time points involved.

It was shown that the repeated cross sections (RCS) Markov model resembles the dynamic panel model (Amemiya 1985, Diggle, Liang and Zeger 1994) which uses individual panel data to estimate individual transition probabilities over time. The model also resembles the aggregate proportions models (Lee, Judge and Zellner 1970, King, Rosen and Tanner 2004) in that it lacks the actual observations of the quantities of interest, being the numbers of individuals that make each possible transition at each time point. In the introductory Chapter 1 an overview was given of some important developments in the field of statistical Markov models for individual panel and aggregated proportions data.

We described the RCS Markov model structure, the core of which is formed by the first order Markov accounting identity, which for binary $\Upsilon$ is given by $p_{i,t} = p_{i,t-1}(1-\lambda_{i,t}) + (1-p_{i,t-1})\mu_{i,t}$. The identity relates the probability $p_{i,t}$ for case $i$ to be in state 1 of $\Upsilon$ at time point $t$ to the probability $p_{i,t-1}$ to have been in state 1 at $t-1$, the entry probability $\mu_{i,t}$ and the exit probability $\lambda_{i,t}$. It was shown how time-constant and time-varying predictor variables can be used in logit link functions over $p_{i,1}$, $\mu_{i,t}$ and $\lambda_{i,t}$ to yield transition probabilities that can vary between individual cases and within cases over time. The log-likelihood of the parameters of these predictors was also shown, along with it's first and second order derivatives. We briefly explained the Fisher scoring algorithm used to obtain ML estimates of the model parameters.

Several important extensions were added to Moffitt's basic model. These include accounting for time-dependency of the parameters, dealing with unobserved heterogeneity, using nonbackcastable (or limitedly backcastable) predictor variables and using separate parameters for the first cross section. Further, we showed how to deal with the problem of respondents being too young at the time point(s) of the first cross section(s) and how to model inflow of young birth-cohorts in the Markov process.

Next to the more traditional ML approach, the model was examined in a Bayesian framework. We showed how an MCMC procedure like Metropolis sampling offers the possibility to enquire into the (small) sample properties of the parameter estimates of a given model. Alternatively, parametric bootstrapping was used to this end. Further, two different data types, individual panel data and cross-sectional data, were combined in a Bayesian analysis, to yield more well defined posterior parameter distributions.

In each chapter the model was illustrated with an example application. Two of these used repeated cross sections, three used individual panel data and one example used both data types. The panel data examples allowed us to compare the panel transitions and model estimates with the outcomes of the RCS model. In general we found that the RCS model results fitted well to the observed panel transitions and panel model results. Based on the outcomes of all six applications, we conclude that the RCS Markov model can be a useful tool for dynamically modelling longitudinal cross-sectional data. However, more work on this model is needed. Below we mention a number of topics that we will be focusing on in our future work.

### *Identification rules*

We already discussed the relevance of putting constraints on the model parameters. The need for parameter constraints is caused by the fact that in the RCS model, the predicted state 1 probabilities $p_t$ for $t > 1$ are affected twice by the same predictor, i.e., through the entry and exit probability, while there is only a single $\Upsilon$ proportion or value observed. When models become complex it is not always easy to find out whether the constraints employed are sufficient or not. If not, the model is overspecified, resulting in an infinitely large number of equally well fitting solutions for the parameters. To avoid such situations rules are needed on which one can rely to make sure that the parameter restrictions used are strong enough to guarantee a unique ML solution.

A necessary condition for the existence of unique parameter estimates that applies to all kinds of statistical models, is that the number of parameters should not exceed the total number of observational units. For the RCS Markov model discussed here, with separate parameters for the first observed time point, this rule can more precisely be formulated as: the number of parameters used for the initial probability $p_1$ should not exceed the number of predictor patterns (combinations of values of predictors) observed at $t = 1$; the number of parameters used for the entry and exit probabilities together should not exceed the total number of predictor patterns summed over $t = 2...T$, with $T$ being the number of cross sections observed. It should be noted that, for a time-varying predictor, different histories may exist in the data. Each such history is maximally made up of $t$ values, i.e., $t-1$ backcast values and one observed value, with $t$ being the time point at which the case is observed. For such predictor, each history counts as a separate predictor 'value' which is combined in the data with the values of other predictors. However, this rule is merely a very rough 'sine qua non' condition that will often be satisfied when applying the RCS model. Obviously, more detailed rules are required if, for complex models, one wants to be sure that a unique ML solution for the parameters exists.

### *Multimodality*

Even if a unique ML solution for the model parameter exists, it is not always simple to find that particular solution, because the log-likelihood function may be multimodal. Depending on the starting values chosen for the parameters, the Fisher scoring algorithm used to find the maximum value of the log-likelihood function, may get stuck in a local maximum instead of the global one. On the other hand, it may also happen that the global maximum is actually found but it appears to be uninterpretable from a theoretical point of view, while there exists a local maximum that is very good interpretable but not found by the algorithm. Therefore, trying different starting values for the model parameters is a way to avoid ending up 'on the wrong hilltop'.

Instead of Fisher's scoring method, other algorithms may be better suited to deal with a multimodal log-likelihood surface. So-called 'stochastic algorithms' start from the idea that many good solutions may exist, that are close to each other in the parameter space. During the search process, such algorithm can be stopped at any point in time to provide the best solution up that time point. In general, these algorithms are more likely to yield the different modes of a function than the more classical gradient methods like Fisher scoring. Another class of algorithms is formed by the 'evolutionary algorithms'. These typically maintain a population of potential solutions instead of a single one. A brief explanation of the most important stochastic and evolutionary algorithms is given by Omran (2005).

*Information loss*

Due to not observing the actual transitions in the case of repeated cross-sectional data as opposed to individual panel data, one may be inclined to think that the RCS Markov model estimates are less informative than the ones obtained from a model using individual panel data. For the example data of Table 7.1 this is true.

Table 7.1 Data observed at three time points

| | | $\Upsilon_2$ | | | | | $\Upsilon_3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | | | | 0 | 1 | |
| $\Upsilon_1$ | 0 | 36 | 44 | 80 | $\Upsilon_2$ | 0 | 18 | 22 | 40 |
| | 1 | 4 | 16 | 20 | | 1 | 12 | 48 | 60 |
| | | 40 | 60 | | | | 30 | 70 | |

Table 7.1 shows the marginal distributions (80,20), (40,60) and (30,70) of $\Upsilon_1$, $\Upsilon_2$ and $\Upsilon_3$. These would be the only data available in repeated cross sections. In panel data, the inner cell frequencies would also be observed. Table 7.2 shows the results of two models applied to the data in Table 7.1, the RCS Markov and the dynamic panel model, assuming time invariant entry and exit probabilities $\mu$ and $\lambda$, respectively. Both models yield the same point estimates but with different standard errors for the logits.

Table 7.2 RCS and Panel model estimates and standard errors

| model | $\hat{\mu}$ | $\hat{\lambda}$ | logit $\hat{\mu}$ | logit $\hat{\lambda}$ | se(logit $\hat{\mu}$) | se(logit $\hat{\lambda}$) |
|---|---|---|---|---|---|---|
| RCS Markov | .55 | .20 | .201 | -1.386 | .340 | .737 |
| Dynamic panel | .55 | .20 | .201 | -1.386 | .183 | .280 |

The standard errors for the RCS model are largest and hence, there is information loss when using the cross-sectional data (the margins of Table 7.1) instead of the panel data (the inner cell frequencies). However, in other situations there can be information gain, at least for some of the parameters used. We already noted this somewhat counterintuitive result in chapter 5 when discussing the PC ownership example. For these data the entry parameters had smaller standard errors for the RCS model than for the panel data model. The question then arises as to how the information differences between the two models can be explained. Note that, if the only data available are repeated cross sections, one cannot compare the standard errors of the RCS model with those of the panel model and ask where the differences come from. However, what can be asked instead is: "How large would the information loss (or gain) of the RCS model be, compared to the information of the panel model, if there *would have been* panel data for the same time period?".

An important difference between the RCS and the panel model is that, for the RCS model, Fisher's information matrix contains off-diagonal cells related to the correlations between entry and exit parameters. For the panel model, these cells and correlations are zero, since the entry and exit probabilities are estimated independently from one another. In general, for the RCS model it holds that the stronger these entry/exit correlations, the more inflated the variances of the corresponding parameter estimates and the greater the information loss of the RCS model, compared to the panel model. For the data in Table 7.1, the correlation between the two logit estimates using the RCS model is rather strong (-.81); for the PC ownership data of Chapter 5 the highest correlation of the six entry parameters with the single exit parameter was -.37. The precision of each separate entry parameter does not suffer much from such weak correlation. On the other hand, the exit parameter is, albeit weakly, correlated with six entry parameters and this accumulates to less precision for the exit parameter.

An approach to the issue of information loss is offered by Steel, Beh and Chambers (2004). These authors deal with a similar problem: the information loss in an ecological inference study in which only the aggregated marginal frequencies of a set of contingency tables are observed but the individual cell frequencies are not. They show how, in general, Fisher's information for a model that uses aggregated data is related to the expected information for a model using the underlying individual data; here, 'expected' is to be understood as: expected over all possible occurrences of the individual data that would fit to the observed aggregated data. (For example, for the observed aggregated marginal (40, 60) at $t = 2$ in Table 7.1, the expectation is over all possible 2x2 panel tables of $\Upsilon_1$ by $\Upsilon_2$ with cell frequencies such that the column totals are 40 and 60.) By taking the expectation over all individual panel data that could possibly have occurred, given the aggregated cross-sectional data, an average information loss of the RCS model compared to the panel model can be obtained.

*Overdispersion due to unobserved heterogeneity*

Due to unknown or unobserved influences affecting the outcome of $\Upsilon$ there may be more variation in $\Upsilon$ that can possibly be reproduced by the known and observed influences that are incorporated in the model equations as predictors. To deal with such overdispersion in the data, many strategies have been proposed; see for example Collett (1981) for an explanation of the phenomenon and a discussion of some approaches to deal with it. One of the methods shown by Collett was pursued in Chapter 4, where a normal random effect was added to the logit equations of both entry and exit probabilities to represent the joint effect of all omitted predictor variables. This way of dealing with heterogeneity can even be applied to ungrouped binary data, with each case having it's own unique predictor pattern. Instead of the likelihood, the expected (or marginal) likelihood is maximized, expected over all normally distributed values that the omitted predictors can possibly take.

Another method of dealing with overdispersion is one in which a hierarchical model specification is adopted. The method can only be applied to grouped binary data in which groups of individual cases are observed, with each group having a unique predictor pattern. On the highest level, each group's probability $p_j$ to be in state $j$ of $\Upsilon$ is taken to be drawn from some theoretical distribution, typically a beta or Dirichlet distribution. On the lowest level, the number of group members in a particular $\Upsilon$ state is taken to be drawn from a binomial or multinomial distribution. For dichotomous $\Upsilon$, this two-level sampling procedure results in a beta-binomial distribution for the number of group members in both states. For polychotomous $\Upsilon$ the resulting joint distribution of the numbers of group members in all $\Upsilon$ states is Dirichlet-multinomial. See Brown and Payne (1986) and King, Rosen and Tanner (1999) for examples of such hierarchical approaches in the field of ecological inference research. For the two state RCS model, the binomial likelihood can easily be replaced by the beta-binomial one. Estimating such model is computationally less intensive than the random effect approach proposed in Chapter 4, where integration over the normally distributed random effect can draw heavily on computer time. Also, for the random effect model, at least two random effect parameters, one for entry and one for exit, have to be estimated, while for the beta binomial model a single parameter related to the variance of the beta distribution can be used. Further, the random effect model assumes that for each predictor pattern the same amount of unobserved variance is generated by the omitted variables; the beta binomial model is more flexible in that it allows the existence of different variances for the marginal state 1 probabilities $p_{it}$ and offers the possibility to let these variances depend on predictor variables used. Prentice (1986) showed how this can be accomplished for a simple binary regression model.

### *Multi-state models*

Throughout this text, we elaborated on the two-state RCS Markov model only. The extension to three or more states is relatively straightforward. However, with more states involved the number of independent transition probabilities grows rapidly. For $k$ states there are $k(k-1)$ independent transition probabilities, meaning an almost quadratic increase of the number of probabilities and corresponding parameters to be estimated. Obviously, to obtain an acceptable precision for the many parameters one could be faced with, the total number of cases of all cross sections should keep up with the number of parameters. Also, the need for an optimization algorithm that is able to cope with multimodal log-likelihood functions becomes even more urgent since, in general, we expect more modes as more effects of the same predictor affect each separate marginal probability. Considering these issues, the first logical step is to examine the possibilities of applying three-state $\Upsilon$ models to actual data.

### *Second order Markov*

As with multi-state $\Upsilon$ variables, the extension to higher order Markov models is relatively straightforward and also increases the total number of parameters to be estimated. For binary $\Upsilon$, a second-order model has four independent transitions as opposed to two for a first-order model. Hence, the number of parameters to be estimated will in general be twice as large.

## References

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

Brown, P.J. and C.D. Payne. 1986. "Aggregate Data, Ecological Regression an Voting Transitions," *Journal of the American Statistical Association*, 81: 452-460.

Collett, D. 1991. *Modelling Binary Data.* London: Chapman and Hall.

Diggle, P.J., K. Liang and S.L. Zeger. 1994. *Analysis of Longitudinal Data.* Oxford: Clarendon Press.

King, G., O. Rosen, and M.A. Tanner. 1999. "Binomial-beta Hierarchical Models for Ecological Inference," *Sociological Methods and Research*, 28: 61-90.

King, G., O. Rosen, and M.A. Tanner. 2004. *Ecological Inference New Methodological Strategies.* Cambridge: Cambridge University Press.

Lee, T.C., G.G. Judge, and A. Zellner. 1970. "Maximum Likelihood and Bayesian Estimation of Transition Probabilities," *Journal of the American Statistical Association*, 63: 1162-1179.

Moffitt, R. 1990. "The Effect of the U.S. Welfare System on Marital Status," *Journal of Public Economics,* 59:99-123.

Moffitt, R. 1993. “Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-sections,” *Journal of Econometrics,* 59:99-123.

Omran, M.G.H. 2005. *Partical Swarm Optimization Methods for Pattern Recognition and Image Processing.* Pretoria: University of Pretoria, available from http://upetd.up.ac.za/thesis/available/etd-02172005-110834/.

Prentice, R.L. 1986. "Binary Regression Using an Extended Beta-Binomial Distribution, With Discussion of Correlation Induced by Covariate Measurement Errors," *Journal of the American Statistical Association*, 81: 321-327.

Steel, D.G., E.J. Beth, and R.L. Chambers. 2004. "The Information in Aggregate Data." In G. King, O. Rosen, and M.A. Tanner (eds.), *Ecological Inference New Methodological Strategies*, Cambridge: Cambridge University Press, pp. 51-68.

# Appendix CrossMark 2.0.0 User Manual

The program *CrossMark* is designed to estimate transition probabilities using data from repeated cross sections. Given a dichotomous $Y$ variable, *CrossMark* estimates the effects of predictor variables $X$ on the entry and exit probabilities using a Markov model.

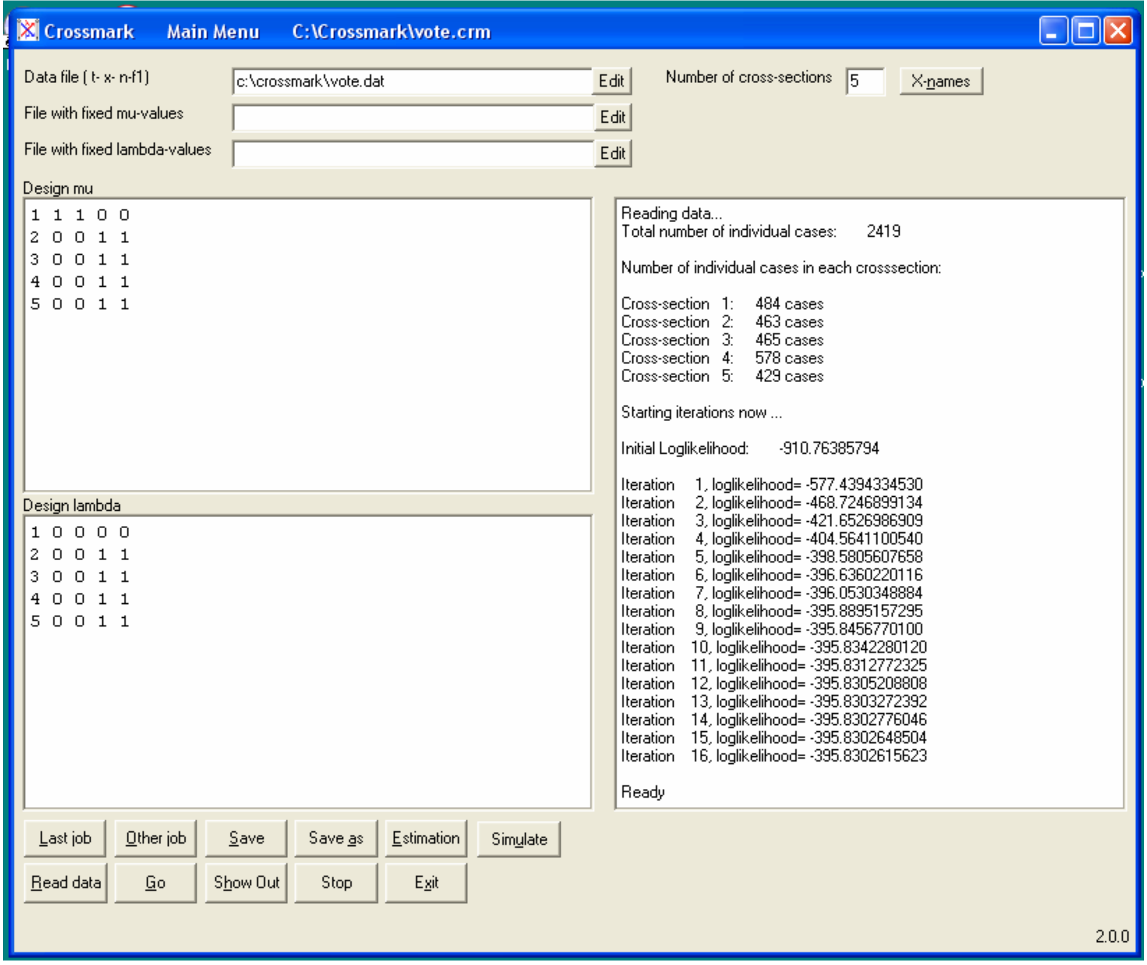
*CrossMark* is available for Windows 95, 98, 2000 an XP. The program needs not be installed: simply place file *CrossMark.exe* in a directory of your choice and double-click on this file (in Windows Explorer) to start CrossMark. The **Main Menu** then appears on the screen. This menu looks like the one in Figure 1, except that all fields are still empty.

## 1 Standard analysis

We shall describe how a standard analysis with *CrossMark* proceeds using a fictitious example on vote intention. To highlight all the options of the program, we use **bold face** characters for buttons that must be clicked and fields or menu's that have to be filled in.

Suppose the data to be analyzed are from 5 cross-sections, gathered in consecutive years, i.e., from 1996 to 2000. The dependent variable is the 'intention to vote for political party A' (code 1 = 'vote for', 0 = 'not vote for') and the independent variable is the respondent’s age (ranging from 18 to 70 years). The file containing the data is named 'c:\crossmark\vote.dat'. This filename has to be entered on the Main Menu in the field **Data file (t-x-n-f1)**. The data file can be inspected by clicking the **Edit** button, which opens the data file in WordPad. The total number of cross-sections (5) has to be

Figure 1 Main Menu



Data file ( t- x- n-f1) c:\crossmark\vote.dat

File with fixed mu-values

File with fixed lambda-values

Number of cross-sections 5

Design mu

```
1 1 1 0 0
2 0 0 1 1
3 0 0 1 1
4 0 0 1 1
5 0 0 1 1
```

Design lambda

```
1 0 0 0 0
2 0 0 1 1
3 0 0 1 1
4 0 0 1 1
5 0 0 1 1
```

```
Reading data...
Total number of individual cases:      2419

Number of individual cases in each crosssection:

Cross-section  1:    484 cases
Cross-section  2:    463 cases
Cross-section  3:    465 cases
Cross-section  4:    578 cases
Cross-section  5:    429 cases

Starting iterations now ...

Initial Loglikelihood:      -910.76385794

Iteration    1, loglikelihood= -577.4394334530
Iteration    2, loglikelihood= -468.7246899134
Iteration    3, loglikelihood= -421.6526986909
Iteration    4, loglikelihood= -404.5641100540
Iteration    5, loglikelihood= -398.5805607658
Iteration    6, loglikelihood= -396.6360220116
Iteration    7, loglikelihood= -396.0530348884
Iteration    8, loglikelihood= -395.8895157295
Iteration    9, loglikelihood= -395.8456770100
Iteration   10, loglikelihood= -395.8342280120
Iteration   11, loglikelihood= -395.8312772325
Iteration   12, loglikelihood= -395.8305208808
Iteration   13, loglikelihood= -395.8303272392
Iteration   14, loglikelihood= -395.8302776046
Iteration   15, loglikelihood= -395.8302648504
Iteration   16, loglikelihood= -395.8302615623

Ready
```

entered in the field **Number of cross-sections**. The abbreviation 't-x-n-f1' behind 'Data file' stands for t=time index, x=$X$ or predictor variables, n=number of cases and f1=number of cases in category $Y=1$, respectively, and reflects the order in which the data must appear in the data file. The first three lines of the example data of each cross-section are shown below:

```
1    1 18       1   18 19 20 21 22       9      2
1    1 19       1   19 20 21 22 23       5      0
1    1 20       1   20 21 22 23 24       3      0
2    1 18       1   18 19 20 21 22       4      1
2    1 19       1   19 20 21 22 23      13      5
2    1 20       1   20 21 22 23 24       8      0
3    1 18       1   18 19 20 21 22       4      2
3    1 19       1   19 20 21 22 23       5      1
3    1 20       1   20 21 22 23 24       8      3
4    1 18       1   18 19 20 21 22       4      2
```

```
4   1 19       1  19 20 21 22 23       12     11
4   1 20       1  20 21 22 23 24        8      5
5   1 18       1  18 19 20 21 22        7      5
5   1 19       1  19 20 21 22 23        4      1
5   1 20       1  20 21 22 23 24        2      1
```

The first data column is the time index *t.* As there are five cross-sections the time index has to have the values 1, 2, 3, 4, and 5 denoting the years 1996, 1997, 1998, 1999, 2000, respectively. *CrossMark* expects the data to be ordered in time, with the data of the first cross section located at the top of the file, those of the second cross-section following underneath and so on until the data of the last cross-section which must be located at the end of the file.

The next 8 data columns of the data file in this example, i..e., column 2 through 9, contain the values of the predictor variables $X$. There are 4 predictor variables here:

1. An intercept, having the value 1 for each case. It is located in column 2 of the data file. In the sequel we will refer to it as ‘intercept 1’.
2. The respondents age in 1996, located in column 3. For the respondents of the cross sections 1997 and following, the age in 1996 has been computed by ‘backcasting’ their age to the year 1996. We shall explain below why we use ‘age in 1996’ as a separate predictor, which we call ‘age 1996’.
3. A second intercept in column 4, which is called ‘intercept 2’.
4. The respondents age in each of the five years, located in columns 5 through 9. These five age values together constitute a single predictor variable, the values of which change over time. We call this predictor 'age'.

We will return to the characteristics of the 4 predictors and the way they affect the transition probabilities in more detail below. The last two columns, 10 and 11, of the data file concern the total number of cases and the number of cases in $Y$ category 1, respectively. For example, the first record of the cross section at $t$=5, i.e., the record

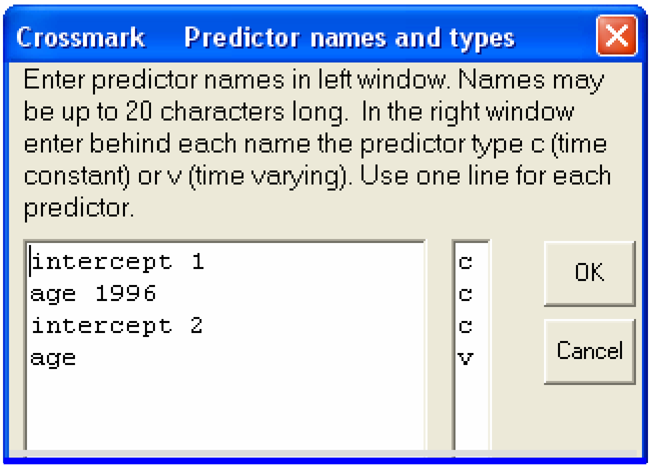
```
5   1 18       1  18 19 20 21 22        7      5
```

specifies that there are 7 cases in this cross section who were 18 years old in 1996, 19 years old in 1997 etc., and that 5 of them are in $Y$ category 1 (at $t = 5$) while the other 2 are in category 0. If each row in the datafile would contain data for just a single case, then the last but one column (here column 10) would be 1 for all cases while the last column would be either 1 or 0. There is no need to aggregate over $t$, $X$ and $Y$. However, aggregating the data, as is done in this example, can speed up the estimation process considerably.

We now return to the predictor variables $X$. The predictors numbered 1, 2 and 3 above are constant over time, while predictor 4 takes a different value in each of the five years. Time constant predictors occupy a single column in the data file, while time varying predictors occupy as many columns as there are cross-sections, i.e. five in the example. The names and types (constant or varying) of the predictors have to be specified in the submenu **Predictor names and types,** which shows up after clicking the **X-names** button of the Main Menu and is shown in Figure 2. The left field of the submenu **Predictor names and types** contains the predictor's name and the right field the predictor's type. For a time constant predictor enter the character **c**, and for a time varying predictor enter **v**. Having done so, click OK to return to the Main Menu.

To understand why we use two intercepts and two age predictors (instead of just one intercept and one age predictor, which would be possible too)

Figure 2 Predictor names and types



Crossmark Predictor names and types

Enter predictor names in left window. Names may be up to 20 characters long. In the right window enter behind each name the predictor type c (time constant) or v (time varying). Use one line for each predictor.

```
intercept 1      c
age 1996         c
intercept 2      c
age              v
```

we take a closer look at the model equations for $p_1$, $p_2$, $p_3$, $p_4$ and $p_5$ or, in words, the probabilities to vote for political party A in each of the five years. In general, the basic equations *CrossMark* uses are, with five cross-sections:

$$\begin{aligned}
p_1 &= \mu_1 \\
p_2 &= p_1\ (1-\lambda_2) + (1-p_1)\ \mu_2 \\
p_3 &= p_2\ (1-\lambda_3) + (1-p_2)\ \mu_3 \\
p_4 &= p_3\ (1-\lambda_4) + (1-p_3)\ \mu_4 \\
p_5 &= p_4\ (1-\lambda_5) + (1-p_4)\ \mu_5
\end{aligned}$$

In the example, the transition probabilities $\mu$ and $\lambda$ depend on the respondents ages as follows:

$$\begin{aligned}
&\text{logit}(\mu_1) = \beta_1 + \beta_2\, Age_{1996} && \\
&\text{logit}(\mu_2) = \beta_3 + \beta_4\, Age_{1997} && \text{logit}(1-\lambda_2) = \beta_1^* + \beta_2^*\, Age_{1997} \\
&\text{logit}(\mu_3) = \beta_3 + \beta_4\, Age_{1998} && \text{logit}(1-\lambda_3) = \beta_1^* + \beta_2^*\, Age_{1998} \\
&\text{logit}(\mu_4) = \beta_3 + \beta_4\, Age_{1999} && \text{logit}(1-\lambda_4) = \beta_1^* + \beta_2^*\, Age_{1999} \\
&\text{logit}(\mu_5) = \beta_3 + \beta_4\, Age_{2000} && \text{logit}(1-\lambda_5) = \beta_1^* + \beta_2^*\, Age_{2000}
\end{aligned}$$

$\text{Age}_{1996}$ refers to the respondent's age in 1996, $\text{Age}_{1997}$ to the age in 1997, etcetera. The symbol $\lambda$ indicates the exit probability: $\lambda_3$ is the probability not to vote for party A in 1998 given a 'vote for A' in 1997. For the complement of $\lambda$, or the probability to stay in state $Y = 1$, the term '1-exit' probability is used in the sequel. The symbol $\mu$ indicates the entry probability: $\mu_3$ is the probability to vote for A in 1998 given a 'not vote for A' in 1997.

Speaking of $\mu_1 = p_1$ as an entry probability can be problematic. Generally spoken, $p_1$ is the probability to be in state $Y = 1$ at $t = 1$ and this need not to be the same as the probability to be in state $Y = 1$ given that the previous state was $Y = 0$. Only if one knows that each respondent's previous state was $Y = 0$, one may truly consider $p_1$ an entry probability. This would e.g. be the case if political party A did not exist before 1996. In many applications, of course, the $Y = 1$ state does exist prior to $t = 1$ and respondents could have been in that state. In such situations, one may prefer to model $p_1$ as a state probability, rather than an entry probability. This is accomplished by estimating different sets of parameters for $\mu_1$ and for $\mu_2$ and following, as is done in the model above, where the parameters $\beta_1$ and $\beta_2$ only apply to $\mu_1$.

In *CrossMark* the model equations can be specified in the **Design mu** and **Design lambda** fields of the Main Menu. In **Design mu** we indicate which predictor variable acts upon which entry probability $\mu$. For the example this is done as follows:

```
 1   1 1   0 0
 2   0 0   1 1
 3   0 0   1 1
 4   0 0   1 1
 5   0 0   1 1
```

The first column is the time index $t$ and the other four columns correspond to the four predictor variables in the model. The second column corresponds to 'intercept 1', and the value 1 for $t = 1$ indicates that ‘intercept 1’ has an effect on $\mu_1$; the 0 scores in the second column for $t = 2,\ 3,\ 4$ and $5$ indicate that ‘intercept 1’ does not have an effect on $\mu_2$, $\mu_3$, $\mu_4$ and $\mu_5$. The rightmost column is related to the time varying predictor ‘age’; the 0 value for $t = 1$ indicates that 'age' does not occur in the equation for $\mu_1$ while the 1 values for $t = 2,\ 3,\ 4$ and $5$ indicate that 'age' does occur in the equations for $\mu_2$, $\mu_3$, $\mu_4$ and $\mu_5$.

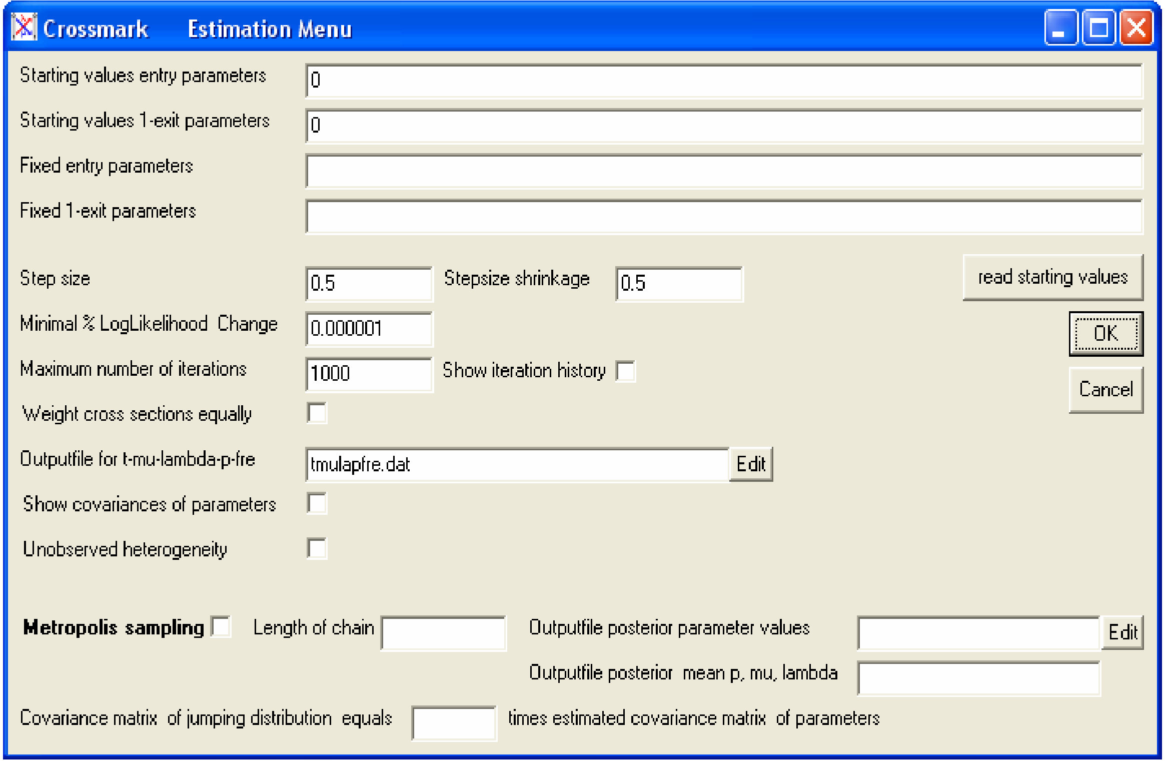
In general, the **Design mu** matrix must have as many rows as there are cross-sections. Each row starts with the time index $t$ and is followed by a 1 or 0 value for each predictor variable indicating whether (1) or not (0) the predictor acts upon entry probability $\mu_t$. In the same way a **Design lambda** matrix has to be specified indicating which predictor acts upon which exit probability $\lambda$. For the present example the lambda matrix is specified as:

```
 1   0 0   0 0
 2   0 0   1 1
 3   0 0   1 1
 4   0 0   1 1
 5   0 0   1 1
```

Note that the first row of the **Design lambda** matrix contains the value 1 for the time index $t = 1$ and else only 0 values to indicate that none of the four predictor variables has an effect on $\lambda_1$. This is just to specify that $\lambda_1$ does not play a part in the model equations.

We proceed by clicking the **Estimation** button of the Main menu to invoke the Estimation Menu as shown in Figure 3. The upper two fields in this Estimation Menu specify the **starting values** for the iterative Fisher scoring scheme. The default values are 0 for all $\beta$ and $\beta^*$ parameters of the entry and 1-exit probabilities respectively. Good starting values, i.e., values close to the final ML estimates, speed up the estimation process. Starting values far removed from the final estimates slow down this process or may cause the estimates to be caught in a local maximum or not to reach convergence at all. When convergence has been reached, it is advisable to choose other starting values and let *CrossMark* run again to check whether the same parameter estimates are found. If this turns out to be the case, one can be more confident that the estimates are indeed the true global ML estimates instead of estimates associated with a local maximum.

Figure 3   Estimation Menu



Crossmark    Estimation Menu

Starting values entry parameters: 0
Starting values 1-exit parameters: 0
Fixed entry parameters:
Fixed 1-exit parameters:

Step size: 0.5    Stepsize shrinkage: 0.5    read starting values
Minimal % LogLikelihood Change: 0.000001    OK
Maximum number of iterations: 1000    Show iteration history □    Cancel
Weight cross sections equally □
Outputfile for t-mu-lambda-p-fre: tmulapfre.dat    Edit
Show covariances of parameters □
Unobserved heterogeneity □

**Metropolis sampling** □    Length of chain:    Outputfile posterior parameter values:    Edit
Outputfile posterior mean p, mu, lambda:
Covariance matrix of jumping distribution equals    times estimated covariance matrix of parameters

When analyzing complex models, in the sense of having many predictors, starting values become more of an issue. The final estimates of a previous, relatively simple model can be used as starting values for a new model having additional predictors. To this end the button **read starting values** can be helpful. After clicking, the final estimates of the previous model are filled in as starting values in both fields. The starting values for the additional predictors in the second model are defined to be zero and automatically added to the list. If a predictor that was present in the previous model does not appear in the second, the user has to remove the relevant starting values from both lines.

If, for some reason, one would like to fix the parameters of one or more predictors to certain predefined values instead of having them estimated by CrossMark, one can be proceed as follows. In the field named **Fixed entry parameters** enter a value 0 or 1 for each predictor parameter that has to be estimated (enter 0) or not (enter 1). Be sure to enter a value 0 or 1 for *all* predictors and to use the same *order* for the predictors as was used in the menu Predictor names and types. For predictors that have a value 0 specified, CrossMark will estimate a parameter starting from the starting value. For predictors that have a value 1 specified, CrossMark will not estimate a parameter but substitutes the given starting value as the parameter value to be used for this predictor's effect on the entry probability. In CrossMark's output, fixed parameters are denoted by the character 'f' and have a Wald Significance and Std. error of 1.0. In the same manner, one can fix parameters for the 1-exit probability.

The **Step size** field in the Estimation Window refers to the step size $\varepsilon$ of the Fisher scoring algorithm employed for iteratively updating the parameter estimates. The algorithm is given by $\hat{\theta}_{k+1} = \hat{\theta}_k + \varepsilon \ \hat{I}_k^{-1} \ (\delta LL / \delta\theta)_k$, where $\hat{\theta}_k$ and $\hat{\theta}_{k+1}$ are the parameter estimates at the iterations $k$ and $k+1$, $\hat{I}_k^{-1}$ is the inverse of the Fisher information matrix evaluated at $\theta = \hat{\theta}_k$, and $(\delta LL / \delta\theta)_k$ are the derivatives of the log likelihood with respect to the parameters, evaluated at $\theta = \hat{\theta}_k$. By default, the value of the step size $\varepsilon$ is 0.5. If the log likelihood function has a single mode, the optimal value for the step size would be 1. It is not unusual, however, for the log likelihood function to have multiple modes in which case a step size of 1 could easily cause the algorithm to jump over the parameter region with the highest mode. For this reason, a default step size of 0.5 is chosen. A much smaller step size value may slow down the algorithm too much. There is no rule of thumb given here as to the choice of the most efficient step size value.

The **Step size shrinkage** ($s$) also deals with the problem of the step size being too large. If the log likelihood based on $\hat{\theta}_{k+1}$ is lower than the one based on $\hat{\theta}_k$, the current step size has apparently been too large. In that case *CrossMark* produces the message "Not converging, back to parameter estimates of previous iteration" and takes as the new step size the product $s \cdot \varepsilon$. If this smaller step size also leads to $\hat{\theta}_{k+1}$ estimates with a lower log likelihood than the one based on $\hat{\theta}_k$, the step size $s \cdot s \cdot \varepsilon$ is tried. In short, the step size is multiplied by $s$ as many times as needed to produce an increase in log likelihood.

The iterative estimation process ends if either the percentage change in log likelihood is less than the **Minimal % LogLikelihood Change** specified, which by default is 0.000001%, or the **Maximum number of iterations** has been reached, which by default is 1000. Also by default, CrossMark only shows the parameter estimates of the final iteration and not those of previous iterations. To force CrossMark showing the estimates of each iteration, check the **Show iteration history** option.

By default CrossMark applies caseweights resulting in the same weighted number of cases for each cross section. The sum of all caseweights is equal to the total number of cases in all cross sections. To prevent this weighting procedure uncheck the option **Weight cross sections equally**.

*CrossMark* produces an output file, the name of which can be specified in the field **Outputfile for t-mu-lambda-p-fre**. By default it is labeled 'tmulapfre' and placed in the directory where the 'crossmark.exe' resides. The output file contains one line for each case in the data file. For case $i$, this line has the following information from left to right:

- the time index of the cross-section case $i$ belongs to,
- the predicted values of $\mu_{i1}$ to $\mu_{iT}$,
- the predicted values of $\lambda_{i1}$ to $\lambda_{iT}$,

- the predicted values of $p_{i1}$ to $p_{iT}$,
- the frequency of case $i$, equal to the frequency specified in the rightmost column of the data file.

Predicted $\mu$, $\lambda$ and $p$ values that do not apply to a particular case (e.g., $\mu_3$ for a case of cross-section 2, or $\lambda_1$ for all cases) are assigned the ‘missing value’ 9.

By default, in CrossMark's output no (co)variances of parameter estimates are shown. They will be, if the option **Show covariances of parameters** is checked before running the model.

The options for **Unobserved heterogeneity** and **Metropolis sampling** will be discussed below in separate sections.

After clicking the **OK** button of the Estimation Menu the Main Menu reappears. To save all the specifications entered, click the **Save** button and specify a file name, e.g. 'vote.crm' which then appears in the top line of the Main Menu. Using the **Save as** button enables saving the job under a different name. The most recently saved job can be opened by clicking on the button **Last job** while older jobs may be opened with **Other job**.

To start the analysis the data have to be read first. This is done by clicking on **Read data**. When finished reading, *CrossMark* presents the total number of cases as well as the number of cases for each cross-section in the rightmost window of the Main Menu. After reading the data, the estimation can be carried out by clicking on **Go.** The initial log likelihood, based on the starting values of the parameters, appears on the screen after a few moments, as does the log likelihood of each subsequent iteration. When the last iteration is finished, a 'Ready' message is delivered. The estimation may take some time, especially when many cases and/or predictor variables are involved. In the mean time the user may want to look at intermediate results by clicking the **Show Out** button or pressing **Ctrl+Tab** on the keyboard. The **Output window** then appears, with the parameter estimates of each iteration scrolling over the screen, accompanied by the log likelihood and, possibly, messages concerning corrective actions undertaken by the estimation algorithm. Pressing **Ctrl+Tab** again (or clicking the cross X in the upper right corner of the screen) closes the Output window.

Back in the Main Menu the estimation process - if still running - can be stopped by using the **Stop** button. This may be useful if e.g. the log likelihood does not change substantially anymore. Another reason to stop the iterations is that the algorithm does not converge, which may happen if the model contains too many (i.e., not uniquely identified) parameters.

To leave *CrossMark* click **Exit** or the cross X in the upper right corner of the screen.

## 2 Nonbackcastable variables

It may be that the respondent's value on a predictor variable at time $t$ is known, but the values at $t-1$, $t-2$ and so on are not. Take e.g. the variable ‘monthly income’. Given the income of a respondent of cross-section $t$, usually little, if anything, is know about his or her income at earlier points in time. To put it another way: the variable income cannot be 'backcasted'. Such a nonbackcastable variable can be used as a predictor for the entry and exit probability only at the time the respondent was observed but not at preceding points in time. We will show using a simple example how such variables can be handled in *CrossMark*.

Suppose that we have three cross-sections and the nonbackcastable predictor we would like to use is named $Inc$, representing the monthly personal income of a respondent at the time of observation. Also, we have the backcastable predictor age specified as $Age(t)$, where the $t$ between brackets denotes that there are three age vectors, one for each of the three points in time. For simplicity, we omit the intercept in the equations for $\mu$ below. For any respondent of the second and subsequent cross-sections, the following two equations apply to $\text{logit}(\mu_t)$, depending on whether $t$ relates to the time the respondent is actually observed or to a preceding point in time:

observed: logit($\mu_t$) = $\beta_1 \cdot Age(t)$ + $\beta_3 \cdot$ Inc (1)

preceding: logit($\mu_t$) = $\beta_2 \cdot Age(t)$ (2)

In equation (1) we can use $Inc$ as a predictor, whereas in equation (2) this is not possible. Of course the $Age$ effects $\beta_1$ and $\beta_2$ need not necessarily be the same. In order to estimate $\beta_1$, $\beta_2$ and $\beta_3$ with *CrossMark* a single equation for $\text{logit}(\mu_t)$ must be specified that applies to all points in time. To achieve this we construct three ancillary time varying predictors, which we shall call $Age\_obs(t)$, $Age\_pre(t)$ and $Inc\_obs(t)$ to be discussed below. The construction of these predictors must precede the analysis with *CrossMark* and the user must add the predictors to the data file and treat them like any normal predictor variable: their names and types (v) have to be entered (using the **X-names** button in the Main Menu) and also, three columns, one for each predictor, have to be added to the **Design mu** and **Design lambda** matrices.

The predictor $Age\_obs(t)$ has to be constructed such that $Age\_obs(t) = Age(t)$ for cases observed at time point $t$ and $Age\_obs(t) = 0$ for all other cases. For predictor $Age\_pre(t)$ it must hold that $Age\_pre(t) = Age(t)$ for cases observed *after* time point $t$ and $Age\_pre(t) = 0$ for all other cases.

For 6 randomly chosen cases, two of each cross-section, the values of $Age(t)$, $Age\_obs(t)$ and $Age\_pre(t)$ might be those shown in the upper part of Table 1. Note that, put next to one another, the three $Age\_\ obs(t)$ vectors form a block-diagonal matrix and the $Age\_pre(t)$ vectors a 'sub-block diagonal' one. For $Inc$ and $Inc\_obs(t)$ the values of the 6 cases might be the ones in the lower part of Table 1, with now the $Inc\_obs(t)$ vectors forming a block-diagonal matrix. Instead of the two separate equations (1) and (2), we can write a single equation, holding for time observed as well as preceding points in time:

$$\text{logit}(\mu_t) = \beta_4 \cdot Age\_\ obs(t) + \beta_5 \cdot Age\_pre(t) + \beta_6 \cdot Inc\_obs(t) \qquad (3)$$

Why (1) and (2) are equivalent to (3) becomes clear when equation (3) is worked out for the observed and preceding time points separately:

Table 1 Ancillary predictors for Age

| | $Age(t)$ | | | $Age\_obs(t)$ | | | $Age\_pre(t)$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (1) | (2) | (3) | (1) | (2) | (3) |
| $t=1$ | 19 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 |
| | 45 | 0 | 0 | 45 | 0 | 0 | 0 | 0 | 0 |
| $t=2$ | 37 | 38 | 0 | 0 | 38 | 0 | 37 | 0 | 0 |
| | 21 | 22 | 0 | 0 | 22 | 0 | 21 | 0 | 0 |
| $t=3$ | 42 | 43 | 77 | 0 | 0 | 44 | 42 | 43 | 0 |
| | 66 | 67 | 68 | 0 | 0 | 68 | 66 | 67 | 0 |

| | $Inc$ | $Inc\_obs(t)$ | | |
|---|---|---|---|---|
| | | (1) | (2) | (3) |
| $t=1$ | 1500 | 1500 | 0 | 0 |
| | 7300 | 7300 | 0 | 0 |
| $t=2$ | 3500 | 0 | 3500 | 0 |
| | 9400 | 0 | 9400 | 0 |
| $t=3$ | 1200 | 0 | 0 | 1200 |
| | 2200 | 0 | 0 | 2200 |

$$\begin{aligned} \text{observed:}\quad \operatorname{logit}(\mu_t) &= \beta_4 \cdot Age\_obs(t) + \beta_5 \cdot Age\_pre(t) + \beta_6 \cdot Inc\_obs(t) \\ &= \beta_4 \cdot Age(t) \qquad + \beta_5 \cdot 0 \qquad\quad + \beta_6 \cdot Inc \\ &= \beta_4 \cdot Age(t) \qquad\qquad\qquad\qquad\;\; + \beta_6 \cdot Inc \end{aligned} \tag{3a}$$

$$\begin{aligned} \text{preceding:}\quad \operatorname{logit}(\mu_t) &= \beta_4 \cdot Age\_obs(t) + \beta_5 \cdot Age\_pre(t) + \beta_6 \cdot Inc\_obs(t) \\ &= \beta_4 \cdot 0 \qquad\qquad + \beta_5 \cdot Age(t) \qquad + \beta_6 \cdot 0 \\ &= \qquad\qquad\qquad\;\; \beta_5 \cdot Age(t) \end{aligned} \tag{3b}$$

Thus, equations (3a) and (3b) appear to be equivalent to (1) and (2), respectively. Since *CrossMark* uses a single equation for $\mu$ we employ the generic equation (3). Parameter $\beta_4$ can be interpreted as $\beta_1$, i.e., the effect of age controlled for income, at observation time; $\beta_5$ is interpreted like $\beta_2$ as the effect of age at preceding points in time not controlled for income; $\beta_6$ has the same interpretation as $\beta_3$, i.e., the effect of income controlled for age at the time of observation.

Instead of (3) way may also use another generic equation in *CrossMark*:

$$\text{logit}(\mu_t) = \beta_7 \cdot Age(t) \quad + \beta_8 \cdot Age\_obs(t) \; + \beta_9 \cdot Inc\_obs(t) \tag{4}$$

Working out (4) for observation time and preceding timepoints results in:

$$\begin{aligned} \text{observed:} \quad \text{logit}(\mu_t) &= \beta_7 \cdot Age(t) \quad + \beta_8 \cdot Age\_obs(t) \; + \beta_9 \cdot Inc\_obs(t) \\ &= \beta_7 \cdot Age(t) \quad + \beta_8 \cdot Age(t) \qquad + \beta_9 \cdot Inc \\ &= (\beta_7 + \beta_8) \cdot Age(t) \qquad\qquad + \beta_9 \cdot Inc \end{aligned} \tag{4a}$$

$$\begin{aligned} \text{preceding:} \quad \text{logit}(\mu_t) &= \beta_7 \cdot Age(t) \quad + \beta_8 \cdot Age\_obs(t) \; + \beta_9 \cdot Inc\_obs(t) \\ \text{logit}(\mu_t) &= \beta_7 \cdot Age(t) \quad + \beta_8 \cdot 0 \qquad\qquad + \beta_9 \cdot 0 \\ \text{logit}(\mu_t) &= \beta_7 \cdot Age(t) \end{aligned} \tag{4b}$$

As can be seen (4a) is equivalent to (3a) and (1), while (4b) is equivalent to (3b) and (2). Therefore, both equation (3) and (4) can be used to model $\text{logit}(\mu_t)$. They differ only in parameterization. The sum $\beta_7 + \beta_8$ has the same interpretation as $\beta_4$ (or $\beta_1$); $\beta_7$ is interpreted in the same way as $\beta_5$ (or $\beta_2$). Finally, the interpretation of $\beta_9$ is similar to the one of $\beta_5$ (or $\beta_3$). A minor advantage of using (4) instead of (3), is that (4) needs on construction of the $Age\_pre(t)$ vectors.

## 2.1 Testing the null-hypothesis $\text{H}_0 : \beta_1 = \beta_2$

Looking at the equations (1) and (2) the question arises as to the equality of the two $Age$ effects $\beta_1$ and $\beta_2$. When applying equation (4) the above null hypothesis translates into $\text{H}_0 : \beta_7 + \beta_8 = \beta_7$ or, more simply, to $\text{H}_0 : \beta_8 = 0$. This test is automatically performed by *CrossMark* and the significance level of the related Wald statistic is reported in the Output window. When, on the other hand, equation (3) is applied, the above hypothesis translates into $\text{H}_0 : \beta_4 - \beta_5 = 0$. Given the hypothesis is true, the sample outcome of the statistic $(\hat{\beta}_4 - \hat{\beta}_5)^2 / \text{var}(\hat{\beta}_4 - \hat{\beta}_5)$, with $\text{var}(\hat{\beta}_4 - \hat{\beta}_5)$ being the estimated sample variance of $\hat{\beta}_4 - \hat{\beta}_5$, follows a $\chi^2$ distribution with 1 degree of freedom. The value of $\hat{\beta}_4 - \hat{\beta}_5$ can of course be derived from the ML estimates produced by *CrossMark* in the final iteration. To derive $\text{var}(\hat{\beta}_4 - \hat{\beta}_5)$ the formula $\text{var}(\hat{\beta}_4 - \hat{\beta}_5) = \text{var}(\hat{\beta}_4) + \text{var}(\hat{\beta}_5) - 2\,\text{cov}(\hat{\beta}_4, \hat{\beta}_5)$ can be applied with $\text{var}(\hat{\beta}_4)$, $\text{var}(\hat{\beta}_5)$

Table 2  Ancillary intercept predictors

| $Intercept$ | | $Intercept\_obs(t)$ | | | $Intercept\_pre(t)$ | | |
|---|---|---|---|---|---|---|---|
| | | (1) | (2) | (3) | (1) | (2) | (3) |
| $t=1$ | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| $t=2$ | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| $t=3$ | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| | 1 | 0 | 0 | 1 | 1 | 1 | 0 |

and $\mathrm{cov}(\hat{\beta}_4, \hat{\beta}_5)$ representing the estimated variances of $\hat{\beta}_4$ and $\hat{\beta}_5$ and their estimated covariance respectively. These variances and covariance are given by *CrossMark* by checking the option **Show covariances of parameters** in the **Estimation Menu**.

If the test outcome leads to not rejecting the null hypothesis, the ancillary variables for the predictor in question are no longer needed and the original predictor, $Age(t)$ in the example, can be used, possibly along with ancillary variables of other predictors for which the hypothesis does not hold.

The equations above did not include an intercept, for simplicity. Of course, in most applications an intercept will be present and we will have to decide which type of intercept vector(s) to employ. If we have no nonbackcastable predictors, the intercept is simply a single vector containing the value 1 for all cases of all cross-sections. If, however, nonbackcastable predictors are utilized, we may want to estimate one intercept for time observed and another one for preceding time, just as was done for $Age(t)$ in equations (1) and (2). In that case we would have to construct two ancillary (time varying) intercept predictors, according to the scheme in Table 2.

## 3 Fixed $\mu$ and $\lambda$ values

*CrossMark* has the option of entering fixed $\mu$ and/or fixed $\lambda$ values for some (or all) cases on some (or all) points in time. We start with discussing three situations in which this option can be utilized to adjust the basic equations for the state probabilities $p$. We also explain how the option has to be specified in *CrossMark*.

In some applications, the values for $\mu$ and/or $\lambda$ may be considered fixed and hence need not be estimated. This would e.g. be the case when the (backcasted) age of a respondent is 17 or younger in a study on voting behavior, given that the voting age is 18. Suppose, in the example given earlier, a respondent is 18 years old at the time that the third cross-section was observed (i.e., on $t = 3$). For this respondent we would like $p_1$ and $p_2$ to be zero; also, since $p_3$ is an entry probability (the respondent could not have voted for party A at $t = 2$) we would like $p_3$ to equal the entry probability $\mu_3$. To implement these restrictions in the model equations, we fix $\mu_1 = \mu_2 = 0$ for this respondent, which implies the following adjusted equations for $p_1$ to $p_5$:

$$
\begin{aligned}
p_1 &= \mu_1 = 0 \\
p_2 &= p_1(1-\lambda_2) + (1-p_1)\mu_2 = 0(1-\lambda_2) + 1\cdot 0 \quad = 0 \\
p_3 &= p_2(1-\lambda_3) + (1-p_2)\mu_3 = 0(1-\lambda_3) + 1\cdot \mu_3 = \mu_3 \\
p_4 &= p_3(1-\lambda_4) + (1-p_3)\mu_4 \\
p_5 &= p_4(1-\lambda_5) + (1-p_4)\mu_5
\end{aligned}
$$

The equations for $p_4$ and $p_5$ have the usual Markov form, while those for $p_1$, $p_2$ and $p_3$ are adjusted in the sense specified above. We shall explain below how the fixed $0$values for the $\mu$ probabilities in question for respondents younger than 18 have to be entered in *CrossMark*.

A second example of adjusting the basic equations for $p$ is the following. Suppose all predictor variables we would like to use are constant over time, but only for a short time period. To be more specific, we assume that the predictor values for a case observed at time $t$ also apply to $t-1$ and $t-2$, but not further back in time. Therefore, we let the Markov chain for each case start two time points preceding to the one the case was observed, instead of starting at time point $t = 1$ as we would have done, had the predictors been perfectly stable. This implies that the first state probability estimated for the cases of the cross-section at $t = 5$ will be $p_3$. For the cases of the cross-section at $t = 4$, $p_2$ will be the first estimated state probability, and for those of the cross-section at $t = 3$, $t = 2$ and $t = 1$, $p_1$ will be the first estimated state probability. This is different from the more general situation where, for all cases of all cross-sections, $p_1$ is the first estimated state probability. Remember that for $p_1$ we used a logistic equation, $p_1 = \mu_1$, with specific $\beta$ parameters, different from the ones of $\mu_2$ through $\mu_5$. Here, we would like the same to hold for $p_2$ and $p_3$, as far as the cases of the cross-sections at $t = 4$ and $t = 5$ respectively are involved. To achieve this, we shall again use the equation $p_1 = \mu_1$ to estimate $p_1$ as the first estimated state probability for all cases of all cross-sections and then (i) let $p_2$ have the same value as $p_1$ for the cases of the cross-section at $t = 4$ and (ii) let $p_3$ have the same value as $p_1$ for the cases of the cross-section at $t = 5$. By doing so, we estimate three first state probabilities, $p_1$, $p_2$ and $p_3$, using the logistic equations $p_1 = \mu_1$, $p_2 = \mu_1$ and $p_3 = \mu_1$. At the same time $p_2$ and $p_3$ are also estimated by a Markov equation for the cases of the cross-sections at $t = 3$ and $t = 4$ respectively.

To specify the model we exploit fixed $\mu$ and $\lambda$ values. Let us take a look at a case of the cross-section at $t = 5$ for which we want to estimate $p_3$ using the equation $p_3 = \mu_1$. We let $\lambda_2 = \lambda_3 = 0$ and $\mu_2 = \mu_3 = 0$, which results in:

$$
\begin{aligned}
p_1 &= \mu_1 \\
p_2 &= p_1(1-\lambda_2) + (1-p_1)\mu_2 &&= \mu_1(1-0) &&+ (1-\mu_1)\cdot 0 = \mu_1 \\
p_3 &= p_2(1-\lambda_3) + (1-p_2)\mu_3 &&= \mu_1(1-0) &&+ (1-\mu_1)\cdot 0 = \mu_1 \\
p_4 &= p_3(1-\lambda_4) + (1-p_3)\mu_4 \\
p_5 &= p_4(1-\lambda_5) + (1-p_4)\mu_5
\end{aligned}
$$

As can be seen, the equations for $p_5$ and $p_4$ are the usual Markov equations, while for $p_3$ we have $p_3 = \mu_1$. For cases of cross-section at $t = 4$ we proceed in a similar way by fixing $\lambda_2 = 0$ and $\mu_2 = 0$ which leads to $p_2 = \mu_1$. For the cases of the cross-sections at $t = 3$, $t = 2$ and

$t=1$, we automatically have $p_1=\mu_1$, so for these cases we do not need to fix any $\mu$ or $\lambda$.

The last example of using fixed $\mu$ and $\lambda$ values concerns the analysis of discrete panel data. Consider a situation in which we have at our disposal a five wave panel data set without any inflow or outflow. The Markov model for discrete panel data reads as

$$p_t = y_{t-1}(1-\lambda_t) \ + \ (1-y_{t-1})\mu_t, \qquad t=2,\ldots,5,$$

while for cross-sections, it reads as

$$p_t = p_{t-1}(1-\lambda_t) \ + \ (1-p_{t-1})\mu_t, \qquad t=2,\ldots,5,$$

the difference being the use of $y_{t-1}$ in the case of panel data and $p_{t-1}$ when using cross-sectional data. As stated earlier, *CrossMark* uses the second equation since it was designed for the analysis of cross-sectional data. However, the program can simply be tricked to analyze panel data as well and thus to apply the first equation.

To do so, we first have to construct the data file in the way *CrossMark* expects it to be, i.e., according to the t-y-x-fre format. Each 'cross-section' in this data file corresponds to a particular wave of the panel data. The data for the first wave have to be placed at the top of the data file, followed by the data for the second wave, the third wave and so on. The order in which the respondents appear within the data for each wave is irrelevant and need not be the same for each wave.

Second, we need to define $p_{t-1}=y_{t-1}$ for $t=2,\ldots,5$ or, to put it simply, $p_t=y_t$ for $t=1,\ldots,4$. To do so we use fixed $\mu$ and fixed $\lambda$ values. To make sure that $p_1=y_1$, we simply let $\mu_1=y_1$, resulting in $p_1=\mu_1=y_1$. For $p_2$ through $p_4$ we proceed as follows. If for a certain case $y_t=0$ ($t=2,\ldots,4$), we let $\lambda_t=1$ and $\mu_t=0$, which results in $p_t=p_{t-1}(1-\lambda_t)+(1-p_{t-1})\mu_t=p_{t-1}(1-1)+(1-p_{t-1})\,0=0$; thus $p_t=y_t=0$, as was meant to be the case. If, on the other hand, $y_t=1$, we let $\lambda_t=0$ and $\mu_t=1$, so that $p_t=p_{t-1}(1-0)+(1-p_{t-1})=1$; thus $p_t=y_t=1$.

The third and final point concerns the fact that in models for panel data the likelihood is commonly computed for the data of $t\geq 2$, while in

*CrossMark*, the likelihood for $t = 1$ is used as well. To delete the likelihood contribution of the cases for $t = 1$ in *CrossMark*, we assign a very small frequency to the cases of the first wave (i.e., 0.0000000001) in the (t-y-x-fre) data file. We can also delete all cases of the first wave from the data file except one case, and assign the small frequency value to this single case. This single remaining case for $t = 1$ may have any values on the $Y$ and $X$ variables since it only acts as a dummy case, having (virtually) no influence on the parameter estimates.

## 3.1 Specifying fixed $\mu$ and $\lambda$ values in CrossMark

The fields **File with fixed mu-values** and **File with fixed lambda-values** in the Main Menu can be used to enter the names of the data files containing fixed $\mu$ and $\lambda$ values for some or all cases of some or all cross-sections. The 'file with fixed mu-values' must contain one line for each case to which fixed $\mu$ values are assigned. Each line starts with the sequence number the case has in the (t-y-x-fre) data file and is followed by as many values 0, 1 or 9 as there are cross-sections. In the first example given above, where the age of a respondent (say the 316th respondent in the data file) was 18 years at the time point of the third cross-section, the line to enter in the 'file with fixed-mu values' for this respondent is the first of the two following lines:

```
316  0  0  9 9 9
925  0  0  0 0 9
```

Value 316 in the first line refers to the sequence number of the respondent; the two 0 values that follow are assigned to $\mu_1$ and $\mu_2$ and the three 9 values indicate that $\mu_3$, $\mu_4$ and $\mu_5$ are not fixed, but have to be estimated. The second line refers to another respondent with sequence number 925 in the data file, who was 18 years old at $t = 5$. In this example a 'file with fixed lambda values' need not be specified, since only values of $\mu$ are fixed.

The 'file with fixed lambda-values' must contain one line for each case to which fixed $\lambda$ values are assigned. Each line starts with the sequence number of the case in the data file and is followed by as many values 0, 1 or 9 as there are cross-sections minus 1, since these values relate to $\lambda_2$ through $\lambda_T$, $T$ being the total number of cross-sections. The third example given above concerned the analysis of five-wave panel data without inflow and outflow. If we assume there are 500 respondents then the data file consists of 2500 lines, 500 lines for each wave. Suppose a particular respondent has the $Y$ pattern 01100 for $t = 1,\ldots,5$. If the sequence number of the respondent in the first wave is 29, then the other four sequence numbers are 529, 1029, 1529 and 2029. In the 'file with fixed mu-values' and the 'File with fixed lambda-values' we have to enter the lines given in the box below.

| File with fixed mu-values | | | | | | File with fixed lambda-values | | | | | Wave |
|---|---|---|---|---|---|---|---|---|---|---|---|
| seqnr | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\mu_4$ | $\mu_5$ | seqnr | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | |
| 529 | 0 | 9 | 9 | 9 | 9 | | | | | | 2 |
| 1029 | 9 | 1 | 9 | 9 | 9 | 1029 | 0 | 9 | 9 | 9 | 3 |
| 1529 | 9 | 9 | 1 | 9 | 9 | 1529 | 9 | 0 | 9 | 9 | 4 |
| 2029 | 9 | 9 | 9 | 0 | 9 | 2029 | 9 | 9 | 1 | 9 | 5 |

As can be seen, for the data of wave $t$ we specify a fixed $\mu_{t-1}$ value in the 'file with fixed mu-values' equal to value of $Y_{t-1}$; e.g. for wave 3 we specify $\mu_2 = y_2 = 1$. The fixed $\lambda_{t-1}$ value that has to be specified in the 'File with fixed lambda-values' for the data of wave $t$ is equal to the complement of $Y_{t-1}$.

## 4 Unobserved heterogeneity

*CrossMark* offers the possibility to account for the influence of unobserved variables on the entry and exit probabilities. In doing so the assumption is made that the overall contribution of these variables to the logits of the transition probabilities is constant for the time period considered. The logit equations for $\mu$ and $1-\lambda$ including the contributions of unobserved variabels can be written as follows:

$$\text{logit}\,(\mu_t) = x\beta + \delta_1$$

$$\text{logit}\,(1-\lambda_t) = x\beta^* + \delta_2\,,$$

where $x$ is a row vector with the values of the observed (potentially backcasted) predictors, $\beta$ and $\beta^*$ are the column vectors with the parameters associated with $x$, and finally $\delta_1$ and $\delta_2$ represent the total contribution of the unobserved variables. The values of $\delta_1$ and $\delta_2$ for all respondents (or cases) are considered to be drawn from a normal distribution with zero mean and variances $\gamma_1^2$ en $\gamma_2^2$. The above equations therefore can also be written as:

$$\text{logit}\,(\mu_t) = x\beta + \gamma_1 z$$

$$\text{logit}\,(1-\lambda_t) = x\beta^* + \gamma_2 z\,,$$

with $z \sim N(0,1)$ being the standardized contribution of the unobserved variables and $\gamma_1$ and $\gamma_2$ the parameters associated with the ‘predictor’ $z$. Since the $z$ values for all cases are unknown the parameters $\beta$, $\beta^*$, $\hat{\gamma}_1$ en $\hat{\gamma}_2$ cannot be estimated. However, given a set of parameter values and the value of $z$, it is of course easy to determine the log likelihood contribution $\ell\ell$ of that case. Also, for a given set of parameter values, the expected (or marginal) log likelihood contribution $E(\ell\ell)$ of a case can be determined, where the expectation is taken over all possible values of $z$ taken from $N(0,1)$. For a case of e.g. the cross-section at $t = 2$ it holds that:

$$E(\ell\ell) = \int_{-\infty}^{\infty} [\quad p_1(1-\lambda_2) + (1-p_1)\,\mu_2 \quad]\; f(z)\; dz \;\text{ if } y_2 = 1,\text{ and}$$

$$E(\ell\ell) = \int_{-\infty}^{\infty} [\quad (1-p_1)(1-\mu_2) + p_1\,\lambda_2 \quad]\, f(z)\; dz \;\text{ if } y_2 = 0$$

Here, $\mu_2$ and $\lambda_2$ are defined as above (i.e., including $z$), $p_1$ is defined as usual (i.e., $p_1 = \mu_1$) without $z$ (in *CrossMark*, controlling for unobserved variables is only possibly for the transitions probabilities at $t \geq 2$.), and $f(z)$ is the height of the standard normal pdf at $z$. The integrals cannot be derived analytically, but are approximated by CrossMark using Gaussian quadrature with 20 mass points. Utilizing the $E(\ell\ell)$ values of all cases of all cross-sections it is possible to estimate those values $\hat{\beta}$, $\hat{\beta}*$, $\hat{\gamma}_1$ en $\hat{\gamma}_2$ that, averaged over all values that $z$ can take, have the highest expected (or marginal) log likelihood. The criterion to maximize in this estimation is the sum of the $E(\ell\ell)$ values of all cases of all cross-sections. The resulting estimates $\hat{\beta}$ en $\hat{\beta}*$ can be interpreted as the effects of the predictors $x$, corrected for the average influence of the unobserved variables. Using the above equations and estimation procedure has consequences for the standard errors of $\hat{\beta}$ and $\hat{\beta}*$, which can be quite different from the ones estimated without taking into account unobserved heterogeneity. The values of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are the estimates of the standard errors of $\delta_1$ and $\delta_2$ respectively, i.e., of the contributions of the unobserved variables to the logits of the entry and exit transition probabilities.

## 4.1 Testing the hypothesis $H_0 : \gamma_1 = \gamma_2 = 0$

To test this hypothesis we may use a test-procedure described by Snijders and Bosker (1999). We first calculate the value of $A = -2 \cdot \text{loglikelihood}$ for the model including $\gamma_1 z$ and $\gamma_2 z$. Then we compute $B = -2 \cdot \text{loglikelihood}$ for the model without $\gamma_1 z$ and $\gamma_2 z$ and obtain the difference $D = B - A$. Finally we test the difference $D$ to be significant using a $\chi^2$ distribution with 2 degrees of freedom, but halve the right tail probability associated with the value of $D$.

The standard estimation procedure in *CrossMark* does not take into account the possible influence of unobserved heterogeneity. If we wish to perform an analysis as described above, including the $\gamma_1 z$ and $\gamma_2 z$ terms in the equations for the transition probabilities, we have to go the Estimation Menu and click on the option called **Extra Bernoulli variance**. After running the model we will find the estimates $\hat{\gamma}_1$ en $\hat{\gamma}_2$ in the Output window.

# 5 Metropolis sampling

In the **Estimation window** an MCMC procedure can be performed that uses pure Meropolis sampling. To do so, check the option **Metropolis sampling** and specify a filename after **Outputfile posterior parameter values** for the file that the sampled parameter-points are written to. We only implemented this option in a very basic sense. There is e.g. no prior distribution that can be specified for the parameters: the implicit prior used for all parameters is the uniform distribution. After **Length of chain** specify the number of samples that has to be drawn from the posterior distributions of all parameters. Note that no burn-in period can be provided and, hence, the length of the chain must be large enough to also contain the desired burn-in period.

After pushing button **OK** *CrossMark* first performs the usual maximum likelihood (ML) estimation process. Once this is finished, the metropolis sampler is started. Consequently, metropolis sampling begins by default at the ML point. To start metropolis sampling from any other parameter-point, specify the parameter values for this point as the starting values to be used and also set the maximum number of iterations to 0.

It is possible to let *CrossMark*, for each sampled parameter-point, calculate the mean values of $p_{it}$, $\mu_{it}$ and $\lambda_{it}$ over all cases $i$ for each timepoint $t$. To this end, a filename must be entered after **Outputfile posterior mean p, mu, lambda**.

The value to be entered on the Estimation window in the sentence

**Covariance matrix of the jumping distribution equals ...**
**times estimated covariance matrix of parameters**

refers to what is discussed by Gelman, Stern and Rubin in 'Bayesian data analysis', 1995, on page 334 at the bottom where *c* = 2.4/sqrt(d). Value 2.4 for *c* is the default *CrossMark* uses if you don't specify another value in the above sentence. After the metropolis sampler is finished, inspection of the chain of sampled parameter-points (in the file specified after Output posterior parameter values) is always recommended, to make sure that the chain has changed fast enough. If the same parameter-points are resampled many times, a smaller value for *c* is probably more appropriate.

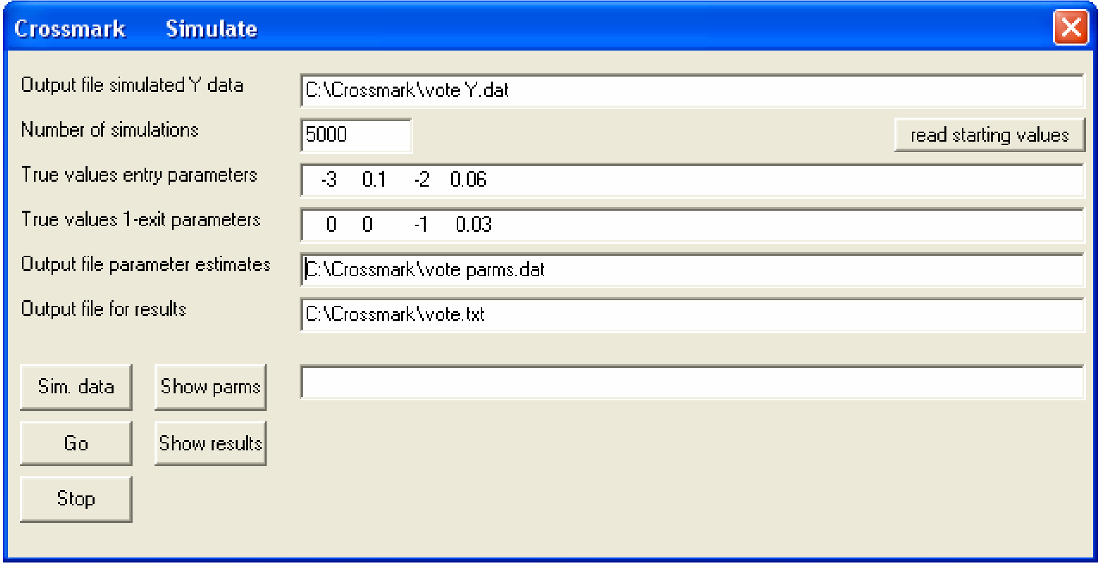
The parameter values that are sampled by the metropolis algorithm, are written out (to the file specified after **Outputfile posterior parameter values**) in the following format: sequence number (1,2,3,4,..., 100000, or more, depending on the length of the chain that was entered) followed by the values of the parameters of all predictors on the entry probabilities, followed by those on the 1-exit probabilities, followed finally by the loglikelihood value associated with these parameter values.

To evaluate the output files with posterior parameter values and/or means of $p_{it}$, $\mu_{it}$ and $\lambda_{it}$, other statistical software must be used. *CrossMark* itself does not perform any chain-evaluation, produces no histogram's of posterior parameter estimates and/or means, nor calculates means or standard deviations of the samples that were taken from the posterior distributions.

# 6 Parametric bootstrap

The **Simulate** button on the **Main Menu** opens the **Simulate** window where a parametric bootstrapping procedure can be performed. This window is shown in Figure 4. In the first step of the parametric bootstrap procedure a number of $Y$ datasets are simulated, based on the observed $X$ values in the data file and a set of true parameter values that must be specified after **True values entry parameters** and **True values 1-exit parameters.** The number $Y$ datasets that have to be simulated is specified after **Number of simulations**. A name is generated automatically (but can be modified) for the output file that will contain the simulated Y data. After clicking the button **Sim. data** the simulation process starts, during which the $Y$ data are generated and written to the file specified. Once the simulation has been finished, the next step can be started, during which the parameters will be estimated for samples that were simulated in the

Figure 4  Simulate window



| | |
|---|---|
| Output file simulated Y data | C:\Crossmark\vote Y.dat |
| Number of simulations | 5000 |
| True values entry parameters | -3  0.1  -2  0.06 |
| True values 1-exit parameters | 0  0  -1  0.03 |
| Output file parameter estimates | C:\Crossmark\vote parms.dat |
| Output file for results | C:\Crossmark\vote.txt |

previous step. Estimation is started by pushing button **Go**. The estimated parameter values for all simulated $Y$ datasets are written to the file specified after **Output file parameter estimates** in the following format: the sample number, the parameter values of all predictors for the entry probability, the parameter values of all predictors for the 1-exit probability, and, finally, the value of the loglikelihood. The file specified after **Output file for results** contains the final results, for all simulated $Y$ datasets, similar to the ones that are generally shown in the Output window. As with the metropolis sampler, here again one will have to evaluate the estimated parameters with other statistical software. The two buttons **Show parms** and **Show results** show the corresponding files in Wordpad.

# Samenvatting
# Summary in Dutch

Dit boek behandelt een discrete tijd, 1e orde Markov model. Het model onderscheidt zich van andere Markov modellen door het feit dat het toegesneden is op de analyse van individuele cross-sectionele gegevens, verzameld op een aantal opeenvolgende tijdstippen. Uitgangspunt voor het model vormt het gegeven dat voor het toepassen van dynamische c.q. transitie modellen het geen conditio sine qua non is om te beschikken over dynamische data, zoals repeated measures of individuele panel data. Hoewel bij herhaalde cross secties ieder individu slechst éénmaal wordt geobserveerd, handelt het gepresenteerde model over de kans dat een individu door de tijd heen verandert van de éne naar de andere toestand van een afhankelijke variabele.

De kern van het Markov model voor herhaalde cross secties (RCS) bestaat uit de zogeheten '1e orde Markov vergelijking' die voor een binaire variabele $\Upsilon$ als volgt luidt: $p_{i,t} = p_{i,t-1}(1-\lambda_{i,t}) + (1-p_{i,t-1})\mu_{i,t}$. De vergelijking relateert de statuskans $p_{i,t}$ dat case $i$ op tijdstip $t$ in staat 1 van $\Upsilon$ is aan de kans $p_{i,t-1}$ om op het vorige tijdstip $t-1$ eveneens in staat 1 te zijn geweest, en voorts aan de beide transitiekansen $\lambda_{i,t}$ en $\mu_{i,t}$. Deze transitiekansen zijn conditionele kansen: $\lambda_{i,t}$ drukt de kans uit dat case $i$ zich op tijdstip $t$ in staat 0 bevindt, gegeven dat case $i$ zich op het vorige tijdstip in staat 1 bevond; $\mu_{i,t}$ geeft de kans weer dat case $i$ zich op tijdstip $t$ in staat 1 bevindt, gegeven dat case $i$ zich op het vorige tijdstip in staat 0 bevond.

Het RCS Markov model hanteert predictoren $X$ om heterogeniteit in statuskansen en transitiekansen toe te staan. Daarbij wordt de logit functie toegepast om de kansen $p_{i,1}$ (zijnde de kans om op het éérste tijdstip van de onderzochte periode in toestand 1 van $\Upsilon$ te zijn) $\lambda_{i,t}$ en $\mu_{i,t}$ te koppelen aan de waarden van de predictoren $X$. Wanneer de waarde van een predictor voor case $i$ door de tijd verandert, resulteert er een Markov model waarvan de transitiekansen niet alleen tussen cases maar ook binnen cases door de tijd heen kunnen veranderen.

In de hoofdstukken 1 tot en met 6 worden de mogelijkheden van het RCS Markov model gedemonstreerd aan de hand van toepassingen op concrete onderzoeksgegevens. In deze toepassingen wordt, behalve van cross-sectionele data, ook gebruik gemaakt van individuele panel data die echter worden behandeld als betrof het cross-sectionele data. Dat maakt het mogelijk om de op basis van het model verwachte transities te vergelijken met de geobserveerde transities in de panel data. Nu volgt een samenvatting van de afzonderlijke hoofstukken.

Hoofdstuk 1 geeft een introductie in relevante begrippen en bespreekt een aantal Markov modellen die de afgelopen 50 jaar binnen de sociale wetenschappen zijn ontwikkeld en toegepast. Het RCS model wordt o.a. gecontrasteerd met het dynamische panel model dat individuele panel data hanteert om transitie kansen door de tijd te schatten. Ook wordt een vergelijking gemaakt met modellen voor geaggregeerde proporties. Daarmee kunnen, gegeven louter de relatieve $\Upsilon$ frequenties op meerdere tijdstippen van een aantal macro-eenheden (regio's, gemeenten, kiesdistricten), de niet geobserveerde transities van de micro-eenheden worden geschat. Het RCS model wordt geïntroduceerd in een variant zonder en een mèt gebruikmaking van predictoren. Tenslotte volgt een toepassing op cross-sectionele data van vijf surveys 'Sociale en Culturele Ontwikkelingen in Nederland' (SOCON) uitgevoerd in 1979, 1985, 1990, 1995 and 2000. De afhankelijke variabele betreft de houding van Nederlanders ten opzichte van abortus.

In hoofdstuk 2 wordt toegelicht hoe met behulp van het maximum likelihood (ML) criterium de effectparameters van de predictoren $X$ op de kansen $p_{i,1}$ , $\lambda_{i,t}$ en $\mu_{i,t}$ geschat kunnen worden. Het daartoe gehanteerde Fisher scoring algorithme leidt niet alleen tot puntschattingen van de betreffende parameters maar ook tot schattingen van de standaardfouten. De assumptie dat parameters voor elk tijdstip dezelfde waarde hebben kan worden afgezwakt door parameterwaarden te laten variëren als een polynome functie van de tijd. Het hoofdstuk sluit af met een toepassing op cross secties betreffende arbeidsparticipatie van vrouwen in Nederland en het voormalige West-Duitsland gedurende de periode 1987-1996 in Nederland en 1989-1994 in West-Duitsland.

In hoofdstuk 3 gaat de toepassing eveneens over arbeidsparticipatie van vrouwen, en wel van Nederlandse vrouwen gedurende de periode 1986-1995. De gebruikte data zijn individuele panel data. De panel transities worden vergeleken met de voorspelde transities op basis van het RCS Markov model.

Hoofdstuk 4 beschrijft enkele uitbreidingen van het basismodel. De eerste uitbreiding betreft het opnemen van tijd-variërende predictoren waarvan de waarden bekend zijn voor slechts een beperkt aantal tijdstippen voorafgaande aan de observatie. Dit type predictoren, zoals het inkomen van een respondent, konden in het door Moffitt gepresenteerde basismodel niet worden opgenomen, hetgeen een ernstige beperking vormde. De tweede uitbreiding betreft een modelvariant die rekening houdt met de aanwezigheid van niet geobserveerde heterogeniteit in de data. Het gevolg daarvan kan zijn dat de ware verdeling van de afhankelijke variabele niet de binomiale is waarop de ML schatting van de modelparameters is gebaseerd. Voorts wordt in dit hoofdstuk het RCS model vergeleken met zogeheten Ecologische Inferentie methodes. De getoonde toepassing is gebaseerd op individuele panel data en handelt over stemintentie in de aanloop naar de presidentsverkiezingen van 1976 in de Verenigde Staten. De uitkomsten van een dynamisch panel model worden vergeleken met die van het RCS model.

Hoofdstuk 5 gaat in op de kwaliteit van de ML schattingen van het RCS Markov model. De parameterschattingen en standaardfouten van een toepassing van het model op panel data worden vergeleken met de overeenkomstige grootheden resulterende uit een Markov Chain Monte Carlo procedure, een parametric bootstrapping procedure en 'cross-sectional subsampling'. Bij deze laatste procedure worden, voor elk tijdstip, uit de panel data at random respondenten geselecteerd en wel zodanig, dat voor elk tijdstip andere respondenten worden geselecteerd. Zo ontstaat een set van evenzovele onafhankelijke cross secties als er panel tijdstippen zijn. Het RCS Markov model wordt toegepast op 5000 aldus verkregen sets van cross secties. De toepassing in dit hoofdstuk heeft betrekking op de ontwikkeling van het personal-computer bezit in Nederlandse huishoudens over de periode 1986-1998.

Hoofdstuk 6 laat aan de hand van een simpel data voorbeeld zien dat de likelihood functie van het RCS Markov model niet altijd ééntoppig is. Met behulp van een Bayesiaanse benadering worden de twee modi van de posterior verdeling van de parameters gevisualiseerd. Voorts wordt gedemonstreerd hoe, in een Bayesiaanse analyse, individuele panel data en cross sectionele data gecombineerd kunnen worden om de onbekende transitiekansen te schatten.

Hoofdstuk 7 geeft naast een samenvatting tevens een vooruitblik op toekomstig werk. Dat laatste betreft onder meer het ontwikkelen van identificatie-regels aan de hand waarvan men vooraf kan uitmaken of een bepaald model wellicht overgespecificeerd is en daarom niet tot unieke parameterschattingen zal leiden. Voorts is onderzoek nodig naar alternatieve computer algorithmen die meer toegesneden zijn op multimodale likelihood functies dan het tot nu toe gebruikte Fisher scoring algorithme. Een zaak die eveneens aandacht verdient betreft de vraag naar de hoeveelheid informatieverlies (of winst) van het werken met herhaalde cross secties in vergelijking met individuele panel data. Een verder aandachtspunt betreft de ontwikkeling van een alternatieve methode, dan die welke beschreven werd in hoofdstuk 4, om niet geobserveerde heterogeniteit te modelleren. Tenslotte is onderzoek nodig naar twee voor de hand liggende model uitbreidingen: $\Upsilon$ variabelen met drie (of meer) statussen en een 2e orde Markov model.

De Appendix bevat een handleiding van het standalone computerprogramma *CrossMark* waarin het RCS Markov model is geïmplementeerd.

# Curriculum Vitae

Ben Pelzer was born in Heerlen, the Netherlands on July 22, 1951. From 1980 to 2002 he worked as a statistical analist at the Radboud University Nijmegen. He studied Social Sciences at this university and received his Master's degree in 1990. Since 2002 he works at the Department of Social Science Research Methodology of the Radboud University Nijmegen where he conducted this dissertation and gave several courses in quantitative research methods. Currently, he works as an assistant professor at the Department of Social Science Research Methodology at the Radboud University Nijmegen.