# Inferring Transition Probabilities from Repeated Cross Sections

**Ben Pelzer**
*Research Technical Department, University of Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands*
*e-mail: b.pelzer@maw.kun.nl*

**Rob Eisinga**
*Department of Social Science Research Methods, University of Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands*
*e-mail: r.eisinga@maw.kun.nl*

**Philip Hans Franses**
*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands*
*e-mail: franses@few.eur.nl*

This paper 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.

*Authors' note:* We would like to thank sincerely the anonymous reviewers for their most helpful comments on an early version of this paper. 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 paper 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.

## 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 cross-sectional 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 time-constant 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 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.[1]

[1] It is assumed in this paper that the responses are observed at evenly 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 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.

## 2 Estimating Transition Probabilities with RCS Data

### 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 $2 \times 2$ 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 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 $p_{it} = \mu/(\mu + \lambda)$.[2] 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 in Eq. (1), it is straightforward to show that the reduced form for $p_{it}$ is

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

[2]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 $p_{it} = \mu(1 + \eta + \cdots + \eta^{t-1}) + \eta^t p_{i0} = \mu\,(1 + \sum_{\tau=1}^{t-1} \eta^{t-\tau}) + \eta^t p_{i0} = (\mu/(\mu + \lambda))(1 - \eta^t) + \eta^t p_{i0}$. As $t \to \infty$, 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, pp. 261–262; Ross 1993, pp. 152–153).

where $\eta_{is} = 1 - \lambda_{is} - \mu_{is}$, assuming either $p_{i0} = 0$ or $t \to \infty$.[3] By explicitly allowing for time dependence and unit heterogeneity, this dynamic version of Eq. (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 Eq. (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 Eq. (2). The hazards themselves are specified as $\mu_{it} = F(\mathbf{x}_{it}\boldsymbol{\beta})$ and $\lambda_{it} = 1 - F(\mathbf{x}_{it}\boldsymbol{\beta}^*)$, where $F$—in the current paper—is the logistic link function [Moffitt (1993) uses the probit]. Hence, it is assumed that

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

where $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ 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 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 $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ can be obtained by maximization of the log-likelihood function

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

with respect to the parameters, where $T$ is the number of cross sections and $n_t$ the number of units of cross section $t$.[4] As Moffitt (1993) notes, obtaining $p_{it}$ by means of Eq. (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

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

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

binary recurrent events the contribution can be written as

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

(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

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

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 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).

## 2.2 *Modifications and Extensions of the Model*

### 2.2.1 Infinite Time Horizon and Initial Condition

The Markov model presented in Eq. (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 $P(y_{i1} = 1) = F(\mathbf{x}_{it}\boldsymbol{\delta})$, where $\boldsymbol{\delta}$ is a set of parameters to be estimated and $F$ is the logistic link function. The $\boldsymbol{\delta}$ parameters for the first observed outcomes at $t = 1$ are estimated simultaneously with the entry and exit parameters of interest at $t = 2, \ldots, T$. 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 $t = 1$.

### 2.2.2 ML Estimation

Maximum likelihood estimation requires the (analytic or numerical) derivatives of the log-likelihood 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 $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ are

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

where $\partial \mu_t / \partial \boldsymbol{\beta} = x_t \mu_t (1 - \mu_t)$ and $\partial \lambda_t / \partial \boldsymbol{\beta}^* = -x_t \lambda_t (1 - \lambda_t)$. 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).

### 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 $\boldsymbol{\zeta}$ as the associated parameter vector. One can then write

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

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 $\boldsymbol{\beta}^{**} = \boldsymbol{\beta}$. 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.

### 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.

### 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 $(\mu^*_{it}) = \mathbf{x}_{it}\boldsymbol{\beta} + \gamma_0\varepsilon_i$ and logit $(1 - \lambda^*_{it}) = \mathbf{x}_{it}\boldsymbol{\beta}^* + \gamma_1\varepsilon_i$, 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., logit $(\mu^*_{it}) \sim N(\mathbf{x}_{it}\boldsymbol{\beta}, \gamma_0^2)$. 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 $f(\varepsilon)$ is 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., $\int_{-\infty}^{\infty} f(z)\, e^{-z^2} dz \approx \sum_{j=1}^{q} w_j f(z_j)$, where $z_j$ are the nodes of the quadrature formula and $w_j$ the associated weights. The integrated log-likelihood then becomes

$$LL = \sum_{t=1}^{T}\sum_{i=1}^{n_t}\sum_{j=1}^{q}\pi^{-\frac{1}{2}} w_j\left[y_{it}\log\left(p^{j^*}_{it}\right) + (1 - y_{it})\log\left(1 - p^{j^*}_{it}\right)\right], \tag{10}$$

where $p^{j^*}_{it} = \mu^{j^*}_{it} + \sum_{\tau=1}^{t-1}\mu^{j^*}_{i\tau}\prod_{s=\tau+1}^{t}(1 - \lambda^{j^*}_{is} - \mu^{j^*}_{is})$, logit$(\mu^{j^*}_{it}) = \mathbf{x}_{it}\boldsymbol{\beta} + \gamma_0 z_j\sqrt{2}$, logit $(1 - \lambda^{j^*}_{it}) = \mathbf{x}_{it}\boldsymbol{\beta}^* + \gamma_1 z_j\sqrt{2}$, $w_j$ 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 $\pi^{-\frac{1}{2}} w_j$ 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.

## 2.3 *Related Lines of Research*

### 2.3.1 Shrinking Logical Bounds

The partition Eq. (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 Eq. (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

$$\begin{aligned} L\mu_{it} &= \max\left(0, \frac{p_{it} - p_{it-1}}{1 - p_{it-1}}\right) \leq \mu_{it} \leq \min\left(\frac{p_{it}}{1 - p_{it-1}}, 1\right) = U\mu_{it}, \\ L\kappa_{it} &= \max\left(0, \frac{p_{it} - (1 - p_{it-1})}{p_{it-1}}\right) \leq \kappa_{it} \leq \min\left(\frac{p_{it}}{p_{it-1}}, 1\right) = U\kappa_{it} \end{aligned} \tag{11}$$

(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 Eq. (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 $p_{it} = \mu_{it}$.

### 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 $2 \times 2$ 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.

## 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).[5] 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, i.e., to study life cycle history issues (see Moffitt 1990). Our program *CrossMark* may be used to do the computations.[6]

## 3 Application

### 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 $\text{cov}(y_t, y_{t-1})$ 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 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, 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.

[5] 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.

[6] 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 the Windows 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 1** Marginal fraction of Democratic vote intention and observed entry and exit transition rates

| *Year . month* | $n_t$ | *Inflow* | *Outflow* | $\bar{y}_t$ | $\bar{y}_t \mid y_{t-1}=0$ | $\bar{y}_t \mid y_{t-1}=1$ |
|---|---|---|---|---|---|---|
| 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 |

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."

## 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 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 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 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 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 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

**Table 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 (0.03) | −1.36 (0.04) | |
| Voted McGovern in 1972 | 1.30 (0.03) | 1.58 (0.11) | −0.56 (0.28) |
| Black | 0.96 (0.07) | | −2.29 (0.38) |
| Education | −0.29 (0.00) | −0.23 (0.01) | |
| Age | −0.01 (0.00) | −0.08 (0.00) | −0.10 (0.02) |
| Female | | | 0.73 (0.21) |
| Constant | 0.82 (0.09) | 3.47 (0.45) | 2.67 (0.65) |
| 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 $\lambda_t$.

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 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 2 and 3 that the enlarged model provides a much better fit. The second column in Table 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 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 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

**Table 3** Markov repeated cross-section estimates of 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 as Democrat | 1.87 (0.19) | 2.59 (0.54) | 2, 3 | −3.15 (0.79) | 3 |
| | | 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 | | | | |

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

$\kappa_{i5}$. 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 Fig. 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.

### 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

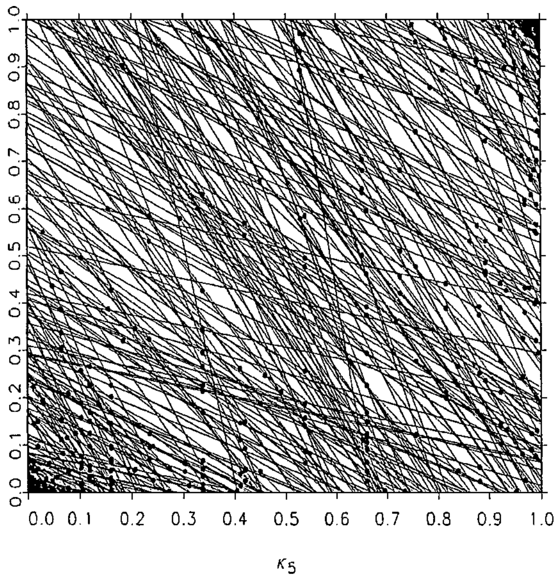


**Fig. 1** Tomography lines (691) for current entry and 1 − exit transitions at sample period $t = 5$.

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 $P(y_{it} = 1 \mid y_{it-1} = 0, 1) = \mathbf{x}_{it}\boldsymbol{\beta} + y_{it-1}\mathbf{x}_{it}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = \boldsymbol{\beta}^* - \boldsymbol{\beta}$. 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 $\mathbf{x}_{it}$. Because $y_{t-1}$ is missing for some respondents, the estimates of the two models reported in Table 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 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 5 indicates that the MSEs are remarkably low and that over time they gradually lean to the 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.[7]

[7]The estimated expected frequencies were computed as follows. With $T$ sample periods, we have $\sum_{t=1}^{T} 2^t$ different (0,1) sequences (which in the present application equals 62) ranging in length from 1 (e.g., "0")

**Table 4** Markov repeated cross-section (RCS) and standard dynamic panel 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 as Democrat | 2.94 (0.43) | 1.75 (0.21) | 2, 3 | −2.43 (0.71) | −1.78 (0.37) | 3 |
| | 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 feelings | −0.26 (0.10) | −0.20 (0.08) | 2 | 2.91 (0.88) | 1.17 (0.23) | 5 |
| | −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 feelings | 0.26 (0.09) | 0.46 (0.08) | 2, 3 | −0.75 (0.18) | −0.43 (0.09) | 3, 4 |
| | 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[a] | | −885.97 | | | | |
| Log likelihood panel | | −798.66 | | | | |

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

[a]The log likelihood of the RCS Markov model is obtained after excluding the contribution of the 856 observations at $t = 1$ of −381.44.

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

to $T$ (e.g., "11111") . We define the probability of a sequence of length $t$ for observation $i$ of cross section $t$ as $\tilde{p}_i(\tilde{y}_1, \ldots, \tilde{y}_t) = P(y_{i1} = \tilde{y}_1 \cap \cdots \cap y_{it} = \tilde{y}_t)$, 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 $\tilde{p}_i(\tilde{y}_1, \ldots, \tilde{y}_t) = \tilde{p}_i(\tilde{y}_1) \prod_{\tau=2}^{t} (p_{00} + p_{01} + p_{10} + p_{11})$, 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 $\tilde{f}(\tilde{y}_1, \ldots, \tilde{y}_t) = \sum_{i=1}^{n_t} \tilde{p}_i(\tilde{y}_1, \ldots, \tilde{y}_t)$.

**Table 5** Mean squared errors

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

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

**Table 6** Frequencies of observed (*Obs*) and estimated expected (Exp) (non-)Democratic vote intention sequences

| *Sequence*[a] | *Obs* | *Exp* | Δ | *Sequence*[a] | *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]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.

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 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 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 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 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 $2 \times 2$ 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. "A Markovian 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.