

Inferring Transition Probabilities from Repeated Cross Sections: A Cross-level Inference Approach to US Presidential Voting

Ben Pelzer

*Research Technical Department, University of Nijmegen,
P.O. Box 9104, 6500 HE Nijmegen, The Netherlands
email: 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
email: r.eisinga@maw.kun.nl*

Philip Hans Franses

*Econometric Institute, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
email: franses@few.eur.nl*

Econometric Institute EI 2001-21

Abstract

This paper outlines 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 (1993) and shares features with standard statistical methods for ecological inference. We show how ML estimates of the parameters can be obtained by the method-of-scoring, how to estimate time-varying covariate effects, and how to include non-backcastable variables in the model. The latter extension of the basic model is an important one as it strongly increases its potential application in a wide array of research contexts. The example illustration uses survey data on American presidential vote intentions from a five-wave panel study conducted by Patterson (1980) in 1976. We treat the panel data as independent cross sections and compare the estimates of the Markov model with the observations in the panel. Directions for future work are discussed.

Authors' note: 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 bear any responsibility for the analysis or interpretation presented in this paper. The program *CrossMark* to do the ML estimation reported here is, although not completed documented yet, available upon request.

1 Introduction

Surveys that trace the same units across occasions provide the most powerful sorts of data for dynamic electoral analysis. However, on many political issues repeated observations are simply unavailable and those panel data sets that do exist are typically of limited time coverage. This shortcoming combined with potential drawbacks like nonrandom attrition and conditioning limits 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 may provide a viable alternative. There exists an abundance of high-quality RCS data and many are available for relatively long time periods. Given the importance of dynamics in electoral studies and the paucity of panel data, it would be of great advantage if such data could 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 discuss 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 (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 hazards. Previous brief discussions of specific versions of the model include Mebane and Wand (1997) and Pelzer, Eisinga, and Franses (2001).

The following section presents the basic Markov model for RCS data along with parameter estimation and various extensions of Moffitt's approach. 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 the predictions of the Markov model for RCS data with the actual transitions in the panel. The calibration results suggest that the model can provide a useful tool for inferring individual-level transition probability estimates in the absence of transition data in cross-sectional samples. A discussion of intended extensions of the model concludes the paper.

2 Estimating transition probabilities with RCS data

It is assumed in the sequel that the population is closed with respect to in- and out-migration, that the responses are observed at evenly spaced discrete time intervals

$t = 1, 2, \dots$, 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 . To simplify notation, this paper uses the symbol it to also index individuals in RCS samples.

Suppose we have a two-state Markov matrix of transition rates in which the cell probabilities sum to unity across rows. For this 2x2 table, we define the following three terms, where Y_{it} denotes the value of the binary random variable Y for unit i at time point t :

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

These marginal and conditional probabilities respectively give rise to the well-known flow equation

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

where $\eta_{it} = 1 - \lambda_{it} - \mu_{it}$. This accounting identity 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 (μ_{it}) and exit (λ_{it}) transition probabilities. Clearly, the major difficulty with using RCS data for dynamic analysis is that the surveys are ‘incomplete’ in the sense that they do not assess directly the state-to-state transitions over time for each individual unit. In RCS data one only observes at each of a number of occasions a different sample of units and their current states, that is y_{it} is observed but y_{it-1} is not. This information gap implies that some identifying restrictions over i and/or t must be imposed to estimate the unobserved transitions.

A rather restrictive approach frequently applied in the statistical literature is to a priori assume that the transition probabilities are time-stationary 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 outcome of p_{it} is $p_{it} = \mu/(\mu + \lambda)$ as t goes to infinity. Some early references relating to this steady-state model include those that estimate transition rates from aggregate frequency data (Kalbfleish and Lawless 1984 1985, Lawless and McLeish 1984, Lee, Judge, and Zellner 1970). The formulation has also been applied in various economic

studies (Topel 1983, McCall 1971), in the famous mover-stayer model of intra-generational job mobility (Bartholomew 1996, Goodman 1961), and in electoral studies on voter transitions (Firth 1982). However, the assumption that individual differences in transitions are not present in the population lacks plausibility in many applications. Consequently, as noted by Hawkins and Han (2000), studies that assume a time-homogeneous Markov evolution with a common transition probability matrix have found their estimates to be extremely inefficient.

A flexible approach that facilitates a more accurate representation of the transition probabilities without imposing some presumed structure is provided by the reduced-form dynamic version of eqn. (1). If we let the initial probability $p_{i0} = 0$ (or $t \rightarrow \infty$), it is straightforward to show that the reduced form for p_{it} is

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

where $\eta_{is} = 1 - \lambda_{is} - \mu_{is}$. By explicitly allowing for time-dependence and unit-heterogeneity, this reduced-form dynamic model is better suited to yield an informative representation of the transition probability estimates. It will therefore be maintained in the ensuing approach.

The framework Moffitt (1993) proposed to estimate eqn. (2) is based on a simple observation. While RCS data lack direct information on transitions in opinions, preferences, choices and other individual behaviors, they often do provide a set of time-invariant and time-varying covariates X_{it} that affect the hazards. If so, the history of the covariates (i.e., $X_{it}, X_{it-1}, \dots, X_{i1}$) 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 key idea is to model the current and past μ_{it} and λ_{it} in a regression setting as functions of current and backcasted values of time-invariant and time-varying covariates X_{it} . Parameter estimates for the covariates are obtained by substituting the hazard functions into eqn. (2).

The hazard functions themselves are specified as $\mu_{it} = F(X_{it}\beta)$ and $\lambda_{it} = 1 - F(X_{it}\beta^*)$, where F - in the current paper - is the logistic link function. Hence, it is assumed that

$$\text{logit}(\mu_{it}) = X_{it}\beta \text{ and } \text{logit}(1 - \lambda_{it}) = X_{it}\beta^*,$$

where β and β^* are two potentially different sets of parameters associated with two potentially different sets of covariates X_{it} . This regression setup offers the opportunity to estimate transition probabilities that vary across both individuals and - if the model includes time-varying covariates - time periods. Maximum likelihood estimates of β and β^* can be obtained by maximization of the log likelihood function

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

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. As Moffitt (1993) notes, obtaining p_{it} by means of eqn. (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 by the i th case at time point t in panel data with the likelihood contribution in RCS data. For a first-order transition model of binary recurrent events the contribution can be written as

$$L_{it}(\beta, \beta^*) = \mu_{it}^{y_{it}(1-y_{it-1})} (1 - \lambda_{it})^{y_{it}y_{it-1}} (1 - \mu_{it})^{(1-y_{it})(1-y_{it-1})} \lambda_{it}^{(1-y_{it})y_{it-1}}$$

(e.g., Stott 1997). Hence, conditional on y_t and y_{t-1} , the likelihood contribution simplifies to a single transition probability estimate. For RCS data with a binary outcome, however, the contribution from the i th case is given by

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

In this formulation the likelihood contribution does not collapse to a single rate estimate but rather to a weighted sum of two hazards. Also note from this comparison that estimates of the parameters of the hazard functions in RCS data are likely to be less efficient than they would be in a comparable panel data set. To summarize the model a graphical presentation is given in Figure 1, omitting the subscript i for clarity.

Figure 1 about here

The marginal probability p_{it} depends on the set of all possible transition histories for each individual i up to time t . The unobserved transition probabilities in their turn are modeled as functions of current and backcasted values of time-invariant and time-varying covariates X_{it} . As Mebane and Wand (1997) point out, an important characteristic of the model is that the transition probabilities are estimated as a function of all the available cross-sectional samples rather than simply the observations from the current time period. This full information strategy expresses the notion that in RCS data different groups of individuals are observed over time, but individuals with similar covariate values are exchangeable in the sense their transition histories are assumed to be identical.

2.2 Extensions and modifications of the basic model

ML estimation. Moffitt (1993) offers no discussion of the computation of the maximum likelihood parameter estimates. A convenient optimization technique, implemented in our program *CrossMark*, is Fisher's method-of-scoring (Amemiya 1981). If we suppress the subscript i for the moment to avoid cumbersome notation and define $p_0 = 0$, the first order partial derivatives of LL with respect to the parameters β and β^* are easily established as

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

and

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

where $\partial \mu_t / \partial \beta = x_t \mu_t (1 - \mu_t)$ and $\partial \lambda_t / \partial \beta^* = -x_t \lambda_t (1 - \lambda_t)$. The ML estimators are the values of the parameters for which the efficient scores are zero, i.e., $\partial LL / \partial \beta = \partial LL / \partial \beta^* = 0$. Let θ denote the stacked column vectors β and β^* , then the method-of-scoring uses the iterative estimation algorithm $\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \varepsilon [\hat{I}(\hat{\theta}^{(k)})]^{-1} (\partial LL(\hat{\theta}^{(k)}) / \partial \theta)$. The parameter ε denotes an appropriate step length that scales the parameter increments and $\hat{I}(\hat{\theta}^{(k)})$ is an estimate of the Fisher information matrix $I(\theta) = -E[\partial^2 LL(\theta) / \partial \theta]$ evaluated at $\theta = \hat{\theta}^{(k)}$, where $\partial^2 LL(\theta) / \partial \theta$ is the Hessian. The method-of-scoring also provides, by design, an estimate of the asymptotic variance-covariance matrix of the model parameters, given by the inverse of the estimated Fisher information matrix evaluated at the values of the maximum likelihood estimates.

Non-backcastable 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, completed education, etcetera are candidates and time-specific aggregates measurable in the past may also enter the model. But variables like 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. Given current information, each age and time-invariant variables relevant for preceding years are known.

In many applications settings, however, we have time-dependent covariates that the basic model would omit because the past histories are unknown. To incorporate these ‘non-backcastable’ variables, we adopt a model with two different sets of parameters for both μ_{it} and λ_{it} , i.e., one for the current transition probability estimates and a separate one for the preceding ones. Define Z_{it} as a vector of non-backcastable variables with $Z_{it} = Z_{it}$ for cross section t and $Z_{it} = 0$ for the cross sections $t-1, \dots, 1$ and ζ as the associated parameter vector. One can then write

$$\text{logit}(\mu_{it}) = \begin{cases} X_{it}\beta^{**} + Z_{it}\zeta & \text{for } t \\ X_{it}\beta & \text{for } t-1, \dots, 1, \end{cases}$$

where $\beta^{**} = \beta + \beta^+$. A similar model with non-backcastable covariate effects on the exit rates may be specified for λ_{it} . This specification offers the opportunity to express the current transition probability estimates as a logistic function of both backcastable and non-backcastable variables. The expression obviously also affords a test – useful for efficiency gains - of the hypothesis $H_0 : \beta^+ = 0$, using the restriction $\beta^{**} = \beta$.

Time-varying covariate effects. Another potential drawback of the basic model is that it assumes that the effects of the covariates are fixed over time. This restriction may not be valid for long time periods and thus potentially biases the estimated effects. Of course, modifying continually the values of the parameters - so as to allow the model to adapt itself to ‘local’ conditions - produces problems of overparametrization. We aim to avoid such problems by assuming the parameters to be constant across a limited number of time periods. An alternative specification, not pursued in this paper, is to allow the regression coefficients to become polynomials in time using the expression $\beta_t = \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \dots + \gamma_d t^d$, where d is a positive integer specifying the degree of the polynomial. For this parametric setup, too, it will be desirable to have models with low degree polynomials that avoid nonexistence of unique ML estimates.

First observed outcome. Moffitt (1993) defines the first observed outcome of the process, $P(Y_{i1} = 1)$, to equal the transition probability μ_{i1} . However, in many applications it will be more plausible to take $P(Y_{i1} = 1)$ to equal the state probability p_{i1} . That is, one conveniently assumes that the Y_{i1} 's are random variables with a probability distribution $P(Y_{i1} = 1) = F(X_{it}\delta)$, where δ is a set of parameters to be estimated and F is the logistic link function. The δ -parameters for the first observed outcome at $t = 1$ are estimated simultaneously with the entry and exit parameters of interest at $t = 2, \dots, T$. Note once again 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 $t = 1$.

Unequal sample sizes. 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

$$LL^* = \sum_{t=1}^T \sum_{i=1}^{n_t} w_i [y_{it} \log(p_{it}) + (1 - y_{it}) \log(1 - p_{it})], \quad \text{where} \quad w_i = \bar{n} / n_t, \quad \text{with}$$

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

Shrinking logical bounds. The partition equation (1) implies the familiar restriction, customarily attributed to Duncan and Davis (1953),

$$\mu_{it} = \frac{p_{it}}{(1 - p_{it-1})} - \frac{p_{it-1}}{(1 - p_{it-1})} \kappa_{it} \quad \text{and} \quad \kappa_{it} = \frac{p_{it}}{p_{it-1}} - \frac{(1 - p_{it-1})}{p_{it-1}} \mu_{it},$$

where $\kappa_{it} = 1 - \lambda_{it}$. These identities were used by King (1997) to construct a so-called tomography plot. The axes of this plot represent the parameters κ_{it} and μ_{it} and the linear constraint on each individual i inherent in eqn. (1) is represented by a tomography line with intercept $p_{it} / (1 - p_{it-1})$ and slope $-p_{it} / (1 - p_{it-1})$ that goes through the point (κ_{it}, μ_{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 that $\mu_{it} \in (L\mu_{it}, U\mu_{it})$ 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

$$L\mu_{it} = \max\left(0, \frac{p_{it} - p_{it-1}}{1 - p_{it-1}}\right) \leq \mu_{it} \leq \min\left(\frac{p_{it}}{1 - p_{it-1}}, 1\right) = U\mu_{it}$$

and

$$L\kappa_{it} = \max\left(0, \frac{p_{it} - (1 - p_{it-1})}{p_{it-1}}\right) \leq \kappa_{it} \leq \min\left(\frac{p_{it}}{p_{it-1}}, 1\right) = U\kappa_{it}$$

(King 1997). Hence the estimated values of μ_{it} and κ_{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

$$U\kappa_{it} = \frac{p_{it}}{p_{it-1}} - \frac{(1-p_{it-1})}{p_{it-1}} L\mu_{it} \text{ and } L\kappa_{it} = \frac{p_{it}}{p_{it-1}} - \frac{(1-p_{it-1})}{p_{it-1}} U\mu_{it},$$

the tomography lines correspond to the main diagonal of the rectangular region defined by the lower and upper bounds. Because the estimates produced are restricted to lie on the diagonal they satisfy $\kappa_{it} = a_{it} - b_{it}\mu_{it}$, where $a_{it} = (U\mu_{it}U\kappa_{it} - L\mu_{it}L\kappa_{it})(U\mu_{it} - L\mu_{it})^{-1}$ and $b_{it} = (U\kappa_{it} - L\kappa_{it})(U\mu_{it} - L\mu_{it})^{-1}$ (Chambers and Steel 2001).

Our estimation procedure implicitly takes into account the bounds and thereby restricts the range of feasible estimates of μ_{it} and κ_{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 μ_{it} and κ_{it} would allow one to narrow the bounds beyond the logical limits. For example, in studies of US 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 eqn. (1) that $\eta_{it} > 0$, we would not recommend using this assumption universally. Also note that if the entry and 1-exit transitions are equal to each other (i.e., $\mu_{it} = \kappa_{it}$), identity (1) reduces to $p_{it} = \mu_{it}$.

Quantities of interest. 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 variate, 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 panel data the model is obviously designed for estimating transition probabilities from repeated cross sections covering relatively long-term periods. An example of when such a formulation is most relevant includes an analysis of the labor force participation decisions of Dutch women over the 1986-1995 period by Pelzer, Eisinga and Franses (2001).

3. Application

Our empirical illustration employs election-year panel data on US 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. Obviously, 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 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. Consequently, in this particular application the variance-covariance matrix of the first derivatives may not be a consistent estimator of the Hessian and hence the parameter standard errors. 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 about here](#)

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 transitions rates in the panel. The table shows that, despite substantial bimonthly turnover with values ranging from .138 to .248, almost half of the respondents continue to prefer the Democratic presidential candidate. The bottom part of the table presents the (non)participation patterns across the five waves of data and the number of sample members attriting from the panel. Because some nonrespondents from

one wave are recruited back into the sample at subsequent waves, both monotone and nonmonotone attrition patterns arise. It is important to note that the analysis includes both attritors and nonattritors.

Next to voting intention, the survey provides information on socio-demographic characteristics and attitudes towards presidential candidates. The analysis presented here uses only covariates 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 these covariates are assumed to be fixed over the surveys' duration. In addition to these time-constant variables the analysis includes several non-backcastable covariates. These include (i) whether the respondents identify themselves 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 towards 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 seven-point Likert-type scales, running from "strongly disagree" to "strongly agree", from "unfavorable" to "favorable", etcetera.

3.1 Model estimation

First a time-stationary Markov model with constant terms only was applied to the data. This model produced the parameters $\beta(\mu_{t>1}) = -.238$ and $\beta^*(\lambda_{t>1}) = .034$ and a corresponding maximum log likelihood value of $LL^* = -2643.56$. These estimates imply constant transition rates of $\mu = .44$ and $\lambda = .51$; hence implausibly high values that amply exceed the observed rates reported in Table 1. The model was thereupon extended to a nonstationary, heterogeneous Markov model (model 1) by including the backcastable covariates reported above. The results are shown in Table 2.

Table 2 about here

The parameters in the first column show the effects of the backcastable variables on the probability of a Democratic vote at $t = 1$, p_{i1} , estimated for all observations. 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 1972 election. The second column of Table 2 presents the effects of the backcastable 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, education, age, and a 1972 vote for Nixon negatively affect the entry decision. The third 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).

The right-hand side of the table (model 2) reports the regression estimates of a transition model that also includes the non-backcastable variables with unknown covariate history. 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 substantive arguments to anticipate that the effects of the non-backcastable 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 2 describes the data best 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 non-backcastable variables. But it is clear from the log likelihood values in Table 2 that the enlarged model provides a much better fit.

The columns pertaining to model 2 again show the estimated effects on the state probability p_{it} . 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 towards Ford and indifference towards the future president's leaning are negative. The last two columns of Table 2 provide the effects on the entry and exit rates respectively with respect to a Democratic vote. The columns labeled t indicate the time periods pertaining to the (time-varying) parameters. For example, favorable feelings towards Carter has an effect of .35 at $t = 2,3$ and an effect of 1.14 at $t = 4,5$. Most of the parameters are again well determined and consistent with those commonly reported in the literature. In short, positive attitudes towards the Republican (Democratic) candidate Ford (Carter) decrease

(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 tomography lines for one time period are singled out for discussion purposes. Figure 2 shows for all i at $t = 5$ the lines $\mu_{i5} = (p_{i5}/1 - p_{i4}) - (p_{i4}/1 - p_{i4})\kappa_{i5}$, where $\kappa_{i5} = 1 - \lambda_{i5}$.

Figure 2 about here

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 each line onto the horizontal (for κ_{i5}) and vertical (for μ_{i5}) axes. Note that while most of the point estimates are below the 45° line, for a substantial number of the estimates μ_{i5} exceeds κ_{i5} . In fact, almost 25% of the observations fail to conform to the restriction that $\kappa_{it} > \mu_{it}$. Hence incorporating the external substantive assumption that party loyalty rates are greater than defection rates would most likely lead to incorrect conclusions. Visual inspection of the figure also suggests a strong relationship between μ_{i5} and κ_{i5} , with low (high) entry rates corresponding with high (low) exit rates. Also note that most of the predictions tend to the basically ideal situation of either extremely high or extremely low transition probability estimates.

3.2 Model validation

To understand how well the model reproduces the panel observation we may examine its efficacy in a variety of ways. One is to assess the fit of the model in terms of prediction errors, using the panel data and various summary measures, i.e., the mean squared error (MSE), the mean value of minus log likelihood error (MML), and the mean probability of correct allocation (MCA). Details are given in Table 3.

Table 3 about here

The prediction error measures can be seen as analogues to the R-squared measure in OLS regression. The MSE and MML tend to zero if $\mu_{it}(\lambda_{it})$ tends to 0 or 1 and the smaller the

error rate, the better the model predicts. Table 3 indicates that the mean squared errors and the mean minus log likelihoods are remarkably low and gradually lean to the ideal situation of perfect separation between the $y_{it} = 0$ and $y_{it} = 1$ groups. The average probability of correct allocation also reveals that the ability of the model to recover the observed transitions is very good, ranging from a low of .736 to a high of .899. 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. The latter 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') to T (e.g., '11111'). We define the probability of a sequence of length t for each observation i of cross section t as $\tilde{p}_i(\mathcal{Y}_1, \dots, \mathcal{Y}_t) = P(Y_{i1} = \mathcal{Y}_1 \cap \dots \cap Y_{it} = \mathcal{Y}_t)$, where $\mathcal{Y}_1, \dots, \mathcal{Y}_t = 0, 1$. Hence

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

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

$$\tilde{p}_i(\mathcal{Y}_1, \dots, \mathcal{Y}_t) = \tilde{p}_i(\mathcal{Y}_1) \prod_{\tau=2}^t (p_{00} + p_{01} + p_{10} + p_{11}),$$

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

An initial examination of the frequencies is to compare the expected with the observed first-order transitions (i.e., y_{t-1}, y_t) over the time period of our data. Before embarking on the findings it is important to note that while model 2 predicts the current probabilities at time point t (i.e., p_{it} , μ_{it} and λ_{it}) very well, it does not in general

reproduce the past probabilities at $t-1$, $t-2$, etcetera equally well. The reason is that the past probabilities are predicted by the backcastable variables only and these 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 the two consecutive cross sections at $t-1$ and t - and subsequently compute the expected frequencies. Table 4 shows the relative frequencies of the observed and the estimated expected first-order voter transitions between parties.

Table 4 about here

As can be seen, both the observed and the predicted frequencies are concentrated in the continuous Democratic vote (11) and the continuous non-Democrat vote (00) categories. Also note that the partisan changes seem to decline over time leading up to the presidential election. Further, the discrepancies between the predicted and the observed frequencies are all relatively small and not significant at the .05 level. This implies that both loyal and defection categories are predicted well.

A final examination of the goodness-of-fit reported here is to compare the estimated expected and actually observed absolute frequencies of all 62 (0,1) voting sequences. They are tabulated in Table 5.

Table 5 about here

The longitudinal voting profiles indicate that most voters remain loyal to their initial preference and that 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 relatively extreme patterns of vote switching. Table 5 indicates quite clearly that for most sequences the estimated expected frequencies predicted by the RCS transition model match the observed frequencies in the panel data well. The only notable exceptions are the highly populated consecutive Democratic vote categories (i.e., the sequences of 1’s). However, even for these sequences model performance is quite good. Hence overall our findings illustrate that the model is well able to recover the actual transitions in the panel.

4. Conclusion

The Markov model for cross-level inference presented here can help us better understand binary transitions when it is either impossible or impractical to collect panel information on the exact sequences. Our example application shows that the model captures voters with very different entry and exit transitions probabilities. More important, it yields transition frequency estimates remarkably consistent with the observations in the panel. The results thus demonstrate that the proposed model can be used to accurately identify transition probabilities solely on the basis of repeated cross sections and hence to coax panel conclusions out of non-panel data.

Although the above model promises to be useful in different settings, there are some extensions that we are currently exploring that may further enhance its applicability. One next step is to allow for unobserved heterogeneity. The model specification assumes that individual heterogeneity is due to the observed variables. It is likely, however, that unobserved and possibly unobservable variables are also a source of heterogeneity. Ignoring this over-dispersion is unlikely to change point estimates in any radical way, but estimates of standard errors will be underestimated and tests will be in error. It is thus important to try to account for it. Another extension of interest is to use Bayesian methods, in the spirit of King, Rosen, and Tanner (1999) and Rosen, Jiang, King, and Tanner (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 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 present model to yield reliable inferences. One approach to study this small sample problem is to analyse the data by MCMC using Gibbs or Metropolis sampling.

Finally, it has frequently been argued that King's ecological inference solution can fruitfully be adapted to repeated cross sections (Penubarti and Schuessler 1998, King, Rosen, and Tanner 1999, Davies Withers 2001). Despite the steady development in ecological analysis in the direction of more sophisticated statistical modeling, little has been done to date on developing models that draw panel inference from non-panel data (Sigelman 1991 and Penubarti and Schuessler 1998 are notable exceptions). It is our believe that the approach presented here 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: Chicago University Press.
- Amemiya, Takeshi. 1981. "Qualitative Response Models: A Survey." *Journal of Econometric Literature* 19:1483-1536.
- Bartholomew, David J. 1996. *The Statistical Approach to Social Measurement*. San Diego: Academic Press.
- Chambers, R.L., and D.G. Steel. 2001. "Simple Methods for Ecological Inference in 2x2 Tables." *Journal of the Royal Statistical Society. Series A* 164:175-192.
- Davies Withers, Suzanne. 2001. "Quantitative Methods: Advancement in Ecological Inference." *Progress in Human Geography* 25:87-96.
- Duncan, Otis Dudley, and Beverly Davis. 1953. "An Alternative to Ecological Correlation." *American Sociological Review* 18:665-666.
- Firth, David. 1982. *Estimation of Voter Transition Matrices from Election Data*. M.Sc. thesis. London: Department of Mathematics, Imperial College London.
- Goodman, L.A. 1961. "Statistical Methods for the Mover-Stayer Model." *Journal of the American Statistical Association* 56:841-868.
- Hawkins, D.L., and C.P. Han. 2000. "Estimating Transition Probabilities from Aggregate Samples Plus Partial Transition Data." *Biometrics* 56:848-854.
- 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.
- Kay, Richard, and Sarah Little. 1986. "Assessing the Fit of the Logistic Model." *Applied Statistics* 35:16-30.
- 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. 1993. "Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-sections." *Journal of Econometrics* 59:99-123.
- Patterson, Thomas E. 1980. *The Mass Media Election: How Americans Choose Their President*. New York: Praeger.
- 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 LA.
- 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.
- Stott, David. 1997. "Sabre 3.0: Software for the Analysis of Binary Recurrent Events." <http://www.cas.lancs.ac.uk:80/software/> (June 2001) .
- 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.
- Topel, Robert H. 1983. "On Layoffs and Unemployment Insurance." *American Economic Review* 73:541-559.
- Van Houwelingen, J.C., and S. Le Cessie. 1990. "Predictive Value of Statistical Models." *Statistics in Medicine* 9:1303-1325.

Table 1: Marginal fraction of Democratic vote intention, observed entry and exit transition rates and panel attrition

year	month	n_t	inflow	outflow	\bar{y}_t	$\bar{y}_t y_{t-1} = 0$	$\bar{y}_t y_{t-1} = 1$
1976	02	856			.384		
	04	790	142	208	.460	.248	.178
	06	792	153	151	.471	.170	.176
	08	727	90	155	.465	.203	.229
	10	691	80	116	.457	.140	.138

panel attrition patterns and *number* of observations across waves*

11111	412	10111	26	01111	57	00111	56
11110	50	10110	6	01110	8	00110	22
11101	33	10101	2	01101	9	00101	8
11100	56	10100	13	01100	14	00100	20
11011	31	10011	10	01011	12	00011	7
11010	10	10010	7	01010	6	00010	7
11001	9	10001	6	01001	1	00001	12
11000	47	10000	138	01000	35		

*1=observed, 0=missing. The figures were obtained after listwise deletion (for each time point separately) of respondents who exhibit item nonresponse.

Table 2: Markov repeated cross-section estimates for transitions into and out of Democratic vote intention *

	<i>model 1</i>			<i>model 2</i>				
	$\delta(p_{t=1})$	$\beta(\mu_t)$	$-\beta^*(\lambda_t)$	$\delta(p_{t=1})$	$\beta(\mu_t)$	<i>t</i>	$-\beta^*(\lambda_t)$	<i>t</i>
<i>Backcastable variables</i>								
Voted Nixon in 1972	-1.14 (.00)	-1.36 (.00)		-0.92 (.00)	-0.51 (.04)	2,4	1.59 (.02)	2,4
Voted McGovern in 1972	1.30 (.00)	1.58 (.00)	-.56 (.04)	0.57 (.01)	0.88 (.00)	2		
Black	.96 (.00)		-2.29 (.00)		1.29 (.00)	2		
Education	-.29 (.00)	-.23 (.00)					0.71 (.00)	2,3,4
Age	-.01 (.01)	-.08 (.00)	-.10 (.00)					
Female			.73 (.00)					
Constant	.82 (.00)	3.47 (.00)	2.67 (.00)	-1.37 (.00)	-1.11 (.00)	2,3,4,5	-4.60 (.00)	2,3,4,5
<i>Non-backcastable variables</i>								
Self-identification as Democrat				1.87 (.00)	2.38 (.00)	2,3	-3.09 (.00)	3
					1.44 (.01)	5	-2.95 (.00)	4
Indifferent towards Democratic or Republican president				-0.19 (.00)			0.45 (.00)	2,3,4
Ford:								
- good job as president					-0.36 (.02)	4,5	0.64 (.00)	2,3,4
- favorable feelings				-0.28 (.00)	-0.29 (.00)	2	0.99 (.00)	5
					-1.21 (.00)	4		
- trustworthiness					-1.04 (.00)	5	1.43 (.00)	4
- leadership					-0.35 (.00)	3		
- ability							1.34 (.00)	2,5
Carter:								
- favorable feelings					0.35 (.00)	2,3	-0.69 (.00)	3,4
					1.14 (.00)	4,5	-1.81 (.00)	5
- trustworthiness					1.26 (.00)	4,5		
- leadership							-0.78 (.01)	4
- ability							-1.28 (.02)	2
Constant					-1.12 (.00)	2,3	3.28 (.00)	3
					-1.85 (.01)	4	2.82 (.00)	4
					-1.57 (.04)	5		
Log likelihood (<i>LL*</i>)			-2142.48					-1431.17

* *p*-values in parentheses. The β -parameters represent the effect on μ_t , β^* the effect on $(1 - \lambda_t)$ and thus $-\beta^*$ the effect on λ_t . The columns labeled *t* indicate the discrete time periods pertaining to the parameters.

Table 3: Prediction error measures*

t		2	3	4	5
MSE	$\mu : n_t^{-1} \sum_{i=1}^{n_t} ((y_{it} y_{t-1} = 0) - \mu_{it})^2$.146	.123	.068	.049
	$\lambda : n_t^{-1} \sum_{i=1}^{n_t} ((y_{it} y_{t-1} = 1) - \lambda_{it})^2$.155	.121	.126	.069
MML	$\mu : -n_t^{-1} \sum_{i=1}^{n_t} (y_{it} y_{t-1} = 0) \ln \mu_{it} + (1 - (y_{it} y_{t-1} = 0)) \ln(1 - \mu_{it})$.437	.378	.235	.162
	$\lambda : -n_t^{-1} \sum_{i=1}^{n_t} (y_{it} y_{t-1} = 1) \ln \lambda_{it} + (1 - (y_{it} y_{t-1} = 1)) \ln(1 - \lambda_{it})$.607	.412	.495	.250
MCA	$\mu : n_t^{-1} \sum_{i=1}^{n_t} (y_{it} y_{t-1} = 0) \mu_{it} + (1 - (y_{it} y_{t-1} = 0))(1 - \mu_{it})$.736	.756	.879	.899
	$\lambda : n_t^{-1} \sum_{i=1}^{n_t} (y_{it} y_{t-1} = 1) \lambda_{it} + (1 - (y_{it} y_{t-1} = 1))(1 - \lambda_{it})$.788	.803	.817	.867

* MSE is the mean squared error, MML is the mean value of minus log likelihood error (Van Houwelingen and Le Cessie 1990), and MCA is the mean probability of correct allocation (Kay and Little 1986).

Table 4: Frequencies of observed (*obs*) and estimated expected (*exp*) (non-)Democratic vote transitions (y_{t-1} y_t) at sample period t *

t	n_t	(00)		(01)		(11)		(10)		χ^2
		<i>obs</i>	<i>exp</i>	<i>obs</i>	<i>exp</i>	<i>obs</i>	<i>exp</i>	<i>obs</i>	<i>exp</i>	
2	670	309	296	102	104	213	219	46	51	1.2
3	643	279	270	57	73	253	253	54	47	5.0
4	642	271	269	69	70	233	239	69	64	0.7
5	617	288	305	47	42	243	236	39	34	2.7

*1=Democratic, 0=non-Democratic. The frequencies were only obtained for respondents with a valid score on both y_t and y_{t-1} .

Table 5: Frequencies of observed (*obs*) and estimated expected (*exp*) (non-)Democratic vote intention sequences *

sequence	<i>obs</i>	<i>exp</i>	Δ									
0	527	524	-3	0110	15	12	-3	01011	4	2	-2	
1	329	332	3	0111	43	40	-3	01100	10	7	-3	
00	309	296	-13	1000	12	19	7	01101	4	3	-1	
01	102	104	2	1001	5	2	-3	01110	4	5	1	
10	46	50	4	1010	5	6	1	01111	33	29	-4	
11	213	219	6	1011	4	5	1	10000	9	18	9	
000	223	207	-16	1100	12	11	-1	10001	3	2	-1	
001	37	40	3	1101	4	6	2	10010	1	1	0	
010	26	20	-6	1110	23	13	-10	10011	4	1	-3	
011	66	69	3	1111	114	132	18	10100	3	5	2	
100	25	26	1	00000	140	138	-2	10101	2	1	-1	
101	13	14	1	00001	7	9	2	10110	0	2	2	
110	20	20	0	00010	9	3	-6	10111	4	2	-2	
111	160	174	14	00011	14	13	-1	11000	9	6	-3	
0000	160	157	-3	00100	9	13	4	11001	0	1	1	
0001	30	24	-6	00101	2	2	0	11010	1	0	-1	
0010	12	18	6	00110	2	3	1	11011	3	3	0	
0011	14	14	0	00111	11	8	-3	11100	9	7	-2	
0100	13	13	0	01000	8	10	2	11101	11	3	-8	
0101	10	4	-6	01001	5	1	-4	11110	9	12	3	
				01010	3	0	-3	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 only obtained for respondents with a valid score on y_1 through y_t in the panel.

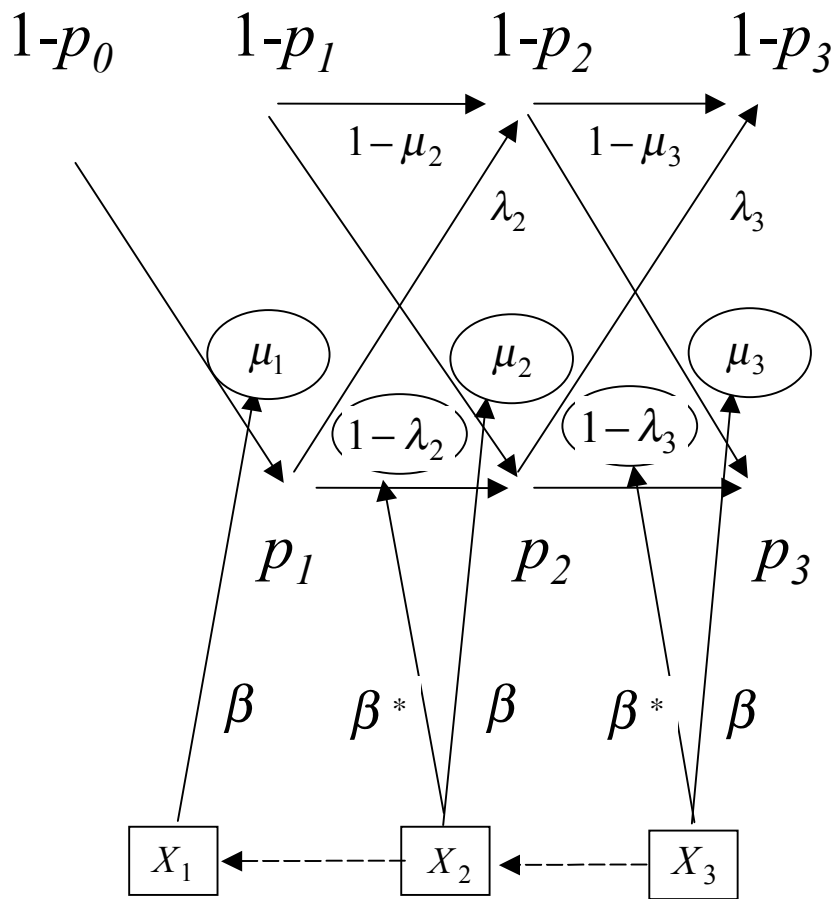


Figure 1: Graphical illustration of Markov model for RCS data

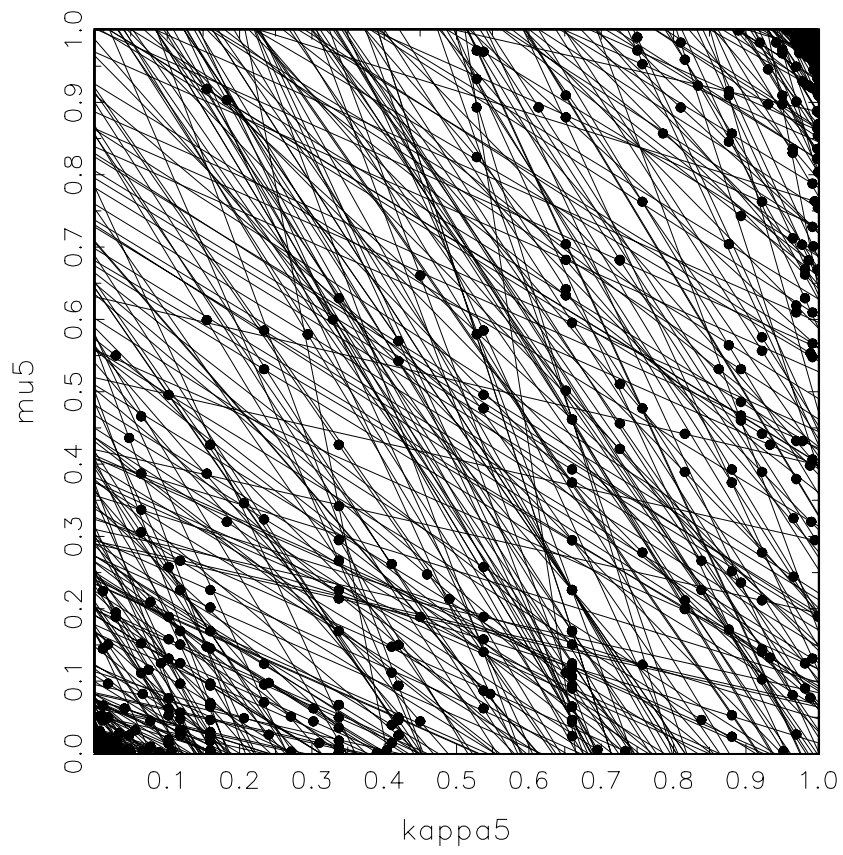


Figure 2: Tomography plot for current entry and 1-exit transitions at sample period $t = 5$