CATEGORICAL AUTO-REGRESSIVE MODELS FOR ANALYSING PANEL SURVEY DATA

A.C. Singh and G. Lemaitre, Statistics Canada
A.C. Singh, SSMD, Statistics Canada, Ottawa K1A 0T6

ABSTRACT

A discrete time series model for categorical arising from longitudinal (or panel) surveys is proposed. The model is expected to be useful in making adjustments for biases due to classification error and nonresponse in order to provide smoothed estimates of transitions (or gross flows) between categories from one point to another. Neither the reinterview data nor the assumption of independent classification errors would be required for the classification bias adjustment procedure under the proposed model. Some potential applications to monthly Canadian Labour force Survey (LFS) are considered.

KEYWORDS: Multivariate binary time series; Autoregressive model; Transition probabilities; Complex surveys; Nonresponse and classification errors.

1. INTRODUCTION

Time dependent categorical data obtained by means of repeated surveys of a sample of individuals over time arise quite often in longitudinal surveys, e.g. the Canadian monthly Labour Force Survey (LFS), the U.S. monthly current population survey (CPS), and national crime surveys. Generally survey data is obtained using a multistage stratified cluster sampling of households with a rotating panel design. In the LFS, for example, approximately $5/6^{th}$ of the households are common to two consecutive months and each sampled household is interviewed consecutively for six months before being dropped. For each sampled individual belonging to civilian noninstitutional population with 15 years of age or over, data on labour force status in the week before the survey week is collected. Individuals are classified as either E (employed), U (unemployed), N (not in the labour force), M (missing information due to no contact or refusal), and O (out of population of interest or rotated out of the sample).

With cross-sectional data (i.e. at a point in time), one can obtain estimates of total counts or proportions of individuals belonging to the given categories in the population of interest. With longitudinal or panel survey data (i.e., repeated measurement over time on some individuals), it is also possible to estimate proportions of individuals making transitions between categories at one time point to another. In the case of labour force participation, for example, one can estimate labour force stocks at a point in time (e.g. proportion unemployed) and gross flows between two consecutive points in time (e.g. what parts of the proportion unemployed are due to persons losing their jobs in the last month, and due to persons entering the labour force? How many unemployed persons leave the labour force?)

In this paper, we present a model based approach to the problem of gross flow estimation in the context of labour force participation status. The main problems in modelling of such data include presence of period to period nonresponse, response or classification error, period to period difference in sample based weights, inflows to or outflows from the population of interest, and inconsistency with external population counts. Little (1985), Fay (1986), Stasny (1986), and Stasny and Fienberg (1985) among others considered modeling in the presence of nonrandom nonresponse, i.e. the distribution of variable of interest for nonrespondents was not necessarily assumed to be the same as that for respondents. Abowd and Zellner (1985) proposed, under certain regularity conditions, a two-stage adjustment procedure to correct for nonresponse as well as classification errors by using gross flow data from several points in time. The assumption of independent classification errors was made in order to use reinterview data for estimating the proportions of persons misclassified in any given month. The independence assumption means that an individual's observed classification at time t depends stochastically on his true classification at t but not on his true or observed classification at time t-1. Classification errors potentially can cause serious upward biases in the off-diagonal cell proportions of the (E,U,N) by (E,U,N) gross-flow table and downward biases in the diagonal cells. Chua and Fuller (1987) and Poterba and Summers (1985) also assumed independent errors in correcting for classification error using reinterview data.

The model proposed in this paper takes advantage of autocorrelations over time in view of the longitudinal nature of the survey data. The model is motivated from time series methods for modeling dependence in observations via autoregressive parameters. There is however one point of departure from the usual time series when dealing with panel survey data. It is not assumed that the number of observation points in time for an individual is large, rather there is a large number of individuals, each contributing a series of observations over only a few points in time. The proposed model is based on a multivariate generalization of first order discrete autoregressive (DAR-1) models of Jacobs and Lewis (1978) and Zeger, Liang, and Self (1985) for binary data. It can also be seen as a generalization of models for Markov dependent Bernoulli trials as considered by Lindquist (1978).

This paper mainly addresses the problems of nonresponse and classification errors. The method considered here does not require reinterview data nor does it assume the independent response error structure in adjusting for classification errors. It is assumed instead that the response error biases in the gross flow table for the current time t+1 are of the same magnitudes as the corresponding ones for the table at time t. We then model the differences of the corresponding cell proportions to eliminate unknown classification error biases. Thus, gross-flow data for two time intervals (t-1,t) and (t,t+1), are required to compute differences in proportions. Another possibility would be to use gross-flow data from the previous year for the same time interval (t-1,t), if available. We also assume that the response biases in survey estimates of levels (or stocks) are negligible for both respondents and nonrespondents with partial information at one of the two time points. The partial information from nonrespondents is used to provide nonrandom nonresponse adjustments.

The proposed model allows for separate modeling of respondents and nonrespondents if there is partial information for nonrespondents from months in which they responded. More specifically, each sampled individual is classified into one of two types - (i) Partial respondent (PR) which means that he failed to respond on at least one of the time points due to refusal, temporary absence, or because of a move out of the sampled dwelling. (ii) Complete respondent (CR) which means that he responded on all occasions, except when by design, he was rotated out of the sample, or was outside the population of interest. Then separate models for the PR type and the CR type can be developed by means of additional parameters. This approach seems desirable in view of the findings of Paul and Lawes (1982) and Fienberg and Stasny (1983) which indicated a relationship between nonresponse and labour force status.

Besides adjustments for nonresponse and classification errors, the gross-flow estimates can be constrained so that the marginal distributions of the categorical variable constructed from each point in time are consistent with the comparable proportions estimated by the full panel survey data. Although this paper describes and considers applications of the proposed model with respect to labour force status, it may be noted that the model provides a general categorical autoregressive parametrization for longitudinal categorical data which might have other applications as well.

2. THE PROPOSED MODEL FOR GROSS-FLOW DATA

Suppose an individual's labour force status over time is denoted by a multivariate binary time series $\{(Y_1(t), Y_2(t), Y_3(t))': t=0,1,\dots T\}$ where the binary variables $Y_k(t)$'s, $k=1,2,3$ are simply indicators for the three possible categories (E,U,N) of an individual at time t. Here we assume for simplicity that there is no nonresponse. Suppose the summary data from individuals over a few time points are given in the form of tables of transition counts for each time interval (t-1,t). The proposed model is motivated by assuming first order nonstationary Markov chain for the disctete time series. This in turn implies independent multinomial sampling for the table of one step transition counts. Following the time series parametrization of Jacobs and Lewis (1978), and Zeger, Liang, and Self (1985), for a univariate binary series, we consider a generalization to the multivariate case, namely, the conditional mean of the vector $(Y_1(t),Y_2(t)'$ given $(Y_1(t-1), Y_2(t-1)'$, $t=1,\dots,T$ can be expressed as

$$E\left[\begin{pmatrix}Y_1(t)\\Y_2(t)\end{pmatrix}\middle|\begin{pmatrix}y_1(t-1)\\y_2(t-1)\end{pmatrix}\right]=\begin{pmatrix}\pi_1(t)\\\pi_2(t)\end{pmatrix}$$

$$+\begin{pmatrix}\beta_{11}(t-1,t) & \beta_{21}(t-1,t)\\\beta_{12}(t-1,t) & \beta_{22}(t-1,t)\end{pmatrix}\begin{pmatrix}y_1(t-1)-\pi_1(t-1)\\y_2(t-1)-\pi_2(t-1)\end{pmatrix} \qquad (2.1)$$

where $E(Y_i(t)) = \pi_i(t)$, β_{ij}'s are partial regression coefficients which can be expressed in the standardized form ϕ_{ij}'s as

$$\beta_{ij}(t-1,t) = \phi_{ij}(t-1,t)\,\frac{\sigma_j(t)}{\sigma_i(t-1)} \qquad (2.2)$$

Here $\sigma_i(t)$ is defined as $\sqrt{\pi_i(t)(1-\pi_i(t))}$. Further if we let $P_{ij}(t-1,t)$ denote the conditional probability $P[Y_j(t) = 1|Y_i(t-1) = 1]$, then (2.1) can be rewritten as

$$\begin{pmatrix}P_{11}(t-1,t) & P_{21}(t-1,t) & P_{31}(t-1,t)\\P_{12}(t-1,t) & P_{22}(t-1,t) & P_{32}(t-1,t)\end{pmatrix}$$

$$=\begin{pmatrix}\pi_1(t) & \pi_1(t) & \pi_1(t)\\\pi_2(t) & \pi_2(t) & \pi_2(t)\end{pmatrix}+$$

$$\begin{pmatrix}\phi_{11}(t-1,t)\sigma_1(t)/\sigma_1(t-1) & \phi_{21}(t-1,t)\sigma_1(t)/\sigma_2(t-1)\\\phi_{12}(t-1,t)\sigma_2(t)/\sigma_1(t-1) & \phi_{22}(t-1,t)\sigma_2(t)/\sigma_2(t-1)\end{pmatrix}$$

$$\times\begin{pmatrix}1-\pi_1(t-1) & -\pi_1(t-1) & -\pi_1(t-1)\\\pi_2(t-1) & 1-\pi_2(t-1) & -\pi_2(t-1)\end{pmatrix} \qquad (2.3)$$

Note that $P_{i3}(t-1,t)$'s can be obtained from (2.3) by using the constraints

$$\sum_{j=1}^{3} P_{ij}(t-1,t)=1,\ i=1,2,3. \qquad (2.4)$$

Let $\pi_{ij}(t-1,t)$ denote the joint probability $P[Y_i(t-1)=1, Y_j(t)=1]$ and for convenience let us denote $\pi_i(t-1)$, $\sigma_i(t-1)$ by π_i, σ_i and $\pi_i(t)$, $\sigma_i(t)$ by π_i', σ_i'. Further, let $\overline{\pi}_i$ stand for $1-\pi_i$. Then, multiplying $P_{ij}(t-1,t)$ by $\pi_i(t-1)$, the model (2.3) reduces to

$$\begin{pmatrix}\pi_{11} & \pi_{21}\\\pi_{12} & \pi_{22}\end{pmatrix}=\begin{pmatrix}\pi_1\pi_1' & \pi_2\pi_1'\\\pi_1\pi_2' & \pi_2\pi_2'\end{pmatrix}+$$

$$\begin{pmatrix}\phi_{11}\sigma_1'/\sigma_1 & \phi_{21}\sigma_1'/\sigma_2\\\phi_{12}\sigma_2'/\sigma_1 & \phi_{22}\sigma_2'/\sigma_2\end{pmatrix}\begin{pmatrix}\pi_1\overline{\pi}_1 & -\pi_1\pi_2\\-\pi_1\pi_2 & \pi_2\overline{\pi}_2\end{pmatrix}. \qquad (2.5)$$

The remaining π_{ij}'s can be obtained from the marginal probabilities π_i, π_i'. Thus the joint probabilities π_{ij}'s are parametrized by the above (saturated) model in terms of eight parameters π_1, π_2, π_1', π_2', ϕ_{11}, ϕ_{21}, ϕ_{12}, and ϕ_{22}. It follows directly from (2.5) that the standardized partial regression coefficients ϕ_{ij}'s can be expressed in terms of the simple autocorrelations ρ_{ij}'s between Y(t-1) and Y(t) as

$$\phi_{11}=\rho_{11}\overline{\pi}_1-\rho_{31}\overline{\pi}_3\sigma_1/\sigma_3,\quad \phi_{21}=\rho_{21}\overline{\pi}_2-\rho_{31}\overline{\pi}_3\ \sigma_2/\sigma_3$$

$$\phi_{12}=\rho_{12}\overline{\pi}_1-\rho_{32}\overline{\pi}_3\sigma_1/\sigma_3,\quad \phi_{22}=\rho_{22}\overline{\pi}_2-\rho_{32}\overline{\pi}_3\ \sigma_2/\sigma_3. \qquad (2.6)$$

A simple interpretation of ϕ_{ij}'s follows from the usual multiple regression with indicator variables as predictors in that they are defined as differences of conditional expectations. This is so because the conditional probabilities p_{ij}'s can be written as

$$P_{ij}(t-1,t)=\pi_j(t)+\rho_{ij}(t-1,t)\ \frac{\sigma_j(t)}{\sigma_i(t-1)}\ (1-\pi_i(t-1)),$$

or

$$\pi_{ij} = \pi_i\pi_j' + \rho_{ij}\ \sigma_i\sigma_j' \ . \qquad (2.7)$$

Note that ρ_{ij}'s satisfy the constraints

$$\rho_{i3}(t-1,t)\sigma_3(t)=-[\rho_{i1}(t-1,t)\sigma_1(t)+\rho_{i2}(t-1,t)\sigma_2(t)]$$

$$\rho_{3j}(t-1,t)\sigma_3(t-1)=$$

$$-[\rho_{1j}(t-1,t)\sigma_1(t-1)+\rho_{2j}(t-1,t)\sigma_2(t-1)]. \qquad (2.8)$$

The parametrization (2.7) for π_{ij}'s in terms of π_i, π_i' and ρ_{ij}'s is a special case of Bahadur representation (see Kedem, 1980, p.116). It provides a simple alternative to the model (2.5) which may not be convenient in practice in view of the constraints (2.8).

The model (2.5) gives a reparametrization of flow probabilities π_{ij}'s in terms of easily interpretable parameters using time series motivation as described above. This provides flexibility in making suitable assumptions when dealing with gross flow probabilities at several time points. For instance, the autoregression parameters ϕ_{ij}'s may be considered fixed over time points $1,2,\ldots,T$ while $\pi(t)$ is allowed to vary. Further, if we assume that the level or marginal proportions $\pi(t)$ can be estimated with negligible response error bias and that classification error biases in observed cell proportions $\hat{\Pi}(t-1,t)$ are the same over the time points $1,\ldots$ T; then the differences $\hat{\Pi}(t-1,t)-\hat{\Pi}(t,t+1)$ would have no bias. These differences can be linearly modelled using (2.5) in terms of ϕ parameters assuming that the level proportions are known and using a generalized least squares (GLS) approach for inference.

For nonrespondents, it is usually possible to estimate $\pi(t)$ from partial information about an individual's classification but not ϕ because the cells for transitions are empty. If we take the working assumption that ϕ's are the same for both respondents and nonrespondents, then the model (2.5) can be applied in estimating gross flows for nonrespondents. This will correspond to a nonrandom nonresponse model (i.e. not missing at random in the sense of Rubin, 1976) because $\pi(t)$ is not the same for the two cases. Furthermore, one can take advantage of the longitudinal character of the database to classify individuals into two types - CR (complete respondent) and PR (partial respondent) as defined in the introduction. In this approach, we assume that for partial respondents there are some data about gross flows between the three labour force categories at time intervals $(t-1,t)$ and $(t,t+1)$. It is no longer necessary to assume that ϕ's are the same for the two types CR and PR. We can again use (2.5) to model separately the two types by employing the given stock estimates and the gross flow data for each of the two types.

3. STATISTICAL INFERENCE WITH CATEGORICAL AUTO-REGRESSIVE MODELS FOR LFS DATA

We shall make the following assumptions:

(A1) Stock proportions for both respondents and non-respondents are free from bias due to classification error and are given.

(A2) The differences between observed flow proportions at time intervals $(t-1,t)$ and $(t,t+1)$ are free from response error biases.

(A3) The autoregressive parameters ϕ's are time homogeneous for the intervals under consideration for each group of respondents and non-respondents.

(A4) This assumption is required for inference purpose only. Let the vector η denote the cell proportions π_{ij}'s when rows are stacked vertically and $\hat{\eta}$ denote the corresponding survey estimates. Then,

$$\hat{\eta}(t-1,t)-\eta(t-1,t) \stackrel{a}{\sim} N(0,\Gamma(t-1,t)/n(t))$$

where $n(t)$ denotes the total sample size at time t and Γ is an appropriate covariance matrix depending on the underlying survey design. The symbol "$\stackrel{a}{\sim}$" denotes "asymptotically distributed as".

3.1 Adjustment for Classification Error with No Nonresponse about Individual's Categories at the Given Time Points

Suppose we have data about transition counts (inflated by design sub-weights) for three time points $t+1$, t, and $t-1$ i.e. current and two proceeding points in time. The flow data can be summarized as shown in Table 1.

The symbols E, U, N and O denote respectively employed, unemployed, not in labour force and out of sample due to rotation or out of population of interest. The total F_{++} denotes the estimated number of individuals who were represented by sample individuals responding on at least one of the two occasions $t-1$ and t. This is not an estimate of the size of the population at t or $t-1$ because it includes counts corresponding to the individuals who were either rotated in or out of the sample from time point $t-1$ to t. Let $N(t-1,t)$ denote the estimated number of individuals responding on both occasions $t-1$ and t. Thus $N(t-1,t)$ is the sum of F_{ij}'s for $i,j=1,2,3$. The quantities G_{++} and $N(t,t+1)$ can be similarly interpreted. Now define for $i,j=1,2,3$

$$\hat{\pi}_{ij}(t-1,t) = \frac{F_{ij}}{N(t-1,t)},\ \hat{\pi}_{ij}(t,t+1) = \frac{G_{ij}}{N(t,t+1)} \qquad (3.1)$$

and let π_i^*, $\pi_i^{*'}$, $\pi_i^{*''}$ denote the given stock proportions at times $t-1$, t, and $t+1$ respectively estimated from the full LFS data i.e. from all those who responded at the corresponding points in time.

Consider now the following linear model in the four ϕ parameters for flow differences based on the model (2.5). We have

Table 1. Transition Counts for Time Intervals (t-1,t) and (t,t+1)

t-1/t	E	U	N	O	Row Total	t/t+1	E	U	N	O	Row Total
E	F_{11}	F_{12}	F_{13}	F_{14}	F_{1+}	E	G_{11}	G_{12}	G_{13}	G_{14}	G_{1+}
U	F_{21}	F_{22}	F_{23}	F_{24}	F_{2+}	U	G_{21}	G_{22}	G_{23}	G_{24}	G_{2+}
N	F_{31}	F_{32}	F_{33}	F_{34}	F_{3+}	N	G_{31}	G_{32}	G_{33}	G_{34}	G_{3+}
O	F_{41}	F_{42}	F_{43}	-	F_{4+}	O	G_{41}	G_{42}	G_{43}	-	G_{4+}
Col.Total	F_{+1}	F_{+2}	F_{+3}	F_{+4}	G_{++}	Col.Total	G_{+1}	G_{+2}	G_{+3}	G_{+4}	G_{++}

$$\hat{\pi}_{11}(t-1,t)-\hat{\pi}_{11}(t,t+1)=(\pi_1^*\pi_1^{*\prime}-\pi_1^{*\prime}\pi_1^{*\prime\prime})+\phi_{11}(\pi_1^*\bar{\pi}_1^*\sigma_1^{*\prime}/\sigma_1^*$$
$$-\pi_1^{*\prime}\bar{\pi}_1^{*\prime}\sigma_1^{*\prime\prime}/\sigma_1^{*\prime})-\phi_{21}(\pi_1^*\pi_2^*\sigma_1^{*\prime}/\sigma_2^*-\pi_1^{*\prime}\pi_2^{*\prime}\sigma_1^{*\prime\prime}/\sigma_2^{*\prime})+\varepsilon_{11}, \tag{3.2}$$

and so on, where ε_{ij}'s denote random errors with mean 0 and some covariance Σ. The matrix Σ depends on the underlying complex survey design and can be estimated consistently by a suitable replication method (see Wolter 1985). The σ's are as before functions of π's and are known for fixed values of π's.

The flow table for (t-2,t-1) could also be used to have more flow differences if deemed appropriate. The four parameters of the model (3.2) can be estimated and the model fit can be tested by the method of generalized least squares (GLS) and then use (2.5) to get smoothed estimates of the cell proportions for the 3×3 table for (t,t+1). Now distribute the estimated population total at time (t+1) into 9 cells according to cell proportions estimated. This implies that the individuals in 'O' category are treated as "missing at random". The resulting gross flow estimates will ensure consistency with the stock estimates $\pi^*(t+1)$ from the full LFS data.

It may be remarked that the coefficients of ϕ's in the model (3.2) could be very small (near zero) if stock estimates π^* do not change much over time. This would then cause instability in the ϕ estimates and could lead to inadmissible values for the flow estimates. This problem of instability should be carefully investigated in order to have useful applications of the proposed model.

3.2 Adjustments for Classification Error and Nonresponse with Partial Information about Nonrespondents at one of the Time Points in Each Pair

In this case, there will be an extra row and column in table 1 labelled 'M' which denotes missing information about the individual's category. Let $(F_{15},F_{25},F_{35},F_{51},F_{52},F_{53},F_{55},)$ denote the supplemental counts (row and column) for the table at (t-1,t). Here F_{55} stands for the number of individuals with missing information at both time points. Similarly, we can define row and column supplements for table at (t,t+1).

By making an additional assumption to the effect that the ϕ's are the same for respondents and nonrespondents, we can estimate ϕ's from the table of respondents only, as in subsection 3.1. Estimation of flows will, however, be modified in this case. The individuals labelled 'O' are distributed over 9 cells in the same relative proportions as the estimates of cell proportions obtained in the previous subsection. But individuals labelled 'M' are distributed over 9 cells differently by using a separate set of $\pi^*(t)$, $\pi^*(t+1)$ for calculating smoothed estimates of cell proportions for nonrespondents. The $\pi^*_M(t)$ is given by

$$\pi_M^*(t) = (G_{15},G_{25},G_{35})'(G_{15}+G_{25}+G_{35})^{-1}.$$

Similarly $\pi_M^*(t+1)$ is calculated. Now the total number of individuals in category 'M' are distributed over the 9 cells according to cell proportions calculated from (2.5) using $\pi_M^*(t)$, $\pi_M^*(t+1)$, and $\hat{\phi}$ obtained before. The resulting grow flow estimates will be consistent with the comparable stock estimates from the full LFS data (i.e. stock estimates not inflated for nonrespondents at time t+1).

3.3 Adjustments for Classification Error and Nonresponse with Partial Information about Nonrespondents at more than one of the Previous Time Points

In this case, we can construct two sets of table 1, one for each of CR and PR. The types CR (complete respondents) and PR (partial respondents) were defined in the introduction. We then proceed as in subsection (3.2) to develop two models, (i) for CR with $\pi_{CR}^*(t)$, $\pi_{CR}^*(t+1)$, and $\hat{\phi}_{CR}$, and (ii) for PR with $\pi_{PR}^*(t)$, $\pi_{PR}^*(t+1)$, and $\hat{\phi}_{PR}$. This case uses the full scope of the longitudinal character of LFS data base because partial information about nonrespondents in the past is used to model their behaviour.

4. SUMMARY AND REMARKS

It was shown that a categorical autoregressive model of gross flow proportions from t-1 to t and from t to t+1 in terms of stock proportions at time points t-1, t, t+1 and first order autoregressive parameters provides a simple parametrzation for the differences in cell proportions for gross flow tables corresponding to (t-1,t) and (t,t+1). By assuming that these cell differences and the given stock estimates are free from biases due to classification errors, the problem reduces to the estimation of autoregressive parameters and the testing of model fit by the generalized least squares approach. The method could be used to produce "smoothed" estimates of gross flow proportions. The model could also provide, under fairly weak assumptions about the autoregressive parameters, adjustments for nonrandom nonresponse using partial information about nonrespondents. Apart from some of the possible applications outlined above, the proposed autoregressive model for longitudinal categorical data might be of independent interest and future theoretical development of this model would be desirable.

ACKNOWLEDGEMENT: The first author's research was supported in part by NSERC of Canada.

REFERENCES

Abowd, J.M. and Zellner, A. (1985). Estimating gross labour force flows. Journal of Business and Economic Statistics, 3, 254-283.

Chua, T.C., and Fuller, W.A. (1987). A model for multinomial response error applied to labour flows. Jour. Amer. Statist. Assoc., 82,46-51.

Fay, R.E. (1986). Causal models for patterns of nonresponse. Jour. Amer. Statist. Assoc.,81, 354-365.

Fienberg, S.E. and Stasny, E. (1983). Estimating monthly gross flows in labour force participation. Survey Methodology, 9, 77-102.

Jacobs, P. and Lewis, P.A.W. (1978). Discrete time series generated by mixtures. 1: correlation and runs properties. Jour. Roy. Statist. Soc. Ser. B, 40, 222-228.

Kedem, B. (1980). Binary Time Series, Marcel Dekker, New York.

Little, R.J.A. (1985). Nonresponse adjustment in longitudinal surveys. Models for categorical data. Invited paper at 45th session of the Int. Statist. Inst., Amsterdam.

Lindquist, B. (1978). A note on Bernoulli trials with dependence. Scand. J. Statist. 5, 205-208.

Paul, E.C. and Lawes, M. (1982). Characteristics of respondent and nonrespondent households in the Canadian labour force survey. Survey Methodology, 8, 48-85.

Poterba, J.M., and Summers, L.H. (1985). Adjusting the gross changes data: Implications for labour market dynamics (with discussion). Proceedings of the Conference on Gross Flows in Labour Force Statistics, U.S. Dept. of Commerce and U.S. Dept. of Labour, 81-99.

Rubin, D.B. (1976). Inference and missing data. Biometrika, 63, 581-592.

Stasny, E. (1986). Estimating gross flows using panel data with nonresponse: an example from the Canadian labour force survey. Jour. Amer. Statist. Assoc., 81, 42-47.

Stasny, E., and Fienberg, S.E. (1985). Some stochastic models for estimating gross flows in the presence of nonrandom nonresponse (with discussion). Proceedings of the conference on Gross Flows in Labour Force Statistics, U.S. Dept. of Commerce and U.S. Dept. of Labour, 25-39.

Wolter, K.M. (1985). Introduction to Variance Estimation. Springer-Verlag, New York.

Zeger, S.L., Liang, K., and Self, S.G. (1985). The analysis of binary longitudinal data with time independent covariates. Biometrika,72, 31-38.