MODELING NON-IGNORABLE NON-RESPONSE IN PANEL DATA

Elizabeth A. Stasny, Ohio State University

1. INTRODUCTION

A number of large-scale sample surveys, such as the Current Population Survey and the National Crime Survey in the United States and the Labour Force Survey in Canada, are designed using rotating panel structures so that subjects are interviewed a number of times before being dropped from the sample. Although such surveys are used mainly to obtain cross-sectional estimates, it has long been recognized that information from the repeated interviewing of subjects provides an additional longitudinal data base that could be exploited to give estimates of gross change over time for a small additional cost (see, for example, Kalachek, 1979, and Fienberg and Tanur, 1983 and 1984). Naturally, there are problems associated with the use of such data to estimate change. The problems include how to handle period-to-period non-response, response error, period-to-period differences in sample-based weights, and inconsistencies between the estimates of change over time and point-in-time estimates. This paper considers the problem of handling period-to-period non-response in using panel data to provide estimates of gross change over time.

In this paper we consider the case where survey responses are categorical as would be the case, for example, for the Current Population Survey where subjects are classified as employed, unemployed, or not in the labor force, and the National Crime Survey where subjects are classified according to the type of crime committed against them. We further restrict the problem to the case of estimating gross changes or gross flows from one time period to another rather than over several periods.

One possible approach to the problem of gross flow estimation is to use only the information from individuals who are respondents in both of the interview periods. In order to use this approach, we must assume that individuals who do not respond in both periods are a random sample of all individuals (Rubin, 1976). However, in most cases, we do not believe that non-response occurs at random. For example, Saphire (1984) gives evidence that non-response in the National Crime Survey is related to victimizations. Paul and Lawes (1982) and Fienberg and Stasny (1983) give evidence that non-response in the Canadian Labour Force Survey is related to labor force classification.

Since there is evidence that non-response does not occur at random, we would like to consider models for estimating period-to-period gross flows that allow us to treat non-response as related to survey classification. In this paper, we develop five such models.

The models proposed in this paper are fit using maximum likelihood estimation to employment status data from the Current Population Survey (CPS) and the Labour Force Survey (LFS). Thus, we describe the models in terms of this data. The extension of the models to categorical data from other panel surveys is straightforward.

In Section 2 of this paper, we set up notation for the problem of estimating month-to-month gross flows in labor force participation when some individuals are observed in only one of the two months. In Section 3, we consider estimating gross flows using ideas developed by Chen and Fienberg (1974) for maximum likelihood estimation in contingency tables with some partially cross-classified data. The method allows for some flexibility in modeling the processes that produce the observed gross flow data. We consider five models that have natural interpretations for the gross flow problem. The data analysis is given in Section 4. Conclusions and some extensions are given in Section 5.

2. A MODEL FOR THE OBSERVED PANEL DATA

The CPS is based on monthly interviews with respondents in approximately 60,000 households. The survey currently uses a 4-8-4 rotation scheme in which each of eight rotation groups is interviewed for four months, is dropped from the sample for the next eight months, and is reincluded for the final four months. In this CPS scheme, the month-to-month overlap is 75% while there is a 50% overlap in the sample location for the same month in successive years. The LFS is based on monthly interviews with respondents in approximately 56,000 households. Sampled households are retained in the sample for six months before being rotated out of the sample. In this LFS scheme, the month-to-month overlap is 83%.

Persons interviewed for the CPS or LFS in a given month are classified as employed, unemployed, or not in the labor force. Here we do not consider individuals who are members of the armed forces or not in the population of interest although such additional classifications could easily be used.

We estimate gross flows among the three labor force classifications using records of individuals matched over two consecutive months. Typically, before this matching is done, the monthly data are edited and records of individuals who failed to respond within that month are removed from the data file. It is still not possible to match records from one month to the next for individuals who rotated into or out of the sample and for individuals who only responded in one of the two months. Thus, as a result of the matching of two consecutive months of data, we have a group of records for completely cross-classified individuals, that is individuals whose labor force status in both months is available, and a group of records for partially cross-classified individuals whose labor force status is reported in only one month.

Since we do not in general accept the assumption of non-random non-response, we would like to use the information from both completely and partially cross-classified individuals to estimate gross flows. However, we do not feel that all non-reponse violates the assumption of missing at random. For example, non-reponse due to panel rotation is designed non-response and, hence, should satisfy the definition of missing at random with respect to labor force status. Thus, in the following, we handle non-response due to panel rotation differently from non-response due to other reasons.

The labor force classification data for individuals who responded in two consecutive months, t-1 and t, can be summarized in a 3×3 matrix. The available information for individuals who rotated out of the sample after the month t-1 interview may be summarized in a rotation column while the available information for individuals who rotated into the sample before the month t interview may be summarized in a rotation row. Information for individuals who were non-respondents in one of the two months for reasons other than rotation may be given in row and column supplements. Thus, the observed gross flow data is as shown in Table 1.

Using the method proposed by Stasny (1984), we extend the ideas of Chen and Fienberg (1974) for maximum likelihood estimation in contingency tables with partially cross-classified data and take the observed gross flow data to be the end result of a three-stage process. In the unobserved first stage, individuals are allocated to the nine cells of a 3×3 matrix according to a single multinomial distribution. Let

ω_{ij} = probability that an individual has labor force classification i in month t-1 and j in month t.

At the second stage of the process, which is also unobserved, the sampling plan is determined and each individual may be chosen to either rotate out of the sample after the interview for month t-1 or rotate into the sample before the month t interview. Let

π_t = probability than an individual does not respond in month t due to panel rotation and

π_{t-1}= probability that an individual does not respond in month t-l due to panel rotation.

Note that we have modelled the probabilities of non-response for CPS both panels rotating into the sample as the same even though one panel is entering for the first time while the other is returning to the sample after eight months. Similarly, the probabilities of non-response are the same for the panel rotating out of the sample for the final time and the panel rotating out for a period of eight months. In view of the problem of rotation group bias, (e.g. see Bailar 1975, 1979) it would be of interest to consider different probabilities of non-response due to rotation for each of the panels. Our models could easily be extended to allow for this distinction.

Finally, in the third stage of the process, each remaining observation in the (i,j) cell of the gross flow matrix may either be a non-respondent for other reasons in month t-1 and lose its row classification, or be a non-respondent for other reasons in month t and lose its column classification. Let

ϕ_{ij} = probability that an observation in the (i,j) cell of the matrix loses its row classification and

ψ_{ij} = probability that an observation in the (i,j) cell of the matrix loses its column classification.

We assume that the probability that an individual is a non-respondent in both months for any reason is zero. There are, in fact, individuals who are missing in both months. Models that allow for non-response in both months have been considered in Stasny (1983).

TABLE 1: OBSERVED GROSS FLOW DATA

		Month t				
		E	U	N	*Rotation*	*Row Supp.*
Month	E	x_{EE}	x_{EU}	x_{EN}	Q_E	R_E
t-1	U	x_{UE}	x_{UU}	x_{UN}	Q_U	R_U
	N	x_{NE}	x_{NU}	x_{NN}	Q_N	R_N
Rotation		B_E	B_U	B_N		
Col. Supp.		C_E	C_U	C_N		

where E = employed,
U = unemployed,
N = not in the labor force,
x_{ij} = # of sampled individuals with labor force status i in month t-l and j in month t,
Q_i = # of individuals who rotated out of the sample after month t-1 and had employment status i in month t-1,
B_j = # of individuals who rotated into the sample before month t and had employment status j in month t,
R_i = # of individuals who did not respond for other reasons in month t and had labor force status i in month t-1, and
C_j = # of individuals who did not respond for other reasons in month t-1 and had labor force status j in month t.

The data are observed after this third stage. From the observed data, we want to make inferences about the cell probabilities, $\{\omega_{ij}\}$, of the unobserved first stage of the allocation process. That is, our goal is to make inferences about the underlying cell probabilities as if there had been no loss of information.

In the context of this three stage model, the probability that an individual with labor force classification i in month t-1 and j in month t is observed in the (i,j) cell of the gross flow matrix is $(1-\phi_{ij}-\psi_{ij})(1-\pi_{t-1}-\pi_t)\omega_{ij}$. Thus, the underlying probabilities for the observed gross flow matrix are as given in Table 2.

TABLE 2: PROBABILITIES FOR OBSERVED GROSS FLOW DATA

		Month t				
		E	U	N	*Rotation*	*Row Supp.*
Month	E					
t-1	U	$\{(1-\phi_{ij}-\psi_{ij})(1-\pi_{t-1}-\pi_t)\omega_{ij}\}$			$\{\pi_t\omega_{i+}\}$	$\{\Sigma_j\psi_{ij}(1-\pi_{t-1}-\pi_t)\omega_{ij}\}$
	N					
Rotation		$\{\pi_{t-1}\omega_{+j}\}$				
Col. Supp.		$\{\Sigma_i\phi_{ij}(1-\pi_{t-1}-\pi_t)\omega_{ij}\}$				

If m_{ij} is the expected count in the $(i,j)^{th}$ cell under the original multinomial sampling scheme, then the likelihood function for the observed data is proportional to

$$\begin{aligned}&\{\Pi_i\Pi_j[(1-\phi_{ij}-\psi_{ij})(1-\pi_{t-1}-\pi_t)m_{ij}]^{x_{ij}}\}\\ &\quad\times\{\Pi_i[\Sigma_j\psi_{ij}(1-\pi_{t-1}-\pi_t)m_{ij}]^{R_i}\} \qquad (1)\\ &\quad\times\{\Pi_j[\Sigma_i\phi_{ij}(1-\pi_{t-1}-\pi_t)m_{ij}]^{C_j}\}\\ &\quad\times\{\Pi_i[\pi_t m_{i+}]^{Q_i}\}\times\{\Pi_j[\pi_{t-1}m_{+j}]^{B_j}\}.\end{aligned}$$

Obviously, we cannot obtain MLE's for all of the 29 parameters that appear in this likelihood using only the 21 available observed counts. In order to estimate parameters, we will reduce the number of parameters by considering models for the ϕ_{ij} and ψ_{ij}, the probabilities of non-response for reasons other than panel rotation. Five such models are developed in the following section.

3. MODELS FOR NON-RANDOM NON-RESPONSE

The probability, ϕ_{ij}, that an individual's month t-1 identity is lost depends on both the month t-1 and month t labor force classifications. Similarly, the probability, ψ_{ij}, depends on the labor force classifications for both months. We can reduce the number of parameters that must be estimated by using simpler models for the probability that a given month's labor force classification is lost. We consider five models for these parameters. The first two are of the type described by Chen and Fienberg (1974). The models are as follows:

A. $\phi_{ij} = \lambda_{t-1(j)}$ $\psi_{ij} = \lambda_{t(i)}$
B. $\phi_{ij} = \lambda_{t-1}$ $\psi_{ij} = \lambda_t$
C. $\phi_{ij} = \lambda_{(j)}$ $\psi_{ij} = \lambda_{(i)}$
D. $\phi_{ij} = \lambda_{t-1(i)}$ $\psi_{ij} = \lambda_{t(j)}$
E. $\phi_{ij} = \lambda_{(i)}$ $\psi_{ij} = \lambda_{(j)}$.

Under model A, the probability that an individual's labor force classification for a given month is lost depends on the month and the individual's classification in the observed month. Under model B, the probability that an individual's labor force classification is lost in a given month depends only on the month. Under model C, the probability that an individual's labor force classification is lost in a given month depends only on the labor force classification in the observed month. Under model D, non-response depends on the month and on labor force status in the month when the individual does not respond. Under model E, non-response depends only on labor force status in the month when the individual does not respond.

Intuitively, models D and E are preferable to models A and C respectively since the probabilities of non-response under D and E depend on the labor force classifications in the month when an individual does not respond while under models A and C the probabilities of non-response depend on the classification in the observed month. An advantage of models A, B, and C is that under those models the likelihood function of equation (1) separates into a factor involving the m parameters alone, a factor involving the λ parameters alone, and a factor involving the π parameters alone. Thus, the MLE's for the m, λ, and π parameters may be obtained separately under models A, B, and C. Under models D and E, the likelihood function only separates into two factors: one involving the m and λ parameters and a second involving the π parameters. Since the MLE's for the m and λ parameters must be obtained simultaneously, models D and E are more difficult to fit.

3.1 Models A, B, and C

In general, iterative methods must be used to obtain the MLE's under models A, B, and C. Since these models have previously been fit to CPS data (see Stasny and Fienberg, 1985) and LFS data (see Stasny 1983, 1984), the following discussion does not include the formulas and details for obtaining the MLE's under these models. Note that under model A, there are 4 degrees of freedom while model B has 8 degrees of freedom and model C has 7 degrees of freedom.

3.2 Model D

The likelihood function for the observed data under model D may be written as the product of two factors:

$$f_{D1} = \{\Pi_i\Pi_j(1-\pi_{t-1}-\pi_t)^{x_{ij}}\} \times \{\Pi_i\pi_t^{Q_i}(1-\pi_{t-1}-\pi_t)^{R_i}\}$$
$$\times \{\Pi_j\pi_{t-1}^{B_j}(1-\pi_{t-1}-\pi_t)^{C_j}\} \text{ and}$$

$$f_{D2} = \{\Pi_i\Pi_j[(1-\lambda_{t-1(i)}-\lambda_{t(j)})m_{ij}]^{x_{ij}}\}$$
$$\times \{\Pi_i m_{i+}^{Q_i}[\Sigma_j\lambda_{t(j)}m_{ij}]^{R_i}\}$$
$$\times \{\Pi_j m_{+j}^{B_j}[\Sigma_i\lambda_{t-1(i)}m_{ij}]^{C_j}\}.$$

The factor f_{D1} involves only the π parameters and is the same as the factor of the likelihood functions involving only the π parameters under all five models described here. The π parameters may be considered to be known since the panel rotation structure is built into the design of the sample. However, in both the CPS and LFS, housing units are the units rotating into and out of the sample. Thus, the sample design does not tell us exactly how many individuals will rotate into and out of the sample and we may want to treat the π parameters as unknown. In that case, simple closed form estimates of the π parameters are

$$\hat{\pi}_{t-1} = B_+/(x_{++}+R_++C_++Q_++B_+) \text{ and} \tag{2}$$
$$\hat{\pi}_t = Q_+/(x_{++}+R_++C_++Q_++B_+).$$

Factor f_{D2} is maximized using a Lagrange multiplier to impose the constraint that $\Sigma_i\Sigma_j m_{ij} = x_{++}+Q_++R_++B_++C_+$. In general, the MLE's for the m and λ parameters must be found iteratively. The iterative procedure used for the data analysis of Section 4 is as follows:

1. $m_{ij}^{(0)} = x_{ij}(x_{++}+Q_++R_++B_++C_+)/x_{++}$

$\lambda_{t-1(i)}^{(0)} = C_+/(x_{++}+R_++C_+)$ and

$\lambda_{t(j)}^{(0)} = R_+/(x_{++}+R_++C_+)$.

2. $$m_{ij}^{(v+1)} = x_{ij} + R_i\{\lambda_{t(j)}^{(v)}m_{ij}^{(v)}/\Sigma_k\lambda_{t(k)}^{(v)}m_{ik}^{(v)}\}$$
$$+ C_j\{\lambda_{t-1(i)}^{(v)}m_{ij}^{(v)}/\Sigma_k\lambda_{t-1(k)}^{(v)}m_{kj}^{(v)}\}$$
$$+ Q_i\{m_{ij}^{(v)}/m_{i+}^{(v)}\} + B_j\{m_{ij}^{(v)}/m_{+j}^{(v)}\}$$
$$\lambda_{t-1(i)}^{(v+1)} = \{\Sigma_jC_j[\lambda_{t-1(i)}^{(v)}m_{ij}^{(v)}/\Sigma_k\lambda_{t-1(k)}^{(v)}m_{kj}^{(v)}]\}$$
$$\times \{\Sigma_j[x_{ij}/(1-\lambda_{t-1(i)}^{(v)}-\lambda_{t(j)}^{(v)})]\}^{-1} \text{ and}$$
$$\lambda_{t(j)}^{(v+1)} = \{\Sigma_iR_i[\lambda_{t(j)}^{(v)}m_{ij}^{(v)}/\Sigma_k\lambda_{t(k)}^{(v)}m_{ik}^{(v)}]\}$$
$$\times \{\Sigma_i[x_{ij}/(1-\lambda_{t-1(i)}^{(v)}-\lambda_{t(j)}^{(v)})]\}^{-1}.$$

Step 2 is repeated for $v = 0, 1, 2, \ldots$ until the m and λ parameter estimates converge to the desired degree of accuracy. The initial estimates in step 1 above, and in the iterative procedure for model E given below, are merely suggested values. They may be replaced by any positive values satisfying $\Sigma_i\Sigma_j m_{ij}^{(0)} = x_{++}+Q_++R_++B_++C_+$ for the m parameters and any values between 0 and 1 for the λ parameters.There are 4 degrees of freedom associated with model D.

3.3 Model E

The likelihood function for the observed data under model E may be written as the product of two factors:

$f_{E1} = f_{D1}$ and

$$f_{E2} = \{\Pi_i\Pi_j[(1-\lambda_{(i)}-\lambda_{(j)})m_{ij}]^{x_{ij}}\}$$
$$\times \{\Pi_i m_{i+}^{Q_i}[\Sigma_j\lambda_{(j)}m_{ij}]^{R_i}\}$$
$$\times \{\Pi_j m_{+j}^{B_j}[\Sigma_i\lambda_{(i)}m_{ij}]^{C_j}\}.$$

Factor f_{E1} is maximized using equation (2). Factor f_{E2} is maximized using a LaGrange multiplier to impose the constraint that $\Sigma_i\Sigma_j m_{ij} = x_{++}+Q_++R_++B_++C_+$. In general the MLE's for the m and λ parameters must be found iteratively. The iterative procedure used for the data analysis of Section 4 is as follows:

1. $m_{ij}^{(0)} = x_{ij}(x_{++}+Q_++R_++B_++C_+)/x_{++}$ and

$\lambda_{(i)}^{(0)} = (R_++C_+)/2(x_{++}+R_++C_+)$.

2. $$m_{ij}^{(v+1)} = x_{ij} + R_i\{\lambda_{(j)}^{(v)}m_{ij}^{(v)}/\Sigma_k\lambda_{(k)}^{(v)}m_{ik}^{(v)}\}$$
$$+ C_j\{\lambda_{(i)}^{(v)}m_{ij}^{(v)}/\Sigma_k\lambda_{(k)}^{(v)}m_{kj}^{(v)}\}$$
$$+ Q_i\{m_{ij}^{(v)}/m_{i+}^{(v)}\} + B_j\{m_{ij}^{(v)}/m_{+j}^{(v)}\} \text{ and}$$
$$\lambda_{(i)}^{(v+1)} = \{\Sigma_jR_j[\lambda_{(i)}^{(v)}m_{ji}^{(v)}/\Sigma_k\lambda_{(k)}^{(v)}m_{jk}^{(v)}]$$
$$+\Sigma_jC_j[\lambda_{(i)}^{(v)}m_{ij}^{(v)}/\Sigma_k\lambda_{(k)}^{(v)}m_{kj}^{(v)}]\}$$
$$\times \{\Sigma_j[(x_{ij}+x_{ji})/(1-\lambda_{(i)}^{(v)}-\lambda_{(j)}^{(v)})]\}^{-1}.$$

Step 2 is repeated for $v = 0, 1, 2, \ldots$ until the m and λ parameter estimates converge to the desired degree of accuracy. There are 7 degrees of freedom associated with model E.

4. DATA ANALYSIS

The five models described in Section 3 were fit to CPS data from December, 1981 through December, 1982 and to LFS data from the single panel that rotated into the sample in August, 1979 and remained in the sample through January, 1980. The CPS data have been weighted using sample-based weights. The models proposed in this paper are suitable for unweighted data from simple random samples of individuals (see Stasny and Fienberg, 1985). Although the available CPS data are not in exactly the correct form for our models, we will use the data for illustrative purposes. The LFS data used in this analysis is unweighted micro-data. However, since only a single panel is available, we do not have information on non-response due to panel rotation. Thus, we cannot estimate the π parameters from the LFS data. This, in effect, reduces the degrees of freedom by 4 for each of the five models. Hence, models A and D will fit the LFS data exactly.

As an example of the forms of the data, a single observed gross flow matrix from both the CPS and LFS is given in Appendix I. The entire CPS data set may be found in Stasny and Fienberg (1985) and the LFS data is given in Stasny (1983).

The five models were fit to the CPS and LFS data using the procedures described in Section 3. For the iterative procedures, the criterion used for stopping the iteration was that the maximum difference between estimates of the m parameters at two consecutive steps was less than 0.5 and the maximum difference between estimates of the λ parameters at two consecutive steps was less than 0.0005.

The iterative procedure for obtaining the MLE's of the m parameters under models A, B, and C converged in 7 or fewer iterations in all cases. The λ parameter estimates for models A and C were obtained in 6 or fewer iterations. Up to 16 iterations were required to obtain the MLE's of the m and λ parameters under models D and E.

The MLE's of all the π and λ parameters and an example of the m parameter estimates under all five models for both the CPS and LFS data are given in Appendix II. The fits of the models, as measured by the X^2 statistic, are given in Table 3 and Table 4 for the CPS and LFS data respectively.

TABLE 3: X^2 VALUES FOR FITTING THE WEIGHTED CPS DATA

Months	Model A (4 d.f.)	Model B (8 d.f.)	Model C (7 d.f.)	Model D (4 d.f.)	Model E (7 d.f.)
12/81-1/82	81	5158	106	88	111
1/82-2/82	15	5073	18	16	36
2/82-3/82	6	5111	29	5	39
3/82-4/82	12	5871	75	20	102
4/82-5/82	44	5042	47	54	65
5/82-6/82	48	5448	68	53	74
6/82-7/82	24	5982	28	36	54
7/82-8/82	13	5669	32	10	45
8/82-9/82	63	6170	97	44	97
9/82-10/82	11	5422	27	16	45
10/82-11/82	23	5086	38	29	50
11/82-12/82	20	4772	38	34	57

TABLE 4: X^2 VALUES FOR FITTING THE LFS DATA

Months	Model A (0 d.f.)	Model B (4 d.f.)	Model C (3 d.f.)	Model D (0 d.f.)	Model E (3 d.f.)
8/79-9/79	--	19	337	--	333
9/79-10/79	--	38	17	--	17
10/79-11/79	--	49	1	--	1
11/79-12/79	--	74	11	--	10
12/79-1/80	--	43	19	--	18

Keeping in mind the fact that we have very large cell counts in both the CPS and LFS data, we are pleased with the fits of models A, C, D, and E for the CPS data as well as for the LFS data except in the case of the months 8/79-9/79 data. (Models A and D, of course, fit the LFS data exactly.) Recall that under model B, the probability of losing a month's employment classification depends only on the month. The fact that this model provides a better fit than either model C or E to the 8/79-9/79 LFS data is consistent with the finding that, for LFS panels, non-response is high the first time a panel is included in the survey. In addition, since many people go on vacation in August, we expect more non-response in August than in September.

Except for the 8/79-9/79 LFS data, the fits of models A, C, D, and E are fairly good, given the large cell counts. Under these four models, the probability of non-response is related to the employment classification either in the observed month (under models A and C) or the unobserved month (D and E). For both the CPS and LFS data, the fits of models A and D are similar and the fits of models C and E are similar. We may prefer to use a model with fewer parameters and hence would select either model C or E. However, the fits of models C and E are similar and it is not clear which of the two models should be prefered on that basis.

The choice of either model C or E does make a difference in the values of the parameter estimates obtained from the model. For example, the estimated expected cell counts in the unemployed to unemployed cell of the unobserved first stage are always larger under model E than under model C. In addition, except for two cases where the counts are equal, the estimated expected cell counts in any of the cells involving the unemployment classification are higher under model E. Also under model E, the $\hat{\lambda}_{(U)}$'s, the estimated probabilities that an unemployed person is a non-respondent for reasons other than panel rotation, are higher than the corresponding estimates under model C.

It is not clear which of the models C or E is preferable based on the parameter estimates or the fit of the models. Intuitively, we feel that model E is more realistic since under that model non-response is related to employment status in the month when the non-response occurs. However, information other than that in the CPS and LFS data used here is needed to support that belief.

5. EXTENSIONS AND CONCLUSIONS

In this paper we have developed five models for estimating gross flows from categorical panel data having non-random non-response. The models were fit using maximum likelihood estimation to labor force status data from the CPS and LFS. The models under which the probability of non-response is related to labor force status in either the observed or unobserved month provided an adequate fit to the data.

Intuitively, the model under which non-response is related to labor force status in the month when the individual is observed would be the prefered model. However, the model under which non-response is related to employment status in the unobserved month provides a similar fit to the data. The fact that this model fits the data as well as the model we would intuitively prefer suggests that employment status in the observed month may serve as a surrogate for employment status in the unobserved month. Additional information is needed, however, in order to choose between the two models.

In this paper, we have only considered estimating gross flows using two time periods. An important extension of these models would be to the case where more time periods could be used to estimate flows. Some extensions of the Chen-Fienberg type of models to higher-dimensional contingency tables are given in Chen (1972) and Chen and Fienberg (1976). If such models were used for estimating gross flows, it would be possible to include individuals who were non-respondents in more than a single time period.

REFERENCES

Bailar, B. A. (1975). "The Effects of Rotation Group Bias on Estimates for Panel Surveys". *Journal of the American Statistical Association,* **70**, 23-30.

__________(1979). "Rotation Sample Biases and Their Effects on Estimates of Change". *Proceedings of the Conference on Recent Developments in Statistical Methods and Applications,* 13-29.

Chen, T. (1972). "Mixed-up Frequencies and Missing Data in Contingency Tables", unpublished Ph.D. dissertation, Department of Statistics, University of Chicago.

Chen, T. and Fienberg, S. E. (1974). "Two-Dimensional Contingency Tables With Both Completely and Partially Cross-Classified Data." *Biometrics*, **30**, 629-642.

Chen, T. and Fienberg, S. E. (1976). "The Analysis of Contingency Tables with Incompletely Classified Data." *Biometrics,* 32, 133-144.

Fienberg, S. E. and Stasny, E. A. (1983). "Estimating Monthly Gross Flows in Labour Force Participation". *Survey Methodology,* **9**, 77-102.

Fienberg, S. E. and Tanur, J. M. (1983). "Large-Scale Social Surveys: Perspectives, Problems, and Prospects". *Behavioral Science*, **28**, 135-153.

__________(1984). "The Design and Analysis of Longitudinal Surveys: Controversies and Issues of Cost and Continuity". Technical Report No. 289, Department of Statistics, Carnegie-Mellon University.

Kalachek, E. (1979). *Longitudinal Surveys and Labor Market Analysis. Data Collection, Processing and Presentation: National and Local* (Counting the Local Force, Appendix, Volume II), U.S. Government Printing Office, 160-189.

Paul, E. C. and Lawes, M. (1982). "Characteristics of Respondent and Non-Respondent Households in the Canadian Labour Force Survey". *Survey Methodology*, **8**, 48-85.

Rubin, D. B. (1976). "Inference and Missing Data". *Biometrika*, **63**, 581-592.

Saphire, D. G. (1984). "Estimation of Victimization Prevalence Using Data From the National Crime Survey". *Lecture Notes in Statistics*, **23**, edited by Brillinger, D. et. al., New York: Springer-Verlag.

Stasny, E. A. (1983). "Estimating Gross Flows in Labor Force Participation Using Data From the Canadian Labour Force Survey". Unpublished Ph.D. dissertation, Department of Statistics, Carnegie-Mellon University.

__________(1984). "Estimating Gross Flows in Labor Force Participation Using Information From Individuals With Incomplete Classifications. Technical Report No. 272, Department of Statistics, Carnegie-Mellon University.

Stasny, E. A. and Fienberg, S. E. (1985). "Some Stochastic Models for Estimating Gross Flows in the Presence of Nonrandom Nonresponse". *Proceedings of the Conference on Gross Flows in Labor Force Statistics*, U.S. Department of Commerce and U.S. Department of Labor, 25-39.

Appendix I. Examples of CPS and LFS Data

A. Example of Weighted CPS Data (Given in thousands)

January 1982

		E	U	N	Rotation	Row Supp.
December 1981	E	65445	1089	1691	24457	4976
	U	1870	3658	1277	2279	733
	N	2202	1273	40803	15123	3059
Rotation		24569	2631	15634		
Col. Supp.		5135	797	3262		

B. Example of Observed LFS Data

September 1979

		E	U	N	Row Supp.
August 1979	E	9223	128	662	473
	U	221	322	151	59
	N	256	164	5941	292
Col. Supp.		997	69	677	5016

Appendix II. Parameter Estimates

A. Estimated Probabilities of Non-Response Due to Rotation - CPS Data

Months	π_{t-1}	π_t
12/81 - 1/82	.193	.189
1/82 - 2/82	.192	.192
2/82 - 3/82	.192	.191
3/82 - 4/82	.192	.191
4/82 - 5/82	.192	.192
5/82 - 6/82	.191	.191
6/82 - 7/82	.191	.190
7/82 - 8/82	.191	.191
8/82 - 9/82	.191	.191
9/82 - 10/82	.192	.192
10/82 - 11/82	.192	.192
11/82 - 12/82	.192	.192

B. Estimated Expected Cell Counts After Unobserved First Stage - CPS Data (Given in thousands)

		Models A, B, C			Model D			Model E		
		E	U	N	E	U	N	E	U	N
12/81 - 1/82	E	121612	2232	3167	121415	2234	3145	121453	2244	3162
	U	3447	7431	2372	3532	7672	2419	3490	7648	2404
	N	4023	2561	75118	4034	2576	74935	4016	2566	74980

C. Estimated Expected Cell Counts After Unobserved First Stage - LFS Data

		Models A, B, C			Model D			Model E		
		E	U	N	E	U	N	E	U	N
8/79 - 9/79	E	10606	148	760	10598	158	759	10590	148	760
	U	263	386	179	256	401	174	275	402	188
	N	294	190	6809	293	202	6793	293	189	6790

D. Estimated Probabilities of Non-Response - CPS Data

Months	Model A $\hat{\lambda}_{t-1(E)}$	$\hat{\lambda}_{t-1(U)}$	$\hat{\lambda}_{t-1(N)}$	$\hat{\lambda}_{t(E)}$	$\hat{\lambda}_{t(U)}$	$\hat{\lambda}_{t(N)}$	Model B $\hat{\lambda}_{t-1}$	$\hat{\lambda}_{t}$	Model C $\hat{\lambda}_{(E)}$	$\hat{\lambda}_{(U)}$	$\hat{\lambda}_{(N)}$
12/81 - 1/82	.064	.108	.065	.064	.089	.060	.067	.064	.064	.098	.063
1/82 - 2/82	.063	.097	.057	.061	.095	.059	.063	.062	.062	.086	.058
2/82 - 3/82	.064	.087	.061	.069	.099	.063	.064	.068	.066	.093	.062
3/82 - 4/82	.073	.108	.072	.067	.113	.061	.075	.068	.070	.110	.066
4/82 - 5/82	.066	.089	.062	.064	.089	.062	.066	.065	.065	.089	.062
5/82 - 6/82	.069	.107	.063	.070	.087	.065	.069	.069	.069	.097	.064
6/82 - 7/82	.076	.106	.066	.077	.102	.069	.074	.076	.077	.104	.067
7/82 - 8/82	.074	.102	.067	.072	.099	.060	.073	.069	.073	.100	.063
8/82 - 9/82	.079	.108	.075	.076	.111	.066	.080	.075	.078	.110	.070
9/82 - 10/82	.070	.102	.063	.069	.088	.060	.070	.067	.070	.095	.061
10/82 - 11/82	.066	.093	.058	.064	.084	.055	.065	.062	.065	.089	.056
11/82 - 12/82	.059	.091	.057	.061	.077	.054	.061	.060	.060	.084	.055

Months	Model D $\hat{\lambda}_{t-1(E)}$	$\hat{\lambda}_{t-1(U)}$	$\hat{\lambda}_{t-1(N)}$	$\hat{\lambda}_{t(E)}$	$\hat{\lambda}_{t(U)}$	$\hat{\lambda}_{t(N)}$	Model E $\hat{\lambda}_{(E)}$	$\hat{\lambda}_{(U)}$	$\hat{\lambda}_{(N)}$
12/81 - 1/82	.063	.122	.064	.063	.111	.059	.063	.107	.063
1/82 - 2/82	.062	.113	.055	.060	.114	.057	.062	.101	.057
2/82 - 3/82	.063	.097	.060	.068	.115	.061	.066	.096	.061
3/82 - 4/82	.072	.124	.070	.066	.134	.058	.070	.116	.066
4/82 - 5/82	.066	.101	.060	.064	.101	.061	.065	.093	.061
5/82 - 6/82	.068	.124	.061	.070	.101	.064	.069	.103	.063
6/82 - 7/82	.075	.124	.063	.077	.115	.067	.077	.109	.066
7/82 - 8/82	.073	.120	.065	.071	.112	.058	.072	.104	.063
8/82 - 9/82	.079	.122	.073	.075	.137	.063	.077	.117	.070
9/82 - 10/82	.070	.118	.061	.069	.099	.058	.070	.098	.061
10/82 - 11/82	.066	.103	.057	.063	.096	.053	.065	.093	.056
11/82 - 12/82	.058	.103	.055	.061	.085	.052	.060	.088	.055

E. Esimated Probabilities of Non-Response - LFS Data

Months	Model A $\hat{\lambda}_{t-1(E)}$	$\hat{\lambda}_{t-1(U)}$	$\hat{\lambda}_{t-1(N)}$	$\hat{\lambda}_{t(E)}$	$\hat{\lambda}_{t(U)}$	$\hat{\lambda}_{t(N)}$	Model B $\hat{\lambda}_{t-1}$	$\hat{\lambda}_{t}$	Model C $\hat{\lambda}_{(E)}$	$\hat{\lambda}_{(U)}$	$\hat{\lambda}_{(N)}$
8/79 - 9/79	.089	.095	.087	.041	.071	.040	.089	.042	.065	.083	.064
9/79 - 10/79	.049	.079	.044	.042	.063	.054	.048	.048	.046	.071	.049
10/79 - 11/79	.038	.069	.034	.036	.071	.036	.038	.037	.037	.070	.035
11/79 - 12/79	.027	.069	.028	.035	.062	.030	.029	.034	.031	.066	.029
12/79 - 1/80	.024	.055	.023	.032	.049	.030	.025	.032	.028	.052	.027

Months	Model D $\hat{\lambda}_{t-1(E)}$	$\hat{\lambda}_{t-1(U)}$	$\hat{\lambda}_{t-1(N)}$	$\hat{\lambda}_{t(E)}$	$\hat{\lambda}_{t(U)}$	$\hat{\lambda}_{t(N)}$	Model E $\hat{\lambda}_{(E)}$	$\hat{\lambda}_{(U)}$	$\hat{\lambda}_{(N)}$
8/79 - 9/79	.089	.095	.087	.040	.103	.038	.065	.105	.062
9/79 - 10/79	.049	.108	.042	.041	.073	.055	.045	.090	.048
10/79 - 11/79	.037	.092	.033	.035	.093	.035	.036	.092	.034
11/79 - 12/79	.026	.100	.026	.034	.077	.029	.030	.089	.028
12/79 - 1/80	.023	.078	.022	.032	.059	.030	.027	.068	.026