ECONOMIC GROWTH CENTER

YALE UNIVERSITY

Box 1987, Yale Station
New Haven, Connecticut 06520

CENTER DISCUSSION PAPER NO. 540

THE ISSUE OF WEIGHTS IN PANEL SURVEYS OF INDIVIDUAL BEHAVIOR

Jan M. Hoem

University of Stockholm

July 1987

Notes: Center Discussion Papers are preliminary materials circulated to stimulate discussion and critical comments. References in publications to Discussion Papers should be cleared with the authors to protect the tentative character of these papers.

Professor Hoem was a Visiting Fellow at the Economic Growth Center in the fall of 1984.

Abstract

This paper argues that sample-based analyses of individual longitudinal behavior can normally do well without sampling weights. Instead of worrying about such weights, it pays to concentrate on the modelling of behavior and on drawing inference about features of the model. One should not feel confined to finite population totals and means, finite population regression coefficients, and other finite population statistics. Also, some of the claims about the good properties of conventional weighting seem exaggerated.

Jan M. Hoem is Professor of Demometry, Section of Demography, University of Stockholm, S-106 Stockholm, Sweden.

He has benefited from discussions with Ralph Folsom, D. Holt, Graham Kalton, Jan Lanke, T. M. F. Smith and T. N. Srinivasan, and Rick Williams on topics in this paper; as well as from editorial advice from Anders Klevmarken.

Contents

1. INTRODUCTION

Many social scientists and most survey statisticians display great unease when anyone analyzes survey data without using conventional sampling weights. Such unease is certainly warranted when the analysis consists in estimating population-level statistics in the finite population from which the sample was drawn. Currently, however, the feeling also extends to cases where it is misplaced, for instance to situations where the analysis is evidently based on probabilistic models of human behavior, say, and where you must really force your imagination to find counterparts to model parameters and other characteristics in finite population statistics. This is particularly clear in sample-based event-history analyses and panel studies. It is hard to see what meaningful finite population statistics are estimated by occurrence/exposure rates, Nelson-Aalen plots, logistic regression coefficients, or the estimated coefficients of hazard regression analysis, all of which are used with samples as well as with complete sets of life history segments. Introducing sampling weights in such analyses can complicate matters, for the standard statistical theory of inference procedures of this nature then collapses and special theory must be applied. Such theory has been developed for models of contingency tables (see Rao and Scott 1984; Smith 1984, Section 4.2; and their references; for a computer program, see Fay 1982) and indeed for generalized linear models (Binder 1983; Chambless and Boyle 1985). Some conscientious empirical investigators have computed values for estimators and test statistics both with and without weights, and have often been relieved to find that the outcomes have largely been the same (Schirm et al. 1982; Rindfuss, Swicegood, and Rosenfeld 1986, footnote 1; and others). The position taken in this paper is that such worries and such computing exercises can be superfluous in situations where the investigator is really involved in modelling human behavior rather than in calculating descriptive statistics for the finite population, and that they may divert attention away from more important concerns of modelling and analysis.

In the considerable disagreement on the issue of weighting in the current statistical literature, there is a standing dispute between those who would apparently really like to see standard weights (reciprocal selection probabilities) applied in "most" circumstances and others who feel that «if the analyst wishes to use the [sample] data to estimate a properly specified model, then the case for weighting is much weaker, since the model presumably "controls for" the effects of the factors which lead to the need for weights in the first place» except apparently for particular dependent variables in the model. (This typical quotation from the PSID User Guide, 1983, p. A13, catches the spirit of many other formulations in the literature.) Some weighting protagonists seem to see the latter position as quite lenient in this spectrum, perhaps as a concession to model builders. To me (and to some others), even that standpoint is too much in favour of the use of sampling weights in the situations of this paper. Our approach is to follow general statistical notions, to regard the sampling mechanism as part of the total model of the "random experiment" which produces the sample data, and to incorporate it into the likelihood, with normal consequences for the statistical analysis.

Since I take issue with much that is found in contributions from sample theory practitioners and teachers (Kalton 1981; O'Muircheartaigh and Wong 1981; Hansen, Madow and Tepping 1983; the PSID User Guide 1983; Kish 1981; and others), I should perhaps make clear from the outset that I realize that they do not all have the same position on all relevant issues, that opinions may develop over time, and that no one can be expected to present the full breadth of his reasoning on any single occasion, let alone the reasoning of others. There is no collective responsibility for arguments and recommendations presented. Let me also state unequivocally that I do not question the appropriateness of common weighting procedures in inference to finite population statistics. However, I want to line up with those who feel that the thinking on the matters raised here has been unduly dominated by the spirit of finite population descriptions. One cannot allow comments on the modelling of behavior to be confined to brief asides (see Section 5.1 of Hansen, Madow and Tepping, 1983, for a typical example), nor is it sensible to relegate infinite population modelling concepts to the role of motivators of definitions of new finite population parameters to be studied by the design-based approach, the way Binder (1983) and Chambless and Boyle (1985) as well as Folsom, LaVange, and Williams (1986) invite us to do.

The modelling approach induces us to focus on issues that have received insufficient attention or less than lucid treatment in the sampling literature. For instance, it is important to distinguish between the various elements of the comprehensive model of the real-life phenomenon investigated by means of the sample data. Likewise, one needs to keep these elements separate from the various statistical procedures available and from the functions which the procedures have in the analysis. We have in mind collections of records of segments of individual life histories, so one part of the total model is the sub-model of individual behavior. Sub-model misspecification is one issue and the use of sampling weights is another, and an operative connection between them remains to be demonstrated. Sampling weights have not been devised to correct for or protect against such misspecification in model-based analysis, and we know of no proof that they can serve this function in general, as seems implied by many formulations in the literature. Whether the sampling mechanism is informative (i.e., whether it depends on the random outcome of the real-life phenomena under analysis) is a separate question again. For instance, the examples offered by Hansen, Madow and Tepping (1983, Section 2) and Duncan (1982, Appendix 2) in support of the supremacy of the design-based approach have sampling mechanisms that are manifestly informative and therefore are not relevant for the issue of robustness against behavioral model misspecification nor for the separate issue of robustness against informatory status.

The informatory status of the sampling plan depends on the model, for the model defines which variables are seen as stochastic. Model misspecification at this level may result in an unrealistic declaration that the sampling plan is noninformative when it is not, which may lead into the well-known dangers of outcome-based sampling. In this connection, weighting does seem to have a function as a guard against model misspecification in certain cases. Holt, Smith and Winter (1980) and Nathan and Holt (1980) have shown that weighting may produce a robust though inefficient estimator for a linear regression coefficient when the sampling plan is informative. (See also Jewell 1985.) Unfortunately, one does not seem to really know why this is so nor to what extent current results can be generalized.

Outcome-based sampling does not appear as much of a problem in the selection of the initial target sample for prospective panel surveys of individuals, for the sample is typically drawn at the beginning of the observational period (at "time 0") and therefore just cannot be influenced by the later behavior of the (potential) respondents. On the other hand, since the extent of nonresponse may depend on such behavior, it may introduce an element of outcome-dependence in the effective sample, as is well recognized. (See for instance Fay 1986.) This gives leeway to all the usual ingenuity of survey samplers in estimating nonresponse probabilities and applying their reciprocals to the various response groups, but in itself it gives no opening for the reciprocals of the sampling probabilities of the target sample.

This reserved attitude to the use of sampling weights is not weakened by the fact that it is sensible to use all information available about the members of the target population at the time of sample selection. If we condition on whatever has happened up to and including time 0 and concentrate on investigating whatever happens after that point, the sampling plan can be outcome-independent even if population characteristics at time 0 is used extensively when the sample is drawn. To the extent that such "starting data" have a bearing on the topic of the investigation, the information should be included in the model and thus be one of the guides of subsequent analysis. Post-stratification into behaviorally distinct groups can be sensible, across prior stratum boundaries if this is suitable. Concomitant variables should be exploited as usual. (See Sugden and Smith 1984, for a discussion of problems which arise if the analyst has less information than the sampler.)

It is important to be careful about features like post-stratification according to behavior or other outcomes after time 0, however. Such procedures are prone to introduce selection biases of a form characteristic of outcome-dependent sampling. In an example below, we discuss the role of sampling weights in counteracting such biases in panel studies.

One must exercise similar care in the analysis of data for respondents who are included in the sample after time 0. In general, it will be unproblematic to utilize individual-level data for periods after entry of such individuals provided they are homogeneous with those who have been in the study from the outset. (Technically speaking, entry should be representable by a non-informative left-truncation mechanism; see Wellek 1986, and Keiding and Gill 1987, Section 5a.) The use of retrospective data for periods before entry may be another matter when entry into the sample is part of the life course outcome under analysis, as Lillard (1986, final page) evidently suspects. In the PSID, for instance, non-sample individuals may enter the sample because they become members of existing sample households. When such entry is through marriage to a sample household member, then the analysis of the process leading up to marriage needs to take into account that it actually ends in the upcoming marriage. Sample entry then represents a form of outcome-based sampling. (See Hoem 1969, and Keiding and Gill 1987, for some discussion of technical aspects.) If one is able to compute the real inclusion probability of such an entrant, there is a legitimate place for reciprocal probability weights in the analysis (Hoem 1985, Section 2.2). This probability will be a much more complex entity than merely the initial inclusion probability of the household entered, however.

The panel medium is typically geared to the collection of data about the respondents' situation at fixed times during the observational period. This sets the stage for a particularly simple model presentation, and we use the time-discrete Markov chain as an uncomplicated prototype of panel models to convey the essentials of our notions with a minimum of disturbance by mathematical or circumstantial complexity. We extend this simplicity further by mostly assuming that data collection is by a two-wave panel only. In practical applications, more complex models and more extensive observational plans are bound to be needed. (A broader catalogue over various issues of design and usages of panel data has been given by Duncan and Kalton, 1985.) Indeed, much richer analyses can be made if data are obtained for continuous life histories, for then the whole tool-bag of event-history studies is available. This is particularly useful if the timing of events or the duration of spells are important for an understanding of the dynamics of behavior, as it is bound to be in most fields. (See Allison, 1982, for a discussion of the pros and cons of continuous-time and discrete-time methods.) The panel vehicle can be used to obtain retrospective information for periods before time 0 and between other times of data collection, and is used in this manner by some major data agencies. The issue of weighting does not change in character by such an extension.

Much of the reasoning presented here has basically been given before, though usually with different emphases and not with the special issues of panel studies in mind. Some recent references are Fienberg (1980, pp. 335-338), Little (1982), Smith (1983, 1984), and Hoem (1985). Lest I be accused like so many others of being a meddling theorist with no ground contact and therefore nothing useful to tell sampling practitioners (cf. O'Muircheartaigh and Wong 1981, p. 487; Kish 1981), let me note as a credential that I got interested in the weighting issue in connection with my own <u>empirical</u> research in demography based on panel and event-history data, and as a discussion partner with a number of colleagues working in demography, sociology, economics, and epidemiology.

2. SAMPLING MARKOV CHAIN SAMPLE PATHS

2.1 Framework

Here is the very simple mathematical framework in which our arguments will be dressed. Assume that the N individuals of a population move independently between states in the finite state space $\mathcal{J}$ of a time-homogeneous Markov chain model. Suppose that the state x_t occupied by individual i at time t is observed at discrete times t = 0, 1, 2, $\cdots$, and let the unit time transition probability be

$$p_{jk} = P\{x_{t+1} = k | x_t = j\}.$$

Suppose to begin with that x_t is observed for each individual only at times 0 and 1, and regard x_0 as exogenous, i.e., as determined before the "experiment" whose outcome is observed by the investigator. Let $\chi_i(j) = 1$ if individual i is in state j at time 0, and let $\chi_i(j) = 0$ otherwise. Then the $\chi_i(j)$ are nonstochastic indicators, and the number

$$N_j = \sum_{i=1}^{N} \chi_i(j)$$

of population members in state j at time 0 is also nonstochastic and perhaps known to the investigator. Let N_{jk} of the latter be in state k at time 1, and suppose for now that the $\{p_{jk}\}$ have no particular structure, i.e., that all that is essentially known about these probabilities is that $\Sigma_{k\in\mathcal{J}}\, p_{jk} = 1$ for all $j\in\mathcal{J}$, with the possible exception that some transitions may be impossible and the corresponding p_{jk} may equal 0. If all population data $\{N_{jk}\}$ were available, then the maximum likelihood estimator of a (nonzero) p_{jk} would of course be the multinomial proportion

$$\overset{*}{p}_{jk} = N_{jk}/N_j, \qquad (2.1)$$

and we would have

$$\text{var}\ \overset{*}{p}_{jk} = p_{jk}\ (1-p_{jk})/N_j. \qquad (2.2)$$

Now suppose, however, that a sample S of individuals is drawn at time 0 according to some known sampling plan $p(s) = P\{S=s\}$. (If sampling is with replacement, then our S is the sample after the removal of any doubles.) The probability that individual i is a member of S is $\pi_i = \Sigma_{\{s:\, i\in s\}}\ p(s)$, which we assume to be positive for all i. Suppose that the estimation of the $\{p_{jk}\}$ is to be based on the sample data

$$\mathcal{D} = \{S;\ [N_{ijk}:\ i\in S,\ j\in\mathcal{J},\ k\in\mathcal{J}]\},$$

where $N_{ijk} = 1$ if individual i is in state j at time 0 as well as in state k at time 1, with $N_{ijk} = 0$ otherwise. Assume that the individual transitions after time 0 are independent of membership in the sample, i.e., that the sampling (and subsequent observation) does not affect individual behavior. Then strict adherence to conventional sampling theory would lead to the estimation of N_{ij} by

$$\tilde{N}_{jk} = \sum_{i\in S} N_{ijk}/\pi_i,$$

and correspondingly to $\tilde{\overset{*}{p}}_{jk} = \tilde{N}_{jk}/N_j$ as an estimator (predictor) of the population statistic $\overset{*}{p}_{jk}$ (if N_j is known).

In a superpopulationist vein one may note that

$$E\{N_{ijk}|S=s\} = \chi_i(j)\ p_{jk}\ \text{ if } i\in S. \qquad (2.3)$$

If we define

$$\tilde{N}_j = \sum_{i\in S} \chi_i(j)/\pi_i,$$

therefore,

$$E(\tilde{p}^{*}_{jk}|S=s) = p_{jk}\,\tilde{N}_j/N_j,$$

so $\tilde{p}^{*}_{jk}$ is not an unbiased estimator for the parameter p_{jk} when the sample is given, in contrast to, say,

$$\tilde{p}_{jk} = \tilde{N}_{jk}/\tilde{N}_j, \tag{2.4}$$

which may be used even if N_j is unknown to the investigator. On the other hand,

$$E\tilde{N}_j = \sum_s p(s) \sum_{i\in s} \chi_i(j)/\pi_i = \sum_{i=1}^{N} \chi_i(j)\, \pi_i^{-1} \sum_{\{s:\, i\in s\}} p(s) = N_j,$$

so

$$E\tilde{p}^{*}_{jk} = E\tilde{p}_{jk} = p_{jk}$$

when E denotes the expectation operator in the model which accounts for the randomness of <u>both</u> S <u>and</u> the population data $\{N_{ijk} : i = 1, \cdots, N; j\in\mathcal{J}; k\in\mathcal{J}\}$, unless a conditional expectation is indicated explicitly. In the total model, therefore, (i.e., when all currently random elements are included), both estimators $\tilde{p}^{*}_{jk}$ and $\tilde{p}_{jk}$ are unbiased.

The two-wave set-up does not really exploit the Markovian properties of the chain model. All we have used so far is its notation. In reality, we are only dealing with a set of related contingency tables, one for each starting state $j\in\mathcal{J}$. As we noted at the beginning of Section 1, the mathematics for dealing with inference in such models, with or without weighting, is already available.

2.2 The Likelihood Approach

These estimators are based on survey sampling notions and their superpopulationist extension. A classical statistical approach would be to establish the total likelihood corresponding to the observed data $\mathcal{D}$, and to maximize it. The likelihood is $\Lambda' = p(S)\,\Lambda$, where

$$\Lambda = \prod_{i\in S} \prod_{j\in\mathcal{J}} \prod_{k\in\mathcal{J}} [p_{jk}]^{N(i,j,k)}, \tag{2.5}$$

with $N(i,j,k) = N_{ijk}$ in the exponent. If $N_S(j,k) = \sum_{i\in S} N_{ijk}$ is the number of

$j \to k$ transitions observed in the <u>sample</u>, and if

$$n_S(j) = \sum_{k \in \mathcal{J}} N_S(j,k) = \sum_{i \in S} \chi_i(j) \qquad (2.6)$$

is the number of <u>sample</u> members who start out in state j at time 0, then the MLE of p_{jk} is

$$\hat{p}_{jk} = N_S(j,k)/n_S(j), \qquad (2.7)$$

provided that $p(\cdot)$ is functionally independent of the $\{p_{jk}\}$, as we will assume throughout. Note that (2.7) has precisely the same structure as (2.1), i.e., $\hat{p}_{jk}$ is the estimator which elementary statistical theory will lead to if the sample is treated as if it where the whole population. Beyond the fact that the sample is the vehicle which provides the data, the form of the estimator is not influenced by the sampling mechanism. In particular, no sampling weights are involved.

By (2.3) and (2.6), we easily derive the unbiasedness results

$$E\{\hat{p}_{jk} | S=s\} = E\{\hat{p}_{jk}\} = p_{jk}.$$

In parallel with (2.2),

$$\operatorname{var}\{\hat{p}_{jk} | S=s\} = p_{jk}(1-p_{jk})/n_S(j),$$

and consequently

$$\operatorname{var} \hat{p}_{jk} = p_{jk}(1-p_{jk})\, E\{1/n_S(j)\}, \qquad (2.8)$$

with similar results for covariances. Formula (2.8) shows that the <u>properties</u> of $\hat{p}_{jk}$ are certainly influenced by the sampling mechanism, for $p(\cdot)$ determines the final item in the formula. The likelihood approach allows us to construct estimators whose <u>form</u> is not influenced by the sampling, but we cannot ignore the fact that a sample has been drawn when we study (unconditional) estimator <u>properties</u>.

2.3 A Weighted "Likelihood"

The likelihood approach is sometimes interpreted in a manner different from the one which lead to (2.5), namely as follows. (See , e.g, Chambless and Boyle 1985.) If all population data were available, then the likelihood would be

$$\Lambda_{tot} = \prod_{j \in \mathcal{J}} \prod_{k \in \mathcal{J}} [p_{jk}]^{N(j,k)},$$

with $N(j,k) = N_{jk}$. When one is restricted to the sample data $\mathcal{D}$, then each member i in the <u>sample</u> "represents" $1/\pi_i$ of the members in the <u>population</u>. It seems logical then to maximize an "estimate" of Λ_{tot} given by

$$\tilde{\Lambda}_{tot} = \prod_{i \in S} (\Lambda_i)^{1/\pi(i)},$$

where $\pi(i) = \pi_i$, and where

$$\Lambda_i = \prod_{j \in \mathcal{J}} \prod_{k \in \mathcal{J}} [p_{jk}]^{N(i,j,k)}$$

is the likelihood contribution of individual i. Such maximization leads to the estimator $\tilde{p}_{jk}$ of (2.4), which thus gets a kind of legitimization as a pseudo-maximum-likelihood estimator. However, given S, $\hat{p}_{jk}$ of (2.7) is known from general theory to have minimal variance among unbiased estimators, and this property carries over to the unconditional variance in (2.8). Using $\tilde{p}_{jk}$ instead of $\hat{p}_{jk}$ must entail some loss in efficiency.

2.4 Concomitant Information

We have stuck to the very simple situation above to minimize the effort of presentation. The main outcome of our argument is retained in cases with more extensive observational plans or a more complex structure in the transition probabilities. Assume for instance that for individual i, the $j \to k$ transition probability is $p_{ijk} = \varphi_{jk}(z_i, \theta)$, where z_i is this individual's value on a vector of concomitant variables; θ is some unknown multidimensional parameter; and the function φ_{jk} may perhaps specify a logistic regression, it may have a simple form as in our Section 4 below, or it may be of some quite different complexity.

Assume that the z_i are exogenous and that the sampling mechanism is independent of θ, though it may depend on $z_1, \cdots, z_N$, perhaps through some system of stratification of the members of the population. Then the likelihood of $\mathcal{D}$ continues to have the form $\Lambda' = p(S)\Lambda$, where Λ is given in (2.5), except that p_{jk} is now replaced by $\varphi_{jk}(z_i, \theta)$. Maximization again proceeds without regard to any sampling weights $\{1/\pi_i\}$ and the MLE for θ is constructed as if the sample members constitute the whole population of interest.

Some investigations will oversample certain minorities. This in itself is hardly a sufficient reason to use sampling weights in the estimation of parameters of behavioral models. Let us distinguish three situations:

(i) If the behavior of the minority is the same as that of other people, then applying reciprocal sampling probabilities just gives more weight to some observations than to others <u>of the same kind</u>, which cannot be efficient under any approach.

(ii) If minority behavior differs from other behavior (and of course knowledge or a suspicion of this is the reason why they were oversampled in the first place), then it should be reflected in sufficiently accurate modelling. Either the model has one or more parameters whose values are different for the minority, in which case likelihood maximization (or something similar) will pick up these differentials. Or alternatively a different model is needed for the minority, in which case separate analysis is more sensible. Who has a proof that weighting can overcome the inferential errors of an inadequate model which tries to account for behavioral differences?

(iii) If minority behavior differs from other behavior and a model is fit which does <u>not</u> have features to pick this up, then the model is misspecified and weighting will not of be much use. Instead of parameter values (or separate models) which identify the behavioral differences between population groups, one gets fitted parameter values which represent some fictitious "mean behavior" which no group has. One has then lost sight of interesting behavioral differentials. Modelling them is more sensible.

Note that we discuss the role of the sampling weights as an issue separate from the question whether the model φ_{jk} is correctly specified. The latter question is certainly important for the empirical analysis, but it must be addressed directly. There is no a priori reason why the use of sampling weights can be expected to compensate for an incorrect specification of individual behavior. Anyone who feels that weights may give some protection against model misspecification of this kind should demonstrate it and explore why it works of it does.

2.5 Three Waves

The character of these arguments does not change if the observational plan is more extensive than the one above. To make a single step in such a direction, let us revert to non-specified transition probabilities p_{jk} of a Markov chain but let the state x_t of each individual be observed at time 2 as well as at time 0 and 1. For sample member i, let $N(i, j, k, \ell) = 1$ if this individual has the state sequence $x_0=j$, $x_1=k$, and $x_2=\ell$, and let $N(i, j, k, \ell) = 0$ otherwise. With a time-homogeneous Markov chain model, the likelihood of the sample data now becomes

$$p(S) \prod_{i \in S} \prod_{j \in ?} \prod_{k \in ?} \prod_{\ell \in ?} \{p_{jk} \, p_{k\ell}\}^{N(i, j, k, \ell)}$$

which is maximized by

$$\hat{p}_{jk} = \frac{N_S(j,k) + \Sigma_{i \in S} \, \Sigma_{\ell \in ?} \, N(i, \ell, j, k)}{n_S(j) + \Sigma_{i \in S} \, \Sigma_{\ell \in ?} \, \Sigma_{k \in ?} \, N(i, \ell, j, k)} . \qquad (2.9)$$

Properties of this kind of estimator were studied by Anderson and Goodman (1957). There is no role for sampling weights $\{1/\pi_i\}$ in this estimation procedure either.

In what follows, we revert to the case where individuals are observed in two waves only.

3. UNIT NONRESPONSE.

Let us now address the issue of characteristic-dependent nonresponse. Consider the simple Markov chain model again, let sample individuals be observed at times 0 and 1, and let us make the assumptions

(i) that whether an individual responds at time 1 is independent of the outcome at time 0 as well as of the transition behavior between times 0 and 1,

(ii) that both at time 0 and 1 each individual in state j has a response probability of β_j, and

(iii) that individuals choose to respond or abstain independently of each other.

The response model above is sufficient to serve as an illustration for our purposes. In practice, a more complex response model will surely be needed. For instance, one must often expect the response outcome at time 1 to depend on what has happened before that time. For a more complete model in a three-state set-up, see Stasny (1986a, b) and her references. Marini, Olsen, and Rubin (1979) study a situation with normally distributed variables. Both papers use the maximum likelihood approach and no weighting.

To establish a likelihood, we introduce $A(i,t)$, which equals 1 if individual i responds when approached at time t, and which equals 0 if this individual is a nonrespondent at time t. (A is for "answer".) Suppose that the state at time 0 is known for all (sample) members; the state at time 1 is obtained only for respondents. Then $\{X_i(j) : i \in S,\ j \in \mathcal{J}\}$ is exogenous and the sample data consist of S, $\{A(i,t) : i \in S;\ t=0,1\}$, and $\{N(i,j,k) : i \in S;\ j \in \mathcal{J};\ k \in \mathcal{J};\ A(i,1) = 1\}$. The likelihood of these data is

$$p(S) \prod_{i \in S} \prod_{j \in \mathcal{J}} \Big[[\beta_j]^{A(i,0)} \, [1-\beta_j]^{1-A(i,0)} \cdot$$

$$\cdot \Big(\prod_{k \in \mathcal{J}} [p_{jk} \, \beta_k]^{A(i,1)N(i,j,k)} \Big) \, [\gamma_j]^{1-A(i,1)} \Big]^{\chi_i(j)},$$

where

$$\gamma_j = 1 - \sum_{k \in \mathcal{J}} p_{jk} \, \beta_k$$

is the probability that an individual who is in state j at time 0 will be a nonrespondent at time 1. We introduce $\alpha_{jk} = p_{jk} \, \beta_k$, see that $\gamma_j = 1 - \Sigma_k \, \alpha_{jk}$, and reorganize the likelihood, which becomes

$$p(S) \Big[\prod_{j \in \mathcal{J}} [\beta_j]^{R(j,0)} \, [1-\beta_j]^{n(j)-R(j,0)} \Big] \cdot$$

$$\cdot \prod_{j \in \mathcal{J}} \prod_{k \in \mathcal{J}} \Big[[\alpha_{jk}]^{T(j,k)} \, [\gamma_j]^{n(j)-R(j,1)} \Big],$$

where

$$R(j,t) = \sum_{i \in S} \chi_i(j) \, A(i,t)$$

is the number of respondents at time t among sample members who were in state j at time 0; where

$$T(j,k) = \sum_{i \in S} A(i,1) \, N(i,j,k)$$

is the number of $j \to k$ transitions actually observed; and where we have written n(j) for $n_S(j)$ to facilitate the typing of exponents. Since $\Sigma_k \, T(j,k) = \Sigma_{i \in S} \, A(i,1) \, \chi_i(j) = R(j,1)$, maximizing the likelihood is straightforward, and we get the MLEs

$$\hat{\beta}_j = R(j,0)/n_S(j) \qquad (3.1)$$

and

$$\hat{\alpha}_{jk} = T(j,k)/n_S(j),$$

which makes $\hat{\alpha}_{jk}/\hat{\beta}_k$ the current maximum likelihood estimator of p_{jk}. Unfortunately, these estimators do not add up to 1 when we take the sum $\Sigma_{k \in \mathcal{J}}$, so the adjusted estimator

$$\hat{P}'_{jk} = \frac{T(j,k)/\hat{\beta}_k}{\sum_{\ell \in \mathcal{J}} T(j,\ell)/\hat{\beta}_\ell} \tag{3.2}$$

is perhaps preferable. As the population size and sample size go to infinity together, the denominator of (3.2) will converge to 1 in probability under any reasonable asymptotic.

Adjustment by means of reciprocal response probabilities is of course an old trick in survey sampling analysis. It appears so easily above because we have made things simple for ourselves through our assumptions. More complex response models will lead to results of a similar nature, however, and the main message conveyed again is that sampling probabilities just do not enter into the formula for the estimator.

4. OUTCOME-DEPENDENT SAMPLING

4.1 Basic Notions

We have assumed that the sample was drawn at time 0, and of course that it could only be based on information available at that point. This is the natural situation in prospective panel surveys. In retrospective surveys, by contrast, one has the option of using whatever information is available when the sample is drawn at the end of the study period. (In a retrospective panel study, information would be obtained concerning the situation of the respondents at fixed times prior to the time of selection.) If any information concerning the respondents' behavior during the study period is used in the sample selection, then the sample S is outcome-dependent, and subsequent data analysis must be made with great care to avoid the many pitfalls inherent in such a set-up. Even if the original sample S is outcome-independent, subsequent post-stratification according to the value of an outcome-variable may introduce similar effects.

Properties of the sampling plan will generally enter into a likelihood analysis if the sampling is outcome-dependent, and they may help provide a guard against selection biases otherwise produced. In some situations, the influence of the sampling plan then works via the (reciprocal) sampling fractions in outcome-based strata.

We discuss a simple example in Sections 4.2 to 4.4 below. The formal model there goes back to Colding-Jørgensen and Simonsen (1940), and it has been used for purposes similar to ours by Aalen et al. (1980) and by Hoem and Funck Jensen (1982, Section 5.3). It is a time-continuous Markov chain model used for statistical inference from panel data. A review of such issues has been given recently by Kalbfleisch and Lawless (1985), who also address computational aspects as well as the incorporation of covariates. Among the references that they do not give, are Singer and Spilerman (1977), Singer (1981), and Allison (1982). Formulas given by Funck Jensen (1982) for transition probabilities in terms of transition intensities will be useful in such analyses.

Outcome-dependent observational plans and the biases they produce appear in many shapes in most fields of statistics and have correspondingly many names, such as length biased sampling, prevalence sampling, selection biases, restriction biases, selection by virtue of survival, purged sampling, anticipatory observation, and choice based sampling. We have reviewed them in the Markov chain setting elsewhere (Hoem and Funck Jensen 1982, Section 6; Hoem 1985, Sections 2.2 and 2.3). Cohen and Cohen (1984) recently discussed them for clinical trials. For an account of their appearance in sociology, building mainly on previous work in economics by A. S. Goldberger and J. J. Heckman, see Berk and Ray (1982) and Berk (1983). Some further references are Hoem (1969), Cosslett (1981), Manski and McFadden (1981), Vardi (1985), Rao (1985), Rindfuss, Bumpass, and Palmore (1985), and Hoem, Rennermalm, and Selmer (1986).

4.2 Example 1: Childbearing and Promotion

We now turn to the simple example in Figure 1, which (in one of its guises) reflects some central features of the interaction between childbearing and promotion of women in a bureaucratic hierarchy where the employer is not permitted to let promotion to a higher grade job be influenced by the employee's private life.

[Figure 1 about here.]

In this model, the states are denoted by a two-digit code (x, y), where x indicates whether she has a lower grade job (code $x=0$) or a higher grade job (code $x=1$). The second element y indicates whether a woman has had a child (code $y=1$) or not (code $y=0$). Thus a woman is in state $(0, 0)$ if she works in a lower grade job and has not had a child yet. At the birth of her first child, she moves to state $(0, 1)$ if she still has a lower grade job, and so on. A woman cannot have a child and move to a higher grade job at the same time. Otherwise, she can have her first child or be promoted at any time. Let (x_t, y_t) be a woman's state at time t. The transition intensities in our time-continuous Markov chain model for a particular group of women are the constant parameters λ, ψ, and φ indicated in the figure. Thus, ψ and φ are birth intensities for women in lower and higher grade jobs, respectively, and λ is the rate at which women are promoted to higher grade jobs. We assume that the bureaucratic rules ensure that the latter is not influenced by the presence or arrival of a child. (In the terminology introduced by Schweder, 1970, x_t is locally independent of y_t.) On the other hand, suppose that women in higher grade jobs may have specific motives for reducing their natality, so $\psi \geq \varphi$. We take (x_0, y_0) to be exogenous and let

$$p_{ab}(x, y) = P\{x_1=x, \ y_1=y | x_0=a, y_0=b\} \text{ for } a, b, x, y \in \mathcal{J},$$

with $\mathcal{J}=\{0,1\}$. If $\delta=\psi-\varphi$, it follows readily that

$$p_{00}(0,0) = e^{-(\lambda+\psi)}, \ p_{00}(0,1) = e^{-\lambda} \ (1-e^{-\psi}),$$

$$p_{00}(1,0) = e^{-\varphi} \ (1-e^{-(\lambda+\delta)}) \ \lambda/(\lambda+\delta),$$

$$p_{01}(0,1) = e^{-\lambda}, \ p_{10}(1,0) = e^{-\varphi},$$

and so on. In particular,

$$P\{x_1=1 | x_0=0, \ y_0=b\} = 1-e^{-\lambda} \text{ for } b = 0, 1, \qquad (4.1)$$

i.e., the probability of getting promoted to a higher grade job by time 1 for a woman who is not there at time 0 equals $1-e^{-\lambda}$, irrespective of her childbearing status at time 0.

Assume that a Scandinavian type population register is available, so that the target population may be stratified by childbearing status whenever needed, and suppose that information on job status at times 0 and 1 is collected from the members of a sample. Assume that the respondents are grouped according to childbearing status at <u>time 1</u>, either because the sample was selected this way in the first place or through post-stratification when the data are prepared for analysis. As part of the investigation, one may then estimate transition probabilities for the promotion variable x_t, given the outcome on the childbearing variable, i.e., conditional probabilities of the form

$$\eta_{by} = P\{x_1=1 | x_0=0, \ y_0=b, \ y_1=y\} = p_{0b}(1,y)/[p_{0b}(0,y)+p_{0b}(1,y)].$$

As we demonstrate below, it turns out that

$$\eta_{01} < \eta_{11} = 1-e^{-\lambda} < \eta_{00} \text{ when } \psi > \varphi. \qquad (4.2)$$

This looks as if the arrival of your first child $(y_0=0, y_1=1)$ reduces your chances of getting promoted to a higher grade job over the unit time period, and as if <u>not</u> having a child $(y_0=0, y_1=0)$ improves those chances, despite the fact that we have <u>postulated</u> no influence of childbearing status on the promotion variable in our model. The childbearing-status <u>in</u>dependence of the promotion variable is well reflected in the unconditional probabilities in (4.1), but it gets garbled in the conditional transition probabilities, as is apparent from (4.2) The analysis based on such post-stratification easily induces the investigator to conclude that the arrival of a first child is a hindrance to further promotion even when it is not. Conditioning on the outcome of one variable in the investigation of another in life course analysis is a risky business.

To prove (4.2), it simplifies matters to study $\bar{\eta}_{by} = 1-\eta_{by}$ and to demonstrate the equivalent relation

$$\bar{\eta}_{00} < \bar{\eta}_{11} = e^{-\lambda} = \bar{\eta}_{01} \quad \text{when } \delta>0.$$

First note that $\bar{\eta}_{00}=e^{-\lambda}\, v/w$, $\bar{\eta}_{11}=e^{-\lambda}$, and $\bar{\eta}_{01}=e^{-\lambda}\,(1-v)/(1-w)$, with $v=e^{-\psi}$ and $w=p_{00}(0,0)+p_{00}(1,0)=e^{-\psi}\, f_{\lambda}(\delta)$, where

$$f_{\lambda}(\delta) = e^{-\lambda} + (e^{\delta}-e^{-\lambda})\ \lambda/(\lambda+\delta).$$

Since both w and 1-w are positive probabilities, $0<w<1$. Because $f_{\lambda}(0)=1$ and $\partial f_{\lambda}(\delta)/\partial\delta>0$, we get $f_{\lambda}(\delta)>1$ and therefore all in all $0<v<w<1$ when $\delta>0$. From this our inequalities follow directly.

The above formulas for the $\bar{\eta}_{by}$ also show that $\eta_{00}=\eta_{11}=\eta_{01}=1-e^{-\lambda}$ if $\psi=\varphi$. This is the unconditional promotion probability in (4.1). When child-bearing is <u>not</u> influenced by job grade, therefore, one is relieved of the dangers of systematic errors in conclusions (within this model) otherwise inherent in outcome-dependent analysis. For the particular value 0 of the parameter δ, outcome-dependence is still present but the selection biases it causes vanish.

Klevmarken (1986) has an example in which not only do the selection biases vanish when a particular parameter has the value 0, but the entire outcome-dependence disappears at the same time.

4.3 Example 1 continued: Likelihood Analysis

The selection biases just described arise because the investigator supposedly has not used the full information contained in the data, as summarized by their likelihood. To see how the likelihood approach would work in this particular example, let us assume that all the N members of a target population start in state (0,0) at time 0, and introduce p_{00} = $p_{00}(0,0)$, p_{01} = $p_{00}(0,1)$, p_{10} = $p_{00}(1,0)$, and p_{11} = $p_{00}(1,1)$. Let $q = p_{00} + p_{10}$ be the probability that a woman has had no child by time 1, and let us drop the initial subscript 0 in the ηs now involved and write $\eta_y=\eta_{0y}$ for $y=0,1$. Assume that at time 1, the women are grouped into two strata, Stratum 0 for those who have had no children and Stratum 1 for the rest. Suppose that a simple random sample is then drawn from each of these strata at time 1, and that the job grade status at time 1 is obtained (without misclassification error or nonresponse) from all members of the sample. We now want to establish the likelihood of these data. We usually follow the convention of denoting random variables by capital Latin letters and their values by the corresponding lower case letters, but in the present connection this would lead us to cover up some correspondences with survey sampling theory which we want to display. For the remainder of this single section, therefore, let random variables be capitals or lower case letters typed in boldface and let their values be corresponding capitals or lower case letters given in ordinary typeface. At time 1, then, $\mathbf{M}_k$ members of the target population are in state k, for k = (0,0), (0,1), and so on. Of the members, $\mathbf{N}_0 = \mathbf{M}_{00} + \mathbf{M}_{10}$ have not had a child, and $\mathbf{N}_1=N-\mathbf{N}_0$ of them have had a child. A random sample of $\mathbf{n}_0$ out of the $\mathbf{N}_0$ is drawn from Stratum 0, and another random sample of $\mathbf{n}_1$ is drawn from the $\mathbf{N}_1$ members of Stratum 1. (Since the sample sizes are restricted to not exceed the random number of members of the two strata, they are random variables themselves.) The number $\mathbf{m}_k$ of members of the sample who are in each state k at time 1 are registered. The likelihood of obtaining these data is then the observed value of the probability

$$P\{\mathbf{N}_0=N_0,\ \mathbf{m}_{00}=m_{00},\ \mathbf{m}_{01}=m_{01},\ \mathbf{m}_{10}=m_{10},\ \text{and } \mathbf{m}_{11}=m_{11}\} = \tag{4.3}$$

$$\Sigma_0\,\Sigma_1 \frac{N!}{M_{00}!\,M_{01}!\,M_{10}!\,M_{11}!}\left[\prod_{K=(0,0)}^{(1,1)} (p_K)^{M_K}\right] \frac{\prod_{K=(0,0)}^{(1,1)} \binom{M_K}{m_K}}{\binom{M_{00}+M_{10}}{n_0}\binom{M_{01}+M_{11}}{n_1}},$$

where Σ_0 is the sum over all pairs (M_{00}, M_{01}) for which $M_{00}+M_{01}=N_0$ and Σ_1 is the sum over all pairs (M_{01}, M_{11}) for which $M_{01}+M_{11}=N_1$. We show below that the likelihood can be rewritten as

$$\binom{N}{\mathbf{N}_0} q^{\mathbf{N}_0}(1-q)^{N-\mathbf{N}_0} \binom{\mathbf{n}_0}{\mathbf{m}_{10}} (\eta_0)^{\mathbf{m}_{10}}(1-\eta_0)^{\mathbf{n}_0-\mathbf{m}_{10}} \binom{\mathbf{n}_1}{\mathbf{m}_{11}} (\eta_1)^{\mathbf{m}_{11}}(1-\eta_1)^{\mathbf{n}_1-\mathbf{m}_{11}} \tag{4.4}$$

which is maximized by the natural estimators

$$\hat{\mathbf{q}} = \mathbf{N}_0/N,\quad \hat{\boldsymbol{\eta}}_0 = \mathbf{m}_{10}/\mathbf{n}_0,\quad \text{and } \hat{\boldsymbol{\eta}}_1 = \mathbf{m}_{11}/\mathbf{n}_1.$$

We substitute these into the one-to-one relations which connect the parameters p_K with our current equivalent parameters q, η_0, and η_1, namely the relations $p_{00}=q(1-\eta_0)$, $p_{10}=q\eta_0$, and so on, and get the maximum likelihood estimators

$$\hat{\mathbf{p}}_{00} = \mathbf{m}_{00}/(N\mathbf{f}_0),\quad \hat{\mathbf{p}}_{01} = \mathbf{m}_{01}/(N\mathbf{f}_1), \tag{4.5}$$
$$\hat{\mathbf{p}}_{10} = \mathbf{m}_{10}/(N\mathbf{f}_0),\quad \text{and } \hat{\mathbf{p}}_{11} = \mathbf{m}_{11}/(N\mathbf{f}_1),$$

where $\mathbf{f}_K=\mathbf{n}_K/\mathbf{N}_K$ is the sampling fraction in Stratum K, for K=0, 1. (Let $\mathbf{f}_K=1$ in the unlikely case that $\mathbf{N}_K=0$.) Provided $\mathbf{n}_0$ and $\mathbf{n}_1$ are sensible functions of N and $\mathbf{N}_0$, as we can safely assume, these estimators will be consistent for their respective estimands as $N \to \infty$.

To bring out the close connection between these estimators and what survey sampling theory would suggest without any appeal to the likelihood approach, note that $\mathbf{X}_0=\mathbf{m}_{10}$ is the number of members of Stratum 0 who have the property of being in State (1, 0) at time 1. The corresponding number of members from Stratum 1 is $\mathbf{X}_1=0$. A Horvitz-Thompson estimator of the proportion $\mathbf{M}_{10}/N$ of the target population that is in State (1, 0) at time 1, can then be written as

$\hat{M}_{10}/N = N^{-1} \Sigma_K X_K/f_K$,

which is $\hat{p}_{10}$. (The sum is taken from k=0 to k=1.) By symmetry, the other probability estimators can be represented in a similar manner. It turns out, therefore, that in this particular case, the maximum likelihood estimators and the HT-estimators coincide. The reciprocal sampling fractions serve to balance the sampling biases otherwise inherent in the outcome-dependent sampling plan.

Note, however, that the sampling fractions f_0 and f_1 are not a priori inclusion probabilities. Every member of the target population has the same probability of ending up in the sample, and that probability is

$$q\, E(f_0) + (1-q)\, E(f_1).$$

We need to know what functions n_0 and n_1 are to compute this probability. The sampling fraction f_K is "only" the <u>conditional</u> selection probability, given that a population member has ended up in Stratum k by time 1.

To demonstrate the transition between (4.3) and (4.4), note that

$$P\{N_0 = N_0\} = \binom{N}{N_0} (q)^{N_0} (1-q)^{N-N_0} \tag{4.5}$$

and

$$P\{\bigcap_{K=(0,0)}^{(1,1)} (M_K = M_K) \mid N_0 = N_0\} \tag{4.6}$$

$$= \binom{N_0}{M_{10}} (\eta_0)^{M_{10}} (1-\eta_0)^{N_0-M_{10}} \binom{N_1}{M_{11}} (\eta_1)^{M_{11}} (1-\eta_1)^{N_1-M_{11}}.$$

Therefore, the expression in (4.3) can be written as that of (4.5) multiplied by the product of two sums corresponding to Σ_0 and Σ_1 in (4.3) After some minor rearrangement, the first sum can be written as

$$\binom{n_0}{m_{10}} \sum_{M_{10}=0}^{N_0} \binom{N_0-n_0}{M_{10}-m_{10}} (\eta_0)^{M_{10}} (1-\eta_0)^{N_0-M_{10}}$$

$$= \binom{n_0}{m_{10}} \sum_{x=-m_{10}}^{N_0-m_{10}} \binom{N_0-n_0}{x} (\eta_0)^{x+m_{10}} (1-\eta_0)^{N_0-m_{10}-x}.$$

Items in the latter sum are 0 for $x<0$ and for $x>N_0-n_0$ (remember that $m_{10} \le n_0$), so this expression reduces to

$$\binom{n_0}{m_{10}} (\eta_0)^{m_{10}} (1-\eta_0)^{n_0-m_{10}} . \qquad (4.7)$$

The sum based on Σ_1 is quite similar. Therefore, (4.4) results if you multiply together the expressions in (4.5), (4.7), and the Σ_1-based expression corresponding to (4.7), and subsequently substitute $\mathbf{N}_0$, $\mathbf{n}_0$, $\mathbf{n}_1$, $\mathbf{m}_{10}$, and $\mathbf{m}_{11}$ for N_0, n_0, n_1, m_{10}, and m_{11}.

4.4 Example 2: Maternal and Infant Mortality

The model in Figure 1 may be reinterpreted in a manner which makes it suitable for the analysis of the impact of the death of a mother on the survival of her baby in historical data. Consider a mother-and-infant pair, let x_t be 0 as long as the mother is alive, and let x_t jump to 1 if the mother dies. Similarly, let y_t be an indicator of whether the child is dead at time t. Let the interval between times 0 and 1 be such that both mother and child have (acceptably) constant mortality during it, let λ be the mother's force of mortality, and let ψ and φ be the forces of mortality of the infant before and after any death of the mother. Suppose that the two cannot die simultaneously. Assume that the child is sufficiently dependent on the mother's personal care that the infant's (force of) mortality jumps to a higher level if the mother dies, i.e., $\varphi > \psi$, but that the mother's mortality is not similarly influenced by the death of the child. Then all inequalities in (4.2) are reversed. If the data for the mother-and-infant pairs are sorted according to whether the child is alive at time 1, therefore, the investigator is invited to conclude erroneously that the death of the child adversely affects the mother's chances of survival in this model. Furthermore, the mother's survival chances will be estimated as better than they really are from data on pairs with infants surviving to time 1. Prior stratification or post-stratification according to the infants' survival status at time 1 should be avoided.

Sorting the pairs according to the <u>mother's</u> survival status at time 1 is less dangerous for conclusions about the infant's survival chances. Such grouping corresponds to working with conditional probabilities of the form

$$\zeta_{ax} = P\{y_1=1|x_0=a, y_0=0, x_1=x\} = p_{a0}(x,1)/[p_{a0}(x,0)+p_{a0}(x,1)]$$

instead of $\theta_a = P\{y_1=1|x_0=a, y_0=0\}$. A simple argument similar to the one which established (4.2) shows that

$$\zeta_{00}=1-e^{-\psi} < \theta_0 < \zeta_{01} < \zeta_{11} = \theta_1=1-e^{-\varphi} \quad \text{when } \psi<\varphi. \tag{4.6}$$

Thus, it is less dangerous for the infant to have a surviving mother than to lose her during the unit interval, and the latter event is less adverse than being without the mother throughout the interval, all of which conclusions are correct.

4.5 Intraclass Correlations in Clusters

Papers by weighting advocates contain many admonitions to use weighting procedures to counteract the adverse effects of "intraclass correlations due to cluster sampling". It remains to be demonstrated, however, how weighting can have such a function in the analysis of panel data for individual behavior, or how it can replace direct attention to intracluster interaction, as it surely is intended to do. Again, weighting must be an issue separate from that of model misspecification, which could now reappear in the guise of an "assumption" that individuals in clusters behave independently.

To illustrate interaction behavior, let us return again to Figure 1, and let the two dimensions correspond to the two members of a two-person household in a manner similar to that of our example of mother-and-infant mortality (Section 4.4). Regard person 1 as the head of the household, and disregard household dissolutions for the sake of this argument. In the model of the figure, heads of households behave independently of any partners because the intensities corresponding to the two vertical arrows are the same (namely λ). There is complete within-household independence if $\psi=\varphi$, otherwise not. A no-interaction misspecification would consist in analyzing the data as if $\psi=\varphi$ even if the two intensities really were different from each other. How can conventional sampling weights balance the biases caused by this error? Even if they could, would it not be more interesting to find out what goes on in the clusters?

5. CONCLUDING DISCUSSION

The previous sections embed some very general issues into some very straightforward settings. The question of whether to use weights in the investigation of models for panel data is itself a special case of a more general issue concerning the role of weights in any analysis of sample data which involve the parameters of statistical models. Let us reiterate some general results here for the case of panel studies.

In a population of N independent individuals, let Y_i be a description of the life course of member i over a finite set of time points $t_0=0 < t_1 < \cdots < t_m < \infty$. Since we have sample paths of time-discrete Markov chains particularly in mind, assume for simplicity that for each Y_i there are only a finite number of possible paths, let y be one of them, and introduce $\xi_i(y) = P\{Y_i=y\}$. Note that $y = \{y(t_0), \cdots, y(t_m)\}$, so the probability law ξ_i is a multi-dimensional distribution function. It reflects our notions about the possible behavior of individual i over the finite time set $\{t_k\}$. It may depend in any way whatsoever on individual-specific values of exogenous variables, including the value $Y_i(0)$ of Y_i at time 0, which we take as nonstochastic but not necessarily known to the investigator before the sample is drawn. (If real individuals come in independent clusters with internal interaction, like households, then redefine an "individual" to be such a cluster for the purposes of the general theory. If clusters can recombine over time, we need a more general framework than the present one. Conditioning on the $Y_i(0)$ is convenient and useful in panel studies, but it is not essential for our argument.)

From this population, a sample S is selected according to the sampling mechanism $p(s) = P\{S=s\}$. We allow $p(\cdot)$ to depend on the same exogenous variables as the ε_i do, including the given values $Y_i(0)$ if they are known, but $p(\cdot)$ may not depend on the outcomes of the Y_i after time 0. Essentially, the sample is drawn at time 0, and then the investigator cannot utilize anything that happens later. We assume that observation is unobtrusive, in the sense that membership in the sample does not influence an individual's behavior. (Alternatively, the investigator can have a theory for the influence of observation on behavior. The problem of obtrusive observation is common to all statistical analysis. See Section 4 of Duncan and Kalton, 1985, for a brief review of current experience with it in panel studies.) On the other hand, we allow for nonresponse by letting this feature be an integral part of Y_i, i.e., one of the possible values for $Y_i(t_k)$ at any time t_k may be an indication that data are missing because of nonresponse.

Let $S = \{I_1, \cdots, I_{n(S)}\}$ and $s = \{i_1, \cdots, i_{n(s)}\}$, and let us write $I(j)$ for I_j and $i(j)$ for i_j in subscripts. Then the sample data $\mathcal{D} = \{S;\ Y_i: i \in S\}$ have a probability distribution given by

$$P\Big[\ \{S=s\}\ \&\ \{Y_i = y_i \text{ for all } i \in S\}\ \Big]$$

$$= P\Big[\ \{n(S)=n(s)\}\ \&\ \bigcap_{j=1}^{n(s)} \{I_j = i_j\}\ \&\ \bigcap_{j=1}^{n(s)} \{Y_{I(j)} = y_{I(j)}\}\ \Big]$$

$$= P\Big[\ \{n(S)=n(s)\}\ \&\ \bigcap_{j=1}^{n(s)} \{I_j = i_j\}\ \&\ \bigcap_{j=1}^{n(s)} \{Y_{i(j)} = y_{i(j)}\}\ \Big]$$

$$= P\left[\{n(S)=n(s)\} \ \& \ \bigcap_{j=1}^{n(s)} \{I_j = i_j\} \ \middle| \ \bigcap_{j=1}^{n(s)} \{Y_{i(j)} = y_{i(j)}\}\right] \cdot P\left[\bigcap_{j=1}^{n(s)} \{Y_{i(j)} = y_{i(j)}\}\right]$$

$$= P\{S=s\} \cdot \prod_{j=1}^{n(s)} P\{Y_{i(j)} = y_{i(j)}\} = p(s) \prod_{j=1}^{n(s)} \xi_{i(j)}(y_{i(j)}),$$

when we use the stochastic independence between S and the collection $\{ Y_i : i=1, \cdots, N \}$ to reduce the big conditional probability to $P\{S=s\}$ in the next to final line above. Thus we have just proved that

$$P\{ Y_i = y_i \text{ for all } i \in S \mid S=s \} = \prod_{j=1}^{n(s)} \xi_{i(j)}(y_{i(j)}), \tag{6.1}$$

i.e., given the sample S, the distribution of the sample observations $\{Y_{i(j)}\}$ is the same as it would be in an imagined exhaustive census whose data had the same stochastic properties as those in the sample survey. The likelihood of the sample data $\mathcal{D}$ is

$$p(S) \prod_{i \in S} \xi_i(Y_i). \tag{6.2}$$

Since $p(\cdot)$ must be independent of the unknown parameters and other unknown characteristics of the $\{\xi_i\}$, likelihood maximization does not involve $p(\cdot)$ in any way, and in particular it does not involve any reciprocal inclusion probabilities $1/\pi_i$. The sample S is an ancillary statistic, and inference may be based on (6.1) alone if you follow the ancillarity principle.

The essential role of the sampling plan is to provide a randomizing vehicle to determine which life histories to include in the sample in a way which induces cost-effective analysis and helps make sure that S actually *is* stochastically independent of the life courses Y_i after time 0. Lack of such a randomizing mechanism entails the risk that S becomes informative, as well as the usual problems of generalizability of results. (See Smith 1983, and Royall, 1985, for critical assessments of the need for randomization, and Smith, 1984, for a discussion of its meaning.)

Any likelihood analysis depends of course on the specification of the $\{\xi_i\}$. If the model ξ_i of individual behavior is incorrect or unrealistic, then the outcome of the analysis must be affected unfavourably. Indeed, as everybody professes to realize, <u>any</u> model is incorrect or unrealistic in many respects. This is inherent in all analyses which use statistical models, and the analysis of sample survey data is no exception. One must not let this fact stifle one's ability to use models productively for the analysis of sample data any more than for other kinds of data. The value of a model lies in its ability to pick up important aspects of behavior and to serve as a guide to our inference about reality. Some sampling practitioners display an evident ambivalence (or even aversion) towards behavioral modelling, but an investigator interested in analyzing a particular aspect of behavior by means of sample data should not let this dictate his own choice of method. The preparation of the information in a major data set for publication in a book of official statistics, say, is quite a different operation than the penetration of a sub-area for an analytical purpose. The concerns of data producing agencies are certainly real and important enough, but there is no need let them dominate the picture the way they have done so far. There is little reason why others should feel restricted by the same considerations.

In response to the need for "infinite population" modelling concepts in the analysis of survey samples, some statisticians have begun to provide a kind of half-way house where new finite population statistics are <u>defined</u> in terms of model parameters but the properties of their estimators are <u>studied</u> as if there were no model but only the finite population. For instance, Chambless and Boyle (1985) have suggested that a parametric likelihood

$$\Lambda(\underset{\sim}{\beta}) = \prod_{i=1}^{N} \Lambda_i(\underset{\sim}{\beta})$$

that would apply to the entire finite population (of size N) under a given behavioral model, be estimated by its sample counterpart

$$\hat{\Lambda}(\beta) = \prod_{i \in S} [\Lambda_i(\beta)]^{1/\pi(i)}$$

for individual inclusion probabilities $\{\pi(i)\}$. If $\Lambda(\beta)$ is maximized when $\beta = B$, they suggest that B be regarded as the finite population quantity of interest, and that the value $\hat{B}$ of β which maximizes $\hat{\Lambda}(\beta)$ be regarded as an estimator of B. This is in line with previous suggestions for generalized linear models by Binder (1983) and his predecessors. It has been followed up by Folsom, LaVange, and Williams (1986), who have also extended the theory to the (nonparametric) Kaplan-Meier product limit estimator.

We discussed an application of these ideas to the estimation of Markov chain transition probabilities in Section 2.3. In that setting, B is a set of (unobserved) transition proportions in the finite population, which it may certainly make sense to estimate. Similarly, the (unobserved) Kaplan-Meier product limit function for the finite population describes the distribution of a positive variable over the members of that population, and again it may make sense to estimate it from the sample data. In a case-by-case consideration of other situations, we are bound to find more models where a finite population estimator B is a meaningful statistic in its own right for which the sample counterpart $\hat{B}$ is a sensible estimator. For such situations, the statistical theory developed will be useful. Admitting this is a far cry however from accepting that

(a) a finite population estimator B <u>always</u> is a meaningful descriptive statistic, irrespective of any appeal to an underlying model, or that

(b) one should restrict one's analysis of survey sample data to situations where B is meaningful in this manner, or indeed that

(c) statistical inference from survey samples must be only to the finite population level.

Individual statisticians may hold or reject any one of these views; I disagree with all three of them. I find it particularly puzzling that statisticians should insist on item (c) above, the way Ralph Folsom and Rick Williams have done in conversations during the Washington Symposium on Panel Surveys to which earlier versions of this paper and theirs were contributions.

Be that as it may, it is important to maintain that the specification of a model of individual behavior is an issue separate from the role of sampling weights. Whether the investigator has got ε_i right or not, the likelihood has the form in (6.2) so long as sampling and analysis are not outcome-dependent. Of course, there is no compulsion to rely on the likelihood approach. One is free to use any inference procedure available, subject only to the assessment of the statistical properties of the procedure. Some such procedures may involve sampling weights. In fact, properties of the sampling plan will generally enter into likelihood analysis if the sampling is outcome-based, for then the likelihood has a form like

$$\Lambda = [\prod_{i\in S} \varepsilon_i(Y_i)] \sum_{i\notin S} \sum_{y_i} \left[\{ \prod_{i\notin S} \varepsilon_i(y_i) \} \; p(S|Y_i: i\in S;\; y_i: i\notin S) \right],$$

where the double sum is taken over all $i\notin S$ and over all values y_i that the sample path of target population member i outside the sample can attain, and where $p(s|y_1, \cdots, y_N) = P\{S=s|Y_i=y_i \text{ for } i=1,\cdots,N\}$ is the conditional probability of drawing the sample s when the sample path outcomes are as specified.

In certain cases, the sampling mechanism turns out to enter the likelihood only via the sampling fractions of outcome-dependent strata. The example of our Section 4.3 is a case in point. However, weighting is no panacea which can solve most problems of survey analysis, including model misspecification, nor can it replace modelling and make behavioral models superfluous.

REFERENCES

AALEN, ODD O.; BORGAN, ØRNULF; KEIDING NIELS; and THORMANN, JENS (1980), "Interaction between life history events: Nonparametric analysis for prospective and retrospective data in the presence of censoring," Scandinavian Journal of Statistics 7, 161-171.

ALLISON, PAUL D. (1982), "Discrete-time methods for the analysis of event histories," in Sociological Methodology 1982, ed. Samuel Leinhardt, San Francisco: Jossey-Bass, 61-98.

ANDERSON, T. W. and GOODMAN L. A. (1957), "Statistical inference about Markov chains," The Annals of Mathematical Statistics, 28, 89-110.

BERK, RICHARD A. and SUBHASH C. RAY (1982), "Selection biases in sociological data," Social Science Research, 11, 352-398.

BERK, RICHARD A. (1983), "An introduction to sample selection bias," American Sociological Review, 48, 386-398.

BINDER, D. A. (1983), "On the variances of asymptotically normal estimators from complex surveys," International Statistical Review, 51, 279-292.

CHAMBLESS, LLOYD E. and BOYLE, KERRIE E. (1985), "Maximum likelihood methods for complex sample data: Logistic regression and discrete proportional hazards models," Communications in Statistics - Theory and Methods, 14, 1377-1392.

COHEN, R. and COHEN, J. (1984), "The clinicians's illusion," Archives of General Psychiatry, 41, 1178-1182.

COLDING-JØRGENSEN, H. and SIMONSEN, W. (1940), Statistical appendix in Graviditet og gallestensdannelse, by J.M. Wollesen, Copenhagen: Nyt Nordisk Forlag, 81-101.

COSSLETT, STEPHEN, R. (1981), "Maximum likelihood estimator for choice-based samples," Econometrica, 49, 1289-1316.

DUNCAN, GREG J. (1982), "The implications of changing family composition for the dynamic analysis of family economic well-being," in Panel Data on Incomes: A Selection of the Papers presented at a Conference on the Analysis of Panel Data on Incomes 24/25 June 1982 at the International Centre for Economics and Related Disciplines, 203-239.

DUNCAN, GREG J. and KALTON, GRAHAM (1985), "Issues of design and analysis of surveys across time," Bulletin of the International Statistical Institute, 51, 14.1. Also in International Statistical Review, 55 (1987), 97-117.

FAY, ROBERT (1982), "Contingency table analysis for complex sample designs: CPLX," Proceedings of the American Statistical Association, Section on Survey Research Methods, 44-53.

—— (1986), "Estimating nonignorable nonresponse in longitudinal surveys through causal modeling," Paper contributed to the International Symposium on Panel Surveys, Washington, D. C., November 1986.

FIENBERG, S. E. (1980), "The measurement of crime victimization: Prospects for panel analysis of a panel survey," The Statistician, 29, 313-350.

FOLSOM, RALPH; LAVANGE, LISA; and WILLIAMS, RICK L. (1986), "A probability sampling perspective on panel data analysis," Paper contributed to the International Symposium on Panel Surveys, Washington, D. C., November 1986.

FUNCK JENSEN, ULLA (1982), "The Feller-Kolmogorov differential equation and the state hierarchy present in models in demography and related fields," Stockholm Research Reports in Demography, No. 9, University of Stockholm, Section of Demography.

HANSEN, M. H.; MADOW, W. G.; and TEPPING, B. J. (1983), "An evaluation of model-dependent and probability-sampling inferences in sample surveys," Journal of the American Statistical Association, 78, 776-807, (with discussion).

HOEM, JAN M. (1969), "Purged and partial Markov chains," Skandinavisk Aktuarietidskrift, 52, 147-155.

—— (1985), "Weighting, misclassification, and other issues in the analysis of survey samples of life histories," in Longitudinal analysis of labor market data, James J. Heckman and Burton Singer, Cambridge University Press, 249-293. List of misprints available from the author.

HOEM, JAN M. and FUNCK JENSEN, ULLA (1982), "Multistate life table methodology: A probabilist critique," in Multidimensional Mathematical Demography, eds. Kenneth C. Land and Andrei Rogers, New York: Academic Press, 155-264.

HOEM, JAN M.; RENNERMALM, BO; and SELMER, RANDI (1986), "Restriction biases in the analysis of births and marriages to cohabiting women from data on the most recent conjugal union only," Stockholm Research Reports in Demography, No. 18 (revised), University of Stockholm, Section of Demography.

HOLT, D.; SMITH T. M. F.; and WINTER P. D. (1980), "Regression analysis of data from complex surveys," Journal of the Royal Statistical Society, Series A, 143, 474-487.

JEWELL, NICHOLAS P. (1985), "Least squares regression with data arising from stratified samples of the dependent variable," Biometrika, 72, 11-21.

KALBFLEISCH, J. D. and LAWLESS, J. F. (1985), "The analysis of panel data under a Markov assumption," Journal of the American Statistical Association, 80, 863-871.

KALTON, GRAHAM (1981), "Models in the practice of survey sampling," Bulletin of the International Statistical Institute, 49, 514-531. Also in International Statistical Review, 51 (1983), 175-188.

KEIDING, NIELS and GILL, RICHARD D. (1987), "Random truncation models and Markov processes," Research Report 87/3, Statistical Research Unit, University of Copenhagen.

KISH, L. (1981), "Contribution to discussion on conceptual and theoretical framework for survey sampling," Bulletin of the International Statistical Institute, 49, 536-539.

KLEVMARKEN, ANDERS (1986), "On the estimation of recursive and interdependent models from sample survey data," Paper contributed to the International Symposium on Panel Surveys, Washington, D. C., November 1986.

LILLARD, LEE A. (1986), "Some issues related to the dynamics of sample size in the PSID," Paper contributed to the International Symposium on Panel Surveys, Washington, D. C., November 1986.

LITTLE, RODERICK J. A. (1982), "Models for nonresponse in sample surveys," Journal of the American Statistical Association, 77, 237-250.

MANSKI, CHARLES F. and MCFADDEN, DANIEL (1981), "Alternative estimators and sample designs for discrete choice analysis," in Structural Analyses of Discrete Data, eds. C. F. Manski and D. McFadden, Cambridge, Mass: MIT Press, 1-50.

MARINI, MARGARET MOONEY; OLSEN, ANTHONY R.; and RUBIN, DONALD B. (1979), "Maximum-likelihood estimation in panel studies with missing data," in Sociological Methodology 1980, ed. Karl F. Schuessler, San Francisco: Jossey-Bass, 314-357.

NATHAN, G. and HOLT, D. (1980), "The effect of survey design on regression analysis," Journal of the Royal Statistical Society. Series B, 42, 377-386.

O'MUIRCHEARTAIGH, COLM and WONG, SOON TECK (1981), "The impact of sampling theory on survey sampling practice: A review," Bulletin of the International Statistical Institute, 49, 465-493.

PSID (1983). User Guide for the Panel Study of Income Dynamics. Preliminary draft. University of Michigan, Survey Research Center, Ann Arbor. Final version issued in 1984.

RAO, C. R. (1985), "Weighted distributions arising out of methods of ascertainment: What population does a sample represent?" in A Celebration of Statistics: The ISI Centenary Volume, eds. A. C. Atkinson and S. E. Fienberg, New York: Springer-Verlag, 543-569.

RAO, J. N. K. and SCOTT, A. J. (1984), "On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data," Annals of Statistics, 12, 46-60.

RINDFUSS, RONALD R.; BUMPASS LARRY L.; and PALMORE, JAMES A. (1985), "Analyzing selected fertility histories: Do restrictions bias results?" Demography, 24, 113-122.

RINDFUSS, RONALD R.; SWICEGOOD, C. GRAY; and ROSENFELD, RACHEL A. (1986), "Disorder in the life course: How common and does it matter?" Unpublished manuscript.

ROYALL, RICHARD (1985), "Current advances in sampling theory: Implications for human observational studies," American Journal of Epidemiology, 104, 463-474.

SCHIRM, ALLEN; TRUSSELL, JAMES; MENKEN, JANE; and GRADY, WILLIAM R. (1982), "Contraceptive failure in the United States: The impact of social, economic, and demographic factors," Family Planning Perspectives, 14, 68-75.

SCHWEDER, TORE (1970), "Composable Markov processes," Journal of Applied Probability, 7, 400-410.

SINGER, BURTON (1981), "Estimation of nonstationary Markov chains from panel data," in Sociological Methodology 1981, ed. S. Leinhardt, San Francisco: Jossey-Bass, 319-337.

SINGER, BURTON and SPILERMAN, SEYMOUR (1977), "Fitting stochastic models to longitudinal survey data: Some examples in the social sciences," Bulletin of the International Statistical Institute, 47, 283-300.

SMITH, T. M. F. (1983), "On the validity of inferences from non-random samples," Journal of the Royal Statistical Society, Series A, 146, 394-403.

—— (1984), "Present position and potential developments: Some personal views: Sample surveys," Journal of the Royal Statistical Society, Series A, 147, 208-221.

STASNY, ELIZABETH A. (1986a), "Estimating gross flows using panel data with nonresponse: An example from the Canadian labor force survey," Journal of the American Statistical Association, 81, 42-47.

—— (1986b), "Some Markov chain models for nonresponse in estimating gross labor force flows," Paper contributed to the International Symposium on Panel Surveys, Washington, D. C., November 1986.

SUGDEN, R. A. and SMITH, T. M. F. (1984), "Ignorable and informative designs in survey sampling inference," Biometrika, 71, 495-506.

VARDI, Y. (1985), "Empirical distributions in selection bias models," Annals of Statistics, 50, 178-205, (with discussion).

WELLEK, S. (1986), "A nonparametric model for product-limit estimation under right censoring and left truncation," Technical Report, Institut für Medizinische Statistik und Dokumentation der Universität Mainz.

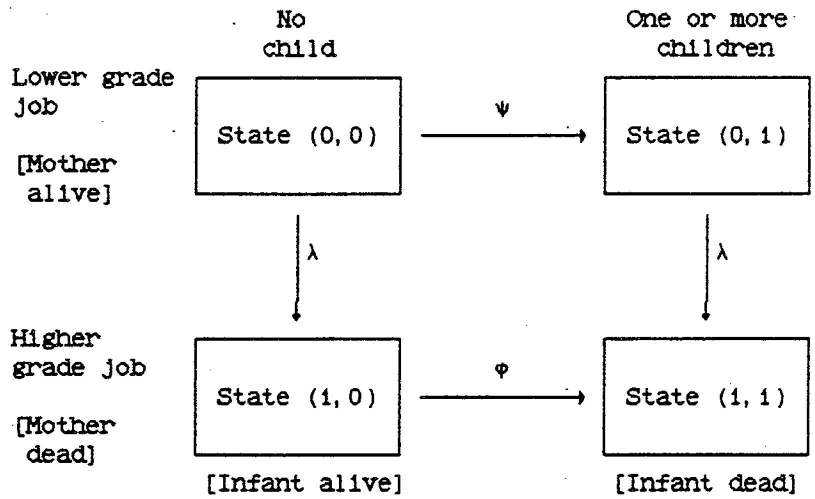

Figure 1. Markov chain caricature of first childbearing and promotion to a higher grade job in a bureaucratic hierarchy [as well as of mother-and-infant mortality].

STOCKHOLM RESEARCH REPORTS IN DEMOGRAPHY

SECTION OF DEMOGRAPHY, UNIVERSITY OF STOCKHOLM

S-106 91 STOCKHOLM, SWEDEN

SRRD-1 Hoem, Jan M. and Randi Selmer: The interaction between premarital cohabitation, marriage, and the first two births in current Danish cohorts, 1975. (March, 1982.)

SRRD-2 Hoem, Jan M. and Ulla Funck Jensen: Multistate life table methodology: A probabilistic critique. (March, 1982.) Pp. 155-264 in K. C. Land and A. Rogers (eds.), Multidimensional Mathematical Demography. New York: Academic Press.

SRRD-3 Hoem, Jan M.: Balancing bias in vital rates due to an informative sampling plan. (April, 1982.) Pp 81-88 in Statistical Review 1983:5. Essays in honour of Tore E. Dalenius.

SRRD-4 Hoem, Jan M.: Distortions caused by nonobservation of periods of cohabitation before the latest. (May, 1982.) Demography 20 (4), 491-506, 1983.

SRRD-5 Hoem, Jan M.: The reticent trio: Some little-known early discoveries in insurance mathematics by Oppermann, Thiele, and Gram. (July, 1982.) International Statistical Review 51, 213-221, 1983.

SRRD-6 Hoem, Jan M. and Bo Rennermalm: Biases in cohabitational nuptiality rates caused by nonobservation of periods of cohabitation before the latest: An empirical note. (Dec., 1982.)

SRRD-7 Funck Jensen, Ulla: An elementary derivation of moment formulas for numbers of transitions in time-continous Markov chains. (Dec., 1982.)

SRRD-8 Hoem, Jan M. and Bo Rennermalm: Cohabitation, marriage, and first birth among never-married Swedish women in cohorts born 1936-1960. (Dec., 1982.) Published as "Modern family initiation in Sweden: Experience of women born between 1936 and 1960". European Journal of Population 1, 81-112, 1985.

SRRD-9 Funck Jensen, Ulla: The Feller-Kolmogorov differential equation and the state hierarchy present in models in demography and related fields. (Dec., 1982.)

SRRD-10 Hoem, Jan M.: Multistate mathematical demography should adopt the notions of event-history analysis. (Feb., 1983.)

SRRD-11 Hoem, Jan M.: Weighting, misclassification, and other issues in the analysis of survey samples of life histories. (Feb., 1983.) Pp. 249-293 in James J. Heckman and Burton Singer (1985), Longitudinal analysis of labor market data. Cambridge: Cambridge University Press.

SRRD-12 Lyberg, Ingrid: Nonresponse effects on survey estimates in the analysis of competing exponential risks. (April, 1983.)

SRRD-13 Hartmann, Michael: Past and recent experiments in modeling mortality at all ages. (Nov., 1983.)

SRRD-14 Lyberg, Ingrid: The effects of sampling and nonresponse on estimates of transition intensities: Some empirical results from the 1981 Swedish fertility survey. (Dec., 1983.)

SRRD-15 Hoem, Jan M.: A flaw in actuarial exposed-to-risk theory. (Jan., 1984.) Scand. Actuarial J. 1984, 187-194.

SRRD-16 Hoem, Jan M.: Statistical analysis of a multiplicative model and its application to the standardization of vital rates: A review (Feb., 1984.)

SRRD-17 Lindberg, Bengt: A numerical algoritm for the calculation of coefficients in moving averages (March, 1984.)

SRRD-18 Hoem, Jan M., Bo Rennermalm, and Randi Selmer: Analyzing cohabitational first birth and marriage rates with data on the most recent union only in current Swedish and Danish cohorts. (March, 1984.) Revised version: Restriction biases in the analysis of births and marriges to cohabiting women from data on the most recent conjugal union only. (April, 1986.)

SRRD-19 Hoem, Jan M.: Marriages connected with first births among cohabiting women in the Danish fertility survey of 1975. (April, 1984.)

SRRD-20 Bernhardt, Eva: Den sociala bakgrundens betydelse för barnafödande och giftermål inom informellt samboende bland svenska kvinnor födda 1936-60. (Oct., 1984.)

SRRD-21 Etzler, Cecilia: Första steget i familjebildningen - Gifta sig, börja sambo eller få barn: Utveckling och skillnader bland svenska kvinnor med olika social bakgrund födda 1936-60. (Nov., 1984.)

SRRD-22 Hoem, Britta: Regional utveckling i det moderna samboendet, analysmetoder och resultat. (Nov., 1984.)

SRRD-23 Bernhardt, Eva and Britta Hoem: Cohabitation and social background: Trends observed for Swedish women born 1936-60. (Feb., 1985.) European Journal of Population 1, 375-395, 1985.

SRRD-24 Holmlund, Bertil: Fertility and the opportunities of children: Evidence on quantity-quality interactions in Sweden. (Sept., 1984.)

SRRD-25 Hoem, Britta: Ett barn är inte nog. Vad har hänt med svenska ettbarnskvinnor födda 1936-60? (June, 1985.)

SRRD-26 Kravdal, Øystein: Flytting fra foreldrehjemmet: Regionale og sosiale forskjeller blant svenske kvinner født 1936-60. (June, 1985.)

SRRD-27 Hoem, Jan M.: The impact of education on modern union initiation (August, 1985) European Journal of Population 2, 113-133, 1986.

SRRD-28 Bernhardt, Eva M.: Women's home attachment at first birth: The case of Sweden. (Oct., 1985.) European Journal of Population 2 (1), 5-29, 1986.

SRRD-29 Etzler, Cecilia: Barnafödande inom äktenskapet: Det föräktenskapliga samboendets betydelse för förstfödslar inom äktenskapet bland svenska kvinnor födda 1936-60. (January, 1986.)

SRRD-30 Hoem, Britta: Tredjebarnsfödslar bland svenska kvinnor födda 1936-50. (August, 1986.)

SRRD-31 Borgan, Ørnulf and Jan M. Hoem: Demographic reproduction rates and the estimation of an expected total count per person in an open population. (December, 1986.)

SRRD-32 Hoem, Britta: Sysselsättningshistoriens betydelse för tvåbarnsmödrars fortsatta barnafödande bland svenska kvinnor födda 1936-50. (January, 1987.)

SRRD-33 Bernhardt, Eva M: Continued home attachment versus transition to employment: An intensity regression analysis of Swedish one-child mothers. (January, 1987.)

SRRD-34 Etzler, Cecilia: Education, Cohabitation and the First Child: Some Empirical Findings from Sweden. (March, 1987.)

SRRD-35 Pleiborn, Maria: Övergångar in i och ut ur utbildning bland unga svenska kvinnor födda 1936-1960. (March, 1987.)

SRRD-36 Hoem, Britta and Jan M. Hoem: The impact of female employment on second and third births in modern Sweden. (April, 1987.)

SRRD-37 Hoem, Jan M. and Per Linnemann: The tails in moving average graduation. (May, 1987.)

SRRD-38 Hoem, Britta: Graviditetens betydelse för barnlösa svenska kvinnors civilståndsändringar. (May, 1987.)

SRRD-39 Hoem, Jan M.: The issue of weights in panel surveys of individual behavior. (May, 1987)99

Support from the Swedish Council for Research in the Humanities and Social Sciences for the publication of the Stockholm Research Reports in Demography is gratefully acknowledged.

Also available on request:

Bernhardt, Eva (1985). Nya samlevnadsformer: Arbetardöttrar, stockholmskor och norrländskor bildade förtrupp. Välfärdsbulletinen, SCB, 1985:1, 18-19.

Hoem, Jan M.: Sveriges hundraåriga fruktsamhetsfall. Pp. 185-212 in Barn? Författare, forskare och skolungdom diskuterar varför det föds så få barn. Stockholm: Liber Förlag, 1983.

Hoem, Jan M.: Forsikringsmatematik. Pp. 229-235 in Københavns Universitet 1479-1979, Bind XII. København: G.E.C. Gads Forlag, 1983.

Hoem, Jan M.: A contribution to the statistical theory of linear graduation. Insurance: Mathematics and Economics 3, 1-17, 1984.

Hoem, Jan M. and Randi Selmer: The negligible influence of premarital cohabitation on marital fertility in current Danish cohorts, 1975. Demography 21 (2), 193-206, 1984.

Gustafsson, Siv: Equal opportunity policies in Sweden, in Schmid, G. and Renate Weitzel (eds). Sex Discrimination and Equal Opportunity. The Labour Market and Employment Policy. Gower, Hampshire England, 1984.

Kravdal, Øystein. Flickorna flyttar hemifrån allt tidigare. Välfärdsbulletinen, SCB, 1985:3, 14-15.

Leighton, Linda and Siv Gustafsson. Differential patterns of unemployment in Sweden. Research in Labor Economics 6, 251-285, 1984.