Survey Methodology, June 1997
Vol. 23, No. 1, pp. 53-60
Statistics Canada

Instrumental Variable Estimation of Gross Flows in the Presence of Measurement Error

K. HUMPHREYS and C. J. SKINNER[1]

ABSTRACT

The problem of estimating transition rates from longitudinal survey data in the presence of misclassification error is considered. Approaches which use external information on misclassification rates are reviewed, together with alternative models for measurement error. We define categorical instrumental variables and propose methods for the identification and estimation of models including such variables by viewing the model as a restricted latent class model. The numerical properties of the implied instrumental variable estimators of flow rates are studied using data from the Panel Study of Income Dynamics.

KEY WORDS: Latent class; Longitudinal; Misclassification; Transition rate.

1. INTRODUCTION

One of the major benefits of longitudinal surveys is that they permit the estimation of gross flows, for example flows out of unemployment into employment (see *e.g.*, Hogue and Flaim 1986). A key problem when estimating flows is the bias induced by measurement error. For the estimation of cross-sectional proportions, misclassification into and out of states may tend to cancel out (Chua and Fuller 1987). Such compensation tends not to occur, however, when estimating longitudinal flows.

The first response to the problem of measurement error should clearly be to attempt to reduce the error in the survey measurement procedures. Relevant approaches are discussed by Biemer, Groves, Lyberg, Mathiowetz and Sudman (1991), but will not be considered here. Even with the "best" survey procedures, however, some measurement error will inevitably arise and there will remain a need to compensate for the effect of error in the survey analysis.

Methods for compensating for measurement error are generally based on some assumed model of the error process. Some models which have been proposed in the literature will be referred to in Section 2. In order to identify and estimate these models it is generally necessary to use additional auxiliary information, such as provided by reinterview studies (*e.g.*, Meyer 1988). Since reinterview studies are costly, however, and since in practice their aim is often not to estimate the characteristics of the measurement error distribution (Forsman and Schreiner 1991), there remains a need for alternative procedures which may be used when no reinterview data is available. For measurement error on continuous variables, a common approach employed in the absence of auxiliary information about the measurement error distribution is the method of instrumental variable estimation (*e.g.*, Fuller 1987, Sect. 1.4). An instrumental variable is a variable included in the survey dataset which is related to the true variable measured with error but is uncorrelated with the measurement error. These and associated assumptions supply information which replaces that provided by reinterview studies and enables parameters of the model involving the true variable to be identified and estimated. The aim of this paper is to investigate how the instrumental variable estimation method may be adapted to estimate flows among discrete states. We find that latent class models (*e.g.*, Bartholomew 1987, Ch. 2) provide a general framework within which the assumptions about the instrumental variable correspond to certain restrictions on the model parameters. Our approach is thus related to other approaches which impose restrictions on latent class models (*e.g.*, van de Pol and de Leeuw 1986; van de Pol and Langeheine 1990).

2. MODELS

We consider only the case of two occasions $t = 1$ and $t = 2$. Let the number of states into which each individual can be classified at each occasion be r. Denote the classified states at $t = 1$ and $t = 2$ by X and Y respectively and the corresponding true states by x and y. We assume a model in which the vectors of values of (X, Y, x, y) are generated as independent outcomes of a common random vector with distribution $\mathrm{pr}(X = i, Y = j, x = u, y = v)$.

The first assumption about this distribution, made by a number of authors (*e.g.*, Abowd and Zellner 1985; Poterba and Summers 1986 and Chua and Fuller 1987) and which we shall also make, is that the classification errors on the two occasions are conditionally independent given the true states, that is

$$\mathrm{pr}(X = i, Y = j \mid x = u, y = v) = \mathrm{pr}(X = i \mid x = u, y = v)\,\mathrm{pr}(Y = i \mid x = u, y = v). \qquad \text{(A1)}$$

[1] K. Humphreys, Department of Psychology, Stockholm University, S-106 91 Stockholm, Sweden; C.J. Skinner, Department of Social Statistics, University of Southampton, Southampton, SO17 1BJ, United Kingdom.

Such an assumption is common in general latent variable models (*e.g.*, Anderson 1959). It seems a reasonable initial assumption when the survey measurement procedures are independent on the two occasions. On the other hand, if X is obtained retrospectively from the same interview in which Y is measured then it seems likely that the tendency for respondents to give over-consistent responses in a single interview may tend to induce positive association between classification errors. See, for example, Marquis and Moore (1990) on evidence from the Survey of Income and Program Participation. A further reason for doubting the conditional independence assumption is the possibility of individual heterogeneity in misclassification probabilities, for example some respondents may be more reliable than others. See Skinner and Torelli (1993) and Singh and Rao (1995). In Section 4 we shall allow for heterogeneity by assuming only that the model holds within cells of a cross-classification of observed variables.

Our next basic assumption is that classification error only depends on current true state so that

$$\mathrm{pr}(X=i \mid x=u, y=v) = \mathrm{pr}(X=i \mid x=u) = K_{xiu}, \text{ say,}$$

$$\mathrm{pr}(Y=j \mid x=u, y=v) = \mathrm{pr}(Y=j \mid y=v) = K_{yjv}, \text{ say.} \quad \text{(A2)}$$

The K_{xiu} and K_{yjv} define $r \times r$ misclassification matrices $\boldsymbol{K}_x = [K_{xiu}]$ and $\boldsymbol{K}_y = [K_{yjv}]$. Letting $\boldsymbol{P}$ denote the $r \times r$ matrix with ij-th element $\mathrm{pr}(X=i, Y=j)$ and Π the $r \times r$ matrix with uv-th element $\mathrm{pr}(x=u, y=v)$ we have the matrix equation

$$\boldsymbol{P} = \boldsymbol{K}_x \Pi \boldsymbol{K}_y'. \quad (1)$$

The matrix Π contains the parameters of interest, whereas it is the matrix $\boldsymbol{P}$ which may be estimated consistently from sample X and Y values. If auxiliary estimates of $\boldsymbol{K}_x$ and $\boldsymbol{K}_y$ are available and these are non-singular then we can solve equation (1) to obtain estimates of Π. If it is possible to ascertain the true states in reinterview studies then $\boldsymbol{K}_x$ and $\boldsymbol{K}_y$ may be estimated directly (Abowd and Zellner 1985). On the other hand, if the reinterview study only provides independent reclassifications then it is only possible to estimate the interview-reinterview matrices

$$\boldsymbol{K}_x \Delta_x \boldsymbol{K}_x' \text{ and } \boldsymbol{K}_y \Delta_y \boldsymbol{K}_y'$$

where $\Delta_x = \mathrm{diag}[\mathrm{pr}(x=u)]$, $\Delta_y = \mathrm{diag}[\mathrm{pr}(y=v)]$ (Chua and Fuller 1987). Each interview-reinterview matrix is symmetric with elements summing to one and so only contains $r(r+1)/2 - 1$ "independent" items of information. Since each column of each $\boldsymbol{K}$ matrix and the diagonal of each Δ matrix sum to one, the number of unknown parameters on each occasion is $r(r-1) + r - 1 = r^2 - 1$. The excess of parameters over items of information is therefore $r^2 - 1 - r(r+1)/2 + 1 = r(r-1)/2$ at each occasion and so the model is underidentified for $r \geq 2$. Chua and Fuller (1987) suggest that a natural extra assumption to make to help achieve identification is to suppose that the measurement errors are unbiased on each occasion in the sense that

$$\mathrm{pr}(x=i) = \mathrm{pr}(X=i), \ \mathrm{pr}(y=i) = \mathrm{pr}(Y=i) \ \ i = 1, \ldots, r. \quad (2)$$

In this case false positives and false negatives tend to compensate for each other in cross-sectional estimates of proportions. This assumption reduces the number of parameters by $r-1$ on each occasion. Even under this assumption the model remains underidentified for $r \geq 3$ and Chua and Fuller (1987) have to introduce further assumptions.

Let us now consider how the model might be identified when no reinterview data is available. For simple linear regression with measurement error in the covariate, the instrumental variable approach (Fuller 1987, Sect. 1.4) assumes the availability of an observed "instrumental" variable W, which is correlated with the covariate, but is independent of the measurement error and independent of the error in the regression equation. We extend this assumption to our framework by defining W to be an *instrumental variable* if it is not independent of x and if

W and (X, Y) are conditionally independent given (x,y), (A3)

W and y are conditionally independent given x. (A4)

In general we shall allow W to be a categorical variable with an arbitrary number s of categories, although since we shall desire W to be closely related to x, we shall usually have $s = r$ in practice. One specific possibility is to take W as the classified state at time $t-1$. This use of a lagged value of a "covariate" as an instrumental variable may be traced back to the earliest discussions of instrumental variable estimation (*e.g.*, Reiersol 1941; Durbin 1954). In this case, assumption A4 follows if the true states obey a Markov process and the classification errors are conditionally independent, as in A1.

The model resulting from assumptions (A1)-(A4) may be represented by the conditional independence graph in Figure 1. Each vertex in the graph represents a variable. Edges between pairs of vertices are absent if the corresponding variables are conditionally independent given the remaining variables.

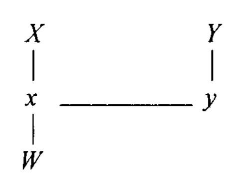

Figure 1. Conditional Independence Graph of Basic Model

The model is an example of a restricted latent class model (Goodman 1974), where the observed variables X, Y and W are conditionally independent given the latent variables x and y, that is they are independent within the r^2 latent classes defined by the pairs of values of (x,y). There are $2(r-1)r^2 + (s-1)r^2 + (r^2-1)$ parameters of this model given by the $(r-1)r^2$ parameters $\mathrm{pr}(X=i \mid x=u, y=v)$, the $(r-1)r^2$ parameters $\mathrm{pr}(Y=j \mid x=u, y=v)$, the $(s-1)r^2$ parameters $\mathrm{pr}(W=k \mid x=u, y=v)$ and the r^2-1 free parameters $\mathrm{pr}(x=u, y=v)$. These parameters are subject to the $2r(r-1)^2$ restrictions in (A2) and the $(s-1)r(r-1)$ restrictions implied by (A4). We first restrict attention to the case $r=2$. In this case there are $4s+7$ parameters subject to $2s+2$ restrictions, leaving $2s+5$ free parameters

$$\{K_{x2u}, K_{y2u}, \varphi_{2u}, \ldots, \varphi_{su}, \theta_u, \pi;\ u=1,2\ v=1,2\},$$

where $\varphi_{ku} = \mathrm{pr}(W=k \mid x=u)$, $\theta_u = \mathrm{pr}(y=2 \mid x=u)$, and $\pi = \mathrm{pr}(x=2)$. The number of "free" cell probabilities in the observed table of X by Y by W is r^2s-1, or $4s-1$ when $r=2$. Hence a necessary condition for identification when $r=2$ is that $4s-1 \geq 2s+5$ or $s \geq 3$. Unfortunately, this is not a sufficient condition. For let

$$R_u = \mathrm{pr}(Y=2 \mid x=u) = \sum_{v=1}^{2} K_{y2v}\, \theta_u^{v-1}(1-\theta_u)^{2-v}. \tag{3}$$

Then

$$\mathrm{pr}(X=i, Y=j, W=k) = \sum_{u=1}^{2} K_{xiu} \varphi_{ku} R_u^{j-1}(1-R_u)^{2-j} \pi^{u-1}(1-\pi)^{2-u}. \tag{4}$$

Hence the $4s-1$ free cell probabilities are determined by just the $2s+3$ parameters

$$\{K_{x2u}, \varphi_{2u}, \ldots, \varphi_{su}, R_u, \pi;\ u=1,2\}$$

so a necessary condition for identification of these parameters is that $4s-1 \geq 2s+3$ or $s \geq 2$. In fact this is also a sufficient condition for identification of these parameters, except for certain exceptional combinations of these parameters. (See Madansky (1960) for the case $s=2$ and Goodman (1974) for the case of general $s \geq 2$.)

However, even though the above $2s+3$ parameters are in general identified for $s \geq 2$ it is not possible to determine the 4 parameters $K_{y21}, K_{y22}, \theta_1$ and θ_2 since they are related to only two identified parameters, R_1 and R_2, via equation (3). In particular the key parameters of interest θ_1 and θ_2 remain underidentified whatever the value of s.

It is therefore necessary to impose at least 2 further restrictions on the model to identify θ_1 and θ_2. Following Chua and Fuller (1987), one idea would be to assume unbiased measurement errors as in (2) which imposes the two constraints

$$\pi = K_{x21}(1-\pi) + K_{x22}\pi \tag{5}$$

$$\theta_1(1-\pi) + \theta_2\pi = R_1(1-\pi) + R_2\pi. \tag{6}$$

Unfortunately the first constraint only applies to the parameters which are already identified for $s \geq 2$ so these constraints are insufficient to identify θ_1 and θ_2. An alternative assumption which we shall make is that the error process is constant over time so that

$$K_{xiu} = K_{yiu} = K_{iu}, \quad \text{say,} \quad \text{for } i,u = 1,2,\ldots,r. \tag{A5}$$

This seems a natural basic assumption if the same survey measurement procedure is used over time. The under-identification problem for the case $r=2$ discussed above is removed by this assumption since, given the identification of $K_{xiu} = K_{iu}$ and R_u, we can determine θ_u from (3) by

$$\theta_u = (R_u - K_{21})/(K_{22} - K_{21}) \tag{7}$$

(excluding the trivial case when the measured variables are independent of the true variables so that $K_{22} = K_{21}$).

In summary, when assumptions (A1) - (A5) hold and $r=2$, our model has $2s+3$ free parameters $\{K_{2u}, \varphi_{2u}, \ldots, \varphi_{su}, \theta_u, \pi;\ u=1,2\}$ which are identified if $s \geq 2$, except in exceptional cases such as discussed by Madansky (1960).

Finally, let us return to the case of general r. Since (A5) imposes $(r-1)r$ restrictions, the number of free parameters becomes $2(r-1)r^2 + (s-1)r^2 + (r^2-1) - [2r(r-1)^2 + (s-1)r(r-1)] - (r-1)r = 2r^2 + sr - 2r - 1$. There are r^2s-1 free cell probabilities in the table of X by Y by W so the model will in general be identified if $r(r-1)(s-2) \geq 0$. Thus the condition for identification of these parameters remains $s \geq 2$, for any value of $r \geq 2$. Furthermore we can write

$$R_{ju} = \Pr(Y=j \mid x=u) = \sum_{v=1}^{r} K_{jv}\theta_{uv}$$

where $\theta_{uv} = \mathrm{pr}(y=v \mid x=u)$. Hence, provided the matrix $[K_{iu}]$ is non-singular, the θ_{uv} may be determined from the R_{ju} and K_{jv} and hence are also identified. Thus for general r, the model is identified under assumptions (A1)-(A5), except for exceptional cases as discussed by Goodman (1974).

3. ESTIMATION

We shall suppose that for a sample of size n we observe counts n_{ijk} in the cells of the $r \times r \times s$ contingency table of $X \times Y \times W$, and that these are multinomially distributed with parameters n and $p_{ijk} = \mathrm{pr}(X=i, Y=j, W=k)$. The implied log likelihood is

$$l = \sum_i \sum_j \sum_k n_{ijk} \log p_{ijk}.$$

Under a complex sampling design, we may take the n_{ijk} to be weighted counts, giving a pseudo log likelihood (Skinner 1989). The estimators of the parameters obtained by maximising l will be called *instrumental variable* (IV) estimators.

For the remainder of this paper we shall only consider the case $r=s=2$ when the model is just identified (except for exceptional values of the parameters). In this case we might attempt to set $p_{ijk} = n_{ijk}/n$ and then solve equations (6) and (7) for the unknown parameters. If the resulting solutions lie within the feasible parameter space, that is probabilities lie in the range [0,1], then these solutions will be the IV estimates. However, in practice we have found that, for moderate sample sizes, infeasible solutions can often arise. Furthermore the solution of these equations is not computationally straightforward. Hence we have found it easier to maximise l directly using the numerical procedures in the package GAUSS (Edlefsen and Jones 1984) or else by using packages which fit latent class models using the EM algorithm such as PANMARK (van de Pol, Langeheine and de Jong 1991). For a latent class package it would be possible to fit an unrestricted two class model and then to estimate θ_1 and θ_2 via (7). However, there would be no guarantee that the resulting estimates would lie in the feasible range [0,1] with this approach. Furthermore there would be the additional complication of determining standard errors for the estimates of θ_1 and θ_2 from the covariance matrix of the estimates of $(R_1, R_2, K_{21}, K_{22})$. Hence we have found it more convenient to fit the model directly as a restricted latent class model. A further advantage of this approach is that it extends naturally to the fitting of similar models across subgroups subject to possible constraints that some parameters are constant across subgroups. This possibility is explored further in Section 4.

Under multinomial assumptions, standard errors may be based on the second derivatives of the log-likelihood evaluated at the IV estimates. This approach becomes problematic, however, if the maximum of l is at the boundary of the parameter space. One approach then is simply to treat the values of the parameters at the boundary as known. However, this is likely to lead to underestimation of uncertainty. Baker and Laird (1988) consider two alternative approaches to obtaining interval estimates for individual parameters in such circumstances: a bootstrap method and a profile likelihood method. The bootstrap method involves drawing repeated multinomial samples with p_{ijk} set equal to n_{ijk}/n and recording the distribution of parameter estimates across repeated bootstrap samples. Interval estimates for given parameters are obtained by the profile likelihood methods as the sets of values of the parameter which are not rejected by a likelihood ratio test. These methods are illustrated at the end of Section 4.

4. NUMERICAL ILLUSTRATIONS

For the purpose of numerical illustration we use data from the equal probability subsample of the US Panel Study of Income Dynamics (PSID). See Hill (1992). We consider the two states employed and not employed, coded 1 and 2 respectively, thus restricting attention again to the binary variable case. For simplicity, we ignore non-response and consider the sample of 5,357 individuals aged 18-64 in 1986 with complete values on the variables: employment status in 1985, 1986 and 1987, car ownership, age, sex and education.

We assess the properties of the IV estimator in two ways. First, in Section 4.1, we compare the bias and standard error of the IV estimator with the "unadjusted" estimator for hypothetical instrumental variables, with a range of different associations with x. Second, in Section 4.2, we consider the impact of using different actual PSID variables as instrumental variables.

4.1 Bias and Standard Error Properties of Estimators for Hypothetical Instrumental Variables

The parameters of primary interest are the joint probabilities $\mathrm{pr}(x = i, y = j)$ or the conditional probabilities $\mathrm{pr}(y = j \mid x = i)$ derived from these. The simple "unadjusted" estimators of these parameters are based on the corresponding sample proportions for the classified variables X and Y and have expectations $\mathrm{pr}(X = i, Y = j)$ under multinomial sampling. Since $\mathrm{Pr}(X = i, Y = j)$ differs in general from $\mathrm{pr}(x = i, y = j)$ the unadjusted estimators are typically biased. Provided the model assumptions (A1)-(A5) hold, the IV estimators of $\mathrm{pr}(x = i, y = j)$ will be asymptotically unbiased although their variances may be larger than those of the unadjusted estimators. The aim of this section is to investigate the extent to which there exists a trade-off in practice between the bias of the unadjusted estimators and the increased variance of the IV estimators. It will be assumed that the model assumptions (A1)-(A5) hold and that the sample is large enough for the IV estimator to be treated as unbiased.

For the numerical investigation in this section we wish to use some "realistic" parameter values. These were determined by rounding the values of estimates for annual flows between the years 1986 and 1987 from analyses in Section 4.2 (reported in Table 3). The values of the five free model parameters not involving W were set to be $K_{21} = 0.03$, $K_{22} = 0.94$, $\mathrm{pr}(x = 2) = \pi = 0.22$, $\mathrm{pr}(y = 2, x = 1) = \theta_1(1 - \pi) = 0.03$ and $\mathrm{pr}(y = 2, x = 2) = \theta_2 \pi = 0.19$. Different values of the remaining two free parameters $\varphi_{11} = \mathrm{pr}(W = 1 \mid x = 1)$ and $\varphi_{12} = \mathrm{pr}(W = 1 \mid x = 2)$ are set in the different columns of Table 1. Cramér's V statistic, which measures the association between two binary variables, essentially by scaling the chi-square statistic to a [0,1] interval, is provided as a summary of the strength of association between the variables W and x. For each of the choices of parameter values, Table 1 displays the estimated standard errors of the IV estimators for the PSID sample size $n = 5{,}357$. Table 1 also contains the biases and standard errors of the unadjusted estimator for the same parameter values K_{21}, K_{22}, π, θ_1 and θ_2 and the same sample size.

To illustrate the calculation of the biases of the unadjusted estimators, consider $\mathrm{pr}(x = 1, y = 1)$. The expectation of the unadjusted estimator of this parameter is $\mathrm{pr}(X = 1, Y = 1)$, which is calculated from the given values of K_{21}, K_{22}, π, θ_1 and θ_2 and assumptions (A1)-(A5) as 0.71. This compares with the assumed value of $\mathrm{pr}(x = 1, y = 1)$ of 0.75. The bias is thus $0.71 - 0.75 = -0.04$. The biases of the IV estimators are, as noted above, assumed to be zero. The standard errors of the unadjusted estimators are obtained from standard binomial

Table 1
Biases and Standard Errors under Alternative Hypothetical IVs

			Parameter Values Assumed for IV estimator						
$\text{pr}(W=1 \mid x=1)$			1.0	0.1	0.1	0.1	0.3	0.1	0.5
$\text{pr}(W=1 \mid x=2)$			0.0	0.9	0.7	0.5	0.7	0.3	0.3
Cramér's V			1.0	0.74	0.59	0.42	0.34	0.24	0.17
		Standard Errors (× 100)							
Parameter Estimated	Bias (× 100) of Unadjusted Estimator	Unadjusted Estimator	IV Estimator						
$\text{pr}(x=1, y=1)$	-4.0	0.62	0.68	0.75	0.88	1.13	1.16	1.82	2.05
$\text{pr}(x=1, y=2)$	3.0	0.32	0.39	0.43	0.51	0.64	0.69	1.03	1.24
$\text{pr}(x=2, y=1)$	3.0	0.32	0.32	0.37	0.44	0.57	0.66	0.95	1.27
$\text{pr}(x=2, y=2)$	-2.0	0.51	0.59	0.65	0.73	0.89	1.06	1.42	1.99
$\text{pr}(y=1 \mid x=1)$	-3.9	0.37	0.50	0.55	0.64	0.81	0.88	1.30	1.58
$\text{pr}(y=1 \mid x=2)$	12.4	0.60	1.40	1.63	1.95	2.56	2.90	4.30	5.55

Note: 1 = employed, 2 = not employed; n = 5,357; multinomial sampling assumed; biases of IV estimators are zero.

formulae. For example, the standard error of the unadjusted estimator of $\text{pr}(x=1, y=1)$ is $\sqrt{0.71 \times 0.29/5{,}357} = 0.0062$, where 0.71 is the value of $\Pr(X=1, Y=1)$. The standard errors of the IV estimators are obtained from the inverse of the expected information matrix, which is given by $n\sum p_{ijk} H_{ijk}$, where H_{ijk} is the 7×7 matrix of second derivatives of $\log p_{ijk}$ with respect to the seven free parameters. Following differentiation, these parameters are set equal to their assumed values, as indicated above. Note that the standard errors obtained from the multinomial information matrix are likely to be under-estimates because of the complex sampling design employed in the PSID.

There is a clear pattern of the standard errors of the IV estimator increasing as the association between W and x decreases. The amount of increase is fairly similar across all parameters, for example the ratio for V = 0.20 versus V = 1.00 lies between 3 and 4 for all parameters. In all cases the standard error of the IV estimator is greater than that of the unadjusted estimator. The loss of efficiency of the "best" IV estimator (with perfect association between W and x) compared to the adjusted estimator varies between parameters. Roughly speaking, the loss is greater for the conditional parameters than for the unconditional parameters. This loss of efficiency might be interpreted as the effect of adjusting for measurement error in y, which is still necessary even when x is perfectly measured by W. Under this interpretation, the greater relative loss of efficiency for the conditional parameters seems plausible since these are "less dependent" on the parameters of the marginal x distribution which the W information helps to estimate.

To examine the trade-off between the bias of the unadjusted estimator and the increased variance of the IV estimator we have calculated the minimum value of the sample size n necessary for the MSE of the IV estimator to be less than that of the unadjusted estimator. For complex designs the sample sizes should be interpreted as effective sample sizes. Table 2 gives these minimum values under a variety of strengths of association between W and x. If there were no misclassification the entries would all be infinity since the unadjusted estimators would always be more efficient than the IV estimators. For the assumed amount of misclassification given by $K_{21} = 0.03$ and $K_{12} = 0.06$, the sample size required increases rapidly as V decreases. The differences between the rows of Table 2 are partly accounted for by the differences between the rows of Table 1 and partly by differences between the biases of the unadjusted estimator. Thus, the bias of the unadjusted estimator of $\text{pr}(x=2, y=2)$ is relatively small and this leads to the large values in the corresponding row of Table 2. Note that the value of 1 for $\text{pr}(x=2, y=1)$ and Cramér's V = 1 arises because in this case the standard errors of the two estimators are equal (see Table 1) and so the bias of the unadjusted estimators implies that the IV estimator has smaller MSE for any $n \geq 1$.

The main conclusion we wish to draw from Table 2, however, is simply that we may expect there to be a number of practical situations where IV estimation will be worthwhile provided the model assumptions hold, even if the necessary sample sizes are inflated somewhat to allow for complex sampling designs.

4.2 Results for Actual Instrumental Variables

The results in the previous section were based on hypothetical instrumental variables. To provide a more realistic illustration we now consider possible real instrumental variables. The key problem is how to choose a variable W which obeys (A3) and (A4). It seems easier to find a variable which satisfies (A3) than (A4), in particular

Table 2
Sample Size Necessary for MSE of IV Estimator to be less than that of Unadjusted Estimator (Multinomial Sampling)

Parameter Estimated	Value of Cramér's V assumed for IV estimators: 1.0	0.74	0.59	0.42	0.34	0.24	0.17
	Sample size n required						
$\mathrm{pr}(x=1, y=1)$	28	59	132	300	320	971	1273
$\mathrm{pr}(x=1, y=2)$	31	50	91	184	219	573	843
$\mathrm{pr}(x=2, y=1)$	1	20	51	129	198	476	811
$\mathrm{pr}(x=2, y=2)$	112	227	366	720	1184	2397	5070
$\mathrm{pr}(y=1 \mid x=1)$	42	60	97	183	219	541	818
$\mathrm{pr}(y=1 \mid x=2)$	57	81	121	216	281	633	1061

measured without error obey (A3). However, it seems more difficult to find variables which one is sure are not related to change in employment status and hence obey (A4).

For illustration, we have considered two possibilities. First we have taken W as car ownership ($W = 2$ if the individual owns a car, $W = 1$ if not). This variable is likely to be measured with some error but it seems a reasonable first assumption that this error is unrelated to errors in measuring employment status. For example, in an analysis of errors in recording car ownership in the 1981 British Census, Britton and Birch (1985, p. 67) conclude that "the main problems associated with the small number of discrepancies were those connected with either vehicles out of use or vehicles temporarily available – for example, those hired..." and it seems at least plausible that such errors need have little relation to the kinds of errors in recording employment status. On the other hand, it is plausible that car ownership acts as a proxy for some kind of social or economic status which is related to change in employment status so assumption (A4) seems more questionable. However, for our illustrative purpose we assume (A3) and (A4) hold.

As a second illustration we have taken W to be the lagged employment status in 1985. A problem here is that (A4) effectively implies that individual employment histories follow Markov processes with common transition rates. In fact, transition rates will vary among individuals and this will invalidate assumption (A4) (*e.g.*, van de Pol and Langeheine 1990). Therefore, to allow for departures from assumption (A4), we disaggregated the sample into 16 groups defined by cross-classifying age (4 groups), sex and education (up to college level or not). We then assumed the model held within subgroups and used likelihood ratio tests to assess what parameters were constant across subgroups. These tests only provide a very rough guide since they ignore the complex sampling design of the PSID. There was no significant evidence of differences in the misclassification probabilities K_{ij} across subgroups. Furthermore, within each of the 8 subgroups defined by age × sex there was no significant evidence of differences in $\Pr(W \mid x, \text{subgroup})$ between the 2 education subgroups. Assuming equality of these parameters gave a non-significant likelihood-ratio goodness-of-fit chi-squared value of 52.9 on 46 df (46 is obtained as the number of cells = $16 \times 8 = 128$, less $2K_{ij}$ parameters, less $16 \times 4 = 64$ $\mathrm{pr}(x, y, \text{subgroup})$ parameters, less $8 \times 2 = 16$ $\mathrm{pr}(W \mid x, \text{ subgroup})$ parameters). Combining the parameter estimates for the disaggregated model appropriately gives estimates of the overall flows $\mathrm{pr}(x, y)$.

Table 3 contains estimates of the key parameters for the two choices of instrumental variable and for the disaggregated version of the second choice. We note first that the standard errors for the IV estimator based on car ownership are relatively high. This may be expected from Table 1 since the association between x and W is low (Cramér's V is 0.12). Even so, the resulting adjustments increasing the estimates for the diagonal entries are plausible and the confidence intervals resulting from this IV estimator seem more realistic than those for the unadjusted estimator.

Table 3
Unadjusted and IV Estimates for PSID Data

Parameter	Unadjusted Estimates	IV Estimates: IV = Car Ownership	IV = Lagged Employment	IV = Lagged Employment (Disaggregated)
$\mathrm{pr}(x=1, y=1)$	0.719 (0.006)	0.773 (0.033)	0.766 (0.008)	0.757 (0.007)
$\mathrm{pr}(x=1, y=2)$	0.055 (0.003)	0.011 (0.020)	0.017 (0.005)	0.025 (0.003)
$\mathrm{pr}(x=2, y=1)$	0.061 (0.003)	0.018 (0.019)	0.024 (0.004)	0.032 (0.003)
$\mathrm{pr}(x=2, y=2)$	0.166 (0.005)	0.198 (0.027)	0.193 (0.007)	0.186 (0.006)

Note: Standard errors under multinomial assumptions in parentheses. Disaggregation is by age (4 groups), sex and education (2 groups).

The standard errors for the second choice of instrumental variable are smaller, as expected since the association with X is now higher (Cramér's V is 0.73). Indeed these standard errors are not much larger than those for the unadjusted estimator. The (2 standard error) confidence intervals now do not overlap with the corresponding intervals for the unadjusted estimator for any of the four parameters.

As noted earlier, assumption (A4) is questionable for the lagged employment variable. The disaggregated version of this estimator makes "weaker" assumptions by only requiring (A4) to hold within subgroups. The resulting estimates are seen to be fairly close to the original IV estimator and to have slightly smaller standard errors, perhaps attributable to the use of the additional information on sex, age and education (but see later discussion). It is interesting that the effect of the disaggregation is to diminish the effect of adjustment by a relatively small amount in each case. It seems plausible that departures from (A4) may tend to lead to overadjustment in the IV estimator and that the disaggregation approach here helps to overcome this bias and, for alternative choices of disaggregating variables, enables an assessment of the sensitivity of results to the model specification.

As noted in Section 3 we have often come across IV estimates on the boundary of the interval [0,1]. Of the analyses reported in Table 3 in fact only the disaggregated analysis involved boundary estimates. For the 64 parameters $\mathrm{pr}(x = i, y = j, \text{subgroup})$ for $i, j = 1,2$, subgroup = 1, ..., 16, five of the estimates were on the boundary (none of the estimates of the remaining 18 parameters, $\mathrm{pr}(W=1 \mid X=1)$ and so forth, were). The standard errors reported in Table 3 treat these parameters as known and hence may underestimate the uncertainty in the estimates of the aggregate $\mathrm{pr}(x = i, y = j)$ parameters.

Table 4
Alternative Estimates of Standard Errors for Males Aged 26-35 with no College Education

Parameter	IV estimates	Estimated Standard Error	
		Standard	Bootstrap
$\mathrm{pr}(W=1 \mid x=1)$	0.947	0.011	0.011
$\mathrm{pr}(W=1 \mid x=2)$	0.107	0.089	0.091
$\mathrm{pr}(X=1 \mid x=1)$	0.969	0.006	0.007
$\mathrm{pr}(X=1 \mid x=2)$	0.084	0.088	0.075
$\mathrm{pr}(x=1, y=1)$	0.953	0.011	0.012
$\mathrm{pr}(x=1, y=2)$	0	*	*
$\mathrm{pr}(x=2, y=1)$	0.006	0.007	0.006
$\mathrm{pr}(x=2, y=2)$	0.041	0.012	0.011
$\mathrm{pr}(x=1)$	0.953	0.011	0.011
$\mathrm{pr}(y=1 \mid x=1)$	1	*	*
$\mathrm{pr}(y=1 \mid x=2)$	0.128	0.139	0.117

Note: n = 455; "standard" estimators based on observed information matrix, treating parameters estimated at the boundary as known; 10,000 replications of bootstrap; multinomial assumptions.

Table 4 presents alternative estimates of the standard errors for one subgroup, males aged 26-35 with no college education. The estimate of $\mathrm{pr}(x=1, y=2)$ as well as derived estimates, such as $\mathrm{pr}(y=1 \mid x=1)$ lie on the boundary. The "standard" estimates of the standard errors are, as in Table 3, based on the observed information matrix, treating parameters estimated at the boundary as known. Bootstrap standard error estimates (for 10,000 replications) are found to be very close to these standard estimates for parameters with estimates not on the boundary. For the IV estimate of $\mathrm{pr}(x=1, y=2)$ at the boundary no standard estimate of the standard error is available. Indeed it seems to make little sense to estimate the standard deviation of the sampling distribution in this case. It seems more sensible to derive a one-sided confidence interval which may be done either using the profile likelihood method, which gives [0, .016], or using the bootstrap percentile method, which gives [0, .009]. The corresponding intervals for $\mathrm{pr}(y=1 \mid x=1)$ are [.983, 1] and [.990, 1].

5. CONCLUSION

The presence of measurement error can induce substantial bias into standard estimates of transition rates from longitudinal data. If external estimates of misclassification rates are available then a variety of adjustment methods exist. If no such information is available then this paper shows how adjustment for measurement error alternatively can be carried out using instrumental variable estimation.

The main problem, as in conventional instrumental variable estimation, is finding a variable which one can be confident satisfies the conditions required of an instrumental variable. Even if the conditions are satisfied then it is desirable, in order to obtain reasonable precision, that there be a fairly strong association between this variable and the true state. If such a variable can be found then instrumental variable estimation may be useful.

ACKNOWLEDGEMENTS

We are grateful to Wayne Fuller for suggesting the basic idea underlying this paper. Research was supported by grant number H519 25 5005 from the Economic and Social Research Council under its Analysis of Large and Complex Datasets programme.

REFERENCES

ABOWD, J.M., and ZELLNER, A. (1985). Estimating gross labor force flows. *Journal of Business and Economic Statistics,* 3, 254-283.

ANDERSON, T.W. (1959). Some scaling models and estimation procedures in the latent class model. *Probability and Statistics,* (Ed. U. Grenander). Stockholm: Wiksell and Almquist.

BAKER, S.G., and LAIRD, N.M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association*, 83, 62-69.

BARTHOLOMEW, D.J. (1987). *Latent Variable Models and Factor Analysis*. London: Griffin.

BIEMER, P.P., GROVES, R.M., LYBERG, L.E., MATHIOWETZ, N.A., and SUDMAN, S. (1991). *Measurement Errors in Surveys* New York: Wiley.

BRITTON, M., and BIRCH, F. (1985). *1981 Census Post-Enumeration Survey.* London: Her Majesty's Stationery Office.

CHUA, T., and FULLER, W.A. (1987). A model for multinomial response error applied to labor flows. *Journal of the American Statistical Association*, 82, 46-51.

DURBIN, J. (1954). Errors in variables. *Review of the International Statistical Institute*, 22, 23-31.

EDLEFSEN, L.E., and JONES, S.D. (1984). Reference Guide to GAUSS. Applied Technical Systems.

FORSMAN, G., and SCHREINER, I. (1991). The design and analysis of reinterview: an overview. In *Measurement Errors in Surveys*. (Eds. Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., and Sudman, S.). New York: Wiley.

FULLER, W.A. (1987). *Measurement Error Models*. New York: Wiley.

GOODMAN, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.

HILL, M.S. (1992). *The Panel Study of Income Dynamics: A User's Guide*. Newbury Park, CA: Sage.

HOGUE, C.R., and FLAIM, P.O. (1986). Measuring gross flows in the labor force: an overview of a special conference. *Journal of Business and Economic Statistics*, 41, 111-21.

MADANSKY, A. (1960). Determinental methods in latent class analysis. *Psychometrika*, 25, 183-198.

MARQUIS, K.H., and MOORE, J.C. (1990). Measurement errors in the Survey of Income and Program Participation (SIPP): Program Reports. *Proceedings of the 1990 Annual Research Conference*. US Bureau of the Census, 721-745.

MEYER, B.D. (1988). Classification-error models and labor-market dynamics. *Journal of Business and Economic Statistics*, 6, 385-390.

POTERBA, J.M., and SUMMERS, L.H. (1986). Reporting errors and labor market dynamics. *Econometrica*, 54, 1319-1338.

REIERSOL, D. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9, 1-24.

SINGH, A.C., and RAO, J.N.K. (1995). On the adjustment of gross flow estimates for classification error with application to data from the Canadian Labour Force Survey. *Journal of the American Statistical Association*, 90, 478-488.

SKINNER, C.J. (1989). Domain means, regression and multivariate analysis. In *Analysis of Complex Surveys*, (Ch. 3) (Eds. Skinner, C.J., Holt, D., and Smith, T.M.F.). Chichester: Wiley.

SKINNER, C.J., and TORELLI, N. (1993). Measurement error and the estimation of gross flows from longitudinal economic data. *Statistica*, 53, 391-405.

VAN DE POL, F., and DE LEEUW, J. (1986). A latent Markov model to correct for measurement error. *Sociological Methods and Research*, 15, 118-141.

VAN DE POL, F., and LANGEHEINE, R. (1990). Mixed Markov latent class models. In *Sociological Methodology 1990*, (Ed. C.C. Clogg). Oxford: Basil Blackwell, 213-247.

VAN DE POL, F., LANGEHEINE, R., and DE JONG, W. (1991). PANMARK User Manual. Panel analysis using Markov chains. Version 2.2. Netherlands Central Bureau of Statistics.