Sociological Methods & Research

http://smr.sagepub.com

Problems in Estimating Measurement Error From Panel Data: An Example Involving the Measurement of Scientific Productivity

Lowell L. Hargens, Barbara F. Reskin and Paul D. Allison

Sociological Methods Research 1976; 4; 439

DOI: 10.1177/004912417600400403

In this paper we apply recently developed models for the estimation of reliability and stability coefficients from panel data to a study of scientific productivity. The models, which assume a first-order autoregressive process among true-score variables, yield either reliability and stability estimates which seem implausible when the statistical fits of the models to the data are good, or poor statistical fits when more plausible estimates are produced. We then examine an alternative model encompassing a latent variable which causes true scores and which is itself governed by a first-order autoregressive process. The results for this model are acceptable, and we conclude that the panel models developed for the estimation of measurement reliability are not appropriate representations of the causal processes involved in scientific productivity. We suggest that this type of misspecification will often occur in the application of models which assume independent disturbances for true-score variables.

PROBLEMS IN ESTIMATING MEASUREMENT ERROR FROM PANEL DATA

An Example Involving the Measurement of Scientific Productivity

LOWELL L. HARGENS
BARBARA F. RESKIN
Indiana University
PAUL D. ALLISON
State University of New York at Stony Brook

Several recent papers (Heise, 1969; Wiley and Wiley, 1970; Werts, Joreskog, and Linn, 1971; Wiley and Wiley, 1974) have presented closely related models for the estimation of reliability coefficients and measurement error

AUTHORS' NOTE: *We gratefully acknowledge the helpful suggestions of Hubert M. Blalock, George W. Bohrnstedt, Herbert L. Costner, Arthur S. Goldberger, Thomas A. Heberlein, Ronald Schoenberg, and H. H. Winsborough.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 4 No. 4, May 1976

variances from univariate panel data collected at three or more time points. The principal innovation of these models is the introduction of a disturbance in the true scores at successive time points. They thus represent a considerable advance over the direct use of test-retest correlations as reliability estimators, since the latter method assumes that true scores are perfectly stable over time. In spite of this wider applicability, the panel models still have some restrictive assumptions, notably that the errors of measurement and the disturbances in the true scores must be auto-uncorrelated over time (Heise, 1969: 98-101; Wiley and Wiley, 1974). Nevertheless, for the four-wave model, Heise (1969: 101) and Werts et al. (1971) have suggested that the violation of these assumptions may be detected by testing the single overidentifying restriction implied by the model.

Here we present an instance in which the four-wave model fits the data very well; yet, when it is elaborated with a tau-equivalent measure at one of the time points, the fit is very poor. We discuss possible violations of the assumptions of the model, propose an alternative model which resolves the apparent difficulty, and conclude that some types of correlated disturbances cannot be detected by the goodness-of-fit test for the basic four-wave model.

ESTIMATING THE RELIABILITY OF SCI ARTICLE COUNTS

In a study of the properties of several measures of scientific productivity, we sought to estimate the reliability and error variance of article counts taken from the *Science Citation Index* (SCI), as a measure of the true article production of individual scientists. Our study was based on a sample of 240 chemists who were members of the population of chemists who obtained the Ph.D. in chemistry during the period from 1955 to 1961.[1] We counted the number of articles attributed to each of the chemists in the Source Index of the SCI for each of the four years from 1965 through 1968. Since the collection of these data was part of a larger study of the career patterns of the cohort of chemists, we had a large amount of information concerning the areas of specialization, past publications, and institutional affiliations of the chemists. This information helped us obtain correct article counts in cases where SCI listings combined information for two or more scientists with the same surname and initials. In ambiguous cases, we consulted the publications listed by SCI in order to ensure that we were correctly attributing authorship. These steps were taken to minimize errors in our transcription of the data from the SCI, and we believe that remaining errors in the enumeration of articles for the members of our sample are largely the result of errors in the listing procedures of the SCI itself. Table 1 gives the correlations between the article counts for the members of our sample over the four-year period above the diagonal, with the variances and covariances on and below the diagonal.

The basic four-wave model proposed by Heise (1969) and Werts et al. (1971) can be represented for a single individual by the following recursive system of equations in unstandardized form (Model Ia):

TABLE 1
Correlations, Variances, and Covariances Among Article Counts in Four One-year Intervals, 240 chemists.[a]

	Science Citation Index				Chemical Abstracts
	1965	1966	1967	1968	1967
	(x_1)	(x_2)	(x_3)	(x_4)	($x_{3'}$)
1965	2.59	.515	.469	.436	.484
1966	1.21	2.15	.613	.520	.585
1967	1.26	1.49	2.73	.552	.937
1968	1.11	1.21	1.45	2.51	.586
1967	1.30	1.43	2.58	1.55	2.77

a. Correlations above the diagonal, variances and covariances on and below the diagonal.

$$x_t = X_t + e_t, \; t = 1, 2, 3, 4$$
$$X_1 = U_1,$$
$$X_t = b_t X_{t-1} + U_t, \; t = 2, 3, 4$$

In this application, t refers to a specific one-year interval, X_t is the true number of articles published in t, x_t is the number of articles counted in SCI, e_t is random measurement error, U_t is a random disturbance, and b_t is a constant for each t. All variables are expressed as deviations about their respective means. The conditions on the disturbance terms are:

$$E(e_t U_s) = 0, \; t, s = 1, 2, 3, 4$$
$$E(e_t e_s) = E(U_t U_s) = 0, \; t, s = 1, 2, 3, 4; t \neq s$$

A path diagram of the model is shown in Figure 1.

This model has 11 parameters—b_t (t = 2, 3, 4), $V(e_t)$, and $V(U_t)$ (t = 1, 2, 3, 4)—but only ten observed moments. Although it is therefore underidentified as a whole, five of the parameters are individually identified: b_3, b_4, $V(e_2)$, $V(e_4)$, and $V(U_3)$. For computational convenience, we reparameterized the model so that the remaining six underidentified parameters were combined into four parameters, all of which are identified. The structural equations for this reparameterized model (denoted Ia*) are:

$$x_1 = a X_2 + e_1{}^*$$
$$x_2 = X_2 + e_2$$
$$x_3 = X_3 + e_3$$
$$x_4 = b_4 X_3 + e_4{}^*$$
$$X_2 = U_2{}^*$$
$$X_3 = b_3 X_2 + U_3$$

where

$$a = \frac{b_2 V(U_1)}{b_2^2 V(U_1) + V(U_2)}$$

$$e_1{}^* = (1 - a b_2) U_1 - a U_2 + e_1$$

$$e_4{}^* = e_4 + U_4$$

$$U_2{}^* = b_2 U_1 + U_2$$

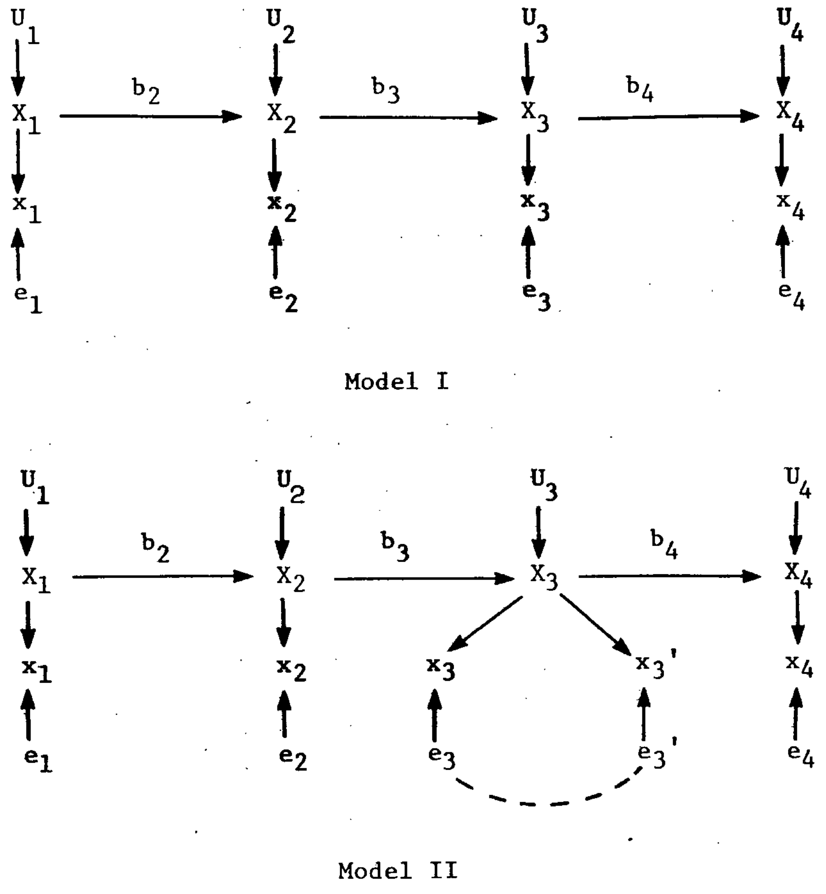

Figure 1: Path diagrams of Models I and II.

It can be shown that this set of equations implies that all the new disturbance and error terms are mutually uncorrelated. The four new parameters are a, $V(e_1*)$, $V(e_4*)$, and $V(U_2*)$.

It is then a straightforward procedure to derive expressions for observed moments (see Wiley and Wiley, 1974: 179 for an example of how this is done). The equations for the observed covariances are:

$$
\begin{aligned}
C_{12} &= a\, V(U_2*)\\
C_{13} &= a\, b_3\, V(U_2*)\\
C_{14} &= a\, b_3\, b_4\, V(U_2*)\\
C_{23} &= b_3\, V(U_2*)\\
C_{24} &= b_3\, b_4\, V(U_2*)\\
C_{34} &= b_3^2\, b_4\, V(U_2*) + b_4\, V(U_3)
\end{aligned}
$$

where $C_{ij} = C(x_i, x_j)$. The equations for the observed variances are

$$
\begin{aligned}
V_1 &= a^2\, V(U_2*) + V(e_1*)\\
V_2 &= V(U_2*) + V(e_2)\\
V_3 &= b_3{}^2\, V(U_2*) + V(U_3) + V(e_3)\\
V_4 &= (b_4 b_3)^2\, V(U_2*) + b_4{}^2\, V(U_3) + V(e_4*)
\end{aligned}
$$

where $V_i = V(x_i)$. The above 10 equations contain 9 parameters. The six covariance equations yield the following parameter solutions:

$$
\begin{aligned}
a &= C_{13}/C_{23} = C_{14}/C_{24}\\
b_3 &= C_{13}/C_{12}\\
b_4 &= C_{14}/C_{13}\\
V(U_2*) &= C_{23}\, C_{12}/C_{13}\\
V(U_3) &= C_{34}\, C_{13}/C_{14} - C_{13}\, C_{23}/C_{12}
\end{aligned}
$$

These expressions may be substituted into the four variance equations to solve for the four error variances. Notice the one overidentifying restriction that $C_{13}/C_{23} = C_{14}/C_{24}$.

To get efficient estimates of these parameters and to test the overidentifying restriction, we used the LISREL program (Joreskog and Van Thillo, 1973) which provides maximum-likelihood estimators and a large-sample chi-square test under the assumption that the observation vector is multinormally distributed. In the first column of Table 2 we give estimates of the parameters of interest, b_3, b_4, $V(e_2)$ and $V(e_3)$. We also give derived estimates of $\rho^2(x^2,X_2)$ and $\rho^2(x_3,X_3)$, the reliability coefficients for the two middle years, and ρ (X_2,X_3), the stability coefficient (Heise, 1969) for the one-year interval.[2] With a chi-square of .47, the model cannot be rejected at the .05 level.

Despite the excellent fit of the model, it seems implausible for two reasons. First, the stability coefficient of .947 seems to high. Given the many contingencies of scientific publication, it is difficult to believe that our sample members' relative article productivity would be almost perfectly stable in two consecutive years. Second, the reliability coefficients of .648 and .635 seem much too low. It is improbable that errors of enumeration and transcription within SCI's collection and publication of these data can amount to over one-third of the observed variance. (See Garfield [1974] for a discussion of possible errors in SCI.)

To further test the model, we used *Chemical Abstracts* (CA) to get a separate count of articles published in 1967 by the members of the sample. CA abstracts articles from both foreign and U.S. journals and is similar to SCI in its coverage of chemistry publications.[3] We examined the 1967, 1968, and 1969 volumes of CA in order to include all the articles published in 1967.[4] Once again, we attempted to eliminate ambiguities in the entries so that the primary source of error would be the enumeration and transcription process of CA itself. The correlations and covariances of the 1967 CA counts with all the SCI counts are shown in Table 1.

Our first test was to substitute the 1967 CA counts for the 1967 SCI counts, or, formally, to substitute the equation

$$x_3' = X_3 + e_3'$$

for

$$x_3 = X_3 + e_3$$

to produce Model Ib*. Estimates based on this model are shown in Table 2. Since the parameter estimates are quite similar for Ia* and Ib*, our reservations about the plausibility of the estimates for Model Ia* also apply to those for Model Ib*. On

TABLE 2
Parameter Estimates and Goodnees-Of-Fit Statistics for Models Ia*, Ib*, IIa*, and IIb*

	Model			
Parameter	Ia*	Ib*	IIa*	IIb*
b_3	1.06	1.07	1.06	1.06
b_4	.833	.850	.588	.842
$V(e_2)$	.757	.821	.769	.789
$V(e_3)$	.996	--	.180	.949
$V(e_3')$	--	.951	.171	.951
$\rho^2(x_2, X_2)$	.648	.619	.643	.634
$\rho^2(x_3, X_3)$	.635	--	.935	.646
$\rho^2(x_3', X_3)$	--	.656	.938	.646
$\rho(X_2, X_3)$	.947	.917	.777	.932
$\rho(e_3, e_3')$	--	--	--	.822
Statistical fit				
X^2	.475	.014	27.6	8.58
d.f.	1	1	5	4
p	.491	.907	.0000	.0725

the other hand, Model Ib* has a chi-square of only .014, which is a better fit than that shown by Ia*.

Our next step was to combine the two models into one by incorporating both the equations for x_3 and x_3', given above, into Model I with the same conditions on e_3' as on e_3. We also assume $E(e_3'e_3) = 0$. This is Model IIa, displayed as a path diagram in Figure 1. The specification of an identical slope (unity) for both x_3 and x_3' on X_3 is an assumption of tau-equivalence (Lord and Novick, 1968). Given that CA and SCI attempt to enumerate the same items, there seems to be sufficient prior justification for imposing this constraint.

The reparameterization of Model I may also be employed for this model, giving IIa*. Adding x_3' to Model Ia* adds five observed moments, but only one new parameter, $V(e_3')$, yielding a total of 5 degrees of freedom. These degrees of freedom correspond to five constraints on the observed covariances. Models Ia and Ib each contribute one degree of freedom by constraining $C_{14}/C_{24} = C_{13}/C_{23}$ and $C_{14}/C_{24} = C_{13}'/C_{23}'$ respectively. Combining the two models and assuming that x_3 and x_3' are congeneric tests (Joreskog, 1971) imply that $C_{34}/C_{3'4} = C_{13}/C_{13}'$. A fourth constraint follows from the tau-equivalence assumption which implies that $C_{13} = C_{13}'$. The fifth comes from C_{33}' which is the only observed moment not included in either Ia or Ib. Specifically, the combined model requires that $C_{33}' = C_{34}C_{13}/C_{14}$.

Estimates for Model IIa* are shown in Table 2. Some of the estimates are markedly different than in the four-variable models and somewhat nearer our expectations. The reliability coefficients of approximately .94 for both x_3 and x_3' are substantially higher, while the stability coefficient of .78 is correspondingly lower. However, the reliability coefficient of x_2 is still only .643 whereas one would expect it to be in the same neighborhood as those for x_3 and x_3'. More important, however, is the chi-square of 27.6 (5 d.f., $p < .0001$) which surely calls for rejection of the model.

This poses a dilemma. The four-variable models, Ia and Ib, fit the data very well indeed, but yield what we judge to be unreasonable estimates. Their natural extension into a five-variable model yields more reasonable estimates for some parameters, but a very poor fit to the data. Is there any way to alter the five-variable model in order to resolve this inconsistency? We considered several possibilities. We first relaxed the tau-equivalence assumption to see if it were the source of the difficulty. Although the resulting estimates are not shown in Table 2, they are quite similar to those for Model IIa*, and the chi-square is still 27.3 (4 d.f.), indicating that this model is also unacceptable. Next, we reasoned that since C_{33}' is the only observed moment not included in Models Ia and Ib, it might be the source of the poor fit. One way to improve the fit is to allow a correlation between e_3 and e_3'. This model (IIb*) is identified, and estimates are given in Table 2. Allowing for the correlation results in a substantial improvement in fit. The chi-square drops to 8.58 (4 d.f.) and the model cannot be rejected at the .05 level. Unfortunately, the parameter estimates are still unreasonable with a stability coefficient of .932 and reliability estimates of around .64 for x_2, x_3, and x_3'. Moreover, the estimated correlation of .822 between e_3 and e_3' seems much too high. CA and SCI are completely distinct organizations, and we cannot imagine any mechanism that would produce so strong a relationship between their errors of enumeration and transcription.[5]

It is curious that the treatment of the 1967 CA and SCI article counts as measures of the true but unmeasured 1967 article counts leads to results for Models IIa* and IIb* which seem untenable on either statistical or theoretical grounds. Since it seemed unreasonable to abandon the assumption that the CA and SCI counts are in fact congeneric measures of the true counts, we decided to explore other possible models for representing the determinants of the true and observed article counts. One possible misspecification of these causal relations is the assumption, contained in all of the above models, that the disturbance terms, U_1 through U_4, are independent of each other. In the case of scientific productivity, there are strong reasons for expecting correlations among these disturbances. Specifically, we can expect the published output of the individual during each time interval to be determined by relatively stable personality or role traits such as ability, motivation, or professional socialization, as well as stable contexts and facilities (Allison and Stewart, 1974). A first approximation to this idea would be to relax the assumption that $E(U_t U_s) = 0$, $t \neq s$, and specify a first-order autoregressive process for the disturbances:

$$\begin{aligned} U_1 &= v_1 \\ U_t &= c_t U_{t-1} + v_t,\ t = 2, 3, 4 \\ E(v_t) &= E(v_t e_s) = 0,\ t,s = 1, 2, 3, 4 \\ E(v_t v_s) &= 0,\ t,s = 1, 2, 3, 4;\ t \neq s \end{aligned}$$

Added to Model IIa, these equations produce Model III which is shown diagrammatically in Figure 2. A very similar model was considered by Heise (1969: Fig. 7).

There are two difficulties with the model. First, as Heise observes, none of the parameters is identified—which brings the analysis to a halt. Second, it is a straightforward but tedious exercise to show that, in general, this model does not imply the overidentifying restriction implied by Model I (and also by Model II). If III is the correct model we should have expected large chi-square values in our estimation of Ia* and Ib*. That they were small suggests that we ought to look further.

Our solution to these difficulties is to choose a variant of Model III which does not violate the overidentifying restriction. Specifically, we set $b_t = 0$, $t = 2, 3, 4$, and incorporate an additional disturbance term into the determination of X_t, $t = 1, 2, 3, 4$. The complete structural equations for Model IV are:

$$\begin{aligned} U_1 &= v_1 \\ U_t &= c_t\, U_{t-1} + v_t,\ t = 2, 3, 4 \\ X_t &= U_t + w_t,\ t = 1, 2, 3, 4 \\ x_t &= X_t + e_t,\ t = 1, 2, 3, 4 \\ x_3{}' &= X_3 + e_3{}' \end{aligned}$$

The disturbance conditions are:

$$E(e_t w_s) = E(e_t v_s) = E(w_t v_s) = 0,\ t,s = 1, 2, 3, 4$$
$$E(e_t e_s) = E(v_t v_s) = E(w_t w_s) = 0,\ t,s = 1, 2, 3, 4;\ t \neq s$$

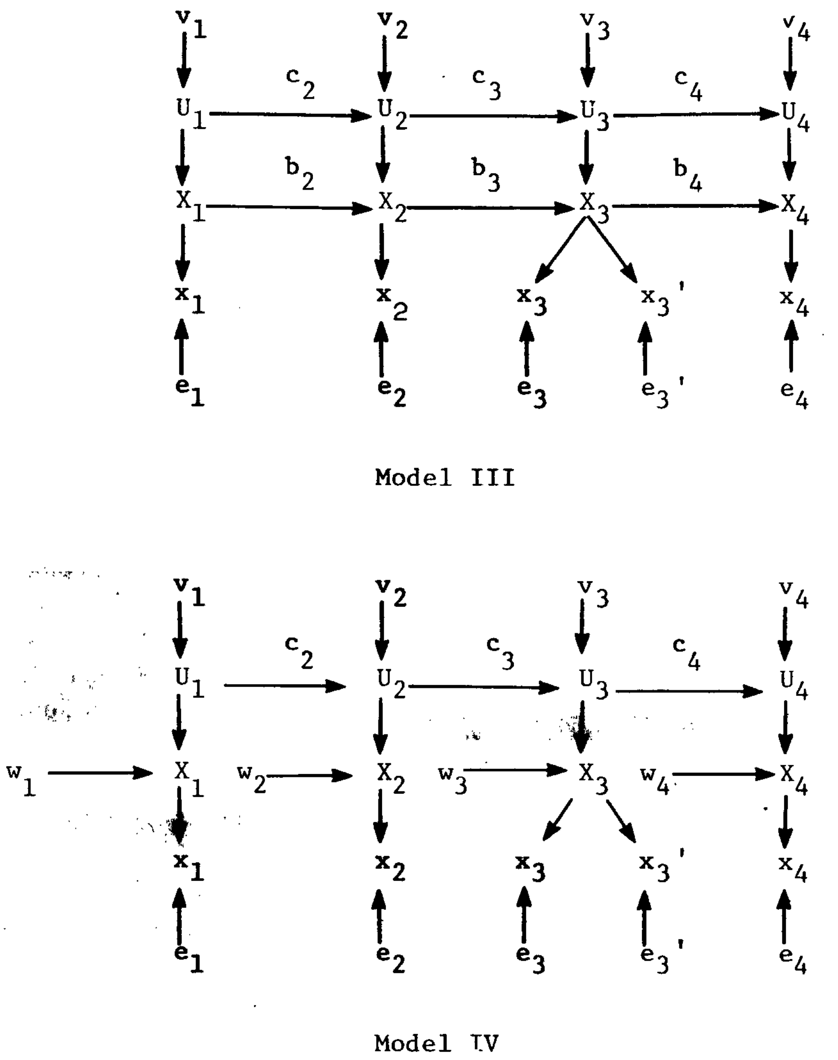

Figure 2: Path diagrams of Models III and IV

A path diagram of the model is shown in Figure 2. Our rationale for setting the $b_t = 0$ is that, like many other variables, publications must be created anew in each time interval. Whatever correlation exists between X_t and X_{t+1} must, therefore, be due to a correlation among their determinants rather than to any direct effect. We have represented what are undoubtedly multiple determinants by a unitary trait U_t, which might be interpreted as the propensity to publish. This may be doing some violence to our understanding of the determinants of scientific productivity over the long run, but, as we shall see, it does seem to be the most parsimonious model consistent with the data for the four-year period of observation.

With 16 parameters and 15 observed moments, Model IV is underidentified. But again, several of the parameters are individually identified. To eliminate any unidentified parameters, we reparameterized the model as follows (Model IV*):

$$
\begin{aligned}
x_1 &= d\,U_2 + e_1{}^* \\
x_2 &= U_2 + e_2{}^* \\
x_3 &= X_3 + e_3 \\
x_3{}' &= X_3 + e_3{}' \\
x_4 &= c_4\,U_3 + e_4{}^* \\
X_3 &= U_3 + w_3 \\
U_2 &= v_2{}^* \\
U_3 &= c_3\,U_2 + v_3
\end{aligned}
$$

Here,

$$
d = \frac{c_2\,V(v_1)}{c_2{}^2\,V(v_1) + V(v_2)}
$$

$$
\begin{aligned}
e_1{}^* &= (1 - d\,c_2)\,v_1 - d\,v_2 + e_1 + w_1 \\
e_2{}^* &= e_2 + w_2 \\
e_4{}^* &= e_4 + w_4 + v_4 \\
v_2{}^* &= c_2\,v_1 + v_2
\end{aligned}
$$

Once again, the disturbances, v_i and w_i, and errors of the reparameterized model are all uncorrelated. This specification reduces the number of parameters to 11, leaving four degrees of freedom. The constraints corresponding to these four degrees of freedom are the same as those of Model IIb*.

Estimates of the parameters of interest, as well as several derived correlational statistics, are given in Table 3.[6] While not as good a fit as Models Ia and Ib, the chi-square of 8.58 (4 d.f.) for Model IV* is a substantial improvement over Model IIa*, and the model cannot be rejected at the .05 level. Moreover, all the parameter estimates are consistent with our expectations. The high reliability of .937 for counts from both CA and SCI in 1967 seems about right, and it is also plausible that the latent trait U_t has a stability coefficient as high as .932. If we assume that $\rho(X_3, U_3) = \rho(X_2, U_2)$, we get an estimate for $\rho(X_2, X_3)$ of .642, which seems more reasonable than the substantially larger values for the year-to-year correlation between the true numbers of articles given by models Ia*, Ib*, and IIb*. We also find that the correlation between x_2 and U_2 (.796) is about the same as $\hat{\rho}(x_3, U_3) = .803$. Thus unlike Model IIa*, Model IV* yields estimates which are fairly stable for t = 2, 3.

It is important to note that in this model, the meanings of reliability and error variance depend on what one is trying to measure. If the object of measurement is X_t, the true number of articles, SCI counts have a reliability of about .94 and a measurement error variance of about .17. On the other hand, if one wants to measure the latent trait U_t, which in this instance is interpreted as the propensity to publish, the reliability goes down to approximately .64 while the error variance increases to around .97. If Model IV is correct, then Models I and II yield incorrect interpretations of the nature of the derived reliability coefficients because they mistakenly hypothesize that the unmeasured variables which meet the assumptions of a lag-1 autoregressive process are the true scores of the yearly article counts when in fact they are, according to our interpretation, the yearly propensities of scholars to publish. As a result, Models I and II suggest that the squared correlation between true and observed article counts tends to be in the neighbor-

TABLE 3
Parameter Estimates and Goodness-Of-Fit Statistics for Model IV

Parameter	Value		
c_3	1.06		
c_4	.841		
$V(e_2^*)$	.788		
$V(e_3)$	.173		
$V(e_3')$	.174		
$V(w_3)$	.801		
$\rho^2(x_2, U_2)$	.634		
$\rho^2(x_3, X_3)$	.937		
$\rho^2(x_3', X_3)$	.937		
$\rho^2(X_3, U_3)$	.689		
$\rho(U_2, U_3)$	.932		
Statistical fit			
χ^2	8.5783		
d.f.	4		
p	.0725		
Additional constraints:			
	χ^2	d.f.	p
$c_3 = c_4$	1.81	1	.10
$V(e_3) = V(e_3')$	.0002	1	.99

hood of .64 while according to Model IV this figure should be interpreted as the squared correlation between observed scores and the propensity to publish.

There is an even more intimate relation between Models IV and IIb. As noted above, both imply the same four constraints on the observed variance-covariance matrix, which means that one model can be viewed as a reparameterization of the other. For this reason, our preference for Model IV can only be justified by our a priori beliefs about the causal process determining productivity and our expectations as to the magnitudes of the coefficients.

We also tested some additional restrictions on Model IV. Since the estimated error variances of e_3 and e_3' were almost identical, we constrained them equal and reestimated the model. Together with the restriction of tau-equivalence this constraint is equivalent to assuming that 1967 SCI and CA counts are parallel measures (Lord and Novick, 1968). Since this additional constraint only increased the chi-square in the fourth decimal place, we conclude that counts from SCI and CA are functionally interchangeable—at least for 1967.

We also tested the hypothesis of constant auto-regression for the latent trait in successive years, specifically that $c_3 = c_4$. Even though the estimates appear to be different (1.06 and .842), the test showed no significant difference ($\chi^2 = 1.81$, 1 d.f.). The constrained estimate was $c_3 = c_4 = .937$. In both the tests, estimates of other parameters hardly varied, and so have not been reported.

CONCLUSION

Recent models for the estimation of measurement error from panel data assume a lag-1 autoregressive process in the true-score variable with uncorrelated disturbances. We believe that this assumption will usually be problematic for sociological variables which typically are determined by other variables having some stability over time. Although it has been suggested that such a violation of the assumptions can be tested when four or more waves of data are present, we have presented a model for which that is not the case. The model assumes a first-order autoregressive process among the *disturbances,* and an absence of any lagged effects of the true-score variable. This model seems particularly appropriate for variables like scientific productivity which must be created or produced anew for each time interval, in contrast to variables which have an internal principle of stability, i.e., which tend to remain the same unless acted upon from without.

For the case of scientific productivity, we showed how the two models could be distinguished by augmenting the data with another measure for one of the time points. While the four-variable model produced an excellent fit to the data, its natural extension to the five-variable model resulted in a very poor fit. On the other hand, our five-variable model incorporating correlated disturbances produced an acceptable fit. Estimates of reliability, stability, and measurement error variance differed markedly for the two models, with the correlated-disturbance model giving much more plausible results. Unfortunately, the correlated-disturbance model cannot be distinguished statistically from a model which allows for correlated measurement errors for the equivalent measures at the same time point. However, the latter model gave parameter estimates which seemed quite unreasonable.

In general, we believe that the application of panel models which assume a first-order autoregressive process among true-score variables can often yield misleading results, even when such models show good statistical fits to the data. It therefore seems prudent to incorporate additional variables into such models so that a more effective test of this assumption can be carried out. This may be accomplished either with multiple indicators, as in the study reported here, or with measures of variables which are exogenous to the true-score variables of interest (Wheaton, Muthen, Alwin, and Summers, forthcoming).

NOTES

1. The sampling frame was the population of all chemists reported in the 1957, 1959, and 1961 editions of the *Directory of Graduate Research* (American Chemical Society, 1957, 1959, 1961) as having completed the Ph.D. between 1955 and 1961. For a more detailed description of the sample, see Reskin (1973: 374-391).

2. For the population, the correlational measures may be derived from the basic parameters as follows:

$$\rho^2(x_2, X_2) = \frac{V(U_2^*)}{V(U_2^*) + V(e_2)}$$

$$\rho^2(x_3, X_3) = \frac{b_3{}^2\, V(U_2^*) + V(U_3)}{b_3{}^2\, V(U_2^*) + V(U_3) + V(e_3)}$$

$$\rho(X_2, X_3) = \left(\frac{b_3{}^2\, V(U_2^*)}{b_3{}^2\, V(U_2^*) + V(U_3)}\right)^{1/2}$$

These formulas also hold for models Ib*, IIa*, and IIb*, with $V(e_3{}')$ substituted for $V(e_3)$ in calculating $\rho^2(x_3{}', X_3)$. Since we have maximum-likelihood estimates of the basic parameters contained in these formulas, the formulas can be used to obtain maximum-likelihood estimates of the correlational measures.

3. There are some differences in coverage between CA and SCI. SCI covers a wide variety of disciplines in addition to chemistry, but CA covers articles in a greater number of chemistry journals than SCI. Thus, we sometimes found that our chemists published work in journals not classified as chemistry journals and that the work was therefore listed in SCI but not CA. Similarly, we sometimes found that the chemists published work in obscure chemistry journals which were not surveyed by SCI. Our experience suggests, however, that the major portion of the published work of chemists is included in both sources. Eighty-four per cent of the papers authored by members of our sample were included in both sources, 90% were in SCI, and 94% in CA.

4. CA abstracts the articles which appear in a given year over a period which extends beyond the year itself. Thus we examined the 1967 through 1969 volumes of CA to obtain article counts for the 1967 calendar year. We found no 1967 articles in the issues of CA for the last half of 1969 (Vol. 71 of CA). We chose 1967 as the year for which both SCI and CA article counts would be collected because the 1967-1969 volumes of CA are indexed in a single cumulative index.

5. We believe that the only possible source of a correlation between the errors of CA and SCI is the misspelling of authors' names on the original articles examined by the two organizations. If this occurred, both CA and SCI counts would be underestimates of the true article counts of affected authors and overestimates of the true article counts of any authors whose names matched the misspelled names. Since authors are especially sensitive to such misprints and usually review galley proofs before actual publication of their articles, the probability of this type of error should be negligible, and any resulting correlation between CA and SCI errors should be small.

6. The correlational statistics are derived from:

$$\rho^2(x_2, U_2) = \frac{V(v_2^*)}{V(v_2^*) + V(w_2) + V(e_2)}$$

$$\rho^2(x_3, X_3) = \frac{c_3{}^2\, V(v_2{}^*) + V(v_3) + V(w_3)}{c_3{}^2\, V(v_2{}^*) + V(v_3) + V(w_3) + V(e_3)}$$

$$\rho^2(X_3, U_3) = \frac{c_3{}^2\, V(v_2{}^*) + V(v_3)}{c_3{}^2\, V(v_2{}^*) + V(v_3) + V(w_3)}$$

$$\rho(U_2, U_3) = \left(\frac{c_3{}^2\, V(v_2{}^*)}{c_3{}^2\, V(v_2{}^*) + V(v_3)}\right)^{1/2}$$

The formula for $\rho^2(x_3, X_3)$ is the same as that for $\rho^2(x_3{}'\, X_3)$ except that $V(e_3{}')$ is substituted for $V(e_3)$.

REFERENCES

ALLISON, P. D. and J. A. STEWART (1974) "Productivity differences among scientists: evidence for accumulative advantage." Amer. Soc. Rev. 39 (August): 596-606.

American Chemical Society (1957, 1959, 1961) Directory of Graduate Research. Washington, D.C.: American Chemical Society.

GARFIELD, E. (1974) "Errors–theirs, ours and yours." Current Contents 6 (19 June): 5-6.

HEISE, D. R. (1969) "Separating reliability and stability in test-retest correlation." Amer. Soc. Rev. 34 (February): 93-101.

JORESKOG, K. G. (1971) "Statistical analysis of sets of congeneric tests." Psychometrika 36 (June): 109-133.

——— and M. VAN THILLO (1973) LISREL: A General Computer Program for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables. Research Bulletin 72-56. Princeton, N.J.: Educational Testing Service.

LORD, F. M. and M. R. NOVICK (1968) Statistical Theories of Mental Test Scores. Reading, Mass.: Addison-Wesley.

RESKIN, B. F. (1973) "Sex differences in the professional life chances of chemists." Ph.D. dissertation, University of Washington.

WERTS, C. E., K. G. JORESKOG, and R. L. LINN (1971) "Comment on 'the estimation of measurement error in panel data'." Amer. Soc. Rev. 36 (February): 110-113.

WHEATON, B., B. MUTHEN, D. F. ALWIN, and G. F. SUMMERS (forthcoming) "Specification and estimation of panel models incorporating reliability and stability parameters." To appear in Sociological Methodology, 1977.

WILEY, D. E. and J. A. WILEY (1970) "The estimation of measurement error in panel data." Amer. Soc. Rev. 35 (February): 112-117.

WILEY, J. A. and M. G. WILEY (1974) "A note on correlated errors in repeated measurements." Sociological Methods and Research 3 (November): 172-188.

Lowell L. Hargens is an Associate Professor of Sociology at Indiana University. In collaboration with Warren O. Hagstrom, he is currently studying relations between levels of social solidarity and status-attainment processes in scientific communities. This project extends a line of inquiry reported in his Patterns of Research, *recently published in the A.S.A. Rose Monograph Series.*

Barbara F. Reskin is an Assistant Professor of Sociology at Indiana University. She is currently studying the determinants of scientific productivity and the impact of doctoral sponsorship on scientists' careers. Some results of a third study on sex differences in the process of status attainment in science will be appearing in the August issue of the American Sociological Review.

Paul D. Allison is a Lecturer in Sociology at the State University of New York at Stony Brook. He has just completed work on a stochastic process model of measurement error (part of which will appear in Psychometrika*), and is now investigating tests for interaction in multiple regression. His recent collaborative study of the consequences of the distribution of scientific productivity is forthcoming in* Social Studies of Science.