Statistical Methods for Comparing Regression Coefficients Between Models

Article *in* American Journal of Sociology · March 1995

DOI: 10.1086/230638 · Source: OAI

4 authors, including:

Eva Petkova
NYU Langone Medical Center
200 PUBLICATIONS **11,710** CITATIONS

SEE PROFILE

SYMPOSIUM ON APPLIED REGRESSION

Statistical Methods for Comparing Regression Coefficients between Models[1]

Clifford C. Clogg
Pennsylvania State University

Eva Petkova
Columbia University

Adamantios Haritou
University of Macedonia

Statistical methods are developed for comparing regression coefficients *between* models in the setting where one of the models is nested in the other. Comparisons of this kind are of interest whenever two explanations of a given phenomenon are specified as linear models. In this case, researchers should ask whether the coefficients associated with a given set of predictors change in a significant way when other predictors or covariates are added as controls. Simple calculations based on quantities provided by routines for regression analysis can be used to obtain the standard errors and other statistics that are required. Results are also given for the class of generalized linear models (e.g., logistic regression, log-linear models, etc.). We recommend fundamental change in strategies for model comparisons in social research as well as modifications in the presentation of results from regression or regression-type models.

I. INTRODUCTION

The most common method used in empirical social research to compare two explanations of some specified dependent variable can be described as follows. A conventional explanation is represented by a linear model

[1] This research was supported in part by grant SBR-9310101 from the National Science Foundation and in part by the Population Research Institute, Pennsylvania State University, with core support from the National Institute of Child Health and Human Development (grant 1-HD28263-01). The authors are indebted to C. Himes for providing the data and some of the calculations used in table 5 and to L. A. Goodman, J. Hagenaars, and four referees for helpful comments. Address correspondence to Clifford C. Clogg, Department of Sociology, Pennsylvania State University, University Park, Pennsylvania 16802.

0002-9602/95/10005-0006$01.50

that includes the set of predictors regarded as important factors in that explanation, as specified from the relevant theory and prior research. Suppose these are denoted as $X = (X_1, \ldots, X_p)$. The alternative explanation is specified by adding to the original model a different set of predictors, say $Z = (Z_1, \ldots, Z_q)$, giving a linear model that includes both X and Z. Both models are then compared in terms of variability explained, the relative magnitudes of the regression coefficients in the two models, and the magnitudes of the coefficients relative to their standard errors in either or both models. Such comparisons are usually made in order to determine whether the original explanation, the alternative explanation, or some combination of the two is most consistent with the data. Model comparisons of this general kind are ubiquitous in social research, as well as in other areas, and they are especially prominent whenever regression techniques are applied for the analysis of nonexperimental data, such as survey data. In this article, we recommend fundamental changes in the logic of model comparisons in social research.

Model comparisons formulated in the above manner generally involve two related but different questions. The first is whether the increment in prediction or explained variability obtained by adding Z is significant or important. Standard procedures that test the significance of the incremental contribution of Z given X can be used to answer this question. For linear models, the usual F-test that compares the more comprehensive model that includes X and Z and the less comprehensive model that includes only X can be applied. For other members of the class of generalized linear models, such as logistic regression, analogous chi-square statistics can be used. Tests such as these for incremental contributions are based on the assumption that the more comprehensive (or "full") model is the model that generated the data (see, e.g., Blalock 1979; Hanushek and Jackson 1977; Neter, Wasserman, and Kutner 1989). That is, in conducting tests for *incremental* variability, the logic for this type of model comparison is predicated on the assumption that the more comprehensive model is true or is conditional on the truth of this model.

A second question is equally important when regression techniques are used to compare explanations. Are the coefficients that describe the relationship between Y and X in the first model different from the coefficients that describe the relationship between Y and X in the second model where Z has been added and has hence been "statistically controlled"? Or is the regression relationship between Y and X *stable* across specifications? This question cannot be answered in a valid way with conventional methods used for the assessment of incremental improvement in prediction. Those procedures have no *direct* relationship to inferential questions pertaining to the comparison of regression coefficients between models,

except in some special cases. In short, the question of incremental improvement in prediction is different from the question of stability in regression relationships, and methods suited for the former question need not be valid for considering the latter question. In spite of this, researchers often compare the coefficients associated with X between the models by examining whether one or more of the regression coefficients associated with X is significantly different from zero in either model; or whether the level of significance of those coefficients differs between the models; or whether the apparent size of those coefficients differs between the two models. Comparisons of this general kind are invalid because (*a*) they ignore the fact that the coefficients in the first model are not independent of the coefficients in the second model and (*b*) they ignore the fact that both models cannot be "true" at the same time, except in the case where the two models are equivalent (i.e., when the Z factors do not contribute to the regression). The latter difficulty implies that the inferential logic used to make statistical assessments of this kind is simply not appropriate.

In this article we give simple procedures that can be used to compare regression coefficients across models when linear models or generalized linear models are used. We suggest modifications in the reporting of regression results that provide direct evidence about the relative plausibility of the two explanations. In our view, these methods ought to be used in practically all empirical research work where linear or generalized linear models are used to compare explanations. These methods utilize quantities that are routinely provided by standard computer programs for regression analysis and thus can be implemented with little extra work (and no new computer programs).

In the next two sections we cover the simplest case involving three variables. This case is covered first to fix concepts and to illustrate the logical and inferential issues that arise. It also gives an alternative framework for the analysis of the case that so often serves as a prototype for regression analysis in social research. For this case it is straightforward to compare a simple regression coefficient to a partial regression coefficient and thereby *test* whether "controlling" for a variable suppresses or enhances the relationship between two variables (see Bollen 1989, chaps. 2, 3; Duncan 1975).[2] In Section IV we give the generaliza-

[2] We hasten to add that the methods presented here assume only that a "single-equation" regression model is available, having a full and a reduced version. This is very different from a "multiple-equation" model, which is featured throughout the literature on structural equation models. The distinction between the two approaches should be clear in the next section; the fact that we use only a single-equation model, with full and reduced versions, should be clear throughout.

tion for multiple regression, where either or both X and Z stand for sets of variables. Results in this section can be applied in all cases where multiple regression is used. Section V gives results for the class of generalized linear models, and includes two examples. Examples throughout the article are designed to illustrate the logic of the suggested procedures and also to indicate how the calculations can be done using quantities routinely provided by software for regression analysis. Section VI summarizes recommendations for applying these methods and suggests several generalizations of the approach useful for different research settings.

II. THE THREE-VARIABLE CASE

It is easiest to illustrate the logic of model comparison using the three-variable case where linear regression and least squares methods would be used. Let Y, X, and Z denote the three variables: Y is the specified outcome variable of interest and X and Z are predictors. Let Y_i, X_i, Z_i denote the values that these variables take on for the ith case, $i = 1, \ldots, n$, where n is the sample size. Two models are considered:

$$H_R: Y_i = \alpha_1 + \beta_1 X_i + \epsilon_i \qquad i = 1, \ldots, n, \tag{1}$$

and

$$H_F: Y_i = \alpha_2 + \beta_2 X_i + \gamma Z_i + \nu_i \qquad i = 1, \ldots, n. \tag{2}$$

We assume that error terms are well behaved, at least in the "full" model, H_F. For example, we assume that ν_i is $N(0, \sigma_\nu^2)$ (normally distributed, homoscedastic errors) and that errors are not correlated with each other (or errors are independent of each other with the normality assumption). We let σ_ϵ^2 denote the error variance under the first model, σ_ν^2 the error variance under the second model; and we denote the standard estimators or estimates of these as $\hat{\sigma}_\epsilon^2$, $\hat{\sigma}_\nu^2$ (i.e., the mean-squared errors for the respective models).

Now let $x_i = X_i - \overline{X}$, the centered values, and let y_i, z_i denote the centered values of the other two variables. The least squares estimator of β_1 is

$$b_{yx} = \hat{\beta}_1 = \Sigma x_i y_i / \Sigma x_i^2 = s_{yx}/s_x^2 = r_{yx} s_y / s_x, \tag{3}$$

where $s_x^2 = \Sigma x_i^2/n$, $s_{yx} = \Sigma y_i x_i/n$, $s_y^2 = \Sigma y_i^2/n$, and r_{yx} denotes the ordinary coefficient of correlation between Y and X. In this section only, we write the least squares estimator of the slope as $b_{yx} = \hat{\beta}_1$, using a popular notation that reminds us of the fact that the estimator (and the model) do not "control" for Z (or any other variable). *Assuming that the*

simple regression model is true, the sampling variance of this estimator is

$$V(b_{yx}|H_R) = \sigma_\epsilon^2/(ns_x^2). \tag{4}$$

Replacing σ_ϵ^2 by $\hat{\sigma}_\epsilon^2$ and taking the square root produces the estimated standard error of b_{yx}, say $s(b_{yx})$.

Using similar notation, the least squares estimator of β_2 in model H_F is

$$b_{yx\cdot z} = \hat{\beta}_2 = (r_{yx} - r_{xz}r_{yz})s_y/[(1 - r_{xz}^2)s_x], \tag{5}$$

where we use the popular notation $b_{yx\cdot z}$ for the partial regression coefficient (Blalock 1979). We see from this expression that $b_{yx} = b_{yx\cdot z}$ if $r_{xz} = 0$. Assuming that model H_F is true, the variance of this estimator is

$$V(b_{yx\cdot z}|H_F) = \sigma_v^2/[ns_x^2(1 - r_{xz}^2)] \tag{6}$$

(see, e.g., Hanushek and Jackson 1977, chap. 2). We note that both models cannot be true simultaneously, except in the special (and usually trivial) case where $\gamma = 0$. The first logical issue that has to be addressed is *which* model of the two being considered is "true," or which can be assumed to be true for inferential purposes, or which ought to be assumed to be true for logically consistent inferences (whether formal or informal).

Next consider the comparison of the simple and the partial regression coefficient given above. In the population, let $\delta = \beta_1 - \beta_2$, and in the sample let

$$d = \hat{\delta} = b_{yx} - b_{yx\cdot z}. \tag{7}$$

(or $d = \hat{\delta} = \hat{\beta}_1 - \hat{\beta}_2$). Note that d is simply the arithmetic difference between the simple regression coefficient and the partial regression coefficient. The quantity d is an important quantity in elementary analyses of causal systems involving three variables. We do not here invoke the causal imagery of three-variable recursive models, however, which would require that the "causal ordering" of X and Z be known, that an additional regression equation with either X or Z as the outcome be specified, that the variables included as predictors in both regressions are uncorrelated with the disturbances in the two models, and so on (Duncan 1975). It is very important to note that the logic involved here posits one outcome variable (Y); there are two competing models for this single outcome: the reduced model and the full model.

If both regression coefficients are positive and their difference is positive, then controlling for Z as in model H_F would be said to partially explain the relationship between Y and X. Of course, if $\beta_2 = 0$ (i.e.,

$b_{yx \cdot z}$ is not significantly different from zero), then we normally say that controlling for Z explains the relationship between Y and X, in which case $\delta = \beta_1$ (or d is approximately the same as b_{yx}). Using (3) and (5), we obtain

$$d = [r_{xz}(r_{yz} - r_{yx}r_{xz})s_y]/[(1 - r_{xz}^2)s_x]. \quad (8)$$

Note that d is identically zero when (*a*) $r_{xz} = 0$ (X and Z are uncorrelated) or (*b*) the coefficient for Z (i.e., $\hat{\gamma}$), which is proportional to $(r_{yz} - r_{yx}r_{xz})$, is zero (Z does not predict once X is included). These results ought to be intuitive, because if the two predictors are uncorrelated then the simple and the partial regression coefficients are identical, and because if Z does not predict at all in model H_F then this model is identical to H_R in the sample. Neither condition will hold true exactly when nonexperimental data are used; condition (*a*) will hold approximately if levels of X have been randomly assigned.[3]

III. INFERENCE ABOUT THE SIZE OF δ IN THE THREE-VARIABLE CASE

The previous section summarized results from least squares analysis of the linear regression involving three variables. Questions of inference—such as hypothesis tests, confidence intervals, and so on—were not addressed. We now elaborate on these issues with an eye on how b_{yx} and $b_{yx \cdot z}$ ought to be compared using the formulae in the preceding section as a guide.

To make inferences about the simple regression (H_R), it is ordinarily assumed that this model is "true." The usual t-statistic ($t = b_{yx}/s(b_{yx})$) is based on this assumption, for example. But if H_F is also considered, then it is normally assumed that this model is true for purposes of making inferences about the parameters of that model. *Both models cannot be true simultaneously,* however, in the common setting where one model represents one causal or structural explanation and the other model represents a different causal or structural explanation. Of course, the reduced model is related to the full model as $H_R = H_F + (\gamma = 0)$, which is the usual way to represent the fact that the reduced model is nested in the full model. Of the two models, H_F can always be assumed true; H_R is true when the full model is true *and* when $\gamma = 0$. This fact has important consequences for any comparison of the two models, not just the inference about $d = b_{yx} - b_{yx \cdot z}$.

[3] For practical purposes, we can assume that d will be nonzero with probability one in all cases where nonexperimental data are used and the sample size is relatively large. In the same situations, r_{xz} and $\hat{\gamma}$ will both be nonzero as well.

We now make the assumption that inferences should be conditional on the full model, H_F, holding true. In this case, we usually regard δ as the *bias* in the estimator of $\beta_{yx} = \beta_1$ when Z is an omitted variable in the former model; d is then the natural estimator of "omitted-variable bias" (Hanushek and Jackson 1977; Arminger 1995). We do not have to make a strong assumption that model H_F is true in some deeply philosophical sense, but we do have to make inferences conditional on *something* in order to maintain logical consistency. In our view, it is most natural to assume that statistical inferences should be conditional on the full model. This is consistent with virtually all research practice where full and reduced models are compared, and so this assumption should not be controversial. For example, this assumption is applied to test for the significance of the incremental variation explained by Z, which would be assessed in this case with a t-statistic, $t = \hat{\gamma}/s(\hat{\gamma})$, or the square of this statistic (the F-statistic for incremental variability explained in this case). The same assumption is now used to draw inferences about the difference in regression coefficients.

Both b_{yx} and $b_{yx \cdot z}$ are linear functions of the Y_i or the y_i. For example,

$$b_{yx} = \Sigma x_i y_i/(ns_x^2),$$

which shows that $b_{yx} = \Sigma w_i^* y_i$, with $w_i^* = x_i/(ns_x^2)$. The difference $d = b_{yx} - b_{yx \cdot z}$ is also a linear function of the y_i,

$$d = \Sigma w_i y_i. \tag{9}$$

Using (7), (3), and (5) and the definition of correlation leads to

$$w_i = x_i/(ns_x^2) - (s_z^2 x_i - s_{xz} z_i)/[n(s_x^2 s_z^2 - s_{xz}^2)], \tag{10}$$

where $s_{xz} = \Sigma x_i z_i/n$. The distribution of d is determined from this expression. For example, the variance of d is $V(d) = \Sigma w_i^2 V(y_i)$, assuming that observations are uncorrelated.[4]

If we assume that model H_F is true and hence condition on X and Z, then this means that we seek $V(d|H_F) = V(d|X, Z)$. This implies that we use the variance of Y given X and Z (or given model H_F), which we write as $V(y_i|H_F)$. Now $V(y_i|H_F) = V(y_i|x_i, z_i) = V(\nu_i) = \sigma_\nu^2$, the error variance under the second model. Direct algebra gives $\Sigma w_i^2 = r_{xz}^2/[ns_x^2(1 - r_{xz}^2)]$, so the standard deviation of the difference d is

$$\sigma(d) = \sigma_\nu |r_{xz}|/[ns_x^2(1 - r_{xz}^2)]^{1/2}. \tag{11}$$

[4] The term $V(y_i)$ actually stands for the conditional variance of y_i given the predictors used (or some other condition), and so the assumption is that, given those predictors (or some other conditions), the observations are conditionally uncorrelated or conditionally independent.

The estimated standard error is obtained by replacing σ_ν with $\hat{\sigma}_\nu$, the square root of the estimate of error variance in the full model. We see that this standard error is zero if $r_{xz} = 0$, in which case $d = 0$ (simple and partial coefficients are identical). Because $r_{xz} \neq 0$ can be assumed when nonexperimental data are analyzed and the sample is relatively large, the fact that $\sigma(d) = 0$ when $r_{xz} = 0$ does not affect the validity of the procedures given here. Note that we have assumed $r_{xz}^2 < 1$ throughout (i.e., X and Z are not colinear). Because we have conditioned on the full model (i.e., on both X and Z), this correlation is not a parameter, although the magnitude of this correlation affects the standard error and hence the power of tests based on it.

Some additional formulae shed light on the logic of comparing coefficients between models in this three-variable case. For model H_F, the standard deviation of $b_{yx \cdot z}$ is $\sigma_\nu/[ns_x^2(1 - r_{xz}^2)]^{1/2}$, and we see from (11) that

$$\sigma(d) = |r_{xz}|\sigma(b_{yx \cdot z}). \tag{12}$$

This expression shows the fallacy of making comparisons between the two coefficients assuming that the coefficients are independent or uncorrelated. If X and Z are correlated with $r_{xz} = .5$, for example, $\sigma(d)$ is one-half of $\sigma(b_{yx \cdot z})$. The sample analogue to (12) is $s(d) = |r_{xz}|s(b_{yx \cdot z})$, using the ordinary standard error of the partial regression coefficient as calculated in the least squares analysis of the full model.

Assuming that H_F is true, $d/s(d)$ follows a t distribution on $n - 3$ *df* under the hypothesis that $\delta = 0$, where $s(d)$ is obtained from (12) by substituting $\hat{\sigma}_\nu$ for σ_ν. It is easy to show also that

$$d = r_{xz}\hat{\gamma}s_z/s_x$$

(see Clogg, Shihadeh, and Petkova 1992, p. 57), and from this it follows that $t = d/s(d) = \hat{\gamma}/s(\hat{\gamma})$. In other words, the *test* for a significant difference between b_{yx} and $b_{yx \cdot z}$ is equivalent to the test for the significance of $\hat{\gamma}$ in model H_F, a result that holds *only* in the three-variable case considered here.

In a situation where this three-variable setup represents a realistic comparison of two explanations, we recommend calculating the difference $d = b_{yx} - b_{yx \cdot z}$ and the standard error of this difference, $s(d)$, as above. In this case only, the significance of d can be determined from the significance of the coefficient of Z, or the "test for omitted variable bias" and the test for incremental explained variability are identical. In this case, we recommend replacing the usual variance estimator (or the standard error) for the quantity b_{yx} by an alternative estimator that recognizes that model H_F, and not H_R, is assumed to be true. We now give simple results that show how an adjustment of this kind can be made.

We seek $V(b_{yx}|H_F)$, not $V(b_{yx}|H_R)$, which is estimated under the simple regression model. Because $b_{yx} = \Sigma\, w_i^* y_i$, with $w_i^* = x_i/(ns_x^2)$, and $V(y_i|H_F) = \sigma_\nu^2$, we obtain the preferred estimator of the standard error,

$$s^*(b_{yx}) = \hat{\sigma}_\nu/(ns_x^2)^{1/2}, \tag{13}$$

taking note of the fact that $V(b_{yx}|H_F) = \sigma_\nu^2\, \Sigma\, w_i^{*2}$. Let $s(b_{yx})$ denote the standard error calculated for model H_R. Then we see that the corrected standard error is

$$s^*(b_{yx}) = s(b_{yx})\hat{\sigma}_\nu/\hat{\sigma}_\epsilon.$$

Because $V(U - W) = V(U) + V(W) - 2\text{cov}(U, W)$ for arbitrary random variables U and W, the covariance between b_{yx} and $b_{yx\cdot z}$ can be obtained easily. That is, $V(d) = V(b_{yx}) + V(b_{yx\cdot z}) - 2\text{cov}(b_{yx}, b_{yx\cdot z})$, where the calculation assumes "variance under H_F" and "covariance under H_F." The first three quantities have already been obtained. Using (12), (13), and (6), we find $\text{cov}(b_{yx}, b_{yx\cdot z}) = \sigma_\nu^2/(ns_x^2)$, that is,

$$\text{cov}(b_{yx}, b_{yx\cdot z}) = V(b_{yx}|H_F). \tag{14}$$

Finally, the correlation between the two regression coefficients is $(1 - r_{xz}^2)^{1/2}$, which is independent of the error variances, and this gives yet another way to see how closely associated the two estimators are.[5]

We conclude this section with another formula for the variance of d that turns out to be suitable for the general case. Using results given immediately above, $V(d) = V[(b_{yx} - b_{yx\cdot z})|H_F]$ can be written as

$$V(d) = V(b_{yx\cdot z}) - V(b_{yx}|H_R)\sigma_\nu^2/\sigma_\epsilon^2. \tag{15}$$

This means that the variance of d can be estimated as $s^2(d) = s^2(b_{yx\cdot z}) - s^2(b_{yx})\hat{\sigma}_\nu^2/\hat{\sigma}_\epsilon^2$, where these quantities are the squared standard errors and error variance estimates produced as standard output by a regression package. The square root of this quantity is $s(d)$.

A. An Example

We illustrate these relationships with an example drawn from the 1990 General Social Survey (GSS). The dependent variable (Y) is occupational prestige as calibrated on the 1980 census occupational classification system. Predictor X is respondent's education in years (EDUC) while the predictor Z is father's education (PAEDUC), both scored as years of

[5] Note that the covariance between the two coefficients is positive and hence the correlation is nonnegative, ranging over the interval [0, 1). That is, the value of the simple regression coefficient and the value of the partial regression coefficient are positively correlated if $r_{xz} \neq 0$.

TABLE 1

COMPARISON OF COEFFICIENTS IN A THREE-VARIABLE CASE: OCCUPATIONAL PRESTIGE REGRESSED ON EDUCATION AND FATHER'S EDUCATION

Predictor	Model H_R	Model H_F	d	$t = d/s(d)$
Intercept	12.946	13.315	−.370	−.605
	(2.912)	(2.977)	(.611)	
EDUC	2.292	2.351	−.059	−.605
	(.210)	(.232)	(.098)	
PAEDUC		−.107		
		(.177)		
$\hat{\sigma}$	11.964	11.974		
R^2	.245	.246		
n	368	368		

SOURCE.—GSS (1990).

NOTE.—Cases pertain to all males ages 20–64 with nonmissing codes for all variables. Note that $\hat{\gamma}/s(\hat{\gamma}) = -.107/.177 = -.605$. SEs are in parentheses. Calculations were carried out to more decimal places than are reported.

schooling completed. The sample is all male respondents ages 20–64 with nonmissing codes on all variables; both regressions were estimated on the same cases, which is necessary for application of the usual model comparisons as well as for the ones suggested here.[6]

Results appear in table 1. The coefficient associated with EDUC in the simple regression is $b_{yx} = 2.292$; for the regression involving PAEDUC (Z) the corresponding coefficient is $b_{yx \cdot z} = 2.351$. Hence, $d = -.059$ and $s(d) = .098 = (.232^2 - .210^2\hat{\sigma}_v^2/\sigma_\epsilon^2)^{1/2}$ using (15) with substitution of the estimated error variances. The ratio of the two variance estimates is 143.3821/143.1338 = 1.0017, which is greater than unity because the PAEDUC variable is unimportant relative to the loss of a degree of freedom. Even in this case where most analysts would delete the additional predictor (i.e., choose the reduced model), the estimated standard error of d is informative. Note that $t = d/s(d) = \hat{\gamma}/s(\hat{\gamma}) = -.605$ (when more decimals are used). All of the preceding formulae can be checked out with this case. For example, (12) gives $s(d) = .0980 = .4224 \times .2320 = r_{xz}s(b_{yx \cdot z})$. (Note that EDUC and PAEDUC have a Pearson correlation of .4224 in the sample of cases used for this regression). Note also that the difference in the intercepts can be studied in exactly the same way, for reasons that will be made clear in the next section. The ratio of the difference in intercepts to the standard error is

[6] A difficulty in applying these formulas to published regression results is that missing-case exclusions are often applied differently from model to model, with the result that different samples are used for each model.

also equal to the t-ratio for the added covariate ($-.605$). If the full model with EDUC and PAEDUC were preferable, we could adjust the standard errors for the parameters in the reduced model by the ratio of the two estimates of standard deviations of the error terms (see [13] and surrounding text).

As shown earlier, the variance (and the standard error) of d was calculated assuming that the full model was true. And in this case, the test of the significance of d turns out to be equal to the test of the significance of $\hat{\gamma}$ in the full model. Under standard assumptions, including normality of the errors v_i in the full model, the ratio $t = d/s(d) = \hat{\gamma}/s(\hat{\gamma})$ follows a t distribution exactly if the null hypothesis is true ($\gamma = \delta = 0$, in this case).

B. Comparison with a Related Procedure

The above procedure is related to several large-sample tests for "model misspecification" as summarized in Godfrey (1988). Hausman's (1978) procedure is the most popular test of this kind. It will be considered briefly in relation to our procedure to illustrate important differences in logic and assumptions. Hausman's test applies to a wide range of model comparisons, not just to the type of model comparison considered in this article (we are not aware of examples in which the Hausman procedure is applied to the class of problems considered here, however). It has a very simple form for the problem of comparing two linear regression models where one is nested in the other.

The Hausman test in this setting gives $V(d) = V(b_{yx \cdot z}) - V(b_{yx})$, which is closely related to our procedure (see [15]). The sample estimator of this variance is used to construct tests similar to those presented here. A different method is used, however, to estimate this variance. Often in *practice,* each variance on the right-hand side is calculated under the model used to obtain each quantity; that is, $V(b_{yx \cdot z})$ is estimated from model H_F and $V(b_{yx})$ is estimated from model H_R. But the logic of the Hausman test is based on estimating the variance of d *under the reduced model,* so $V(d)$ should be understood as $V(d|H_R)$. In other words, Hausman's test assumes that the reduced model is true, whereas we assume that the full model is true. Large-sample approximations are applied to estimate this variance in general settings where the method is applied. (In general settings where the Hausman method is applied, it can be difficult to determine the variance of parameter values in a full model assuming that the reduced model is true, and perhaps this accounts for the common practice of using the estimated variances of those parameters in each model.)[7]

[7] If Hausman's test is applied correctly in this case, in the sense that $V(d|H_R)$ is actually calculated under the reduced model, then this test will be more powerful than our test if the reduced model H_R is true. If the full model is true, however, then

Hausman's procedure as applied to our setting consists of the following logic: (*a*) assume that the reduced model is true so that b_{yx} is both a consistent and an efficient estimator of $\beta_{yx} = \beta_1$; (*b*) under this condition, $b_{yx \cdot z}$ is a consistent estimator of $\beta_{yx} = \beta_1$ also; (*c*) if these assumptions are true, then $V(d)$—and tests of the significance of d—can be based on the above formula, with or without modification. There are many variations of the Hausman test in the literature. For example, Bentler and Chou (1993) present several statistics for assessing model improvement in the context of covariance-structure models that are closely related to Hausman's test and to the related Rao score tests. Their procedures are also based on an assumption that a given reduced model (or a model with one or more restrictions) is true. And their procedures have a different purpose, that of diagnosing lack of fit or of modifying the reduced model. These procedures are different from those presented here, because we assume that the full model is true.

Applying the Hausman method in the usual way gives the variance estimate, $.2302^2 - .2101^2 = .0097$, with square root (SE) .098, which is the same (to three decimal places) as the result derived from our method. But this result obtained because the ratio of mean-squared errors was approximately one in this case. If the variance of the difference is estimated without modification—without adjusting for the mean-squared errors—then in almost all cases where there is significant incremental variability due to Z the variance estimate derived from this approach will be *negative* (i.e., in cases where Z contributes significantly, the estimate of error variance will be reduced and the estimate of the variance $V[b_{yx \cdot z}]$ will be smaller than the estimate of the variance $V[b_{yx}]$ as a consequence).[8] The similarity in this case between the Hausman procedure, as typically applied, and our procedure is due solely to the fact that Z does not contribute significantly to the regression, as reflected in the fact that the ratio of mean-squared errors is nearly unity. The Hausman test could not be used, without modification, in cases where Z contributes significantly and the standard error of $b_{yx \cdot z}$ is smaller than the standard error of b_{yx}. The procedure based on Hausman's framework is very different from the one we have proposed because it actually conditions on the reduced model holding true, whereas we have conditioned on the full model holding true. If the modification described above is not used in the application of the Hausman test, the estimator of the variance (or SE) of d can be negative. This difficulty does not arise with the

---

a test should be based on the full model, and the precision of the d value will be estimated in a valid way with our approach in this case.

[8] An example of this is that the Hausman procedure could not be applied without modification in the example using multiple regression in table 2 below.

method presented here (see, e.g., eq. [12]). Also, under the maintained assumptions (the full model is true, the errors have constant variance, etc.), the test of the significance of d is exact for small as well as large samples, whereas the Hausman procedure is a large-sample test with asymptotic justification.[9]

The conditions assumed in the Hausman test and in related misspecification tests of this general kind are very different from the conditions assumed here. Similar comments apply to the procedures covered next.

IV. MULTIPLE REGRESSION

We now give the generalization to multiple regression. The reduced model is represented as

$$H_R: Y_i = \alpha^* + \beta_1^* X_{i1} + \ldots + \beta_p^* X_{ip} + \epsilon_i, \tag{16}$$

and the full model is represented as

$$\begin{aligned} H_F: Y_i = \alpha &+ \beta_1 X_{i1} + \ldots + \beta_p X_{ip} \\ &+ \gamma_1 Z_{i1} + \ldots + \gamma_q Z_{iq} + \nu_i, \end{aligned} \tag{17}$$

where $i = 1, \ldots, n$, and the same sample is used to evaluate both models.[10] There are p predictors in X and q predictors in Z; the previous sections covered the case where $p = q = 1$. Denote the least squares estimators of the regression coefficients associated with X_k as b_k^* in the former (reduced) model and as b_k in the latter (full) model, $k = 1, \ldots, p$. Call the corresponding vectors of regression coefficients b^* and b, respectively.[11]

Now let $\delta_k = \beta_k^* - \beta_k$ and $d_k = b_k^* - b_k$, quantities that follow from earlier considerations of the three-variable case. In vector terms, we have

[9] If the errors ν_i in the full model are normally distributed with constant variance, then d has an exact t distribution, and in large samples d will be normally distributed without the normality assumption.

[10] Note that we use an asterisk—instead of numbered subscripts as in the previous case—to denote the coefficient from the reduced model. Below we use the symbol b^* (or b_k^*) for the (least squares) estimator in the reduced model, the symbol b (or b_k) for the (least squares) estimator in the full model. We use the notation $\hat{\beta}$ to denote maximum-likelihood estimators. Of course, for linear models with normally distributed errors the least squares estimators are maximum-likelihood estimators.

[11] Note that the intercept can be regarded as the coefficient of the "predictor" defined as "the vector of ones." Then X would have $p + 1$ predictors, and we can examine differences in intercepts in the same way that we examine differences in the other coefficients. Inferences of this kind for intercepts could be of interest in the context of regression standardization (see, e.g., Clogg and Eliason [1986] and references cited therein).

$\delta = \beta^* - \beta$, $d = b^* - b$. In appendix A we show that $V[(b^* - b)|H_F] = V(b) - V(b^*|H_F)$. Note that $V(b^*|H_F) = V(b^*|H_R)\sigma_\nu^2/\sigma_\epsilon^2$, so that the variance-covariance matrix from the reduced model is multiplied by the ratio of error variances, as before, to produce the variance-covariance matrix under the full model. From these facts we obtain the simple expression,

$$s^2(d_k) = s^2(b_k) - s^2(b_k^*)\hat{\sigma}_\nu^2/\hat{\sigma}_\epsilon^2, \tag{18}$$

where $s^2(\cdot)$ refers to the estimated variance of the quantity involved; and $s^2(b_k^*)$ is the sampling variance (squared standard error) of the coefficient in the reduced model, as estimated in that model.[12] (Note that $s^2(b_k^*) \times [\hat{\sigma}_\nu^2/\hat{\sigma}_\epsilon^2]$ is the estimate or estimator of $V[b_k^*|H_F]$.) The scale factor is simply the ratio of the two mean-squared errors. The square root of $s^2(d_k)$ is the standard error of interest; we see that these can be calculated easily from standard output from the two regressions.

A. An Example Using Multiple Regression

To illustrate how these methods can be used, we consider regression models estimated with the 1990 GSS, with Y denoting occupational prestige and six predictors: X_1 and X_2, spline-coded variables for years of education less than or equal to 12 or greater than 12, respectively; X_3, father's occupational prestige score (when the respondent was 16 years old); X_4, age in years; X_5, father's education in years; X_6, mother's education in years; thus $p = 6$ predictors. The predictors added to form H_F are Z_1, a dummy variable for marital status (1 = currently married); Z_2, number of children in respondent's family; Z_3, unemployment history (dummy variable with 1 = ever unemployed); Z_4, a dummy variable indicating satisfaction with current residence; thus $q = 4$ additional predictors. Results appear in table 2. The F-ratio for the comparison of the full and the reduced models is 2.837, which is significant at less than the .05 level with a reference distribution of $F_{4,230}$. The ratio of mean-squared errors is $\hat{\sigma}_\nu^2/\hat{\sigma}_\epsilon^2 = 124.2156/128.1167 = .9696$, and this quantity is used in (18) and could be used to correct the estimates of standard errors in the reduced model.

Even though adding Z contributes to explained variability in a modest way (an increase in R^2 of .032), significant differences between the coefficients of several of the predictors in X are observed. The coefficient for post-high-school education (spline) decreases from 1.817 to 1.548 (d =

[12] Compare (18) with (15). Further insights about this expression can be found in the previous section and in the appendix.

TABLE 2

A COMPARISON OF TWO REGRESSION MODELS PREDICTING OCCUPATIONAL PRESTIGE

Predictor	Model H_R	Model H_F	$d_k = b_k^* - b_k$	$t = d_k/s(d_k)$
Intercept	18.194	21.766	−3.573	−1.45
	(8.273)	(8.508)	(2.457)	
X_1: education ≤ 12 years	1.057	1.176	−.120	−1.24
	(.607)	(.605)	(.096)	
X_2: education > 12 years	1.817	1.548	.269	2.27
	(.757)	(.754)	(.119)	
X_3: father's occupation	.126	.106	.019	2.20
	(.072)	(.072)	(.009)	
X_4: age	.139	.072	.067	2.40
	(.067)	(.072)	(.028)	
X_5: father's education	−.252	−.206	−.046	−1.29
	(.240)	(.239)	(.036)	
X_6: mother's education	−.019	−.028	.009	2.64
	(.034)	(.034)	(.003)	
Z_1: marital status		−3.393		
		(1.692)		
Z_2: no. of children		.090		
		(.561)		
Z_3: unemployment history		−2.803		
		(1.585)		
Z_4: satisfaction with residence		1.236		
		(1.514)		
$\hat{\sigma}$	11.319	11.145		
R^2	.320	.352		

SOURCE.—GSS (1990).
NOTE.—See text for definition of variables. For both models $N = 241$.

.269), which is a difference of more than two standard errors. By controlling for the factors in Z, the coefficients for father's occupation, age, and mother's education have also changed in important ways. As a block, the coefficients for the six predictors have changed significantly as well. Because $p = 6 \geq q = 4$, the test for significance of the vector $d = b^* - b$ is equivalent to the incremental F-statistic ($F = 2.837$, which is significant at less than the .05 level compared to the reference distribution of $F_{4,230}$; see Clogg, Petkova, and Shihadeh 1992).[13]

[13] Note that the special case where $p = q = 1$ was covered in the section on inference in the three-variable case. The squared t-statistic for the significance of d is identical to the incremental F-statistic for the significance of the predictor Z. When $p < q$, a related but different test for block differences has to be used. This is described briefly in the appendix and in Clogg et al. (1992). Finally, note that standard methods can be used to make simultaneous tests for some, but not all, differences. For example, if s differences were of interest, a simple (conservative) method of adjustment is to use $\alpha^* = \alpha/s$ instead of α for each of the s comparisons, which is an application of the Bonferroni method.

To summarize, we find that adding the four predictors in Z changes the relationship between Y and X significantly, and four of the coefficients have been singled out as the source of the difference. Finally, note that the tests based on the d_k values are very different from tests on the individual regression coefficients. For example, the coefficient for mother's education does not appear to be significantly different from zero in either model; however, the difference between these two coefficients is very significant. The explanation offered by the reduced model is thus quite different from the explanation offered by the full model, not only in terms of additional predictors used but also in terms of several coefficients describing the Y–X relationship. Whether these differences are substantively important or not would depend on a variety of factors, but a rigorous and logically consistent assessment of these differences can be based on the d_k, the $s(d_k)$, and their ratios.[14]

B. A Note on Comparing Regression Coefficients between Groups

It is instructive to consider briefly the relationship between the proposed framework and methods for comparing regression relationships between groups. Suppose that the identical regression model corresponding to, say, H_R, were estimated in two groups such as males and females. The separate regressions can be retrieved from a regression model for both groups combined by adding a dummy variable D that denotes the group variable as well as the interaction of D with each of the predictors. If the error variance is the same in each group, then group differences in regression coefficients can be assessed in a valid way by testing the interaction terms, singly or jointly. When the error variances are not the same, different methods have to be used (see Gujarati 1988).

In large samples, the significance of the difference between the coefficient $\hat{\beta}_m$ and the coefficient $\hat{\beta}_f$ (the coefficient in the male sample and the coefficient in the female sample, say) can be assessed with the statistic

$$z = (\hat{\beta}_m - \hat{\beta}_f)/[s^2(\hat{\beta}_m) + s^2(\hat{\beta}_f)]^{1/2},$$

which follows a standard unit normal under the null hypothesis of equality of the two coefficients. The standard error of the difference is the square root of the sum of the two squared standard errors, assuming that the samples are independent.

Our procedures appear at first sight to be similar to the standard

[14] Significance of the incremental contribution of Z need not be associated with significant differences in the d_k values, even if the components of X and the components of Z are moderately correlated. An example demonstrating this appears in Clogg et al. (1992); also see the first example in the next section.

method outlined above, but actually they are very different. First, in our method the comparison of the simple regression coefficient (or any coefficient in a reduced model) and the partial regression coefficient (or any partial regression coefficient for predictors common to both the reduced and the full model) is based on the same sample, not different (independent) samples. The two quantities are correlated, as shown above. We might randomly split the sample into two groups and estimate model H_R for one sample and model H_F for the other, however, and then use a modification of the standard procedure. If the two samples are based on random selection, then we could, say, compare b_{yx} from the first sample with $b_{yx \cdot z}$ in the second sample. But the two models condition on different things, and this has to be taken into account. That is, a z-statistic analogous to that given above for the comparison of coefficients *from the same models* applied to independent samples (with b_{yx} used in place of $\hat{\beta}_m$ and $b_{yx \cdot z}$ used in place of $\hat{\beta}_f$) cannot be used to compare the simple and the partial regression coefficients, even if each is estimated from randomly selected subsamples. For example, the sampling variance of b_{yx} in the first sample needs to be modified to take account of the fact that the comparison should be conditional on the full model. An approximate way of doing this is to take $s^2_*(b_{yx}) = s^2(b_{yx}) \times MSE^*_F/MSE^*_R$, where MSE^*_F denotes the estimate of error variance for the full model as estimated with the randomly selected subsample and MSE^*_R is the estimate of error variance for the reduced model as estimated with the other randomly selected subsample. Although this would be a valid procedure in large samples, it is not efficient, as can be seen from the fact that the full sample is not used to estimate the error variance under the full model. Similar comments would apply to the extensions of our method given next if they were compared to common methods for group comparisons of regression coefficients.[15]

V. GENERALIZED LINEAR MODELS

The above procedures for linear regression will now be extended to the class of generalized linear models (McCullagh and Nelder 1989). That is, procedures are now developed for log-linear models for categorical data (Agresti 1990; Goodman 1978, 1984), logistic regression for binary outcome variables, most logit-type models for polytomous outcome variables, Poisson-regression models for event-count data, and most of the event history regression models used widely in social research. A frame-

[15] We are indebted to a referee for suggesting consideration of this approach.

work for analysis of this class of models was presented in Clogg, Petkova, and Shihadeh (1992); see Petkova and Clogg (1993) for derivations that are summarized here. We consider the general case as well as some specific examples.

The reduced model will be represented as

$$H_R: \eta_i = \alpha^* + \beta_1^* X_{i1} + \ldots + \beta_p^* X_{ip}, \tag{19}$$

and the full model will be represented as

$$H_F: \eta_i = \alpha + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \gamma_1 Z_{i1} + \ldots + \gamma_q Z_{iq}, \tag{20}$$

where $i = 1, \ldots, n$. Here, η_i refers to a function of the expectation of Y, say $\eta_i = g(\mu_i)$, with $E(Y_i) = \mu_i$. The classical regression model in the previous sections is obtained by taking g as the identity "link." If Y_i refers to a count in cell i of a contingency table, then g is the logarithm and (19) and (20) represent two log-linear models for a contingency table with n cells. If Y is a binary dependent variable with $E(Y_i) = \pi_i$, the probability of "success," then taking $g(\pi_i) = \log[\pi_i/(1 - \pi_i)]$ gives the standard logit model. If Y is multivariate, as with multinomial logistic regression, g is then a multivariate link function, usually a set of logits with one category chosen as a base. (If all variables are discrete so that the data can be arranged in a contingency table format, then the log link can be used. The log-frequency model can be reformulated in the usual way as a model for logits; see Goodman 1978.) The framework given here also applies to multivariate outcomes with appropriate modifications. We thus deal with a class of models that is wide enough to cover most applications of regression methods in social research.[16]

For model H_R, denote the "weight matrix" associated with a given generalized linear model as W^*. Denote the corresponding weight matrix for model H_F as W. For the ordinary log-linear analysis of contingency tables, with F_i denoting the expected frequency in cell i, W^* is the diagonal matrix with the expected frequencies under model H_R on the diagonal, and W is the diagonal matrix with the expected frequencies under model H_F on the diagonal. For logistic regression involving a binary dependent variable, the matrices are also diagonal matrices, with the

[16] The conditions leading to $\beta^* = \beta$ ("Z is collapsible") for the class of generalized linear models are different from those in the ordinary linear model. Agresti (1990) summarizes these. The main restriction in our formulation is that we consider only members of this class of models that do not include a special overdispersion parameter for unmodeled heterogeneity.

entry in the ith diagonal defined as $w_i = \pi_i(1 - \pi_i)$ (for the full model), with a corresponding entry for the reduced model (see Agresti 1990, chap. 3). These weight matrices are readily available with standard routines or can be easily calculated from those routines.[17]

Now let β and β^* denote the relevant coefficient vectors where the intercept has been included, both then with $p + 1$ elements. The variance of $\hat{\beta}^*$ under model H_R is the $(p + 1) \times (p + 1)$ matrix,

$$V^*(\hat{\beta}^*) = (X^T W^* X)^{-1}, \tag{21}$$

where the asterisk denotes the reduced model. This quantity is used to calculate the standard errors reported for this model, and this matrix is available from almost all software packages that would be used to estimate this model. The variance of $\hat{\beta}^*$ *under model* H_F is the so-called sandwich estimator,

$$V(\hat{\beta}^*) = (X^T W^* X)^{-1}(X^T W X)(X^T W^* X)^{-1} \tag{22}$$

(see, e.g., White [1980] and Godfrey [1988] for the general form of this estimator). Finally, the covariance between $\hat{\beta}^*$ and $\hat{\beta}$ can be shown to be equal to

$$\text{cov}(\hat{\beta}^*, \hat{\beta}) = V^*(\hat{\beta}^*) = (X^T W^* X)^{-1}. \tag{23}$$

Derivations appear in Petkova and Clogg (1993).[18]

The variance of the estimated difference ($\hat{\delta} = \hat{\beta}^* - \hat{\beta}$) is thus the $p \times p$ matrix,

$$V(\hat{\delta}) = V(\hat{\beta}) + V(\hat{\beta}^*) - 2\text{cov}(\hat{\beta}^*, \hat{\beta}), \tag{24}$$

where $V(\hat{\beta})$ is the variance matrix of the regression coefficients for X in the full model (standard output), and the remaining two quantities are defined in (22) and (23). Note that the variance and the covariance terms in this expression are calculated under the full model, which is consistent with the inferential logic put forth earlier for the linear model. For com-

[17] These weight matrices are used to define the variance estimators, based on the information function, and so are used implicitly to calculate standard errors in routines using maximum likelihood estimation. The upper-left-hand block of the information matrix is $-(X^T W X)$ for the full model, e.g.; and the estimated information is obtained by substituting $\hat{W}$ in this expression.

[18] Note that the covariance under the full model is (asymptotically) equal to the variance of $\hat{\beta}^*$ under the *reduced* model, which is different from the corresponding *exact* variance obtained from the ordinary linear model. In this section all variance and covariance calculations are "large-sample" quantities and must be understood in this light.

parisons of individual coefficients, the standard error is obtained by taking the square root of the relevant diagonal entry in this matrix. This result can be simplified further. Using (21)–(23), we obtain

$$V(\hat{\delta}) = V(\hat{\beta}) + V^*(\hat{\beta}^*)(X^T W X)V^*(\hat{\beta}^*) - 2[V^*(\hat{\beta}^*)], \qquad (25)$$

where $V(\hat{\beta})$ is the variance matrix for the coefficients associated with X in the full model and $V^*(\hat{\beta}^*)$ is the variance matrix for the coefficients associated with X in the reduced model. These quantities will be provided as standard output, perhaps with a user-specified option, in virtually all programs for the analysis of generalized linear models or special cases of them. The only additional quantity to be calculated which is not a by-product of the calculations is the matrix $(X^T W X)$. This matrix can be obtained by inverting the variance-covariance matrix of parameter estimates for the full model, taking the matrix block of the result that corresponds to elements of X and taking care to ensure numerical accuracy. This matrix can be obtained in other ways as well; see below.

It can be noted that the usual Hausman (1978) test, if applied to this particular comparison of nested models, would approximate the variance as $V(\hat{\beta}) - V^*(\hat{\beta}^*)$, using the estimated variance matrices under each model. Both the variance estimator and the suggested test proposed here differ from the Hausman procedure. Estimates obtained from Hausman's formula for the variance can be negative (or nonpositive definite; see also Holly 1988 and Sec. III above), whereas the above estimator is nonnegative. The inferential logic associated with our approach is related to but different from that associated with the Hausman procedure. We do not assume, for example, that the reduced model is true; instead we condition on the full model (or assume that it is true) just as in the standard comparison of the two models using the difference between the two-model chi-square statistics.

A. Example: Collapsibility in Contingency Tables

To illustrate how the above results can be used in the analysis of contingency tables, we analyze the classic Lazarsfeld (1948) panel data analyzed with log-linear models in Goodman (1973) and with other methods by others. While the example illustrates how to apply the techniques in general settings where contingency tables are used, we also develop in this context a method for analyzing the "two faces" of panel analysis (Duncan 1981).

In this example, there are four dichotomous items leading to a four-way cross-classification. By comparing full and reduced models as specified below, the hypothesis of collapsibility can be examined in a rigorous way. Collapsibility with respect to a given parameter or quantity means that

TABLE 3

LAZARSFELD PANEL DATA:
OBSERVED FREQUENCIES AND FITTED FREQUENCIES UNDER TWO MODELS

CELL (*A*, *B*, *C*, *D*)	OBSERVED FREQUENCIES	FITTED FREQUENCIES	
		Model $H_F(\hat{F})$	Model $H_R(\hat{F}^*)$
(1,1,1,1)	129	128.10	72.07
(1,1,1,2)	3	5.45	13.70
(1,1,2,1)	1	.80	7.61
(1,1,2,2)	2	.65	41.62
(1,2,1,1)	11	11.58	18.68
(1,2,1,2)	23	20.87	3.55
(1,2,2,1)	0	.07	1.97
(1,2,2,2)	1	2.48	10.79
(2,1,1,1)	1	1.99	12.81
(2,1,1,2)	0	.08	2.44
(2,1,2,1)	12	12.11	1.35
(2,1,2,2)	11	9.81	7.40
(2,2,1,1)	1	.33	38.44
(2,2,1,2)	1	.60	7.31
(2,2,2,1)	2	2.01	4.06
(2,2,2,2)	68	69.06	22.20

SOURCES.—Lazarsfeld (1948) and Goodman (1973).

NOTE.—*A* and *C* refer to vote intention, Republican vs. Democrat, at the first and second interviews, respectively; *B* and *D* refer to opinion of the Republican candidate, favorable vs. unfavorable, at the first and second interviews, respectively. See text for description of the models. $N = 266$.

the given parameter is the same in the *marginal* table as in the *full* table with no variables collapsed, i.e., that marginal and partial "association" are equivalent (see, e.g., Agresti [1990]; Clogg et al. [1992]; and references in the latter source). For the analysis of collapsibility involving dichotomous variables, relatively straightforward methods can be used (Ducharme and Lepage 1986). But the analysis of collapsibility using existing methods is not so straightforward when polytomous variables are involved, particularly when the table has more than three dimensions. In contrast, the methods given here can be applied to examine collapsibility for arbitrary contingency tables.

Table 3 gives the observed frequencies in the cross-classification of the four items as well as estimated expected frequencies under two models. Variables *A* and *B* refer to vote intention (1 = Republican, 2 = Democrat) and candidate opinion (1 = favorable toward Republican, 2 = unfavorable) in the first interview, and *C* and *D* are similarly defined variables for the second interview. Model H_F is the model fitting the marginals $\{(AB), (AC), (BD), (CD)\}$. This model fits the data well with a

likelihood-ratio chi-square value of 6.74 on 7 *df*.[19] We see that this model includes across-time relationships, or "lagged effects," for the two repeated measures. That is, this model includes the interaction between A and C and the interaction between B and D. But the interaction between opinion at time 1 and intention at time 2 and the interaction between intention at time 1 and opinion at time 2 are both deleted in this model (i.e., the AD and BC interactions are excluded). The model does, however, include the contemporaneous "partial" association of the two variables at time 2, in the C–D interaction term. The reduced model H_R is the model fitting the marginals $\{(AB), (CD)\}$, which says that the joint variable AB is independent of the joint variable CD. This "reduced" model fits poorly, with a likelihood-ratio chi-square value of 497.33, 12 *df*. We would not take this model seriously as a description of the distribution in the four-way table, but this model has a special property that permits the analysis of collapsibility. This property is exploited next.

Collapsing over the first-wave measurements, A and B, gives the 2×2 cross-classification of C and D, the second-wave measurements. The frequencies in this table are

$$\begin{pmatrix} 142 & 27 \\ 15 & 82 \end{pmatrix}$$

(levels of C in the rows, levels of D in the columns). The logarithm of the odds ratio is $\hat{\phi}^* = 3.3587$, with standard error $s(\hat{\phi}^*) = .3506$. The log-linear model for this marginal table, with expected frequencies denoted by F_{++kl}, is

$$\log(F_{++kl}) = \lambda^* + \lambda^*_{C(k)} + \lambda^*_{D(l)} + \lambda^*_{CD(kl)},$$

using the standard definition of parameters. We find (Agresti 1990; Goodman 1973) that $\lambda^*_{CD(11)} = \phi^*/4$, $\hat{\lambda}^*_{CD(11)} = \hat{\phi}^*/4$, $s(\hat{\lambda}^*_{CD(11)}) = s(\hat{\phi}^*)/4$. The asterisk refers to the fact that the parameter or quantity pertains to the marginal $(C \times D)$ table. Now because model H_R for the full table posits independence between AB and CD, the parameters, parameter estimates, and standard errors for the above model can be obtained from the parameters, parameter estimates, and standard errors, respectively, for the log-linear model for the full table corresponding to H_R, that is, from

$$H_R: \log(F^*_{ijkl}) = \lambda^* + \lambda^*_{A(i)} + \lambda^*_{B(j)} + \lambda^*_{AB(ij)} + \lambda^*_{C(k)} + \lambda^*_{D(l)} + \lambda^*_{CD(kl)}.$$

It is also the case that the marginal association between A and B can be obtained directly from the A–B interaction in this expression, and so on. We now examine rigorously whether the marginal C–D interaction given

[19] This model is equivalent to Goodman's (1973, table 5) model H_{10}.

from the collapsed table (or in the model H_R) is different from the *partial* C–D interaction in the model H_F.[20] The full model is

$$H_F: \log(F_{ijkl}) = \lambda + \lambda_{A(i)} + \lambda_{B(j)} + \lambda_{AB(ij)} + \lambda_{C(k)} + \lambda_{D(l)} + \lambda_{CD(kl)} + \lambda_{AC(ik)} + \lambda_{BD(jl)},$$

where we have arranged terms to correspond to those used in H_R above. Except for notational differences, we now have a full and reduced model of the kind introduced at the beginning of this section.

Given the equivalence between the model H_R for the full table and the "model" giving parameters for marginal association in both the A–B and the C–D tables after collapsing, the marginal and partial association can be compared statistically using the same kind of tabular layout as was used earlier for the standard regression problem. That is, we can study rigorously the "two faces" of panel analysis (Duncan 1981) to determine if the contemporaneous marginal association at wave two (the relationship between C and D in the marginal $C \times D$ table) is different from the partial association between C and D in the full table (full model) "controlling" for across-time relationships. The relevant quantities appear in table 4. All results were obtained using (21)–(24), with the diagonal weight matrices estimated from the corresponding fitted values in table 3.

We see from table 4 that the C–D marginal interaction value, .840 corresponding to a log-odds ratio of 3.36 (odds ratio of 28.75), is not significantly different from the C–D partial interaction value, .736 corresponding to a partial log-odds ratio of 2.95 (partial odds ratio of 19.03). The difference (.103) is not significant compared to the standard error (.076). (The standard error is obtained by taking the square root of the relevant entry in the variance matrix defined above.) In summary, "controlling" for across-time relationships included in model H_F does not change the inference about the relationship between intention and candidate opinion at the second wave in a significant way, so there does not need to be a special explanation for how contemporaneous association is produced by lagged "effects" (Duncan 1981). That is, in this example the two faces in the two-wave panel are indistinguishable. Similar comments

[20] Other quantities can be examined in this fashion, but it does not make sense to compare the marginal A–B interaction to the "partial" A–B interaction because variables C and D refer to second-wave measurements and thus should not be "controlled for" when studying the association between variables at the first wave. In H_R we have included the terms relevant for studying the AB marginal table to indicate how such a comparison could be made in a context where this type of comparison is relevant. Exactly the same results would be obtained if the A, B, or AB terms were deleted from H_R, even though the fitted frequencies for the reduced model would change if we deleted those terms to form the reduced model.

TABLE 4

A COMPARISON OF TWO LOG-LINEAR MODELS FOR THE LAZARSFELD PANEL DATA: MARGINAL AND PARTIAL ASSOCIATION COMPARED

Parameter	Model H_R	Model H_F	$d_k = b_k^* - b_k$	$z = d_k/s(d_k)$
C	.284	.081	.203	1.12
	(.088)	(.201)	(.181)	
D	−.009	−.095	.104	1.43
	(.088)	(.114)	(.073)	
CD	.840	.736	.103	1.35
	(.088)	(.112)	(.076)	
A	.251	.210	NR	
	(.076)	(.194)		
B	.063	.113	NR	
	(.076)	(.100)		
AB	.612	.152	NR	
	(.076)	(.108)		
BD		.936		
		(.111)		
AC		1.721		
		(.198)		

NOTE.—Comparisons not relevant (NR) for A, B, and A–B terms; see text.

apply to the C and D main-effect terms, although we would not ordinarily be interested in these terms as such. (Because none of the three terms differs significantly between the models, we can say that the entire CD table "structure" is not affected by controlling for lagged effects.) Note that significant differences might have been expected because the full model is a dramatic improvement over the reduced model in this case.

Straightforward modifications of the above approach can be used to examine any collapsibility question if collapsibility is defined in terms of parameters in log-linear models. If polytomous variables were used, for example, the comparisions between two-factor or multifactor interactions across models formed in the above fashion would lead to chi-square tests with more than one degree of freedom (see, e.g., Clogg et al. 1992). Note that the logic is applicable to contingency-table modeling in general, such as for the analysis of mobility tables, and in even more general circumstances. The reduced model need not correspond to a collapsed table, but it must be nested within the full model in order to apply this framework. For the analysis of a given mobility table, for example, the reduced model might correspond to quasi-perfect mobility having a parameter for each cell on the main diagonal, and the full model might correspond to an association model that also includes parameters for the cells on the main diagonal. This approach could be used to determine

whether adding the parameters for off-diagonal association in the full model changes the inference about the magnitude of "persistence" effects for the cells on the main diagonal. The general suitability of this approach for answering questions of this kind is one advantage of this approach compared to others in the contingency-table area (see, e.g., Ducharme and Lepage 1986; Whittemore 1978). The approach is directly suited for conventional analyses of collapsibility where a *variable* is collapsed, as indicated by the above example.

B. Logistic Regression: Example

Our final example pertains to logistic regression where categorical or continuous predictors are used. Note that the literature on collapsibility in contingency tables is not relevant when one or more of the predictors are continuous. (If predictors are categorical, the previous example can be used as a guide for the analysis, taking into account the standard relationship between a log-linear model for frequencies and the logit model.) The necessary formulae are still (21)–(25), and we simply need to supply the matrix $(X^T W X)$. This matrix can be calculated as follows. For the full model H_F, denote the fitted logit model as

$$\hat{\Phi}_i = \hat{\alpha} + \hat{\beta} X_i + \hat{\gamma} Z_i.$$

The predicted probability for the ith case is thus $\hat{\pi}_i = \exp(\hat{\Phi}_i)/[1 + \exp(\hat{\Phi}_i)]$, and the ith element of $\hat{W}$ is hence $\hat{w}_i = \hat{\pi}_i(1 - \hat{\pi}_i)$. The matrix of interest can then be calculated in a variety of ways, for example, by transforming the variables and obtaining the sum of squares and cross-products for an OLS regression. Generalizations of this procedure can be used for any member of the class of generalized linear models; each is characterized by a different weight matrix, which depends on the link function.[21]

Table 5 presents results from two logistic regression models estimated from the Supplement on Aging (SOA) of the 1984 National Health Interview Survey (NHIS). The models predict whether the respondent reported "fair or poor health" ($Y = 1$) versus "good health" ($Y = 0$). The first model (H_R) predicts the logit of prob($Y = 1$) from self-reports of "symptomatic" conditions or "asymptomatic" conditions summarized in the notes to the table. Model H_F adds $q = 7$ other predictors designed to measure demographic and other characteristics of the respondent, in-

[21] We can also obtain the estimate of the matrix $X^T W X$ by inverting the variance-covariance matrix of parameters for the full model; the upper-left-hand block of this matrix corresponding to the predictors in X, including the intercept, is the desired quantity.

TABLE 5

LOGISTIC REGRESSION MODELS PREDICTING SELF-REPORT OF POOR OR FAIR HEALTH

Predictor	Model H_R	Model H_F	$\hat{\delta}_k$	$z = d_k/s(d_k)$
Constant	−2.136	−.747	−1.379	−7.57
	(.057) [.054]	(.192)	(.182)	
Symptomatic	1.628	.887	.742	37.14
	(.060) [.056]	(.067)	(.020)	
Asymptomatic	.613	.508	.106	7.22
	(.085) [.079]	(.092)	(.015)	
Limitations		.432		
		(.009)		
Age		−.018		
		(.003)		
Female		−.412		
		(.045)		
Nonwhite		.546		
		(.071)		
Not married		−.176		
		(.048)		
Education		−.549		
		(.034)		
Metro/noncentral		−.040		
		(.054)		
Nonmetro		.148		
		(.148)		
Model χ^2	1,125.66	4,677.24		

SOURCE.—Supplement on Aging, 1984 National Health Interview Survey.

NOTE.—The data pertain to N = 15,808 individuals age 55 and older in 1984. Symptomatic conditions include rheumatic heart disease, coronary heart disease, angina pectoris, myocardial infarction, any other heart attack or stroke or cerebrovascular accident, broken hip, rheumatic fever, Alzheimer's disease, cancer of any kind ("ever had"); or arthritis or rheumatism, diabetes, aneurysm, blood clots ("had in past 12 months"). Asymptomatic conditions include hypertension, hardening of the arteries, osteoporosis ("ever had"), or varicose veins ("had in past 12 months"). "Limitations" is number of functional limitations; residence was measured in three categories: metropolitan-central-city (deleted category), metropolitan-not-central-city (metro/noncentral), and nonmetropolitan (nonmetro). SEs are in parentheses; SEs under the full model for the coefficients in the reduced model are in brackets.

cluding residence. Comparing the coefficients for symptomatic conditions and asymptomatic conditions between models indicates how these other status measures influence the relationship between reported conditions and self-reported health; the comparison thus can be used to quantify the degree to which self-reported conditions translate into perceptions of poor health. This relationship has implications for the survey measurement of health status (see Johnson and Wolinsky 1993).

Because the sample size is so large, it is not surprising that the observed differences for the two predictors of interest change significantly. It seems clear that controlling for the status and other characteristics (Z factors) produces a dramatic change in the coefficient for symptomatic conditions (the difference is .74), which is over 37 SEs from zero. The coefficient

for asymptomatic conditions changes less (the reduction is .11), but is still highly significant, over 7 SEs from zero. Controlling for the status, residence, and other factors in Z thus changes the relationship between reported health conditions and self-reported health in a dramatic way. This finding can either be construed as a problem in using (subjective) self-reports of health or as an indication of the extent to which health conditions are mediated by status and other factors and so translate objective health conditions into self-reported perceptions of health.

A block test of no significant differences in any of the three coefficients in the reduced model based on this approach leads to chi-squared statistic 1,794.35, 3 *df.* Obviously, most of the block difference is due to the dramatic change in the coefficient for symptomatic conditions. Note that all variance (or SE) calculations as well as this chi-square value condition on the full model being true. That the full model is an improvement over the reduced model cannot be doubted; the likelihood-ratio statistic comparing the two models is 3,551.58 = 4,677.24 − 1,125.66, 8 *df,* using the "model chi-square" values given in table 5. In this case, we thus find both a significant increment due to adding Z and significant change in the coefficients associated with X. The previous example was a case where the increment by adding Z was very significant, but still there was not significant change in the main parameter of interest.

This procedure can also be contrasted with the Hausman (1978) test, as this procedure would be used for this type of model comparison. This procedure conditions on the reduced model, which is definitely not "true" given the above results. Then $V(\hat{\delta})$ would normally be approximated as $\hat{V}(\hat{\beta}) - \hat{V}^*(\hat{\beta}^*)$, and the Hausman chi-square statistic is $H = \hat{\delta}^T[\hat{V}(\hat{\delta})]^{-1}\hat{\delta}$. The value of H in this case is 882.24, which is very different from the value obtained above.[22] Once again the inferential logic distinguishes this approach from others: as with ordinary tests for incremental contributions, our approach conditions on the full model holding true. Straightforward modifications of the procedure suggested here can be used to develop the estimators and tests for other regression-type models in the class of generalized linear models (Poisson regression for event-count data, multinomial logistic regression, many models for event history data, etc.).

VI. CONCLUSION

Examples used to illustrate the application of this method of comparing regression coefficients were designed mostly to show how calculations

[22] Both statistics lead to rejection, but they are each specified under different conditions, so it is not a surprise that these values differ. Note that the reduced model would be decisively rejected and so a test assuming that the reduced model is true could lead to very different results compared to a test assuming that the full model is true.

can be done easily with standard output from routines for regression analysis. With the exception of the example comparing the two faces of panel analysis (in the previous section), we do not claim that these examples demonstrate well the power or the flexibility of the suggested framework. We thus conclude with some observations on the potential role of these methods in theoretical explanation and in answering some methodological questions that are central in several substantive areas.

The concept of full and reduced models used throughout this article is obviously more than a statistical framework that makes inferences consistent and valid. In sociology as well as other areas, different theoretical explanations or "causal structures" are most often represented in this fashion. If a given theory or causal structure is represented by our equivalent of the reduced model (H_R) with specified predictors X, then it is important to examine the *stability* of the coefficients associated with X when various modifications are made in the explanation or causal structure. The most common way to modify a given explanation is to consider various other models that also include X but include additional factors (Z). This logic is closely related to the discovery and validation of "laws" (Pratt and Schlaifer 1988), and is or ought to be a vital aspect of theoretical analysis where alternative explanations are truly pitted against each other (cf. Jasso 1988). Direct application of this framework would seem to confront this general issue and also give explicit "rules" for elaborating particular theories represented as linear equations. In the vast majority of applications of regression methods for either causal inference or theoretical explanation in social research, analysts are implicitly examining questions of stability or change in regression coefficients (or regression relationships) under alternative specifications. Our framework can be used to examine those questions explicitly, logically, and with valid statistical methods. We do not claim, however, that this framework gives an automatic recipe for causal inference, only that the comparisons developed here provide assistance in this task. (The problem of causal inference with regression models and "nonexperimental data" does not reduce to just stability in coefficients under alternative specifications; see Smith [1990] and references cited therein.)

Although not a panacea for causal inferences using regression methods, there are several important methodological or substantive questions that can be addressed by this general framework. We conclude by mentioning some of these. (1) In an experiment or a quasi-experiment where X denotes the treatment levels "assigned," comparing the coefficient of X in the reduced model with the coefficient of X in various full models with added covariates provides information about the reliability of the causal inference that would be drawn. If the difference d is substantial or significant under alternative full models, for example, then this indicates

that the randomization (or quasi-randomization) was not effective. (2) It is often the case that either X or Z is categorical or can be categorized in various ways. By specifying the categorizations as full and reduced versions of a model of interest, the loss of information or the change in the inference associated with different ways of combining categories can be examined rigorously by means of the framework examined here. An example in a contingency-table setting appears in Clogg and Shihadeh (1994, chap. 2). (3) Many questions related to contextual analysis or group effects (see Manski 1993; Bryk and Raudenbush 1992) can be illuminated by application of the suggested framework. How to aggregate the groups used to define social context is an obvious case in point. If, say, X denotes individual-level factors and Z denotes group-level factors, in the simplest version of an analysis of this kind, then contextual analysis ought to be telling us not just whether Z contributes incrementally but whether the individual-level inferences (coefficients for X) change once the macrolevel factors are included. (4) Finally, at least some questions concerning selectivity bias can be addressed using this approach. Many models for selective samples have been proposed (cf. Manski [1993] and references cited therein). In many cases, selectivity is modeled by introducing the formal equivalent of added covariates (Z). Estimating this model and finding that these factors add "significantly" to the regression is not the same as demonstrating that the inference without an adjustment for selection is wrong for every predictor. Sharper inferences about the role of selection (Which relationships are affected by the selection process?) could be obtained by applying this framework.

The above extensions have not been worked out in sufficient detail so that prescriptions can be made, although these seem to be natural areas in which to extend this approach. But application of the framework put forth here would strengthen the comparison of regression coefficients across or between models that characterizes so much empirical research in sociology and other areas of social research. The modifications suggested for presenting findings from regression models featured in every example in this article would better communicate the evidence for model-to-model differences in regression relationships that figure so prominently in contemporary research practice.

APPENDIX

Derivations for the Linear Model

The main result in Section IV, special cases of which appear in Section III, is that $V(b^* - b) = V(b) - V(b^*)$, where the variance is taken under the full model and the standard assumptions are applied (homoscedastic-

ity, normality of errors). This result does not appear in the literature, to our knowledge, so a complete derivation is provided here.[23]

We let X denote the predictors in the reduced model including the vector of ones for the intercept (so that X is of order $n \times [p + 1]$). The variance of b^* under the full model is

$$V(b^*) = (X^T X)^{-1}\sigma_v^2, \tag{A1}$$

where σ_v^2 is the error variance in the full model, not the reduced model. This expression follows from the fact that the least squares estimator of b^* is $(X^T X)^{-1}X^T Y$. The variance of b (coefficient vector for X in the full model) is

$$V(b) = A\sigma_v^2, \tag{A2}$$

where

$$A = [X^T X - X^T Z(Z^T Z)^{-1}Z^T X]^{-1}. \tag{A3}$$

This matrix is derived in the usual way by partitioning the design matrix for the full model as $Q = [X, Z]$, so that $Q^T Q$ is the partitioned matrix

$$\begin{pmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z \end{pmatrix}.$$

The matrix A in (A3), of order $(p + 1) \times (p + 1)$, is the entry in the inverse of this matrix corresponding to the upper-left position. This quantity times the error variance under the full model gives the variance formula in (A2).

By similar methods, it can be shown (Clogg et al. 1992, p. 59) that the difference in the coefficient vectors is $d = b^* - b = MY$, where

$$M = [(X^T X)^{-1}X^T - AX^T + AX^T Z(Z^T Z)^{-1}Z^T], \tag{A4}$$

where A is defined in (A3). From this it follows that $V(d) = MM^T\sigma_v^2$ (under the full model). We now show that $MM^T = A - (X^T X)^{-1}$, from which it follows that $V(d) = V(b) - V(b^*)$, again with the understanding that the variance is evaluated under the full model.

From the definition of A in (A3), we obtain the relationship,

$$X^T X - A^{-1} = X^T Z(Z^T Z)^{-1}Z^T X, \tag{A5}$$

[23] Although we use results in Clogg et al. (1992), the simplifications given here were not included in this work.

which is used repeatedly below. Term by term multiplication of M and M^T using (A4) gives

$$\begin{aligned} MM^T = {} & (X^TX)^{-1}(X^TX)(X^TX)^{-1} - (X^TX)^{-1}X^TXA \\ & + (X^TX)^{-1}X^TZ(Z^TZ)^{-1}Z^TXA - AX^TX(X^TX)^{-1} + AX^TXA \\ & - AX^TZ(Z^TZ)^{-1}Z^TXA + AX^TZ(Z^TZ)^{-1}Z^TX(X^TX)^{-1} \\ & - AX^TZ(Z^TZ)^{-1}Z^TXA + AX^TZ(Z^TZ)^{-1}(Z^TZ)(Z^TZ)^{-1}Z^TXA. \end{aligned}$$

Using (A5), this simplifies to

$$\begin{aligned} MM^T = {} & (X^TX)^{-1} - A + (X^TX)^{-1}[X^TX - A^{-1}]A - A \\ & + AX^TXA - A[X^TX - A^{-1}]A \\ & + A[X^TX - A^{-1}](X^TX)^{-1} \\ & - A[X^TX - A^{-1}]A + A[X^TX - A^{-1}]A. \end{aligned}$$

Canceling terms leads directly to

$$\begin{aligned} MM^T = {} & (X^TX)^{-1} - A + A - (X^TX)^{-1} \\ & - A + AX^TXA - AX^TXA + A + A - (X^TX)^{-1}, \end{aligned}$$

which simplifies to

$$MM^T = A - (X^TX)^{-1},$$

the result to be shown. This means that

$$V(d) = A\sigma_v^2 - (X^TX)^{-1}\sigma_v^2, \tag{A6}$$

which is $V(b) - V(b^*)$, with the variance calculated under the full model.

The variance is estimated by substituting $\hat{\sigma}_v^2$, the mean-squared error under the full model, with $n - p - q - 1$ degrees of freedom. When the design matrix for the full model is of full rank, the matrix MM^T is positive definite, so negative variance estimates will not arise (see related comments in Sec. III). Because d is a linear combination of the observations (Y), d is normally distributed with mean δ and variance defined above; in small as well as large samples the exact distribution of the standardized value of any d_k value follows the t distribution, a result used in Sections III and IV (see Clogg et al. 1992 for details).

To test the hypothesis that a given $\delta_k = \beta_k^* - \beta_k = 0$, the ordinary t-statistic can be used. To test the hypothesis that the entire block is zero (i.e., $\delta = 0$), the statistic

$$F^* = d^T[\hat{V}(d)]^{-1}d/p^* \tag{A7}$$

can be used, where $p^* = \min(p + 1, q)$. The divisor follows from the fact that the rank of M is p^*, and if $q < p + 1$ then a generalized inverse must be used; however, in the latter case, the statistic in (A7) is equivalent to the usual F-statistic for the incremental contribution of Z given X. The statistic in (A7) follows the $F_{p^*, n-p-q-1}$ distribution under the hypothesis of no differences. If a block is chosen that is a subset, say $p' < p + 1$, of the original elements of X, then $p^* = \min(p', q)$. Subsets of the coefficients can be examined in an analogous manner.

REFERENCES

Agresti, Alan. 1990. *Categorical Data Analysis*. New York: Wiley.

Arminger, Gerhard. 1995. "Mean Structures." In *Handbook of Statistical Modeling in the Social Sciences,* edited by Gerhard Arminger, Clifford C. Clogg, and Michael Sobel. New York: Plenum.

Bentler, Peter M., and Chih-Ping Chou. 1993. "Some New Covariance Structure Improvement Statistics." Pp. 235–255 in *Testing Structural Equation Models,* edited by Kenneth A. Bollen and J. Scott Long. Newbury Park, Calif.: Sage.

Blalock, Hubert M. 1979. *Social Statistics,* 2d ed. rev. New York: McGraw-Hill.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Bryk, Anthony S., and Stephen W. Raudenbush. 1992. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, Calif.: Sage.

Clogg, Clifford C., and Scott R. Eliason. 1986. "On Regression Standardization for Moments." *Sociological Methods and Research* 14:423–46.

Clogg, Clifford C., Eva Petkova, and Edward S. Shihadeh. 1992. "Statistical Methods for Analyzing Collapsibility in Regression Models." *Journal of Educational Statistics* 17:51–74.

Clogg, Clifford C., and Edward S. Shihadeh. 1994. *Statistical Models for Ordinal Variables*. Newbury Park, Calif.: Sage.

Ducharme, G. R., and Y. Lepage. 1986. "Testing Collapsibility in Contingency Tables." *Journal of the Royal Statistical Society,* ser. B, 48:197–205.

Duncan, Otis Dudley. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.

———. 1981. "Two Faces of Panel Analysis: Parallels with Comparative Cross-Sectional Analysis and Time-Lagged Association." *Sociological Methodology* 21: 281–318.

Godfrey, L. G. 1988. *Misspecification Tests in Econometrics*. Cambridge: Cambridge University Press.

Goodman, Leo A. 1973. "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys." *American Journal of Sociology* 78:1135–91.

———. 1978. *Analyzing Qualitative/Categorical Data*. Cambridge, Mass.: Abt Books.

———. 1984. *Analysis of Cross-Classified Data Having Ordered Categories*. Cambridge, Mass.: Harvard University Press.

Gujarati, Damodar N. 1988. *Basic Econometrics,* 2d ed. New York: McGraw-Hill.

Hausman, Jerry A. 1978. "Specification Tests in Econometrics." *Econometrica* 46: 1251–71.

Hanushek, Eric A., and John E. Jackson. 1977. *Statistical Methods for Social Scientists*. New York: Academic Press.

Holly, Albert. 1988. "A Remark on Hausman's Specification Test." *Econometrica* 50:749–59.

Jasso, Guillermina. 1988. "Principles of Theoretical Analysis." *Sociological Theory* 6:1–20.

Johnson, R. J., and F. D. Wolinsky. 1993. "The Structure of Health Status among Older Adults: Disease, Disability, Functional Limitation, and Perceived Health." *Journal of Health and Social Behavior* 34:105–21.

Lazarsfeld, Paul F. 1948. "The Use of Panels in Social Research." *Proceedings of the American Philosophical Society* 92:405–10.

Manski, Charles F. 1993. "Identification Problems in the Social Sciences." *Sociological Methodology* 23:1–56.

McCullagh, Peter M., and John A. Nelder. 1989. *Generalized Linear Models,* 2d ed. London: Chapman & Hall.

Neter, John, William Wasserman, and Michael H. Kutner. 1989. *Applied Linear Regression Models*. Homewood, Ill.: Irwin.

Petkova, Eva, and Clifford C. Clogg. 1993. "Collapsibility and 'Omitted Variables Bias' in Generalized Linear Models." Manuscript. Columbia University, Department of Biostatistics.

Pratt, J. W., and R. Schlaifer. 1988. "On the Interpretation and Observation of Laws." *Journal of Econometrics* 39:23–52.

Smith, Herbert L. 1990. "Specification Problems in Experimental and Nonexperimental Social Research." *Sociological Methodology* 20:59–93.

Whittemore, Alice S. 1978. "Collapsibility of Multidimensional Contingency Tables." *Journal of the Royal Statistical Society,* ser. B, 40:328–40.

White, Halbert. 1980. "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity." *Econometrica* 48:817–38.