CHICAGO JOURNALS

The Impact of Random Predictors on Comparisons of Coefficients Between Models: Comment on Clogg, Petkova, and Haritou
Author(s): Paul D. Allison
Source: *The American Journal of Sociology*, Vol. 100, No. 5 (Mar., 1995), pp. 1294-1305
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/2782278
Accessed: 12/09/2009 18:22

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=ucpress.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *The American Journal of Sociology*.

http://www.jstor.org

The Impact of Random Predictors on Comparisons of Coefficients between Models: Comment on Clogg, Petkova, and Haritou[1]

Paul D. Allison
University of Pennsylvania

As Clogg, Petkova, and Haritou (1995) correctly observe, regression analysts are often intensely concerned with what happens to the coefficient of a predictor variable when additional variables are introduced into a regression model. Unfortunately, this concern is typically expressed in comparisons that lack any measure of statistical reliability. To remedy this deficiency, Clogg, Petkova, and Haritou (hereafter CPH) propose a set of methods for testing whether the change in a regression coefficient (or set of coefficients) is statistically significant. These methods have the virtues of simplicity and applicability to a wide class of generalized linear regression models.

CPH deserve credit for identifying an important but overlooked problem, and their solutions are both clever and elegant. Nevertheless, I believe that their proposed methods suffer from a fundamental flaw: they make unrealistic assumptions about the sampling properties of the predictor variables. Not surprisingly, this strategy leads to a substantial simplification of their methodology. But in doing so, it exposes the analyst to the risk of highly misleading conclusions.

My main objective is to explain why I think their assumptions are problematic and to suggest what may go wrong as a consequence. I shall focus primarily on the three-variable linear model since it embodies all the critical issues without the distracting complications of the multivariable and nonlinear cases. And if the methods are deficient in the three-variable case, there is little point in generalizing them to more complicated situations. I shall also suggest some alternative methods for both three-variable and multivariable linear models.

[1] For helpful suggestions, I am indebted to Nicholas Christakis, Arthur Goldberger, Herbert Smith, and Richard Waterman. Direct correspondence to Paul D. Allison, Department of Sociology, University of Pennsylvania, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6299. E-mail: allison@ssdc.sas.upenn.edu

© 1995 by The University of Chicago. All rights reserved.
0002-9602/95/10005-0007$01.50

THE THREE-VARIABLE CASE

As far as possible, I will utilize the notation of CPH. The basic situation is this: we first regress Y on X. Then we regress Y on X and Z. We want to know if there is a significant difference between the two coefficients for X. Like CPH, I assume that the data are generated by a "full" model:

$$Y_i = \alpha + \beta_{yx \cdot z} X_i + \beta_{yz \cdot x} Z_i + \nu_i, \qquad i = 1, \ldots, n, \tag{1}$$

where ν_i satisfies the usual assumptions of the linear model.[2] All derivations are based on this model, not on the "reduced" model that excludes Z.

The basic aim is to test the null hypothesis that $\delta = 0$, where $\delta = \beta_{yx} - \beta_{yx \cdot z}$, and β_{yx} is the population least squares regression coefficient of Y on X alone. That is, $\beta_{yx} = \sigma_{xy}/\sigma_x^2$, where σ_{xy} is the covariance of X and Y, and σ_x^2 is the variance of X. An unbiased estimator of δ is just $d = b_{yx} - b_{yx \cdot z}$ where the b's are the sample least squares estimators of the β's. It can be shown that

$$\delta = \beta_{yz \cdot x}\left(\frac{\sigma_{xz}}{\sigma_x^2}\right), \tag{2}$$

which is the population analogue of CPH's formula

$$d = b_{yz \cdot x}\left(\frac{s_{xz}}{s_x^2}\right), \tag{3}$$

where s_{xz} is the sample covariance between X and Z and s_x^2 is the sample variance of X. We see then that $\delta = 0$ if $\beta_{yz \cdot x} = 0$ or if $\sigma_{xz} = 0$, that is, if either Z has no effect on Y, controlling for X, or if X and Z are uncorrelated. In other words, there are two different ways that the null hypothesis can be true. But CPH focus only on the first possibility. They conclude that an appropriate test statistic for this null hypothesis is just the usual t-statistic for the null hypothesis that $\beta_{yz \cdot x} = 0$, namely, the ratio of the least squares coefficient $b_{yz \cdot x}$ to its estimated standard error. In this setting, then, one needs only to check the significance of the added variable. If its coefficient is significantly different from zero, we conclude that the coefficient for the initial variable has also *changed* significantly.

[2] Specifically, I assume that $E(\nu_i|X_i, Z_i) = 0$, $V(\nu_i|X_i, Z_i) = \sigma^2$, and $\text{cov}(\nu_i, \nu_j|X_i, Z_i, X_j, Z_j) = 0$ for all i and for all $j \neq i$. For exact inferences, one also needs normality of ν_i.

This is a surprising result, one that CPH claim applies only in the three-variable case. What they do not say is that there is an analogous and even more startling result in the multivariable case: suppose that a single Z variable is added to a model with several X variables. Under CPH's procedure, the t-statistic for the *change* in each X coefficient is identical to the t-statistic for the Z coefficient. In other words, when a single variable is added to a regression model, *either every coefficient changes significantly or none do*. I suspect that most experienced regression analysts will be troubled by this property. When a variableis added to a regression equation, some coefficients may change greatly, others hardly at all. Is it plausible that a procedure for testing changes in coefficients should be insensitive to the magnitudes of the changes?

CPH reach these unusual conclusions because of the way they treat the predictor variables X and Z. In textbook treatments of regression analysis, it is common to assume that the predictor variables are fixed or nonstochastic. That means that, in repeated sampling, all the values of the predictors stay the same from sample to sample, and only the values of the dependent variable change. In fact, however, this assumption is only plausible for experimental designs or for stratified sampling when there is one stratum for every combination of values of the predictor variables. But the vast bulk of social science research is nonexperimental, and no one does stratified sampling in that way. So why do textbook writers make such a patently unrealistic assumption? Because it enormously simplifies the algebra and it is relatively innocuous. For inferences about a correctly specified model, you get the same (or equivalent) results whether you treat the predictor variables as random or fixed. For example, in a two-variable linear model, the standard result for fixed X is that $V(b_{yx}) = \sigma_\epsilon^2/ns_x^2$. When X is random, we have $V(b_{yx}) = \sigma_\epsilon^2 E(1/ns_x^2)$. These results are operationally equivalent because, in the random case, we estimate the expectation by its sample value.

Results are not equivalent, however, when the aim is to compare a full model with a restricted model. When predictor variables are fixed, for example, it is well known that deleting Z from the model reduces the variance of the coefficient of X. But when X and Z are random, Binkley and Abbot (1987) showed that the variance of the coefficient of X may increase substantially when Z is deleted. Similarly, in comparing the performance of alternative variable selection methods, Breiman and Spector (1992) concluded: "There can be startling differences between the x-fixed and x-random situations."

Although CPH do *not* assume that the predictor variables are fixed, they do something equivalent: they make all their inferences *conditional*

on the realized values of the predictor variables.[3] Here's a crucial example: from CPH's formula (11), it is easily shown that the conditional variance of d is

$$V(d|X, Z) = V(b_{yz \cdot x}) \left[\frac{s_{xz}}{s_x^2} \right]^2. \tag{4}$$

Noting that s_{xz}/s_x^2 is the sample regression coefficient of Z on X, we can also write this as

$$V(d|X, Z) = V(b_{yz \cdot x}) b_{zx}^2. \tag{5}$$

This result is correct as far as it goes, but treating (5) as *the* variance of d assumes that s_{xz} and s_x^2 are constants, not random variables that can vary from sample to sample. In the next section I show that when X and Z are random, the *unconditional* variance of d is larger than (5), possibly much larger.

ALTERNATIVE METHODS FOR THE THREE-VARIABLE CASE

In the appendix, I derive the unconditional variance of d in the multivariable case. Specializing to the three-variable case, we have

$$V(d) = V(b_{yz \cdot x}) E(b_{zx}^2) + \beta_{yz \cdot x}^2 V(b_{zx}). \tag{6}$$

Comparing (6) with (5), we see that the first component of the sum in (6) corresponds to (5) except that we have taken the expectation of b_{zx}^2. Since the second component of (6) is nonnegative, the unconditional variance exceeds the conditional variance except when $\beta_{yz \cdot x} = 0$.

In practice, a consistent estimator for this variance is

$$s^2(d) = s^2(b_{yz \cdot x}) b_{zx}^2 + b_{yz \cdot x}^2 s^2(b_{zx}). \tag{7}$$

All of these quantities are available from standard computer output. Here, $s(b_{yz \cdot x})$ is just the estimated standard error of $b_{yz \cdot x}$, and $s(b_{zx})$ is the estimated standard error of b_{zx}. The usual formula for $s(b_{zx})$ depends, of course, on the assumption that Z can be expressed as a linear function of X with an error term satisfying the standard conditions. If this is

[3] CPH appear to confuse conditioning on the *model* with conditioning on the predictor variables when they write $V(d|H_F) = V(d|X, Z)$ where H_F refers to the full model. While they are correct that it is necessary to choose some model as "true" in order to calculate expectations and variances, that does not imply that one must also condition on the variables in that model. In statistical terminology, conditioning on a model has no technical meaning, while conditioning on a random variable (or variables) is a well-defined operation.

implausible (e.g., if Z is dichotomous), one can use robust variance estimates (White 1980) that are now available in many regression programs. Having calculated $s^2(d)$, one can construct the test statistic $d/s(d)$, which will have approximately a standard normal distribution under the null hypothesis that $\delta = 0$. Alternatively, an approximate 95% confidence interval can be calculated as $d \pm 2s(d)$.

For hypothesis testing, another solution is based on likelihood theory. Assuming a trivariate normal distribution for X, Y, and Z, one can maximize the likelihood, subject to the constraint that $\delta = 0$, and construct a likelihood ratio test by comparing the constrained likelihood with the unconstrained likelihood. The chi-square statistic is calculated as twice the positive difference in the two log likelihoods. Notice, however, that we can carry out the constrained estimation in two distinct steps. First constrain $\beta_{yz \cdot x} = 0$, and then constrain $\sigma_{xz} = 0$. Pick whichever likelihood is larger and contrast that with the unrestricted likelihood (which must be larger still). This amounts to doing two likelihood-ratio tests, one for each constraint, and using the chi-square that is smaller.

This logic could be extended to more conventional tests for regression coefficients. First regress Y on both X and Z, and do a standard test for $\beta_{yz \cdot x} = 0$. Then regress Z on X (or, equivalently, X on Z) and test for significance of that coefficient. Reject the null only if both tests lead to rejection. Note that under this procedure, the probability of rejecting the null hypothesis (and concluding that the coefficient of X has changed) is always lower than it is for the CPH test because the rejection region of my proposed test is a proper subset of their rejection region.

A MONTE CARLO STUDY

To get an empirical comparison of the conditional and unconditional methods, I generated random samples from a "population" in which X and Z had a bivariate normal distribution with means of zero, variances of one, and a correlation of ρ, where ρ was assigned values of .00, .30, and .50. I then generated Y according to the equation

$$Y = 1 + 2X + \gamma Z + E, \tag{8}$$

where E, an "unobserved" random disturbance, was normally distributed with a mean of zero, a variance of 100, and was independent of X and Z. The coefficient γ had values of either zero or three. When $\gamma = 0$, it follows that $\delta = 0$. When $\gamma = 3$, we have $\delta = 3\rho$. For each of the six combinations of ρ and γ, I drew 500 random samples of size $n = 100$. In each sample, I calculated d, the conditional and unconditional estimates of the variance of d, and nominal 95% confidence intervals

TABLE 1

COMPARISON OF CONDITIONAL AND UNCONDITIONAL METHODS WITH SIMULATED DATA, $\gamma = 3$

			CONDITIONAL		UNCONDITIONAL	
ρ	δ	$V(d)$	Mean $\hat{V}(d)$	% Covered	Mean $\hat{V}(d)$	% Covered
.00	.00	.102	.014	15	.116	99
.10	.30	.115	.021	51	.124	94
.30	.90	.203	.114	80	.211	94
.50	1.50	.429	.360	92	.439	96

around δ using both the conditional and unconditional variance estimates.

Tables 1 and 2 give the results. I will first examine table 1, in which $\gamma = 3$. The third column gives the actual variance of d across the 500 samples, which is a model-free estimate of the true variance. The fourth column reports the mean of the estimated variances of d using CPH's conditional estimator. When $\rho = 0$, the actual variance is more than seven times the average estimated conditional variance, a truly horrendous performance. As ρ gets larger, the CPH estimator does somewhat better, although even at $\rho = .50$ it is still only 84% of the actual value.

These variance underestimates are reflected in the percentages of confidence intervals that actually include the true value, shown under "% covered." For $\rho = 0$, only 15% of the CPH confidence intervals include the true value $\delta = 0$. Equivalently, a two-tailed test of the null hypothesis (which is true in this case) would be rejected in 85% of the samples. For $\rho = .10$, the coverage improves substantially to 51%, and for $\rho = .50$ it is almost at the nominal level. By contrast, the unconditional methods perform very well. Although the average of the variance estimates is always greater than the actual variance, the overestimate is never greater than 14%. And the coverage of the confidence intervals is fairly close to the nominal value of 95%.

Now turn to table 2, in which $\gamma = 0$. From equation (6), we see that when the true coefficient of Z is zero, the unconditional variance reduces to the conditional variance. Hence, there is no reason in this case to expect the unconditional method to do any better than the conditional method, and the results in table 2 bear that out. In fact, the unconditional estimates do somewhat worse. While the results for the conditional methods are nearly ideal, the unconditional variance estimates are too high, and the confidence intervals are too large. These biases are greatest when $\rho = 0$ and become trivial when $\rho = .50$. Even in the worst case, how-

TABLE 2

COMPARISON OF CONDITIONAL AND UNCONDITIONAL METHODS WITH SIMULATED DATA, $\gamma = 0$

			CONDITIONAL		UNCONDITIONAL	
ρ	δ	$V(d)$	Mean $\hat{V}(d)$	% Covered	Mean $\hat{V}(d)$	% Covered
.00	.00	.009	.010	95	.022	100
.10	.00	.021	.020	93	.032	100
.30	.00	.108	.110	95	.120	99
.50	.00	.367	.365	95	.375	97

ever, the unconditional confidence intervals are conservative in that the probability of rejecting the true null hypothesis is *less* than the nominal α level.

One should not take these simulations as definitive, however. Only a few sets of parameter values and one sample size were considered. The assumption that the predictor variables are normally distributed is also unduly restrictive. Nevertheless, I think the results are sufficiently clear to raise substantial doubts about the appropriateness of CPH's procedures for the kinds of data that are typically analyzed by social scientists.

THE MULTIVARIABLE CASE

The generalization to the multivariable case involves a few new issues. In matrix notation the true model is

$$Y = X\beta + Z\gamma + v, \tag{9}$$

where Y and v are $n \times 1$ vectors, X is an $n \times p$ matrix (including a column of ones for the intercept), Z is an $n \times q$ matrix, β is a $p \times 1$ vector, and γ is a $q \times 1$ vector. I assume further that the usual assumptions of the linear model hold conditionally on X and Z. That is, $E(v|X, Z) = 0$ and $V(v|X, Z) = \sigma_v^2 I$ where I is the $n \times n$ identity matrix. The aim is to make inferences about $\delta = \beta^* - \beta$, where β^* is the $p \times 1$ vector of population least squares regression coefficients of Y on X alone. We estimate δ with $d = b^* - b$, where b^* is the sample least squares estimator of Y on X alone and b is the least squares estimator of Y on X when Z is included. Let g be the least squares estimator of γ when both X and Z are included in the model, and let $H = (X^TX)^{-1}X^TZ$, the $p \times q$ matrix of least squares regression coefficients for Z on X.

In the appendix I show that an appropriate estimator for $V(d)$ is

$$\hat{V}(d) = H\hat{V}(g)H^T + g^T W g(X^T X)^{-1}, \tag{10}$$

where W is a matrix of variances and covariances of the residuals from regressing Z on X. The first component of the sum in (10) is equivalent to CPH's conditional variance estimator. Although the matrix W is *not* standardly available in regression programs, it could be easily computed from the residuals.

No special computation is necessary when there is only a single Z variable, as (10) reduces to

$$\hat{V}(d) = s^2(g)hh^T + g^2\hat{V}(h), \tag{11}$$

where $s^2(g)$ is the squared estimated standard error of g, h is now a $p \times 1$ vector of regression coefficients for Z on X, and $\hat{V}(h)$ is the usual estimated covariance matrix. As in the three-variable case, one may wish to use a more robust estimator for this matrix. From the main diagonal of (11), the variance for the change in the coefficient of a single variable X_i is

$$s^2(d_i) = s^2(g)h_i^2 + g^2 s^2(h_i), \tag{12}$$

where h_i is the coefficient of X_i in the regression of Z on X. Note that (12) has exactly the same form as (7) in the three-variable case, except that the bivariate coefficient of Z on X is replaced by a partial coefficient.

To test the null hypothesis that $\delta = 0$, i.e., that *none* of the coefficients changes with the addition of Z, one may use the statistic

$$F^* = \frac{d^T\hat{V}(d)^{-1}d}{p}, \tag{13}$$

which has approximately an F distribution with p and $n - p - q$ degrees of freedom under the null hypothesis. By contrast, CPH's F-statistic has $\min(p, q)$ as the divisor and the corresponding degrees of freedom. That's because CPH's estimator of $V(d)$ does not have full rank if $q < p$, that is, if the number of Z's is less than the number of X's.[4] But (10) has full rank (except when $g = 0$) with probability 1, because the second component of the sum has full rank.

Likelihood tests of $\delta = 0$ are also possible with LISREL 8 (Jöreskog and Sörbom 1993) because that program allows for testing nonlinear restrictions on the coefficients. The trick is to specify a recursive linear model in which all the X's are causally prior to all the Z's. Then estimate the model under the restriction $Hg = 0$. While earlier versions of

[4] A square matrix has full rank if no column (or row) is a linear function of the other columns (or rows).

LISREL cannot impose nonlinear restrictions, they do report standard error estimates for the "indirect effects" using formulas given by Sobel (1982). Under a recursive model, the indirect effect of X_i on Y through Z is equivalent to d_i. Hence, these standard errors can be used to construct t-statistics and confidence intervals for the individual elements of d. Although Sobel's formulas are equivalent to (10), his results were obtained as asymptotic approximations. Equation (10), on the other hand, is based on exact results in the appendix.

WHEN IS CONDITIONING JUSTIFIED?

I have argued that when predictors are random, making inferences conditional on the sample values of the predictors is justified in some situations but not in others. Is there any way to characterize the distinction between these two kinds of situations? Here are two possible answers to this question:

1. *Ancillarity.*—For conventional inferences about regression coefficients, the practice of making inferences conditional on the predictor variables is well grounded in the theory of ancillarity (Cox and Hinkley 1974). That theory says, roughly, that if the marginal probability distribution of some sample statistic does not depend on the parameters of interest, then inference should be made conditional on that sample statistic, which is termed an ancillary statistic. In standard regression problems, the moments (means, variances, and covariances) of the predictor variables are ancillary for testing hypotheses about the regression coefficients, so it is reasonable to make inferences conditional on those moments.

For the three-variable case considered here, however, we have a sample statistic s_{xz} whose sampling distribution obviously depends on the parameter σ_{xz}. But this is one of the parameters we are trying to make inferences about because the null hypothesis is true if $\sigma_{xz} = 0$. Conditioning on s_{xz} is tantamount to assuming that the only way the null hypothesis (that $\delta = 0$) can be true is if $\beta_{yz \cdot x} = 0$. Hence, conditioning cannot be justified by the principle of ancillarity.

2. *Conditional unbiasedness.*—Consider an arbitrary statistic $\hat{\theta}$ computed from data (X, Y). A plausible claim is that the conditional variance $V(\hat{\theta}|X)$ is acceptable for making inference about $\hat{\theta}$ whenever $V(\hat{\theta}) = E(V(\hat{\theta}|X))$, that is, when the unconditional variance is equal to the expected value of the conditional variance. From (A3) in the appendix, we see that this equality holds if and only if $E(\hat{\theta}|X)$ does not depend on X. For a standard, three-variable linear model, this condition is satisfied because $E[b_{yx \cdot z}|X, Z] = \beta_{yx \cdot z}$, that is, the coefficients are conditionally unbiased. On the other hand,

$$E[d|X, Z] = \delta\left(\frac{b_{zx}}{\beta_{zx}}\right) \neq \delta. \tag{14}$$

This means that, in a particular sample, if b_{zx} happens to be larger (or smaller) than the "true" coefficient due to sampling variation, then d will also tend to be too large (or too small). Using the unconditional variance for inference enables us to take account of this additional source of variation.

GENERALIZED LINEAR MODELS

Everything I have said so far applies only to linear models. For generalized linear models, the situation is rather more complicated because the regression estimators cannot usually be expressed in closed form. Nevertheless, given the arguments just presented, there are strong reasons to suspect that the unconditional variance of d will exceed the conditional variance. While I have not yet attempted to derive the unconditional variance, I am not optimistic that useful formulas will be forthcoming. Bootstrap estimates of the variance offer one potential solution in these situations.

CONCLUSION

I have argued that the methods of CPH for comparing regression coefficients in "full" and "reduced" models depend on treating the predictor variables as though they were fixed from sample to sample, an unrealistic assumption for most social science data. Allowing for randomness in the predictors leads to different, more conservative tests and confidence intervals. This differs from inferences about coefficients in a single, correctly specified model, where methods are the same whether predictors are fixed or random. The alternative methods proposed here are readily implemented with conventional software and appear to outperform the CPH methods on simulated data. While the issues are admittedly subtle, I hope that these arguments are sufficient to forestall the universal and uncritical adoption of the CPH tests as a standard for evaluating changes in regression coefficients.

APPENDIX

Derivation of the Unconditional Variance

Using the notation developed above for the multivariable case, I make use of the well-known result (e.g., Goldberger 1991) that $b^* = b + Hg$. This implies that $d = Hg$. We then have

$$E(d|X, Z) = H\gamma, \tag{A1}$$

and

$$V(d|X, Z) = HV(g)H^T. \tag{A2}$$

Formula (A2) is equivalent to the variance formula given by CPH. To get the unconditional variance, we make use of another well-known formula for arbitrary random vectors U and V,

$$V(U) = E[V(U|V)] + V[E(U|V)], \tag{A3}$$

that is, the unconditional variance is equal to the expectation of the conditional variance plus the variance of the conditional expectation. Applying this to the problem at hand, we have

$$\begin{aligned} V(d) &= E[V(d|X, Z)] + V[E(d|X, Z)] \\ &= E(HV(g)H^T) + V(H\gamma). \end{aligned} \tag{A4}$$

Focusing on the second term in (A4), write $H = [h_1 h_2 \ldots h_q]$, $\gamma = [\gamma_1 \gamma_2 \ldots \gamma_q]^T$ and $Z = [Z_1 Z_2 \ldots Z_q]$. We then have

$$H\gamma = \sum_{j=1}^{q} \gamma_j h_j. \tag{A5}$$

It follows that

$$V(H\gamma) = \sum_j \gamma_j^2 V(h_j) + \sum_j \sum_{k \neq j} \gamma_j \gamma_k \operatorname{cov}(h_j, h_k). \tag{A6}$$

Substituting (A6) into (A4) yields the desired result.

If we are willing to assume that the population regression of Z on X satisfies a standard linear model, then we can further simplify (A6). In that event,

$$V(h_j) = \sigma_j^2 E(X^T X)^{-1}, \tag{A7}$$

and

$$\operatorname{cov}(h_j, h_k) = \sigma_{jk} E(X^T X)^{-1}, \tag{A8}$$

where σ_j^2 is the disturbance variance in the equation for Z_j and σ_{jk} is the covariance between the disturbances in the equations for Z_j and Z_k. Let Ω be a $q \times q$ matrix containing these variances and covariances. We can then write $V(d)$ as

$$V(d) = E(HV(g)H^T) + \gamma^T \Omega \gamma E(X^T X)^{-1}. \tag{A9}$$

To estimate this, we can replace population quantities with sample estimators to get

$$\hat{V}(d) = H\hat{V}(g)H^T + g^T W g(X^T X)^{-1}, \tag{A10}$$

where W is a matrix of variances and covariances of the residuals obtained from regressing Z on X.

REFERENCES

Binkley, James K., and Philip C. Abbot. 1987. "The Fixed X Assumption in Econometrics: Can the Textbooks be Trusted?" *American Statistician* 41:206–14.

Breiman, Leo, and Philip Spector. 1992. "Submodel Selection and Evaluation in Regression: The X-Random Case." *International Statistical Review* 60:291–319.

Clogg, Clifford C., Eva Petkova, and Adamantios Haritou. 1995. "Statistical Methods for Comparing Regression Coefficients between Models." *American Journal of Sociology*. In this issue.

Cox, D. R., and D. V. Hinkley. 1974. *Theoretical Statistics*. Cambridge: Chapman & Hall.

Goldberger, Arthur S. 1991. *A Course in Econometrics*. Cambridge, Mass.: Harvard University Press.

Jöreskog, Karl G., and Dag Sörbom. 1993. *New Features in LISREL 8*. Chicago: Scientific Software.

Sobel, Michael E. 1982. "Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models." Pp. 290–312 in *Sociological Methodology 1982*, edited by Samuel Leinhardt. San Francisco: Jossey-Bass.

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

REPLY TO ALLISON: MORE ON COMPARING REGRESSION COEFFICIENTS[1]

Allison (1994) raises some legitimate questions concerning the interpretation of some of our procedures (Clogg, Petkova, and Haritou 1994; CPH hereafter). Because his comments pertain to the CPH methods for the analysis of linear regression models, in this note we also concentrate on this case.[2]

Virtually all of the literature on specification tests or collapsibility tests

[1] We are indebted to Bing Li and Bruce Lindsay for helpful comments.

[2] The CPH procedures for generalized linear models are different from the CPH procedures for linear models. For example, for the analysis of collapsibility in contingency tables where log-linear models or related models might be used, the CPH procedures are automatically "unconditional" because the models pertain to multivariate (joint or marginal) distributions. For contingency table settings where either log-linear models (unconditional) or logit models (conditional) might be used to address the same question, the inferences are equivalent between the two cases as well. The CPH procedures for generalized linear models are valid unconditionally, for testing the relevant null hypothesis, and so are valid conditionally as well.