7

EXACT VARIANCE OF INDIRECT EFFECTS IN RECURSIVE LINEAR MODELS

*Paul D. Allison**

Using the delta method, Sobel obtained the asymptotic variance of indirect effects in linear structural equation models. Using a reduced-form parameterization and a conditioning argument, I obtain the exact *variance of indirect effects in the special case of recursive linear models with no latent variables. I then show that a consistent estimator for the exact variance is identical to Sobel's estimator.*

Sobel (1982, 1986) used the delta method to obtain formulas for the asymptotic variance of indirect effects in linear structural equation models. These formulas have been incorporated into the LISREL program (Jöreskog and Sörbom 1993) and have proven useful in a variety of applications. However, the fact that they are only large-sample approximations has been cause for some concern (Bollen and Stine 1990). As Bollen (1987) warned, "the accuracy of the delta method for small samples is not known." Stone and Sobel (1990) used Monte Carlo methods to evaluate the performance of the approximate variance estimator in small samples. For a six-variable recursive model with no latent variables, they found that "for the large sample theory to apply, the results suggest that sample sizes of 200 or more . . . are required." For a ten-variable nonrecursive

For helpful suggestions, I am grateful to Arthur Goldberger and Robert Stine.

*University of Pennsylvania

model with two latent variables, samples of 400 were needed to get good results.

Here I derive an *exact* formula for the special but important case of ordinary least squares estimation of recursive models with all variables observed (except for the random disturbance terms). The derivation is accomplished by applying a conditioning argument to a reduced-form parameterization of the indirect effects. I then show that a consistent estimator of the exact variance is identical to the estimator obtained with the delta method.

Unlike Sobel, who worked with all the indirect effects for an entire structural equation system, I focus on a single dependent variable at a time. That is because the derivation involves conditioning on a different set of variables for each equation, making it extremely awkward to treat all the dependent variables at once.

1. THE MODEL

Consider the equation

$$y = \boldsymbol{\beta}'z + \boldsymbol{\gamma}'\boldsymbol{x} + \epsilon, \tag{1}$$

where y is the observed dependent variable, z is an $m \times 1$ vector of observed endogenous variables, $\boldsymbol{x}$ is an $n \times 1$ vector of observed exogenous variables, ϵ is an unobserved disturbance term, $\boldsymbol{\beta}$ is an $m \times 1$ vector of coefficients, and $\boldsymbol{\gamma}$ is an $n \times 1$ vector of coefficients. All variables are expressed as deviations from their respective means, and both $\boldsymbol{x}$ and z are assumed to be independent of ϵ. A second set of equations describes the dependence of z on $\boldsymbol{x}$ and itself:

$$z = \boldsymbol{\Lambda}z + \boldsymbol{K}\boldsymbol{x} + \boldsymbol{\nu}, \tag{2}$$

where $\boldsymbol{\Lambda}$ is an $m \times m$ matrix of coefficients, $\boldsymbol{K}$ is an $m \times n$ matrix of coefficients, and $\boldsymbol{\nu}$ is an $m \times 1$ vector of unobserved disturbances. Again we assume that both $\boldsymbol{x}$ and ϵ are independent of $\boldsymbol{\nu}$.

Our concern is with the indirect effects of $\boldsymbol{x}$ on y through z. These will be defined momentarily. If we want the variance of the indirect effects of $\boldsymbol{x}$ on y alone, the only restriction on $\boldsymbol{\Lambda}$ is that $(\boldsymbol{I} - \boldsymbol{\Lambda})$ must have an inverse.[1] On the other hand, if we also want the

[1]Bollen (1987) argues that for the effects to have a meaningful interpretation, it must be true that $\Lambda k \rightarrow_{\text{O}}$ as $k \rightarrow \infty$.

exact variance of the indirect effects of $\boldsymbol{x}$ on z, we must assume that the system in (2) is fully recursive, meaning that $\boldsymbol{\Lambda}$ is lower triangular (zeros on and above the main diagonal), $V(\boldsymbol{\nu})$ is diagonal (all covariances between disturbances are 0), and z_j is independent of ν_k for all $k > j$. For simplicity, we begin with the case in which there are no overidentifying restrictions on the coefficients. Overidentified models will be considered later.

Solving (2) for z gives reduced-form equations

$$z = \boldsymbol{\Pi x} + \boldsymbol{u}, \tag{3}$$

where $\boldsymbol{\Pi} = (\boldsymbol{I} - \boldsymbol{\Lambda})^{-1}\boldsymbol{K}$ and $\boldsymbol{u} = (\boldsymbol{I} - \boldsymbol{\Lambda})^{-1}\boldsymbol{\nu}$. Let $V(\boldsymbol{u}) = \boldsymbol{\Omega}$. Although $\boldsymbol{x}$ is still independent of $\boldsymbol{u}$, $\boldsymbol{\Omega}$ will not, in general, be diagonal.

In standard terminology, the coefficient vector $\boldsymbol{\gamma}$ in equation (1) is described as the "direct effect" of $\boldsymbol{x}$ on y. The "total effect" of $\boldsymbol{x}$ on y may be found by substituting (3) into (1), yielding

$$\begin{aligned} y &= \boldsymbol{\beta}'(\boldsymbol{\Pi x} + \boldsymbol{u}) + \boldsymbol{\gamma}'\boldsymbol{x} + \epsilon \\ &= (\boldsymbol{\beta}'\boldsymbol{\Pi} + \boldsymbol{\gamma}')\boldsymbol{x} + \boldsymbol{\beta}'\boldsymbol{u} + \epsilon. \end{aligned} \tag{4}$$

We see then that the total effect is $\boldsymbol{\beta}'\boldsymbol{\Pi} + \boldsymbol{\gamma}'$. The "indirect effect" of $\boldsymbol{x}$ on y is just the total effect minus the direct effect, which is apparently $\boldsymbol{\beta}'\boldsymbol{\Pi}$. This parameterization of the indirect effects differs from that of Sobel and Bollen, who expressed them in terms of the structural coefficients. One advantage of the reduced-form expression is that none of the "paths" from $\boldsymbol{x}$ to y involve products of more than two coefficients.

Let $\boldsymbol{b}$ be the OLS estimator of $\boldsymbol{\beta}$ and let $\boldsymbol{P}$ be the OLS estimator of $\boldsymbol{\Pi}$. A natural estimator of the indirect effects is just $\boldsymbol{d} = \boldsymbol{P}'\boldsymbol{b}$. If we assume multivariate normality of ϵ and $\boldsymbol{\nu}$, then $\boldsymbol{d}$ is the maximum likelihood estimator of the indirect effects. We now obtain the variance of $\boldsymbol{d}$.

2. DERIVATION OF EXACT VARIANCE

Assume that we have a simple random sample of N observations on y, z, and $\boldsymbol{x}$. Let $\boldsymbol{Z}$ be the $N \times m$ matrix of data on z and let $\boldsymbol{X}$ be the $N \times n$ matrix of data on $\boldsymbol{x}$. Using a well-known decomposition of the variance (Goldberger 1991), we can write

$$V(\boldsymbol{d}) = E[V(\boldsymbol{d} \mid \boldsymbol{Z}, \boldsymbol{X})] + V[E(\boldsymbol{d} \mid \boldsymbol{Z}, \boldsymbol{X})]. \tag{5}$$

This formula is completely general (so long as the expectations and variances exist), and it does not rest on any of the assumptions made so far. The conditional variance is

$$\begin{aligned} V(\boldsymbol{d} \mid \boldsymbol{Z}, \boldsymbol{X}) &= V(\boldsymbol{P}'\boldsymbol{b} \mid \boldsymbol{Z}, \boldsymbol{X}) \\ &= \boldsymbol{P}'V(\boldsymbol{b} \mid \boldsymbol{Z}, \boldsymbol{X})\boldsymbol{P} \\ &= \boldsymbol{P}'\boldsymbol{\Phi}\boldsymbol{P}, \end{aligned} \tag{6}$$

where $\boldsymbol{\Phi}$ is just the usual variance matrix of $\boldsymbol{b}$.[2] This result follows because $\boldsymbol{P}$ is a constant, given $\boldsymbol{X}$ and $\boldsymbol{Z}$. The conditional expectation is

$$\begin{aligned} E(\boldsymbol{d} \mid \boldsymbol{Z}, \boldsymbol{X}) &= E(\boldsymbol{P}'\boldsymbol{b} \mid \boldsymbol{Z}, \boldsymbol{X}) \\ &= \boldsymbol{P}'E(\boldsymbol{b} \mid \boldsymbol{Z}, \boldsymbol{X}) \\ &= \boldsymbol{P}'\boldsymbol{\beta}. \end{aligned} \tag{7}$$

Substituting (6) and (7) into (5), we have

$$V(\boldsymbol{d}) = E[\boldsymbol{P}'\boldsymbol{\Phi}\boldsymbol{P}] + V(\boldsymbol{P}'\boldsymbol{\beta}). \tag{8}$$

I now focus on the second term on the right-hand side of (8). Let $\boldsymbol{p}_j$ be the jth row of $\boldsymbol{P}$; that is, $\boldsymbol{p}_j$ is the vector of OLS coefficients from the regression of z_j on $\boldsymbol{x}$. We can write

$$\boldsymbol{P}'\boldsymbol{\beta} = \sum_{j=1}^{m} \beta_j \boldsymbol{p}_j. \tag{9}$$

Then

$$V(\boldsymbol{P}'\boldsymbol{\beta}) = \sum_{j=1}^{m} \beta_j^2 \, V(\boldsymbol{p}_j) + \sum_{j=1}^{m} \sum_{k \neq j}^{m} \beta_j \beta_k Cov(\boldsymbol{p}_j, \boldsymbol{p}_k). \tag{10}$$

Drawing on OLS theory for random $\boldsymbol{X}$ (e.g., Goldberger 1991), we have

$$V(\boldsymbol{p}_j) = \omega_{jj} E(\boldsymbol{X}'\boldsymbol{X})^{-1} \quad \text{and} \quad Cov(\boldsymbol{p}_j, \boldsymbol{p}_k) = \omega_{jk} E(\boldsymbol{X}'\boldsymbol{X})^{-1}, \tag{11}$$

where ω_{jk} is an element of $\boldsymbol{\Omega}$, the covariance matrix for the reduced-form disturbances. Substituting (11) into (10) and rearranging terms, we get

$$V(\boldsymbol{P}'\boldsymbol{\beta}) = \boldsymbol{\beta}'\boldsymbol{\Phi}\boldsymbol{\beta} E(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{12}$$

[2]That is, $\boldsymbol{\Phi} = \sigma_\epsilon^2[\boldsymbol{Z}'(\boldsymbol{I} - (\boldsymbol{X}'\boldsymbol{X})^{-1})\boldsymbol{Z}]^{-1}$. This formula arises from expressing the OLS estimator of equation (1) in partitioned matrix form.

Substituting (12) into (8) we get the final result:

$$V(\boldsymbol{d}) = E[\boldsymbol{P}'\boldsymbol{\Phi}\boldsymbol{P}] + \boldsymbol{\beta}'\boldsymbol{\Omega}\boldsymbol{\beta}E(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{13}$$

At no point in this derivation did I assume that the disturbance terms were normally distributed. On the other hand, I did assume that the model has no overidentifying restrictions. Frequently, however, researchers will estimate models that restrict certain structural coefficients to be zero. Although this may complicate the argument somewhat, it does not change the basic result.

First, consider the case in which some elements of $\boldsymbol{\gamma}$, the direct effect of $\boldsymbol{x}$ on y, are set to zero. Since $\boldsymbol{\gamma}$ does not appear in any of the derivations of the exact variance of $\boldsymbol{d}$, restrictions on $\boldsymbol{\gamma}$ cause no problem whatsoever. Neither does any complication arise from setting to zero one or more coefficients in $\boldsymbol{\beta}$, the direct effect of $\boldsymbol{z}$ on y. Although the OLS estimator $\boldsymbol{b}$ does appear in the derivations, the restricted OLS estimator is readily obtained by deleting the appropriate variables from the model, and the variance matrix $\boldsymbol{\Phi}$ is adjusted accordingly.

The situation is slightly more complicated when restrictions are imposed on the structural parameters in equation (2). Linear restrictions on $\boldsymbol{\Lambda}$ and $\boldsymbol{K}$ typically imply nonlinear restrictions on $\boldsymbol{\Pi}$, the reduced form coefficients. Accordingly, the OLS estimator $\boldsymbol{P}$ must also satisfy those restrictions, but estimation with nonlinear restrictions is usually not possible with standard regression programs. Nevertheless, the derivation in (5) through (13) still applies to the restricted estimator.

3. ESTIMATION

By substituting sample for population quantities in (13), we obtain a consistent estimator

$$\hat{V}(\boldsymbol{d}) = \boldsymbol{P}'\hat{\boldsymbol{\Phi}}\boldsymbol{P} + \boldsymbol{b}'\hat{\boldsymbol{\Omega}}\boldsymbol{b}(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{14}$$

Now $\hat{\boldsymbol{\Phi}}$ is just the usual estimated covariance matrix for $\boldsymbol{b}$. To get $\hat{\boldsymbol{\Omega}}$, we can compute the covariance matrix for the residuals from regressing $\boldsymbol{z}$ on $\boldsymbol{x}$. Alternatively, $\hat{\Omega}$ can be gotten directly by using a structural equations program like LISREL or EQS and parameterizing the model in terms of the reduced form in (3).[3] Regarding $(\boldsymbol{X}'\boldsymbol{X})^{-1}$, many regression programs can write out this matrix, although it is necessary to strip off the first row and column corresponding to the intercept. Another approach to getting $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is to note that $\hat{V}(\mathrm{p}_j)$—the estimated covariance matrix of the coefficients for regressing any one of the z_j's on $\boldsymbol{x}$—is just $s^2_j\ (\boldsymbol{X}'\boldsymbol{X})^{-1}$, where s^2_j is the estimated disturbance variance from that regression. Hence $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ is obtained by multiplying $\hat{V}(\mathrm{p}_j)$ by $1/s^2_j$. A third method follows immediately from the fact that $\boldsymbol{X}'\boldsymbol{X} = (N - 1)C\hat{o}v(\boldsymbol{x})$.

When there is a single $\boldsymbol{x}$ variable, (14) reduces to

$$\hat{V}(d) = \mathrm{p}'\hat{\boldsymbol{\Phi}}\mathrm{p} + \boldsymbol{b}'\hat{\boldsymbol{\Omega}}\boldsymbol{b}/(N - 1)s_x^2, \tag{15}$$

where d is now a scalar and p is the single *column* in the P matrix. When there is a single z variable, we have

$$\hat{V}(\boldsymbol{d}) = s_b^2\ \mathrm{pp}' + b^2\hat{V}(\mathrm{p}), \tag{16}$$

where s^2_b is the squared standard error estimate for b, and $\hat{V}(\boldsymbol{p})$ is the estimated covariance matrix for $\boldsymbol{p}$. Note that the main diagonal elements of (16) are just

$$\hat{V}(d_j) = s_b^2 p_j^2 + b^2 s_j^2 \qquad j = 1, \ldots, n. \tag{17}$$

Thus, in this special case, the variance of each indirect effect is easily obtained from standard OLS regression output.

For a fully recursive system, it is straightforward to extend the preceding results to get the variance of the indirect effects of $\boldsymbol{x}$ on each of the z's. For a given z_k, simply redefine z_k as y in equation (1) and delete from the system any z's that are causally consequent to z_k. Then apply the results already obtained. It is also straightforward to get the indirect effect of any z_k on y: Move z_k and any other z's that are causally prior to z_k into the $\boldsymbol{x}$ vector (i.e., treat them as exogenous) and proceed as before. Finally, to get the variance of the "specific" indirect effect of $\boldsymbol{x}$ on y through z_k, delete all z's that are causally consequent to z_k and move all causally prior z's into the $\boldsymbol{x}$ vector. Then use equation (16).[4]

[3] To accomplish this, constrain all direct effects of z on itself to be zero, and let the covariance matrix for the disturbances in the equations for z be unrestricted.

[4] This approach cannot be used to estimate the variance of specific indirect effects that operate through two specified variables.

4. AN EXAMPLE

Like many other authors (Alwin and Hauser 1975; Fox 1980, 1985; Sobel 1982, 1986; Bollen 1987), I illustrate the method with the six-variable model and data on occupational achievement reported by Duncan, Featherman and Duncan (1972, p. 38). The correlations and standard deviations for the 3,214 nonblack, nonfarm men come from their Table 3.1. The variables are father's occupation (x_1), father's education (x_2), number of siblings (x_3), education (z_1), 1961 occupational status (z_2), and 1961 income (y). The model is

$$\begin{aligned} y &= \beta_1 z_1 + \beta_2 z_2 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \epsilon \\ z_1 &= \kappa_{11} x_1 + \kappa_{12} x_2 + \kappa_{13} x_3 + \nu_1 \\ z_2 &= \lambda_1 z_2 + \kappa_{21} x_1 + \kappa_{22} x_2 + \kappa_{23} x_3 + \nu_2, \end{aligned} \tag{18}$$

which is represented by a path diagram in Figure 1. The reduced-form equation for z_2 is

$$z_2 = \pi_{21} x_1 + \pi_{22} x_2 + \pi_{23} x_3 + u_2. \tag{19}$$

Since there are no endogenous independent variables in the structural equation for z_1, it is also the reduced-form equation with $\pi_{1j} = \kappa_{1j}$ for $j = 1, \ldots, 3$.

I estimated both the structural and reduced forms of the model using LISREL8 (Jöreskog and Sörbom 1993) with the covariance matrix as input. The reduced form differed from the structural form only in deleting the path from z_1 to z_2 (i.e., setting $\lambda_1 = 0$) and

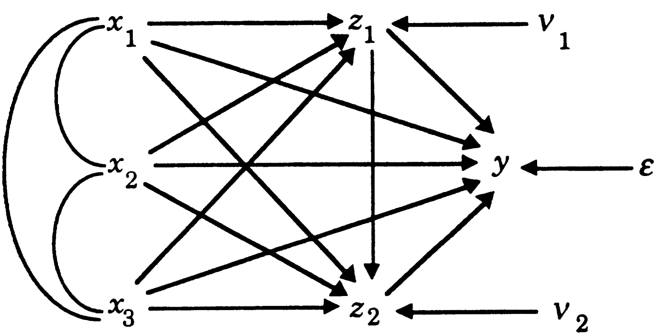

FIGURE 1. Path diagram for occupational achievement example.

allowing a correlation between ν_1 and ν_2. The estimates for the β's were $\boldsymbol{b} = [0.1998\ 0.0704]'$. The estimates for the π's were

$$\boldsymbol{P} = \begin{bmatrix} 0.1707 & 0.0385 & -0.2281 \\ 0.7963 & 0.3035 & -1.4613 \end{bmatrix}.$$

The estimates for the indirect effects are therefore

$$\boldsymbol{d} = \boldsymbol{P}'\boldsymbol{b} = [0.09016\ \ 0.02906\ -0.14845]'.$$

To get the variance of $\boldsymbol{d}$, we also need estimates of $\boldsymbol{\Omega}$, $\boldsymbol{\Phi}$, and $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. In the reduced-form version of the model, LISREL directly reported

$$\hat{\boldsymbol{\Omega}} = \begin{bmatrix} 7.4852 & 32.7605 \\ 32.7605 & 490.9398 \end{bmatrix}.$$

(These estimates are found in what LISREL labels the PSI matrix). The estimate for $\boldsymbol{\Phi}$ was

$$\hat{\boldsymbol{\Phi}} = \begin{bmatrix} 0.0013249 & -0.0000885 \\ -0.0000885 & 0.0000202 \end{bmatrix}.$$

The variances on the main diagonal of $\hat{\boldsymbol{\Phi}}$ were obtained simply by squaring the reported standard errors for b_1 and b_2. The covariance was calculated by multiplying the reported correlation by the two standard errors. Finally, I used the formula $\boldsymbol{X}'\boldsymbol{X} = (N - 1)\hat{Cov}(\boldsymbol{x})$, where the covariance matrix was reported by LISREL

$$\hat{Cov}(\boldsymbol{x}) = \begin{bmatrix} 13.8384 & 45.6228 & -3.0759 \\ 45.6228 & 535.4596 & -16.5009 \\ -3.0759 & -16.5009 & 8.2944 \end{bmatrix}.$$

Plugging these results into equation (14) yields[5]

$$\hat{V}(\boldsymbol{d}) = \begin{bmatrix} 0.001459 & -3.022 \times 10^{-6} & -0.000012 \\ -3.022 \times 10^{-6} & 4.7499 \times 10^{-6} & -6.986 \times 10^{-6} \\ -0.000012 & -6.986 \times 10^{-6} & 0.0002045 \end{bmatrix}.$$

Taking the square root of the main diagonal elements gives the estimated standard errors [0.01208 0.002179 0.01430]. These numbers are identical to those obtained by Sobel (1986) using the delta method, although he reported only three decimal places. They are

[5]Matrix calculations were carried out with PROC IML in SAS (SAS Institute 1990).

also identical to the standard errors of the "total indirect effects" reported by LISREL.

Now let's take z_2 as our dependent variable and estimate the indirect effects of the x's through z_1. We have

$$\boldsymbol{d} = \boldsymbol{p}'b = [0.1707\ 0.0385\ -0.2281]'4.3767 = [0.7471\ 0.1685\ -0.9983]'.$$

Since there is just one intervening variable, we can use equation (17) to get the standard errors. This requires only the reported standard errors for each of the p and b coefficients. Thus

$$\text{s.e.}(d_1) = \sqrt{[(0.1203)^2(0.1707)^2 + (4.3767)^2(0.0156)^2]} = 0.0713$$
$$\text{s.e.}(d_2) = \sqrt{[(0.1203)^2(0.0385)^2 + (4.3767)^2(0.0025)^2]} = 0.0118$$
$$\text{s.e.}(d_3) = \sqrt{[(0.1203)^2(-0.2281)^2 + (4.3767)^2(0.0176)^2]} = 0.0819.$$

Again, these estimates are identical to those reported by Sobel and by LISREL.

The indirect effect of z_1 on y is 4.3767(0.0704) = 0.3082 (the effect of z_1 on z_2 times the effect of z_2 on y). Equation (17) yields

$$\sqrt{[(0.0045)^2(4.3767)^2 + (0.1203)^2(0.0704)^2]} = 0.0214.$$

5. COMPARISON WITH THE DELTA METHOD

How does the exact variance of $\boldsymbol{d}$ in (13) and the corresponding estimator in (14) compare with that obtained by Sobel using the delta method? We have just seen numerical results suggesting that they are very close, if not identical. To answer the question more generally, I now apply the delta method to the reduced-form parameterization in equations (1) and (3). Write the coefficients of the system as a single vector $\boldsymbol{\theta} = [\boldsymbol{\beta}'\ \boldsymbol{\pi}'_1\ \boldsymbol{\pi}'_2\ \ldots\ \boldsymbol{\pi}'_n]'$ with the corresponding estimator $\hat{\boldsymbol{\theta}} = [\boldsymbol{b}'\ \boldsymbol{p}'_1\ \boldsymbol{p}'_2\ \ldots\ \boldsymbol{p}'_n]'$, where the $\boldsymbol{\pi}_j$'s are columns of $\boldsymbol{\Pi}$ and the $\boldsymbol{p}_j$'s are columns of $\boldsymbol{P}$. We know that

$$\boldsymbol{d} = f(\hat{\boldsymbol{\theta}} = [\boldsymbol{b}'\boldsymbol{p}_1\ \boldsymbol{b}'\boldsymbol{p}_2\ \ldots\ \boldsymbol{b}'\boldsymbol{p}_n]'. \tag{20}$$

The delta method says that the asymptotic approximation to the variance of $\boldsymbol{d}$ can be found as[6]

$$\hat{V}(\boldsymbol{d}) = \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) \hat{V}(\hat{\theta}) \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right)\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{21}$$

[6] I follow the convention that if f is $n \times 1$ and $\boldsymbol{\theta}$ is $r \times 1$, $\partial f/\partial \boldsymbol{\theta}$ is $n \times r$.

It can be shown that

$$\left.\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{I}_m \otimes \boldsymbol{b} \end{bmatrix}, \tag{22}$$

where $\otimes$ is the Kronecker product, and

$$\hat{V}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \hat{\boldsymbol{\Phi}} & \boldsymbol{0} \\ \boldsymbol{0} & (\boldsymbol{X}'\boldsymbol{X})^{-1} \otimes \hat{\boldsymbol{\Omega}} \end{bmatrix} \tag{23}$$

where $\hat{\boldsymbol{\Phi}}$ and $\hat{\boldsymbol{\Omega}}$ are consistent estimators of $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$. Substituting (22) and (23) into (21), we get

$$\hat{V}(\boldsymbol{d}) = \boldsymbol{P}'\hat{\boldsymbol{\Phi}}\boldsymbol{P} + \boldsymbol{b}'\hat{\boldsymbol{\Omega}}b(X'X)^{-1}, \tag{24}$$

which is identical to (14). Thus we see that the delta method applied to the reduced-form parameterization leads to an estimator of the variance of the indirect effects that is identical to an estimator based on the exact variance. This equivalence does not rest on any assumptions since the claim is that the *estimators* are the same, regardless of whether the postulated model is correct or not.

But what if the delta method is applied to the structural parameters, as in Sobel's derivation? The first question is whether the estimates of the indirect effects are the same whether they are calculated from the reduced form or the structural form. It is well known that maximum likelihood estimates are invariant to reparameterization and, in the case of fully recursive models, ML may be accomplished by OLS applied to each equation. So in that setting the answer is definitely yes. For nonrecursive models, however, the use of methods other than full-information maximum likelihood (e.g., two-stage least squares) may yield different estimates for indirect effects calculated from the structural and reduced forms.

For ML estimates of the indirect effects, the next question is whether variance estimates obtained by the delta method are invariant to reparameterization. The answer depends on how the original variance estimates are obtained. In the appendix to this chapter I show that if the variance estimates are based on the observed information matrix, the delta method *is* invariant to reparameterization. Therefore the estimates produced by Sobel's formulas (which presume that the variance estimates come from the observed information matrix) will be identical to those obtained with the formulas given here.

6. DISCUSSION

Variance estimates produced by the delta method ordinarily have two sources of error: (1) error in approximating the true variance and (2) sampling error that arises in estimating the approximate variance. I have shown that in the special case of indirect effects in recursive linear models estimated by OLS, the error of approximation is zero. Thus the formulas given by Sobel are not, in fact, large-sample approximations but estimates of the true variance. Nevertheless, sampling errors may still be substantial, and there is no guarantee that test statistics based on the estimated variances will have their postulated distribution. In the abstract to their paper, Stone and Sobel (1990) interpret their Monte Carlo simulations as showing that samples of at least 200 cases are needed to get good standard error estimates for recursive models like the example considered here. But in their discussion section, they state that "confidence can be placed in the point estimates and hypothesis tests in samples as small as $N = 100$." Even with sample sizes as small as 50, moreover, they found that of 3,500 nominal 95 percent confidence intervals, almost exactly 95 percent contained the true value. Thus their results are actually reassuring about the use of the variance estimators for recursive models in small- to moderate-sized samples.

Although the formulas given here are equivalent to Sobel's formulas, they may be simpler to use in some situations because they do not use Kronecker products and they do not require the construction of matrices of first derivatives. The downside is that one must reestimate the model in the reduced-form parameterization. And if some of the structural coefficients are set to zero, the reduced-form model may need to be estimated under nonlinear restrictions.

APPENDIX

Here I show that the delta method is invariant to reparameterization when the parameters are estimated by maximum likelihood and variance estimates are obtained from the observed information matrix. Suppose we have a random vector $\boldsymbol{y}$ with c.d.f. $H(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is $m \times$

1. Let $\hat{\boldsymbol{\theta}}$ be the maximum likelihood estimator of $\boldsymbol{\theta}$. We estimate the variance of $\hat{\boldsymbol{\theta}}$ by

$$\hat{V}(\hat{\boldsymbol{\theta}}) = -\left[\frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]^{-1}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{A.1}$$

where ℓ is the log-likelihood function. Let $\boldsymbol{f}(\boldsymbol{\theta})$ be an $n \times 1$ function that is differentiable in the neighborhood of $\boldsymbol{\theta}_0$, the true value of the parameter vector. We know that $\boldsymbol{f}(\hat{\boldsymbol{\theta}})$ is the maximum likelihood estimator of $\boldsymbol{f}(\boldsymbol{\theta})$. The delta method says that a large-sample approximation to the variance of $\boldsymbol{f}(\hat{\boldsymbol{\theta}})$ is given by

$$\hat{V}_{\boldsymbol{\theta}}[\boldsymbol{f}(\hat{\boldsymbol{\theta}})] = \frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \hat{V}(\hat{\boldsymbol{\theta}}) \frac{\partial \boldsymbol{f}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{A.2}$$

The variance is subscripted by $\boldsymbol{\theta}$ to indicate that it may depend on this particular parameterization.

Now consider the reparameterization $\boldsymbol{\phi} = \boldsymbol{c}(\boldsymbol{\theta})$ and its inverse $\boldsymbol{\theta} = \boldsymbol{g}(\boldsymbol{\phi})$, where $\boldsymbol{g}$ is assumed to be continuous, twice differentiable, and nonvanishing in the neighborhood of the true value $\boldsymbol{\phi}_0$. We also assume that the $m \times m$ matrix $\partial\boldsymbol{\theta}/\partial\boldsymbol{\phi}$ is invertible.

Using the vector version of the chain rule, the first derivative of the log-likelihood with respect to $\boldsymbol{\phi}$ is

$$\frac{\partial \ell}{\partial \boldsymbol{\phi}} = \frac{\partial \ell}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}. \tag{A.3}$$

The second derivative is

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'} = \frac{\partial}{\partial \boldsymbol{\phi}}\left(\frac{\partial \ell}{\partial \boldsymbol{\phi}}\right)' = \frac{\partial}{\partial \boldsymbol{\phi}}\left(\frac{\partial \ell}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}\right)'$$

$$= \frac{\partial}{\partial \boldsymbol{\phi}}\left(\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}'} \frac{\partial \ell}{\partial \boldsymbol{\theta}'}\right). \tag{A.4}$$

Using results in Drhymes (1984, p. 109), we then have

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'} = \left(\frac{\partial \ell}{\partial \boldsymbol{\theta}} \otimes \boldsymbol{I}_m\right) \frac{\partial^2 \boldsymbol{\theta}}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'} + \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}'} \frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}. \tag{A.5}$$

When this expression is evaluated at $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}$, the maximum likelihood estimate, we have $\partial\ell/\partial\boldsymbol{\theta} = \boldsymbol{0}$, so the second derivative reduces to

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}\Bigg|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}} = \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}'} \frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}\Bigg|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}},\, \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{A.6}$$

We therefore have

$$\hat{V}(\hat{\boldsymbol{\phi}}) = -\left[\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}'} \frac{\partial^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}\right]^{-1} \Bigg|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}},\ \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$
$$= \left[\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}\right]^{-1} \hat{V}(\hat{\boldsymbol{\theta}}) \left[\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}'}\right]^{-1} \Bigg|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}},\ \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} . \tag{A.7}$$

Applying the delta method to get the asymptotic variance of $f(\hat{\boldsymbol{\theta}}) = f(g(\hat{\boldsymbol{\phi}}))$ under the new parameterization yields

$$\hat{V}_{\boldsymbol{\phi}}[f(g(\hat{\boldsymbol{\phi}}))] = \frac{\partial f(g(\boldsymbol{\phi}))}{\partial \boldsymbol{\phi}} \hat{V}(\hat{\boldsymbol{\phi}}) \frac{\partial f(g(\boldsymbol{\phi}))}{\partial \boldsymbol{\phi}'} \Bigg|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}}}$$
$$= \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}} \hat{V}(\hat{\boldsymbol{\phi}}) \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}'} \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \Bigg|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}},\ \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} . \tag{A.8}$$

Substituting from (A.7), we get

$$\hat{V}_{\boldsymbol{\phi}}[f(g(\hat{\boldsymbol{\phi}}))] = \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}} \left[\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}}\right]^{-1} \hat{V}(\hat{\boldsymbol{\theta}}) \left[\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}'}\right]^{-1} \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\phi}'} \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \Bigg|_{\boldsymbol{\phi}=\hat{\boldsymbol{\phi}},\ \boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$
$$= \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \hat{V}(\hat{\boldsymbol{\theta}}) \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \hat{V}_{\boldsymbol{\theta}}(f(\hat{\boldsymbol{\theta}})). \tag{A.9}$$

We conclude that the delta method gives the same results under either parameterization. For a statement of a related result, see Bickel and Doksum (1977, p. 147).

REFERENCES

Alwin, D. F., and R. M. Hauser. 1975. "The Decomposition of Effects in Path Analysis." *American Sociological Review* 40:37–47.

Bickel, P. J., and K. A. Doksum. 1977. *Mathematical Statistics.* Oakland, Calif.: Holden-Day.

Bollen, K. A. 1987. "Total, Direct, and Indirect Effects in Structural Equation Models." In *Sociological Methodology 1987,* edited by C. C. Clogg, 37–69. Washington: American Sociological Association.

Bollen, K. A., and R. Stine. 1990. "Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability." In *Sociological Methodology 1990,* edited by C. C. Clogg, 115–140. Oxford: Blackwell Publishers.

Drhymes, P. J. 1984. *Mathematics for Econometrics,* 2nd ed. New York: Springer-Verlag.

Duncan, O. D., D. L. Featherman and B. Duncan. 1972. *Socioeconomic Background and Achievement.* New York: Seminar Press.

Fox, J. 1980. "Effects Analysis in Structural Equation Models." *Sociological Methods and Research* 9:3–28.

———. 1985. "Effects Analysis in Structural Equation Models II: Calculation of Specific Indirect Effects." *Sociological Methods and Research* 14: 81–85.

Goldberger, A. S. 1991. *A Course In Econometrics.* Cambridge: Harvard University Press.

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

SAS Institute Inc. 1990. *SAS/IML Software: Usage and Reference, Version 6.* Cary, N.C.: SAS Institute.

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

———. 1986. "Some New Results on Indirect Effects and their Standard Errors in Covariance Structure Models." In *Sociological Methodology 1986,* edited by N. B. Tuma, 159–86. Washington: American Sociological Association.

Stone, C. A., and M. E. Sobel. 1990. "The Robustness of Estimates of Total Indirect Effects in Covariance Structure Models Estimated by Maximum Likelihood." *Psychometrika* 55: 337–52.