Decomposition of Variables and Correlated Measurement Errors

Saul Lach

International Economic Review, Volume 34, Issue 3 (Aug., 1993), 715-725.

Stable URL:
http://links.jstor.org/sici?sici=0020-6598%28199308%2934%3A3%3C715%3ADOVACM%3E2.0.CO%3B2-E

INTERNATIONAL ECONOMIC REVIEW
Vol. 34, No. 3, August 1993

DECOMPOSITION OF VARIABLES AND CORRELATED MEASUREMENT ERRORS*

BY SAUL LACH[1]

This paper examines the bias in the OLS estimators when the regressors have measurement errors correlated in a particular manner. When a variable is decomposed into two components but only one of them is observed with error, the induced measurement error in the other component is identical but has the opposite sign. This specific correlation pattern enables us to assess the direction of the bias in the OLS estimators from observed data. In the standard EIV case this would require knowledge of the relative variances of the measurement errors. Examples of this type of decomposition in applied work are presented.

1. INTRODUCTION

Regression models that allow for measurement errors in the explanatory variables usually assume that these errors are uncorrelated among them.[2] This is, perhaps, the natural assumption when nothing is known about the data process generating the regressors. In this paper it is argued that in empirical work there are many cases where, due to a trivial manipulation of the data, this assumption cannot be invoked. The bias in the OLS estimators when the explanatory variables are measured with correlated errors is examined. The results both complement and differ from the standard case of uncorrelated measurement errors in interesting ways.

Specifically, the case that motivates the paper is one in which an original explanatory variable, X, is decomposed into two (or more) components because one is interested in testing whether these components have different effects on the dependent variable. This case occurs frequently in applied work. In addition, it is common to have data on the total variable X and on only one of its components; the other is computed as a residual. Provided that X is measured without error, any measurement error in one of its components is transmitted to the second compo-

* Manuscript received June 1991.

[1] I would like to thank Joshua Angrist, Zvi Griliches, Julian Silk and Shlomo Yitzhaki for their helpful comments and especially two anonymous referees whose reports greatly improved the presentation and content of the paper.

[2] This is true in most of the econometric literature (Garber and Klepper 1980, and Klepper and Leamer 1984) but is less common in the statistical literature (Haitovsky 1972, and Fuller 1987). One of the rare exceptions is Duncan and Hill (1985), where explicit recognition and direct measurement of the correlation among measurement errors in labor economics data is presented. Griliches (1986, p. 1480) is also aware of this possibility and refers to the case in which a variable is divided by another erroneous variable. For example, if wages are computed as the ratio of the wage bill to total manhours, then if the latter is measured with a multiplicative error, so will the resulting wage rate (but with the opposite sign). Hence, observed log hours and log wages have correlated measurement errors, but these errors correspond to the *dependent* and independent variables.

nent. The latter measurement error is identical to that of the first component but with the opposite sign.

In the classical EIV (Errors-in-Variables) model, measurement errors are assumed to be uncorrelated among them. In this case, because of the manipulation of the data, this assumption is incorrect: the measurement errors are perfectly (negatively) correlated. In fact, they exactly cancel each other out.

Consider the following examples. In estimating the effect of the composition of R&D expenditures on productivity growth, at the firm level, there are usually data on total R&D and on only one of its components as in the case of company versus federally funded R&D (Lichtenberg 1990) or in applied versus basic R&D (Mansfield 1980). Similarly, the effect of the composition of the labor force on productivity growth fits into this framework. Typically, the decomposition is between production versus nonproduction, or skilled versus unskilled workers. Data for the total labor force and for the component most easily measured are usually available (e.g., the number of production or skilled workers), while the other component equals the residual.[3] In the estimation of earning equations one of the indicators of human capital used as an explanatory variable is work experience. In the vast majority of empirical work it is standard practice to measure work experience as "Age–S–6", where S is years of schooling and is also a regressor in the earning equation. If errors are made in measuring S, then experience will have the same measurement error, but with opposite sign.[4]

Another example is provided by recent empirical work on the importance of market fundamentals and fads on investment decisions by firms, using the q-theory of investment (Blanchard, Rhee, and Summers 1990, Galeotti and Schiantarelli 1990). Log (q), the regressor, is decomposed into two components: (a) the log of the ratio between market value (V) and profits (Π), meant to capture deviations from fundamentals, and (b) the log of profits per unit of capital (K), representing market fundamentals. The decomposition is in log form, i.e., $\log(q) = \log(V/K) = \log(V/\Pi) + \log(\Pi/K)$. The interest lies precisely in testing for differences in the components' coefficients in the investment regression. In this case, neither q nor its components can be directly observed and considerable effort is spent in constructing q and Π/K. If one is willing to assume that measured (constructed) $\log(\Pi/K)$ differs from the true market fundamental term, in logs, by an additive error, then this fits into the case analyzed in this paper.

Another decomposition of interest is the one between expected and unexpected components, as used in the debate over policy ineffectiveness (e.g., Mishkin 1982) or in some empirical investigations of relative price variability (e.g., Lach and

[3] In many of these examples the regressor is actually the *share* of the total attributed to each component. The analysis in this paper is developed in terms of *levels* of components, but it can be easily accommodated to shares by a suitable redefinition of the errors in measurement.

[4] There are many reasons for years of schooling to be erroneously measured: the question may be misunderstood, e.g., do years of formal education refer to the number of grades completed or to the number of years spent in school? Or, more importantly, how does one account for the amount of work or professional training performed while enrolled in school? Sicherman (1990) addresses this issue. Note that, in this context, the issue is whether years of schooling per se, as affecting work experience, are correctly measured and not whether S is a good measure of the level of acquired education.

Tsiddon 1992). Since both parts are not observable they have to be estimated and, in practice, one replaces the true expected and unexpected variables by error-ridden counterparts. These errors are a consequence of replacing true parameters by their estimates and of possible misspecifications in predicting the expected component. Evidently, these errors cancel each other out. It will be shown that this case does not fit into the classical EIV framework since the error is correlated with one of the true variables. It is, nevertheless, interesting enough to warrant investigation.

The paper is organized as follows: the following section presents the (asymptotic) biases in the OLS estimators of the parameters corresponding to the components of an arbitrary decomposition of X. The possibility of inferring the sign of these biases from the data, and of deriving bounds for the true parameters is discussed. This section also explains how to test the hypothesis of equality between the components' effects. Section 3 analyzes the case in which X is decomposed into expected and unexpected components and a misspecification error is committed in predicting the expected component. Conclusions close the paper.

2. AN ARBITRARY DECOMPOSITION OF X

2.1. The Bias in OLS Estimation. In order to concentrate on the fundamentals this section considers the standard linear model with a single explanatory variable measured without error,

$$y_i = \beta_0 + \beta_x X_i + \varepsilon_i \qquad i = 1, \ldots, N,$$

where $\{\varepsilon_i\}$ is an iid sequence of random variables with mean zero and variance $\sigma_{\varepsilon\varepsilon}$, uncorrelated with X. As suggested in the introduction, X is decomposed into two components, X_1^* and X_2^*, each of them uncorrelated with ε, with the objective of estimating and testing their differential effect upon y. Suppose, further, that there exist data on X and X_1^* only. Then X_2^* is computed as $X_2^* = X - X_1^*$. Provided that X is measured without error, a measurement error v on X_1^* generates an identical error in observed X_2^* but with the opposite sign.[5] Letting unstarred letters denote observed variables we have the following model

$$y_i = \beta_0 + \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \varepsilon_i \qquad i = 1, \ldots, N,$$

$$X = X_1^* + X_2^*, \qquad X_1 = X_1^* + v, \qquad X_2 = X - X_1 = X_2^* - v,$$

where v has zero mean and variance σ_{vv} and is uncorrelated with the true variables X_1^* and X_2^* and with ε, implying plim $(1, X_1, X_2)'\varepsilon/N = 0$, and plim $(1, X_1^*, X_2^*)'v/N = 0$. The regression actually being estimated is

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i + (\beta_2 - \beta_1)v_i \qquad i = 1, \ldots, N.$$

[5] If total X is also measured with an error, η, then it will be transmitted to observed X_2. Now, X_2 will differ from the true X_2 by $\eta - v$. The negative correlation between the measurement errors remains but is less than perfect. See the Appendix.

Let W be the observed data matrix, $W = (1, X_1, X_2)$, and let $\beta = (\beta_0, \beta_1, \beta_2)'$. It is assumed that X_1^* and X_2^* are iid sequences.[6] Denote by Ω the limiting cross-product matrix of the observed data W, i.e., $\Omega \equiv \text{plim } W'W/N$. Ω is derived in the Appendix. Similarly, let Ω^* be the corresponding matrix for the true data. Then $\Omega = \Omega^*$ if and only if $\sigma_{vv} = 0$. Both Ω and Ω^* are 3 by 3, symmetric, positive definite matrices. The OLS estimator of β, $b = (b_0, b_1, b_2)'$, is given by $b = \beta + (W'W)^{-1}W'[\varepsilon + (\beta_2 - \beta_1)v]$. Since plim $W'\varepsilon/N = 0$ and plim $W'v(\beta_2 - \beta_1)/N = (\beta_2 - \beta_1)\sigma_{vv}(0, 1, -1)'$, we obtain plim $b = \beta + \Omega^{-1}(0, 1, -1)'(\beta_2 - \beta_1)\sigma_{vv}$. Computing the inverse of Ω (shown in the Appendix), and substituting into the last expression results in the following probability limits for b_0, b_1 and b_2,[7]

$$\text{(1a)} \qquad \text{plim } b_0 = \beta_0 + (\beta_2 - \beta_1)\sigma_{vv}[\mu_2(\omega_{12} + \omega_{11}) - \mu_1(\omega_{22} + \omega_{12})]|\Omega|^{-1}$$

$$\text{(1b)} \qquad \text{plim } b_1 = \beta_1 + (\beta_2 - \beta_1)\sigma_{vv}(\sigma_{22} + \sigma_{12})|\Omega|^{-1}$$

$$\text{(1c)} \qquad \text{plim } b_2 = \beta_2 - (\beta_2 - \beta_1)\sigma_{vv}(\sigma_{11} + \sigma_{12})|\Omega|^{-1}$$

where the σ's are the variances and covariances of the observed variables ($\sigma_{ij} = EX_iX_j - EX_iEX_j$, $i, j = 1, 2$), and $|\Omega| = \sigma_{11}\sigma_{22} - \sigma_{12}^2$ is a positive scalar.

These formulae imply that under H_0: $\beta_1 = \beta_2$, b_0, b_1 and b_2 are consistent estimators of the true parameters. This is so because y can be regressed on $X = X_1 + X_2$ and, since X was assumed to be observed without error, OLS gives consistent estimators under H_0.

In order to sign the biases of b_1 and b_2, the estimators of interest, something must be said about the covariance between X_1 and X_2, σ_{12}.[8] Since the variance of X is positive, σ_{12} is bounded from below by $-(\sigma_{11} + \sigma_{22})/2$. Hence, σ_{12} cannot be too negative. It can, however, be negative enough so as to make either $(\sigma_{11} + \sigma_{12})$ or $(\sigma_{22} + \sigma_{12})$ negative, but not both. In this case, both b_1 and b_2 are biased in the same direction. The direction of the bias is determined by the relative magnitudes of the variances of the observed data and by the true parameters. For example, if the variance of X_2 is smaller than the variance of X_1 and β_2 is the largest parameter then both b_1 and b_2 are biased downward, while if β_1 is the largest parameter then there is an upward bias in both estimators.[9] The other case is the one in which σ_{12} is not negative enough so that both $(\sigma_{11} + \sigma_{12})$ and $(\sigma_{22} + \sigma_{12})$ are positive.[10] In this case, the estimator of the largest parameter is biased downwards, while the estimator of the smallest parameter is biased upwards. That

[6] This is a simplifying assumption; some weak dependence can be incorporated without changing the basic results. All that is required is that a law of large numbers can be applied.

[7] These are clearly not new results as can be seen by writing the model as $y = \beta_0 + \beta_1 X + (\beta_2 - \beta_1)X_2^* + \varepsilon$, with X measured without error and X_2^* measured with error. Thus, the usual attenuation result applies to the OLS estimator of $\beta_2 - \beta_1$ while the OLS estimator of β_1 is also inconsistent with a probability limit given by (1*b*).

[8] Throughout the paper, the term "bias" refers to the difference between the probability limit of an estimator and the corresponding true parameter.

[9] This shows that Result 1 in Garber and Klepper (1980) which states that "The coefficient of a mismeasured regressor is not necessarily attenuated, but at least one of them is attenuated," does not necessarily hold when the measurement errors are allowed to be correlated.

[10] One may even argue that this is the most plausible case.

is, the biases in the estimators are always in opposite directions, as a result of the error term being correlated with each regressor in the opposite direction.

Since σ_{11}, σ_{22} and σ_{12} can be estimated from the observed data, we can know which case applies to each particular sample. But can we know which parameter is the largest? A positive answer can be given if one could consistently estimate the sign of the difference between β_1 and β_2. It turns out that the sign of $(b_2 - b_1)$ consistently estimates the sign of $(\beta_2 - \beta_1)$ and, in fact, a simple algebraic manipulation of (1) gives,

$$(2) \qquad \text{plim}(b_2 - b_1) = (\beta_2 - \beta_1)[|\Omega| - \sigma_{vv} \text{ var}(X)]|\Omega|^{-1} = (\beta_2 - \beta_1)|\Omega^*|/|\Omega|$$

where it can be easily shown that $|\Omega| - \sigma_{vv} \text{ var}(X) = |\Omega^*|$.[11] Thus, $0 \leq |\Omega^*|/|\Omega| \leq 1$. What (2) reveals is that the OLS estimators can identify which of the two coefficients is the largest and then, through (1) and the estimates of σ_{11}, σ_{22} and σ_{12}, one can identify which coefficient estimate is biased upward and which is biased downward. Of course, the magnitude of the biases cannot be estimated since σ_{vv} is not identified.[12]

The expression in (2) also shows that the OLS estimators underestimate the true (absolute) difference between β_2 and β_1, the extent of this bias depending on the severity of the measurement error, σ_{vv}, and on the variance of X. Hence, $b_2 - b_1$ is a lower (upper) bound for $\beta_2 - \beta_1$ when $\beta_2 - \beta_1$ is positive (negative).

2.2. Bounds on β_1 and β_2. Asymptotic bounds on the parameters can be obtained by means of OLS estimates from a direct and a reverse regression.[13] We analyze the bounds on β_1 only; the bounds on β_2 can be found in an analogous way. The direct regression is a reformulation of the original model,

$$y = \beta_0 + \beta_1 X + (\beta_2 - \beta_1)X_2 + \varepsilon + (\beta_2 - \beta_1)v$$

where X is measured without error.[14] The reverse regression is

$$X_2 = \alpha_0 + \alpha_1 X + \alpha_2 y + \eta$$

where $\alpha_1 = \beta_1(\beta_1 - \beta_2)^{-1}$, $\alpha_2 = (\beta_2 - \beta_1)^{-1}$ and $\eta = -\varepsilon(\beta_2 - \beta_1)^{-1} - v$. Denoting by a_1 and a_2 the OLS estimators of α_1 and α_2, the Appendix shows that they satisfy

[11] Recall that Ω^* is the limiting cross-product matrix corresponding to the true variables and is positive definite, $|\Omega^*| = \sigma_{11}^*\sigma_{22}^* - \sigma_{12}^{*2} > 0$.

[12] See footnote 7. Another curious result is that the sum of the estimated coefficients is not always biased towards zero as in the case of uncorrelated measurement errors (see, e.g., Griliches 1986). After some simple algebra we obtain: plim $(b_1 + b_2) = (\beta_1 + \beta_2) + (\beta_2 - \beta_1)\sigma_{vv}(\sigma_{22} - \sigma_{11})/|\Omega|$. If the variable with the highest variance also has the largest coefficient, then the bias is positive, away from zero.

[13] This topic was suggested by one of the referees who kindly outlined the bounding procedure shown here. Klepper and Leamer (1984) derive bounds for the regression coefficients when there are uncorrelated measurement errors in all variables. They extend the well-known "bracketing" result of a single explanatory variable measured with error.

[14] This reformulation makes it clearer that the OLS procedure produces an attenuated estimator of $(\beta_2 - \beta_1)$.

$$\text{plim} - \frac{a_1}{a_2} = \beta_1 - \frac{(\beta_2 - \beta_1)\sigma_{\varepsilon\varepsilon}(\sigma_{22} + \sigma_{12})}{\psi - \sigma_{\varepsilon\varepsilon}\ \text{var}(X)} \tag{3}$$

where $\psi = \text{var}(X)\ \text{var}(y) - \text{cov}(X, y)^2$, and $\psi - \sigma_{\varepsilon\varepsilon}\ \text{var}(X) > 0$.

Combining this result with (1b), and provided that $(\beta_2 - \beta_1)(\sigma_{22} + \sigma_{12})$ is nonnegative, results in the following asymptotic bounds for β_1

$$\text{plim} - \frac{a_1}{a_2} \leq \beta_1 \leq \text{plim}\ b_1. \tag{4}$$

The inequalities are reversed when $(\beta_2 - \beta_1)\ (\sigma_{22} + \sigma_{12})$ is negative.

Similarly, the estimate of the coefficient of X in the direct regression of y on X and X_1, and the estimate of the coefficient of X in the reverse regression of X_1 on X and y, after this reverse regression is solved out so that y is again in the left-hand side, asymptotically bound β_2.

2.3. Testing $\beta_1 = \beta_2$. In many instances it is interesting to test whether the two components of X affect y in a similar way. In order to test the null hypothesis H_0: $\beta_1 = \beta_2$, notice that under H_0 the model behaves as if there are no errors in variables. Thus, under H_0, b has the same properties as the OLS estimator of β in the case where all variables are measured without errors. The test for the equality of the coefficients is, therefore, a standard F or χ^2 test.[15]

2.4. Additional Regressors. Finally, all previous results generalize to the case in which additional, correctly measured regressors are added to the model. Their effects can be controlled by working with the residuals from the regressions of y, X_1, and X_2 on all correctly measured regressors.

As usual, the OLS estimators of the coefficients of the correctly measured variables are also inconsistent. In this case too, their bias, as well as the bias in the OLS estimators of the parameters of the variables measured with error, are proportional to $(\beta_2 - \beta_1)$ and depend, also, on variances and covariances of the observed data.[16] The coefficients of the correctly measured regressors can be asymptotically bounded through the direct and reverse regressions involving y, X_1, X_2, and all the other correctly measured variables.[17]

3. EXPECTED-UNEXPECTED (ORTHOGONAL) DECOMPOSITION OF X

This section departs from the errors-in-variables framework and analyzes the case in which X is decomposed into expected and unexpected components and a

[15] Notice, however, that the above results suggest that the power of such tests may be lower than in the no errors-in-variables case since when departing from the null hypothesis the absolute difference between the estimators is under-estimated.

[16] See the working paper version, Lach (1992), for a detailed derivation of these results.

[17] Let Z be the matrix of K additional regressors. One bound on the coefficient of, say, Z_j is generated by the estimate of the coefficient of Z_j in the direct regression of y on X_1, X_2 and Z. The other bound is obtained from the inverse of the estimate of the coefficient of y in the regression of Z_j on X_1, X_2, $Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_K$, and y.

misspecification error is committed in computing the expected component. The link with the previous analysis is that this misspecification generates two observed regressors that differ from the true regressors by errors that cancel out. These errors, however, are not typical measurement errors. We start with Pagan's (1984) formulation

$$y = \beta_1 X_1^* + \beta_2 X_2^* + \varepsilon$$

$$X = X_1^* + X_2^* = W\alpha + Z\delta + X_2^*.$$

Here, X_1^* and X_2^* are the two components in the orthogonal decomposition of X, representing the expected and unexpected parts of X, satisfying $(W, Z)'X_2^* = 0$.[18] This implies $X_1^{*\prime}X_2^* = 0$, i.e., orthogonal components. Of course, both components of X are unobservable and have to be estimated. The model specifies that an OLS regression of X on the observed predetermined variables in W and Z produces consistent estimators of α and δ, which can then be used to generate predictors of X_1^* and X_2^*. This is the first stage regression. In the second stage, the model is estimated by OLS using the predicted value of X and its residual from the previous stage in place of X_1^* and X_2^* respectively. In this fashion, consistent estimators of β are obtained.

Suppose that a specification error is committed in the first stage regression. Namely, the set of variables in Z is omitted from the regression. Let X_1 be the predictor of X_1^* generated by the procedure described above. In the terminology of EIV models, X_1 is the observed X_1^*. Then, $X_1 = P_w X$, where $P_w = W(W'W)^{-1}W'$. Substituting for X we obtain

$$X_1 = X_1^* + P_w X_2^* - (I - P_w)Z\delta = X_1^* + v$$

$$X_2 = X - X_1 = X_2^* - v.$$

The error, v, is itself composed of two parts. First, there is an error due to our ignorance of the true parameter α and using its OLS estimator in its place, $P_w X_2^*$. Secondly, there is an error due to our omission of Z in the first stage regression, $-(I - P_w)Z\delta$. The first part is always present, but does not lead to inconsistencies as the sample size becomes infinite, whereas the second part constitutes a misspecification error that may have serious consequences.

The error v does not behave like a typical measurement error. Assume, without loss of generality, that W and Z have zero mean. Then, v has zero mean. Its (asymptotic) variance is given by $\sigma_{vv} = \text{plim } v'v/N = \text{plim } (Z\delta)'(I - P_w)Z\delta/N$, since X_2^* is orthogonal to W by assumption. Since v is a function of Z, it is not orthogonal to the true variable X_1^*. In fact, v and X_1^* are negatively correlated with $\text{plim } X_1^{*\prime} v/N = -\sigma_{vv}$. However, since X_2^* is orthogonal to W and to Z, $\text{plim } X_2^{*\prime} v/N = 0$.

Hence, this is not the classical EIV case since orthogonality is required, not only between the true components, but also between the observed (predicted) compo-

[18] That is, if $X = EX + u$, where EX is the linear projection of X on W and Z, then $X_1^* = EX = W\alpha + Z\delta$, and $X_2^* = u$.

nents, i.e., plim $X_1'X_2/N = -(\text{plim } X_1^{*\prime}v/N + \text{plim } v'v/N) = 0$. This can only occur when one of the true variables is correlated with the measurement error. The regression actually being estimated is

$$y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon + (\beta_2 - \beta_1)v.$$

Note that plim $X_1'v/N = 0$ and recall that X_1 and X_2 are orthogonal by construction. These two facts imply that b_1 is a consistent estimator of β_1. Similarly, plim $X_2'v/N = -\sigma_{vv}$, implying plim $b_2 = \beta_2 - (\beta_2 - \beta_1)\sigma_{vv}/(\text{plim } X_2'X_2/N)$. As in previous sections, whether β_2 is under- or over-estimated depends on the sign of $\beta_2 - \beta_1$, but the latter can be consistently estimated since plim $(b_2 - b_1) = (\beta_2 - \beta_1)\theta$, where $0 \leq \theta = (\text{plim } X_2^{*\prime}X_2^*/N)/(\text{plim } X_2'X_2/N) \leq 1$.

Finally, bounds for β_2 are obtained by proceeding as outlined in Section 2.2, while tests of the hypothesis of equality between the effects of expected and unexpected components is carried out in the standard way since, under H_0, $\beta_1 - \beta_2 = 0$.

4. CONCLUSION

This paper analyzed the bias in OLS estimators of parameters corresponding to two regressors measured with errors that are *not* independent. Such a case may arise when a variable is decomposed into two components with the purpose of estimating and testing their differential impact on the dependent variable. If only the total variable and one of its components are observed, and if that component is measured with error, the induced measurement error in the other component is identical but with the opposite sign. Thus, the measurement errors are not only nonindependent but also perfectly negatively correlated. It is argued that this practice is not an uncommon one in applied work, and is common enough to warrant closer examination; several examples are presented in the introduction.

The overall conclusion is that the direction of the bias in the OLS estimators can be assessed from the observed data, i.e., it can be determined which estimator is over- or underestimated without resorting to extraneous information on the (relative) variances of the measurement errors. This is not the case in the classical EIV model with uncorrelated measurement errors. For example, if the covariance between the observed components is not too negative, less than the observed variances in absolute value, the biases in each component's OLS estimator have opposite signs: the smallest parameter is biased upwards, while the largest one is biased downwards. The difference between the OLS estimators, although itself underestimated, can identify the ranking of the true parameters, thereby enabling us to determine which component's parameter is over- or underestimated. In addition, with the help of a "reverse" regression, the components' parameters can very easily be bounded. Finally, the test for equality among the coefficients of the two components proceeds in the standard manner, using an F or χ^2 test, as if there were no errors in variables. In fact, under the null hypothesis the measurement errors cancel out. If the data reject this hypothesis then, in most plausible cases, the

largest parameter estimator's is biased downward while the smallest is biased upward.

These conclusions carry over to the case when additional, correctly measured exogenous variables are added to the regression.

The case in which an explanatory variable is decomposed into two orthogonal and unobservable components, such as expected and unexpected parts, is particularly interesting, although it does not fit into the EIV framework: an omitted variable misspecification in the estimation of the expected part produces a similar misspecification, but with the opposite sign in the generated unexpected part. However, this misspecification only biases the estimation of the parameter corresponding to the unexpected component. Although this bias can be positive or negative, its sign can be consistently estimated from the data. The parameter can also be bounded. Hypothesis testing of equality among coefficients is not affected by the presence of this type of misspecification error.

The results presented in this paper complement and modify in interesting ways the conclusions reached for the classical EIV model with uncorrelated measurement errors. Because of the particular structure of dependence among the errors, the results derived here are sharper and, hopefully, of some applicability.

The Hebrew University of Jerusalem, Israel and National Bureau of Economic Research, U.S.A.

APPENDIX

Section 2.1. The Ω Matrix. Let $W = (1, X_1, X_2)'$. Then,

$$\Omega \equiv \text{plim } W'W/N = \begin{bmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \omega_{11} & \omega_{12} \\ \mu_2 & \omega_{21} & \omega_{22} \end{bmatrix} = \begin{bmatrix} 1 & \mu_1 & \mu_2 \\ \mu_1 & \omega^*_{11} + \sigma_{vv} & \omega^*_{12} - \sigma_{vv} \\ \mu_2 & \omega^*_{21} - \sigma_{vv} & \omega^*_{22} + \sigma_{vv} \end{bmatrix}$$

where the (nonstarred) starred ω's are the expected values of the cross-products of the (observed) true variables, and the μ's are the expected values of both observed true variables.

The inverse and determinant of Ω are

$$\Omega^{-1} = \frac{1}{|\Omega|} \begin{bmatrix} \omega_{11}\omega_{22} - \omega_{12}^2 & \mu_2\omega_{12} - \mu_1\omega_{22} & \mu_1\omega_{12} - \mu_2\omega_{11} \\ \mu_2\omega_{12} - \mu_1\omega_{22} & \sigma_{22} & -\sigma_{12} \\ \mu_1\omega_{12} - \mu_2\omega_{11} & -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

$$|\Omega| = \omega_{11}\omega_{22} + 2\mu_1\mu_2\omega_{12} - \mu_1^2\omega_{22} - \mu_2^2\omega_{11} - \omega_{12}^2 = \sigma_{11}\sigma_{22} - \sigma_{12}^2,$$

where $\omega_{ij} = EX_iX_j$, and $\sigma_{ij} = \omega_{ij} - \mu_i\mu_j$.

Section 2.2. Reverse Regression. Consider the reverse regression $X_2 = \alpha_0 + \alpha_1 X + \alpha_2 y + \eta$, where $\alpha_1 = \beta_1(\beta_1 - \beta_2)^{-1}$, $\alpha_2 = (\beta_2 - \beta_1)^{-1}$ and $\eta = -\varepsilon(\beta_2 - \beta_1)^{-1} - v$. Let $W' = (1, X, y)'$ and note that plim $W'\eta/N = (0, 0, -\sigma_{\varepsilon\varepsilon}\alpha_2)'$.

The (3, 2) and (3, 3) entries in the inverse of the limiting cross-product matrix of W are $-\text{cov}(X, y)/\psi$ and $\text{var}(X)/\psi$ where $\psi = \text{var}(X)\ \text{var}(y) - \text{cov}(X, y)^2$. This implies that the plim of a_1 and a_2, the OLS estimators of α_1 and α_2, are

$$\text{plim}\ a_1 = \frac{\beta_1}{\beta_1 - \beta_2} + \frac{\sigma_{\varepsilon\varepsilon}\ \text{cov}(X, y)}{\psi(\beta_2 - \beta_1)}$$

$$\text{plim}\ a_2 = \frac{1}{\beta_2 - \beta_1} - \frac{\sigma_{\varepsilon\varepsilon}\ \text{var}(X)}{\psi(\beta_2 - \beta_1)}$$

$$\text{plim} - \frac{a_1}{a_2} = \frac{\beta_1\psi - \sigma_{\varepsilon\varepsilon}\ \text{cov}(X, y)}{\psi - \sigma_{\varepsilon\varepsilon}\ \text{var}(X)}.$$

Since $\text{cov}(X, y) = \beta_1\ \text{var}(X) + (\beta_2 - \beta_1)\ \text{cov}(X, X2)$ equation (3), in the text, follows. To show that the denominator of the second term in the right-hand side of equation (3) is positive observe that $\psi - \sigma_{\varepsilon\varepsilon}\ \text{var}(X) = \text{var}(X)\ \text{var}(y - \varepsilon) - \text{cov}(X, y - \varepsilon)^2 = (1 - r^2_{X,y-\varepsilon})V(X)V(y - \varepsilon) \geq 0$, where $r^2_{X,y-\varepsilon}$ is the square of the simple correlation coefficient between X and $y - \varepsilon$.

The Case of Arbitrarily Correlated Measurement Errors. The general case in which the measurement errors are correlated, but not perfectly so, generates less clear-cut results. Let ξ and η be the measurement errors in X_1 and X_2 respectively, with variances $\sigma_{\xi\xi}$ and $\sigma_{\eta\eta}$ and covariance $\sigma_{\xi\eta}$. In Section 2.1 we have $\sigma_{\xi\xi} = \sigma_{\eta\eta} = -\sigma_{\xi\eta} = \sigma_{vv}$. Proceeding in a similar way as before, it can be shown that

$$\text{plim}\ b_1 = \beta_1 - \{\beta_1(\sigma_{22}\sigma_{\xi\xi} - \sigma_{12}\sigma_{\xi\eta}) + \beta_2(\sigma_{22}\sigma_{\xi\eta} - \sigma_{12}\sigma_{\eta\eta})\}|\Omega|^{-1}$$

$$\text{plim}\ b_2 = \beta_2 - \{\beta_2(\sigma_{11}\sigma_{\eta\eta} - \sigma_{12}\sigma_{\xi\eta}) + \beta_1(\sigma_{11}\sigma_{\xi\eta} - \sigma_{12}\sigma_{\xi\xi})\}|\Omega|^{-1}.$$

The resulting biases cannot be signed without additional information on the variances and covariances among the observed variables and among the measurement errors, and on the magnitudes of the true parameters.[19]

REFERENCES

BLANCHARD, O., C. RHEE AND L. SUMMERS, "The Stock Market, Profit and Investment," Working Paper No. 3370, National Bureau of Economic Research, 1990.

DUNCAN, G. AND D. HILL, "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data," *Journal of Labor Economics* 3 (1985), 508–532.

FULLER, W., *Measurement Error Models* (New York: John Wiley & Sons, 1987).

GALEOTTI, M. AND F. SCHIANTARELLI, "Stock Market Volatility and Investment: Do Only Fundamentals Matter?" Working Paper No. 90-15, New York University C. V. Starr Center for Applied Economic Research, 1990.

GARBER, S. AND S. KLEPPER, "Extending the Classical Errors-in-Variables Model," *Econometrica* 48 (1980), 1541–1546.

GRILICHES, Z., "Economic Data Issues," in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Vol. 3 (Amsterdam: North-Holland, 1986), 1465–1514.

HAITOVSKY, Y., "On Errors of Measurements in Regression Analysis in Economics," *International*

[19] The bias of b_1 reported in equation (4.8) of Griliches (1986, p. 1479) results when $\sigma_{11} = \sigma_{22}$ and $\sigma_{\xi\eta} = 0$.

Statistical Review 40 (1972), 23–35.

KLEPPER, S. AND E. LEAMER, "Consistent Sets of Estimates for Regressions with Errors in All Variables," *Econometrica* 52 (1984), 163–183.

LACH, S., "Decomposition of Variables and Corrleated Measurement Errors," Working Paper No. 246, The Hebrew University of Jerusalem, 1992.

——— AND D. TSIDDON, "The Behavior of Prices and Inflation: An Empirical Analysis of Disaggregated Price Data," *Journal of Political Economy* 100 (1992), 349–389.

LICHTENBERG, F., "Aggregation of Variables in Least Squares Regression," *The American Statistician* 44 (1990), 169–171.

MANSFIELD, E., "Basic Research and Productivity Increase in Manufacturing," *American Economic Review* 70(1980), 863–873.

MISHKIN, F., "Does Anticipatory Monetary Policy Matter? An Econometric Investigation," *Journal of Political Economy* 90(1982), 22–51.

PAGAN, A., "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review* 25 (1984), 221–247.

SICHERMAN, N., "The Measurement of On-the-Job Training," *Journal of Economic and Social Measurement* 16 (1990), 221–230.