



Federal Reserve Bank of Chicago

**A Note on the Estimation of Linear
Regression Models with Heteroskedastic
Measurement Errors**

By: Daniel G. Sullivan

WP 2001-23

A Note on the Estimation of Linear Regression Models with
Heteroskedastic Measurement Errors

Daniel G. Sullivan
Federal Reserve Bank of Chicago

December, 2001

The views expressed in this paper are solely those of the author and are not official positions of the Federal Reserve Bank of Chicago or the Federal Reserve System. Thanks are owed to Ken Housinger for very capable assistance and to Federal Reserve Bank of Chicago Micro Lunch participants and especially Dan Aaronson and Joe Altonji for helpful discussions.

Abstract

I consider the estimation of linear regression models when the independent variables are measured with errors whose variances differ across observations, a situation that arises, for example, when the explanatory variables in a regression model are estimates of population parameters based on samples of varying sizes. Replacing the error variance that is assumed common to all observations in the standard errors-in-variables estimator by the mean measurement error variance yields a consistent estimator in the case of measurement error heteroscedasticity. However, another estimator, which I call the Heteroskedastic Errors in Variables Estimator (HEIV), is, under standard assumptions, asymptotically more efficient. Simulations show that the efficiency gains are likely to be appreciable in practice. In addition, the HEIV estimator, which is the ordinary least squares regression of the dependent variable on the best linear predictor of the true independent variables, is simple to compute with standard regression software.

I. Introduction

It is well known that when the independent variables in a regression model are measured with error, the ordinary least squares (OLS) estimator is biased and inconsistent. For example, with a single regressor, the OLS estimator of the regression coefficient tends in probability to the product of the true coefficient and the reliability ratio of the regressor – the latter quantity being the ratio of the variance of the true explanatory variable to the total variance of the measured variable.

Textbooks¹ explain, however, that if the (assumed constant) variance of those errors is known, a consistent estimator can easily be obtained. For example, with a single regressor, the errors-in-variables (EIV) estimator obtained by dividing the ordinary least squares estimate by the reliability ratio is a consistent estimator of the true coefficient.

Perhaps the most common situation in which a researcher actually knows the variance of the measurement errors in a variable, and thus is in a position to use the EIV estimator, is when the variable in question is obtained as the result of an earlier statistical procedure. For example, a regression analysis might relate a dependent variable for a geographic region to the population mean of some other characteristic of the region. If the population mean of the characteristic is unknown, it is common practice to replace it with an estimate based on a finite sample. Relative to the true population mean, this sample estimate will be measured with error, and because the variance of that error can often be obtained from sampling theory, the EIV estimator may be applicable. However, in many, if not most, such examples, the variance of the measurement errors will be known to vary by observation. For instance, when the observations in the regression correspond to geographic regions, it will often be the case that the available samples are larger for more populous regions, and thus that the sampling errors will be larger for small regions.

Numerous applied studies fit into this category. For example, Blanchflower and Oswald (1994), Card and Hyslop (1997) and Blanchard and Katz (1997) relate state-level wage levels or wage growth to state-level unemployment rates where both variables are obtained from subsamples of the Current Population Survey. Aaronson and Sullivan (1999) add a measure of job displacement rates taken from the Displaced Worker Supplements as another independent variable. As another

1. See, for example, Greene (1997).

example, Campbell and Hopenhym (2001) study the dependence of the size distribution of retail establishments across cities on a number of variables, including median rents for commercial real estate, a variable they construct from samples whose size varies across cities. In each of these cases independent variables are better measured for large states or cities than for smaller ones. This paper analyzes this frequently occurring situation, which does not appear to have been formally treated in the literature.

As is shown below, simply replacing the assumed-constant measurement error variance in the standard EIV estimator by the mean error variance across observations results in a consistent estimator in the case of heteroskedastic error variances. Such an estimator has, in fact, been used by researchers.² However, it is also shown that another estimator, that more fully takes account of the varying levels of information in the observations, is, under standard assumptions, asymptotically more efficient. This estimator replaces the error-ridden independent variables in the OLS regression with their best linear predictor based on the observed data, the coefficients of which vary with the extent of measurement error. Simulations suggest that the reductions in variance may be appreciable in practice. The alternative estimator is, moreover, straightforward to compute using standard software packages.

The next section of this paper describes the model and motivates the estimators for the case of a single regressor. In section III, I show that under some standard assumptions, the HEIV estimator is more asymptotically efficient than the EIV estimator. Section IV presents some simulations that suggest the gains in efficiency could be large in practice and that the asymptotic results provide a reasonable approximation to what would be found in finite samples. Section V shows how the analysis can be extended to the case of multiple regression and details how one can compute the estimators using standard software. Finally, some brief conclusions are contained in Section VI.

II. Model and Motivation of Estimators

Consider first the case of a true regression model with a single regressor:

2. Aaronson and Sullivan (1999) is an example.

$$(1) \quad y_i = x_i^* \beta + \varepsilon_i ,$$

where x_i^* is the true, but unobserved, independent variable, assumed here to have zero mean and variance $\omega^2 > 0$ and ε_i is a disturbance term, assumed to have zero mean and variance σ_i^2 and to be uncorrelated with x_i^* . The observed, but error-ridden, explanatory variable is given by

$$(2) \quad x_i = x_i^* + \eta_i ,$$

where η_i is mean zero with variance τ_i^2 and is uncorrelated with both x_i^* and ε_i .

To allow for reasonable forms of heteroskedasticity while keeping the asymptotic analysis simple, assume that $\varepsilon_i = \sigma_i \tilde{\varepsilon}_i$ and $\eta_i = \tau_i \tilde{\eta}_i$, where $\tilde{\varepsilon}_i$ and $\tilde{\eta}_i$ have unit variances and that $(x_i^*, \sigma_i^2, \tau_i^2, \tilde{\varepsilon}_i, \tilde{\eta}_i)$ for $i = 1, \dots, n$ are independent and identically distributed. Also, for a given i , assume that $\tilde{\varepsilon}_i$ and $\tilde{\eta}_i$ are independent of each other and of x_i^* , σ_i^2 , and τ_i^2 . I will also need to assume that $(x_i^*, \sigma_i^2, \tau_i^2, \tilde{\varepsilon}_i, \tilde{\eta}_i)$ has finite fourth moments.

The assumptions above imply a relationship between y_i and x_i with a composite error term,

$$(3) \quad y_i = x_i \beta + [\varepsilon_i + (x_i^* - x_i) \beta] .$$

Because $\varepsilon_i + (x_i^* - x_i) \beta = \varepsilon_i - \eta_i \beta$ is correlated with $x_i = x_i^* + \eta_i$, the OLS estimator

$$(4) \quad \hat{\beta}_{ols} = \frac{n^{-1} \sum x_i y_i}{n^{-1} \sum x_i^2}$$

will be biased and inconsistent. Indeed, substituting (1) and (2) into (4), appealing to the law of large numbers, and letting $E\tau_i^2 = \tau^2$,

$$(5) \quad \text{plim} \hat{\beta}_{ols} = \frac{E[(x_i^* + \eta_i)(x_i^* \beta + \varepsilon_i)]}{E[(x_i^* + \eta_i)^2]} = \frac{\omega^2}{\omega^2 + \tau^2} \beta,$$

so that the OLS estimator is asymptotically biased towards zero.

If ω^2 and τ^2 are known, the obvious generalization of the standard EIV estimator is

$$(6) \quad \hat{\beta}_{eiv} = \frac{n^{-1} \sum x_i y_i}{r n^{-1} \sum x_i^2},$$

where $r = \frac{\omega^2}{\omega^2 + \tau^2}$ is the analog of the standard reliability ratio in the standard model in which

the measurement error variance is τ^2 for all observations. Multiplying the denominator of the OLS estimator by r scales its probability limit down to the “correct” value of ω^2 and thus yields a consistent estimator. The only subtlety involved in this generalization relative to the standard case of a constant measurement error variance is that one needs to substitute the mean value of the error variances in the standard estimator, rather than, for example, the mean value of the observation-specific reliability ratio, $r_i = \frac{\omega^2}{\omega^2 + \tau_i^2}$.

The EIV estimator is usually motivated, as above, as adjusting the sample variance of the independent variable so as to match what is required for the population moment condition for β in terms of the correctly measured variables (i.e. $E x_i^* y_i = \beta E x_i^{*2}$). However, the standard EIV estimator can also be viewed in another way that suggests what turns out to be, under standard assumptions, an asymptotically more efficient estimator. Specifically, the EIV estimator can be viewed as the OLS regression of the dependent variable on a predicted value for the true independent variable given the observed variable. Indeed, r is the value of κ than minimizes the unconditional expectation $E[(x_i^* - \kappa x_i)^2]$. Moreover, the regression of y_i on the linear predictor $r x_i$,

$\frac{n^{-1} \sum r x_i y_i}{n^{-1} \sum (r x_i)^2}$, evidently reduces to the EIV estimator. The corresponding regression model,

$$(7) \quad y_i = r x_i \beta + [\varepsilon_i + (x_i^* - r x_i) \beta] ,$$

also has a composite error term. But in contrast to (3), both parts of the error term in (7) are uncorrelated with the regressor. Indeed, $E[r x_i (x_i^* - r x_i)] = 0$ is the normal equation for the prediction problem that can serve to define r . Thus, as long as ω^2 is positive, the EIV estimator will be consistent.

When the measurement error variances vary across observations, one can better predict the true x_i^* by taking that variation into account. Specifically, the best linear (in x_i) predictor given the available information is $r_i x_i$, the product of the observed data and the observation-specific reliability ratio. That is, r_i is the value of κ that minimizes the conditional expectation

$E[(x_i^* - \kappa x_i)^2 | \tau_i^2]$. This suggests an alternative Heteroskedastic Errors in Variables (HEIV) estimator that is the focus of this paper. Specifically, the HEIV estimator is the OLS regression of the dependent variable on $r_i x_i$, the best linear predictor of x_i^* given the observed data and the actual measurement error variance for the observation:

$$(8) \quad \hat{\beta}_{heiv} = \frac{n^{-1} \sum r_i x_i y_i}{n^{-1} \sum (r_i x_i)^2} .$$

The regression underlying (8),

$$(9) \quad y_i = r_i x_i \beta + [\varepsilon_i + (x_i^* - r_i x_i) \beta] ,$$

again has a composite error term. The lack of correlation between the error term and the regressor follows in this case from the iterated expectations identity,

$$E[r_i x_i (x_i^* - r_i x_i)] = E[E[r_i x_i (x_i^* - r_i x_i) | \tau_i^2]] \text{ and the fact that } E[r_i x_i (x_i^* - r_i x_i) | \tau_i^2] = 0 \text{ for}$$

all τ_i^2 , which again is a first order condition that could be used to define the r_i . Thus $\hat{\beta}_{heiv}$ will also be consistent as long as there is sufficient variation in x_i^* .

Moreover, there is reason to expect that $\hat{\beta}_{heiv}$ will be more efficient than $\hat{\beta}_{eiv}$ since (9) leaves more variation in the regressor than (7). That is

$$Var[rx_i] = r^2 Var[x_i^* + \eta_i] = r^2(\omega^2 + \tau^2) = r\omega^2, \text{ while}$$

$$Var[r_i x_i] = E[E[r_i \omega^2 | \tau_i^2]] = E[r_i] \omega^2. \text{ Because } r_i = \frac{\omega^2}{\omega^2 + \tau_i^2} \text{ is a strictly convex function of}$$

$$\tau_i^2 \text{ and } r = \frac{\omega^2}{\omega^2 + E[\tau_i^2]}, \text{ Jensen's inequality implies that } E[r_i] \geq r \text{ and thus that}$$

$$Var[r_i x_i] \geq Var[rx_i]. \text{ The inequality will be strict as long as there is some dispersion in the } \tau_i^2 \text{ s.}$$

Alternatively, the variance of the second component of the error term in (9) is smaller than that of the corresponding term in (7): $Var[(x_i^* - r_i x_i)\beta] = (1 - E[r_i])\omega^2\beta^2$ while

$$Var[(x_i^* - rx_i)\beta] = (1 - r)\omega^2\beta^2, \text{ implying that } Var[(x_i^* - r_i x_i)\beta] < Var[(x_i^* - rx_i)\beta]. \text{ Thus,}$$

relative to model (7), model (9) has higher variance in its independent variable and lower variance in its disturbance, suggesting that it will yield more precise estimates of parameters.

III. Asymptotic Comparison of Estimators

In this section, I compare the asymptotic variances of the EIV and HEIV estimators, showing that under some standard assumptions, the asymptotic variance of the HEIV estimator is lower. The two estimators have a similar structure. In particular,

$$(10) \quad \hat{\beta}_{eiv} = \beta + \frac{n^{-1} \sum rx_i \varepsilon_i}{n^{-1} \sum (rx_i)^2} + \frac{n^{-1} \sum rx_i (x_i^* - rx_i)}{n^{-1} \sum (rx_i)^2} \beta$$

while

$$(11) \quad \hat{\beta}_{heiv} = \beta + \frac{n^{-1} \sum r_i x_i \varepsilon_i}{n^{-1} \sum (r_i x_i)^2} + \frac{n^{-1} \sum r_i x_i (x_i^* - r_i x_i)}{n^{-1} \sum (r_i x_i)^2} \beta.$$

Thus, given the assumption of an i.i.d. data generating process, deriving the asymptotic distribution is straightforward. In particular, by the law of large numbers, $n^{-1} \sum (rx_i)^2$ and $n^{-1} \sum (r_i x_i)^2$ converge in probability to, respectively, $E[(rx_i)^2] = r\omega^2$ and $E[(r_i x_i)^2] = E[r_i]\omega^2$. And, by the central limit theorem, $\sqrt{nn^{-1}} \sum rx_i \varepsilon_i$, $\sqrt{nn^{-1}} \sum r_i x_i \varepsilon_i$, $\sqrt{nn^{-1}} \sum rx_i (x_i^* - rx_i)$, and $\sqrt{nn^{-1}} \sum r_i x_i (x_i^* - r_i x_i)$ tend in distribution to Gaussian random variables with mean zero and variances, respectively, of $Var[rx_i \varepsilon_i]$, $Var[r_i x_i \varepsilon_i]$, $Var[rx_i (x_i^* - rx_i)]$, and $Var[r_i x_i (x_i^* - r_i x_i)]$, provided those exist, which they will given that x_i^* and η_i are assumed to have finite fourth moments. Thus $\sqrt{n}(\hat{\beta}_{eiv} - \beta)$ and $\sqrt{n}(\hat{\beta}_{heiv} - \beta)$ tend to Gaussian variables with mean zero and variances, respectively, of

$$(12) \quad \Lambda_{eiv} = \Lambda_{eiv}^s + \Lambda_{eiv}^p = \frac{Var[rx_i \varepsilon_i]}{(r\omega^2)^2} + \frac{Var[rx_i (x_i^* - rx_i)]}{(r\omega^2)^2} \beta^2$$

and

$$(13) \quad \Lambda_{heiv} = \Lambda_{heiv}^s + \Lambda_{heiv}^p = \frac{Var[r_i x_i \varepsilon_i]}{(E[r_i]\omega^2)^2} + \frac{Var[r_i x_i (x_i^* - r_i x_i)]}{(E[r_i]\omega^2)^2} \beta^2.^3$$

Thus the asymptotic variance of each estimator has two components. The first component of the asymptotic variance, which I call the structural component because its source is the error term ε_i in the true equation, is denoted above as Λ_{eiv}^s for the EIV estimator and as Λ_{heiv}^s for the HEIV estimator. These terms have a form typical of regression models and that does not depend on the regression coefficient. The second component of the asymptotic variance, which I call the predic-

3. The cross product terms converge to zero.

tion component because its source is the imperfection of the linear predictor for the true independent variable, x_i^* , is denoted above as Λ_{eiv}^P for the EIV estimator and as Λ_{heiv}^P for the HEIV estimator. These have a less standard form, depending, in particular, on the true parameter value. Evidently, the prediction components are relatively more important when $|\beta|$ is larger. I consider in turn the assumptions and arguments needed to compare Λ_{eiv}^S to Λ_{heiv}^S and Λ_{eiv}^P to Λ_{heiv}^P .

Structural Error Components

The relative sizes of Λ_{eiv}^S and Λ_{heiv}^S , the portions of the asymptotic variances due to the presence of the structural error term, ε_i , will, in general depend on the nature of any heteroskedasticity in ε_i . But, given the standard assumptions of homoskedasticity and independence of x_i^* and τ_i^2 , $Var[rx_i\varepsilon_i] = \sigma^2 Var[rx_i] = r\omega^2\sigma^2$ where σ^2 is the common variance of the ε_i , while $Var[r_ix_i\varepsilon_i] = E[r_i]\omega^2\sigma^2$. Thus, $\Lambda_{eiv}^S = \sigma^2/r\omega^2$ while $\Lambda_{heiv}^S = \sigma^2/E[r_i]\omega^2$, which establishes the following proposition.

Proposition 1. If ε_i , the error term in the true regression equation has constant variance and is independent of x_i^* and τ_i^2 , the true independent variable and its measurement error, then

$$\Lambda_{heiv}^S/\Lambda_{eiv}^S = r/E[r_i] \leq 1.$$

With arbitrarily malevolent forms of heteroskedasticity, it is possible for the ratio of $Var[r_ix_i\varepsilon_i]$ to $Var[rx_i\varepsilon_i]$ to be larger than $(E[r_i]/r)^2$, resulting $\Lambda_{heiv}^S > \Lambda_{eiv}^S$. This would be the case, for example, if σ_i^2 was proportional to $(r_ix_i)^2$. But this is not what one would expect in practice. Indeed, it is much more likely that σ_i^2 would be *negatively* correlated with the reliability ratio. For instance, if the dependent variable was itself a sample estimate of a population quantity, a portion of its variance would be due to a measurement error whose variance would likely be about

proportional to τ_i^2 . Thus Proposition 1 may be a lower bound on the improvement to be expected in practice from using the HEIV rather than EIV estimator.

Prediction Error Component

It is also very likely that $\Lambda_{heiv}^P < \Lambda_{eiv}^P$, the portion of the asymptotic variance due to the deviations of the linear predictors from the true independent variable is lower for the HEIV estimator than for the EIV estimator. First, as with Λ_{eiv}^S and Λ_{heiv}^S , the denominator of Λ_{heiv}^P is larger than that of Λ_{eiv}^S . Moreover, it appears that in most cases of practical interest, the numerator is also smaller. Indeed, $Var[rx_i(x_i^* - rx_i)] = E[Var[rx_i(x_i^* - rx_i)|\tau_i^2]] + Var[E[rx_i(x_i^* - rx_i)|\tau_i^2]]$ while $Var[r_ix_i(x_i^* - r_ix_i)]$ is simply $E[Var[r_ix_i(x_i^* - r_ix_i)|\tau_i^2]]$, since $E[r_ix_i(x_i^* - r_ix_i)|\tau_i^2] = 0$ for all τ_i^2 . Thus one would expect that in most cases $Var[r_ix_i(x_i^* - r_ix_i)]$ will be less than $Var[rx_i(x_i^* - rx_i)]$.

It is not, however, necessary that $Var[r_ix_i(x_i^* - r_ix_i)] < Var[rx_i(x_i^* - rx_i)]$. In fact, when the distribution of τ_i^2 is weighted heavily towards values that are large relative to ω^2 ,

$Var[r_ix_i(x_i^* - r_ix_i)]$ is typically greater than $Var[rx_i(x_i^* - rx_i)]$. However, in the cases I have examined, even when $Var[r_ix_i(x_i^* - r_ix_i)] > Var[rx_i(x_i^* - rx_i)]$, the ratio of

$Var[r_ix_i(x_i^* - r_ix_i)]$ to $Var[rx_i(x_i^* - rx_i)]$ has been less than $E[r_i]/r$, so that the ratio of Λ_{heiv}^P to Λ_{eiv}^P remains less than $r/E[r_i]$.

The most straightforward case to analyze is when x_i^* and η_i have Gaussian distributions, in which case the following proposition can be established.

Proposition 2. If x_i^* and η_i , the true independent variable and its measurement error, have Gaussian distributions, then $\Lambda_{heiv}^p / \Lambda_{eiv}^p \leq r / E[r_i] \leq 1$.

To prove proposition 2, note that $r_i x_i (x_i^* - r_i x_i)$ is equal to $(r_i x_i^* + r_i \eta_i)((1 - r_i)x_i^* - r_i \eta_i)$, which can be expanded to give

$$(14) \quad E[r_i^2 x_i^2 (x_i^* - r_i x_i)^2 | \tau_i^2] = r_i(1 - r_i)E[x_i^{*4}] + r_i^2(1 - 6r_i + 6r_i^2)\omega^2 \tau_i^2 + r_i^4 E[\eta_i^4 | \tau_i^2].$$

With Gaussian x_i^* and η_i , $E[x_i^{*4}] = 3(\omega^2)^2$ and $E[\eta_i^4 | \tau_i^2] = 3(\tau_i^2)^2$. Substituting these expressions into (14) and simplifying yields

$$(15) \quad E[r_i^2 x_i^2 (x_i^* - r_i x_i)^2] = (\omega^2)^2 E[r_i(1 - r_i)].$$

Similarly, using the iterated expectations identity,

$$(16) \quad E[r^2 x_i^2 (x_i^* - r x_i)^2] = r(1 - r)E[x_i^{*4}] + r^2(1 - 6r + 6r^2)\omega^2 \tau^2 + r^4 E[E[\eta_i^4 | \tau_i^2]].$$

When η_i has a Gaussian distribution, $E[E[\eta_i^4 | \tau_i^2]] = E[3(\tau_i^2)^2]$ which is greater than or equal to $3(\tau^2)^2$ by Jensen's inequality. It follows that

$$(17) \quad E[r^2 x_i^2 (x_i^* - r x_i)^2] \geq 3r(1 - r)(\omega^2)^2 + r^2(1 - 6r + 6r^2)\omega^2 \tau^2 + 3r^4(\tau^2)^2,$$

and, thus, after some simplification, that

$$(18) \quad E[r^2 x_i^2 (x_i^* - r x_i)^2] \geq (\omega^2)^2 r(1 - r).$$

Comparing (15) and (18), it suffices to show that $\frac{E[r_i(1 - r_i)]}{(E[r_i])^2} \frac{r}{1 - r} \leq \frac{r}{E[r_i]}$ or

$$E[r_i(1 - r_i)] \leq (1 - r)E[r_i] \text{ which can be rewritten as } E\left[\frac{\omega^2 \tau_i^2}{(\omega^2 + \tau_i^2)^2} - \frac{\tau^2}{(\omega^2 + \tau^2)(\omega^2 + \tau_i^2)}\right] \leq 0.$$

Putting the fractions over a common denominator, simplifying and dropping multiplicative constants, this is equivalent to

$$(19) \quad E\left[\frac{\tau_i^2 - \tau^2}{(\omega^2 + \tau_i^2)^2}\right] \leq 0.$$

To see that (19) must hold, define $h(\tau_i^2) = \frac{\tau_i^2 - \tau^2}{(\omega^2 + \tau_i^2)^2}$ if $\tau_i^2 < \omega^2 + 2\tau^2$ and

$h(\tau_i^2) = \frac{1}{4(\omega^2 + \tau^2)}$ if $\tau_i^2 \geq \omega^2 + 2\tau^2$.⁴ Then $\frac{\tau_i^2 - \tau^2}{(\omega^2 + \tau_i^2)^2} \leq h(\tau_i^2)$ for all τ_i^2 . So

$$E\left[\frac{\tau_i^2 - \tau^2}{(\omega^2 + \tau_i^2)^2}\right] \leq E[h(\tau_i^2)]. \text{ But } h(\tau_i^2) \text{ is concave, so } E[h(\tau_i^2)] \leq h(\tau^2) = 0, \text{ which completes}$$

the proof of Proposition 2.

The assumption that the true independent variable and measurement errors have Gaussian distributions does not seem overly strong in the context of the likely applications of the HEIV estimator. The independent variables in those applications tend to be continuous measures, such as the unemployment rate for a geographic region, which, at least after a suitable transformation, have distributions that appear approximately Gaussian. Moreover, in most instances, the measurement error is the result of sampling variability in an estimator that is at least asymptotically Gaussian.

In a number of examples I have examined, the bound $\Lambda_{heiv}^p / \Lambda_{eiv}^p \leq r / E[r_i]$ is not particularly sharp in that the ratio of variances is often considerably less than $r / E[r_i]$. For instance, when, as often seems to be the case, $Var[r_i x_i (x_i^* - r_i x_i)] < Var[r x_i (x_i^* - r x_i)]$,

$\Lambda_{heiv}^e / \Lambda_{eiv}^e \leq (r / (E[r_i]))^2$. In the case of Gaussian x_i^* and η_i , a sufficient condition for

4. The function $h(\tau_i^2)$ is equal to $\frac{\tau_i^2 - \tau^2}{(\omega^2 + \tau_i^2)^2}$ up to the point at which the latter reaches its maximum value and is equal to that maximum value for higher values of τ_i^2 .

$Var[r_i x_i(x_i^* - r_i x_i)] < Var[rx_i(x_i^* - rx_i)]$ is that the support of τ_i^2 is entirely contained in the interval $[0, 2\omega^2)$ – that is, the reliabilities are always greater than 1/3. To see why this is the case, note that, as was shown above, $Var[rx_i(x_i^* - rx_i)] \geq (\omega^2)^2 r(1 - r)$ and

$Var[r_i x_i(x_i^* - r_i x_i)] = (\omega^2)^2 E[r_i(1 - r_i)]$. Thus if $r_i(1 - r_i)$ were a concave function of τ_i^2 then $Var[r_i x_i(x_i^* - r_i x_i)] < Var[rx_i(x_i^* - rx_i)]$. In general, $f(\tau_i^2) = r_i(1 - r_i) = \frac{\omega^2 \tau_i^2}{(\omega^2 + \tau_i^2)^2}$ is

not a concave function of τ_i^2 , but over the relevant domain it may be. Figure 1 shows that $f(\tau_i^2)$ rises sharply from a value of zero when $\tau_i^2 = 0$ to its maximum value of $\omega^2/4$ when $\tau_i^2 = \omega^2$. It then slowly asymptotes to zero. The function is strictly concave for $\tau_i^2 < 2\omega^2$. Thus if the support of τ_i^2 is entirely contained in the interval $[0, 2\omega^2)$ – that is, if the reliability is always greater than 1/3 – then $Var[r_i x_i(x_i^* - r_i x_i)] < Var[rx_i(x_i^* - rx_i)]$. As noted, it would follow that

$$\Lambda_{heiv}^e / \Lambda_{eiv}^e \leq (r / (E[r_i]))^2.$$

It follows from Propositions 1 and 2 that if the structural error term is homoskedastic and the true independent variable and measurement error are Gaussian, then $\Lambda_{heiv} / \Lambda_{eiv}$, the ratio of the full asymptotic variance of the HEIV estimator to that of the EIV estimator, is less than or equal to $r / E[r_i]$, which will be less than one if there is any dispersion in the τ_i^2 .

Weighted HEIV estimator

Even when the variance of the structural portion of the error term is constant, the variance of the prediction component will vary by observation. Thus it may be possible to increase the efficiency of the HEIV estimator by computing a weighted version. Specifically, the variance of the error term for the HEIV regression (9) will be $\sigma^2 + Var[x_i^* - r_i x_i] \beta^2$ where σ^2 is the variance of the structural component of the error. Under the assumption that x_i^* and η_i have Gaussian distributions, $Var[x_i^* - r_i x_i]$ is given by (14). Thus for a given value of β , one can compute the optimal

weights, which are inversely proportional to $\sigma^2 + \text{Var}[x_i^* - r_i x_i] \beta^2$. Since β will initially be unknown, one would need to start from the unweighted estimator and iterate to obtain the final weighted HEIV estimator.

If one assumed that ε_i had a Gaussian distribution and that the distributions of x_i^* and η_i were jointly Gaussian, then such an algorithm would have the character of the EM algorithm of Dempster, Laird and Rubin (1977). Specifically, given those assumptions, the linear predictor of x_i^* given x_i would coincide with the conditional expectation that forms the basis of the typical ‘‘E’’ step of the EM algorithm, while the weighted least squares regression on the linear predictor would correspond to the typical ‘‘M’’ step in which the complete-data likelihood function is maximized. In simulations not reported below, I have found, as expected, that the weighted version of the HEIV estimator has modestly lower variance than the unweighted HEIV estimator analyzed above.

Estimating standard errors

In applications one needs to have estimates of Λ_{eiv} or Λ_{heiv} . For these, one can appeal to the results of White (1980). This requires strengthening slightly the assumptions on the existence of moments of x_i^* , ε_i , and η_i , so as to satisfy the assumptions of his Theorem 1. Given such assumptions, one can consistently estimate the asymptotic variance of the EIV estimator by

$$(20) \quad \hat{\Lambda}_{eiv} = \frac{n^{-1} \sum r^2 x_i^2 e_i^2}{(n^{-1} \sum (r x_i)^2)^2},$$

where $e_i = y_i - r x_i \hat{\beta}_{eiv}$ is the residual from the estimated version of (7). Similarly

$$(21) \quad \hat{\Lambda}_{heiv} = \frac{n^{-1} \sum r_i^2 x_i^2 e_i^2}{(n^{-1} \sum (r_i x_i)^2)^2},$$

where $e_i = y_i - r_i x_i \hat{\beta}_{heiv}$ is the residual from the estimated version of (9), is a consistent estimator of the asymptotic variance of the EIV estimator.

Estimating Reliability Ratios

Up to this point the observation-specific reliability ratios have been assumed known as has the reliability ratio corresponding to the mean level of measurement error. However, in most, if not all, of the examples that motivate this analysis, these will have to be estimated and used to construct “feasible” EIV or HEIV estimators. Reliability ratios depend on the levels of measurement error in the individual observations, which in the examples motivating this analysis will be delivered as part of a prior statistical analysis that also constructs the independent variable, x_i . The details of those calculations will vary from application to application and will not be considered here. However, in all cases, an estimate of ω^2 , the variance in the true explanatory variable also will be required.

The parameter ω^2 can be estimated in a number of ways. For example, $E[x_i^2 - \tau_i^2] = \omega^2$ for each i . Thus, $\hat{\omega}^2 = n^{-1} \sum (x_i^2 - \tau_i^2)$ is an unbiased and consistent estimator of ω^2 . However, $Var[x_i^2 - \tau_i^2]$ will not be constant across observations. In particular, observations for which τ_i^2 is large will also likely be ones for which $Var[x_i^2 - \tau_i^2]$ is high. So it is possible to estimate ω^2 more efficiently. For instance, in the case of Gaussian measurement error, one can show $Var[x_i^2 - \tau_i^2] = 2(\omega^2)^2 + 4\omega^2\tau_i^2 + 5(\tau_i^2)^2$. Thus, using weights proportional to $w_i = [2(\omega^2)^2 + 4\omega^2\tau_i^2 + 5(\tau_i^2)^2]^{-1}$, will yield a more efficient estimator, say $\hat{\omega}^2_w = \sum w_i(x_i^2 - \tau_i^2) / \sum w_i$. The weights used to construct $\hat{\omega}^2_w$ depend on ω^2 , but one may calculate the unweighted estimator to get an initial estimate of ω^2 and use that value to estimate an approximate set of weights.

IV. Simulations

This section quantifies the increase in asymptotic efficiency from using the HEIV rather than EIV estimator for a class of examples designed to correspond closely to conditions found in applications. It also shows that the asymptotic approximations are useful for samples of reasonable size.

The calculations are based on a set of distributions for τ_i^2 that are motivated by the likely applications of the estimators. In particular, I assume that the distribution of τ_i^2 is discrete with 50 points of equal mass at values given by γ/N_i $i = 1, \dots, 50$, where γ is a positive constant and the N_i are the levels of employment in the 50 states.⁵ The true, independent variable, x_i^* , and the structural disturbance, ε_i , are taken to have unit Gaussian distributions. As γ varies, $\tau^2 = E[\tau_i^2]$ can take on any positive value. Setting the variance of x_i^* to unity is just a normalization, since what matters for the calculations is τ^2/ω^2 or, equivalently, the reliability ratio corresponding to

$$\tau^2 = E[\tau_i^2], r = \frac{\omega^2}{(\omega^2 + \tau^2)}. \text{ Neither does the variance of } \varepsilon_i \text{ effect the ratios } \Lambda_{heiv}^s/\Lambda_{eiv}^s \text{ and } \Lambda_{heiv}^p/\Lambda_{eiv}^p.$$

Figure 2 shows the two ratios, $\Lambda_{heiv}^s/\Lambda_{eiv}^s$ and $\Lambda_{heiv}^p/\Lambda_{eiv}^p$, as functions of τ^2/ω^2 . When τ^2/ω^2 is small, say 0.1, so that the typical reliability ratio is above 0.9, the reduction in the portion of the variance deriving from the structural error term is slightly less than 1%. However, even when τ^2/ω^2 is only 0.1, the variance of the prediction component of the HEIV error term is less than 69% of the corresponding variance for the EIV estimator. Both ratios decline initially as the variance of the measurement error increases. When τ^2/ω^2 is equal to one, so that the typical reliability is 1/2, $\Lambda_{heiv}^s/\Lambda_{eiv}^s$ is about 84%, while $\Lambda_{heiv}^p/\Lambda_{eiv}^p$ is about 32%, both represent gains that could matter in practice. As τ^2/ω^2 increases further, $\Lambda_{heiv}^s/\Lambda_{eiv}^s$ continues to decline, while

5. Specifically, they are the average level of payroll employment for the 12 months of 1999.

$\Lambda_{heiv}^P/\Lambda_{eiv}^P$ reaches a minimum of a little less than 30% when τ^2/ω^2 is around two and a half, a point at which the typical reliability ratio is about 0.3. At this point, $\Lambda_{heiv}^S/\Lambda_{eiv}^S$ is under 70%.

The two lines eventually converge at a level of about 38.4%, but it is not until τ^2/ω^2 is over 100 that both ratios are within a percentage point of that level.

The overall variance ratio will lie between the two lines in Figure 2. It will be closer to

$\Lambda_{heiv}^S/\Lambda_{eiv}^S$ when $|\beta|$ is large and σ^2 is small and closer to $\Lambda_{heiv}^P/\Lambda_{eiv}^P$, when the opposite is true.

For the particular choice of $\beta = 1$ and $\sigma^2 = 1$, the overall ratio is about midway between the two lines. For instance when $\tau^2 = \omega^2$, the overall HEIV variance is about 65% that of the EIV estimator.

In the last section, the possibility that $Var[r_i x_i(x_i^* - r_i x_i)] > Var[rx_i(x_i^* - rx_i)]$ was discussed.

Figure 3 shows that this can, indeed, occur for high enough values of τ^2/ω^2 . In the figure, the solid line represents $Var[r_i x_i(x_i^* - r_i x_i)]$, while the dashed line represents $Var[rx_i(x_i^* - rx_i)]$.

For values of τ^2/ω^2 less than three or four, $Var[r_i x_i(x_i^* - r_i x_i)]$ is substantially less than

$Var[rx_i(x_i^* - rx_i)]$. But, when τ^2/ω^2 exceeds seven, $Var[rx_i(x_i^* - rx_i)]$ falls below

$Var[r_i x_i(x_i^* - r_i x_i)]$.

Table 1 shows how well the asymptotic distributions approximate the finite sample variances for the state-data simulation just described for the case in which $\tau^2 = \omega^2$ and $\beta = 1$. Data were generated for sample sizes of 100, 500, and 1000. Each such experiment was replicated 10,000 times. The table shows the means across these replications as well as their associated standard errors.

The first two rows of Table 1 show how well Λ_{eiv}^S and Λ_{heiv}^S do in approximating n times the finite-sample variances due to the structural error term. The next two rows do the same for the error term arising from the prediction error, while the next two rows show the total mean square

errors. The final two rows show how the robust variance estimators perform. In general, the asymptotic approximations are quite reasonable, even for samples of size 100. However, there is, for such samples, some tendency for the variance due to the structural error term to be somewhat higher than the asymptotic limit, both for the EIV and HEIV estimators. In addition, the variance estimator for the EIV estimator is biased downwards by about 5% for samples of size 100. But, by and large, the asymptotic results give a good indication of the finite sample behavior.

V. Multiple Regression

Most applied problems involve more than a single regressor. Nevertheless, I have avoided extending the analysis to multiple regression until this point because the single-variable results given above are simpler to understand and because most applications of which I am aware involve only a single independent variable with measurement error and the results above appear to give a good indication of the performance of the EIV and HEIV estimators for such models. In this section, I show how the EIV and HEIV estimators can be extended to multiple regression, including cases in which more than one variable is measured with error.

The model is still

$$(22) \quad y_i = x_i^* \beta + \varepsilon_i ,$$

where x_i^* is the true, but unobserved, independent variable, but now x_i^* is a vector of independent variables, with covariance matrix Ω . The disturbance term, ε_i , is again assumed to have zero mean and variance σ_i^2 and to be uncorrelated with x_i^* . The observed, but error-ridden, explanatory variable is given by

$$(23) \quad x_i = x_i^* + \eta_i ,$$

where η_i is now a vector of errors with mean zero and covariance matrix Υ_i and is uncorrelated with both x_i^* and ε_i . If $E[\Upsilon_i] = \Upsilon$, then the square matrix $R = (\Omega + \Upsilon)^{-1} \Omega$ is a multivariable extension of the reliability ratio based on the mean measurement error. In particular, the j th col-

umn of R is the value of the vector κ that minimizes the unconditional mean square error $E[(x_{ji}^* - x_i \kappa)^2]$, where x_{ji}^* is the j th component of the vector x_i^* .

The multivariable version of the EIV estimator is the OLS regression of y_i on $x_i R$, the linear prediction of x_i^* . That is,

$$(24) \quad \hat{\beta}_{eiv} = [\sum R x_i' x_i R]^{-1} (\sum R x_i' y_i) .$$

Assuming the inverses exist, $\hat{\beta}_{eiv}$ reduces to $[R \sum x_i' x_i]^{-1} (\sum x_i' y_i)$. The regression model underlying $\hat{\beta}_{eiv}$ is

$$(25) \quad y_i = x_i R \beta + [\varepsilon_i + (x_i^* - x_i R) \beta] .$$

If there is sufficient variation in the x_i^* , the consistency of $\hat{\beta}_{eiv}$ follows from the fact that all elements of $E[x_i R (x_i^* - x_i R)]$ are zero by the construction of R .

As in the single variable case, when there is heteroskedasticity in the η_i , the prediction of x_i^* can be improved by taking that heteroskedasticity into account. In particular, the vector consisting of the best linear predictors of the elements of x_i^* is $x_i R_i$ where $R_i = (\Omega + \Upsilon_i)^{-1} \Omega$. Thus the multivariable version of the HEIV estimator is

$$(26) \quad \hat{\beta}_{heiv} = [\sum R_i x_i' x_i R_i]^{-1} (\sum R_i x_i' y_i) ,$$

which is obtained from OLS estimation of

$$(27) \quad y_i = x_i R_i \beta + [\varepsilon_i + (x_i^* - x_i R_i) \beta] .$$

Again, the error term is uncorrelated with $x_i R_i$ by construction. Thus $\hat{\beta}_{heiv}$ provides consistent estimates of β as long as there is sufficient variation in x_i^* .

As in the single regressor case, the precision of estimates of β based on (27) should be higher than those based on (25) because there will be more variation in the independent variables and less variation in the error term.

Estimating Reliability Ratios

As in the single regressor case, it will ordinarily be necessary to estimate the reliability ratios that underlie the EIV and HEIV estimators. The estimation of the observation-specific measurement error variance, Υ_i , and the mean measurement error variance, Υ , will depend on the specifics of the application and will not be discussed here. Given these, however, it still will be necessary to estimate the covariance matrix of the true independent variables, Ω .

Given that $E[x_i'x_i - \Upsilon_i] = \Omega$, the unweighted estimator $\hat{\Omega} = n^{-1} \sum (x_i'x_i - \Upsilon_i)$ will be unbiased and consistent. However, the variances of the elements of $x_i'x_i - \Upsilon_i$ will vary by observation. In particular, if x_i^* and η_i have Gaussian distributions, then one can show that

$Var[x_{ij}x_{ik} - \Upsilon_{ijk}]$ is equal to $\Omega_{jj}\Omega_{kk} + \Omega_{jk}^2 + \Omega_{kk}\Upsilon_{ijj} + 2\Omega_{jk}\Upsilon_{ijk} + \Omega_{jj}\Upsilon_{ikk} + \Upsilon_{ijj}\Upsilon_{ikk} + \Upsilon_{ijk}^2$.

Thus, if the latter is denoted by γ_{ijk} , then $\hat{\Omega}_{wjk} = \sum \gamma_{ijk}^{-1}(x_{ij}x_{ik} - \Upsilon_{ijk}) / \sum \gamma_{ijk}^{-1}$ will be a more efficient estimator of Ω .⁶ Because the weights will depend on Ω , one would need to first use the unweighted estimator to get an initial set of weights.

Subset of variables measured with error

Frequently, only a subset of the independent variables are measured with error. For instance, suppose

6. In addition to varying by observation, $Cov[x_i'x_i - \Upsilon_i]$ will typically be nondiagonal, implying that even more efficient estimation will be possible. In particular, if x_i^* and η_i have Gaussian distributions, then one can express $Cov[vec(x_i'x_i - \Upsilon_i)]$ as a function of Ω and Υ_i , say $\Gamma[\Omega, \Upsilon_i] = \Gamma_i$. Then

$vec(\hat{\Omega}_{eff}) = \left(\sum \Gamma_i^{-1} \right)^{-1} \left[\sum \Gamma_i^{-1} (vec(x_i'x_i - \Upsilon_i)) \right]$ will be an efficient estimator.

$$(28) \quad y_i = x_i^* \beta + z_i^* \gamma + \varepsilon_i ,$$

where there are no measurement errors in z_i^* . Let $\Omega = \begin{bmatrix} \Omega_{xx} & \Omega_{xz} \\ \Omega_{zx} & \Omega_{zz} \end{bmatrix}$, $\Upsilon_i = \begin{bmatrix} \Upsilon_{xxi} & 0 \\ 0 & 0 \end{bmatrix}$, and

$R = \begin{bmatrix} R_{xxi} & R_{xzi} \\ R_{zxi} & R_{zzi} \end{bmatrix}$ be conformable matrices. Then it is straightforward to show

$$(29) \quad R_{xxi} = (\Omega_{xx} - \Omega_{xz} \Omega_{zz}^{-1} \Omega_{zx} + \Upsilon_{xxi})^{-1} (\Omega_{xx} - \Omega_{xz} \Omega_{zz}^{-1} \Omega_{zx}) .$$

$$(30) \quad R_{zxi} = \Omega_{zz}^{-1} \Omega_{zx} (\Omega_{xx} - \Omega_{xz} \Omega_{zz}^{-1} \Omega_{zx} + \Upsilon_i)^{-1} \Upsilon_{xxi} ,$$

$$(31) \quad R_{xzi} = 0 \text{ and}$$

$$(32) \quad R_{zzi} = I .$$

Letting $\bar{E}[x_i^* | z_i] = z_i \Omega_{zz}^{-1} \Omega_{zx}$ denote the best linear predictor of x_i^* given the vector z_i , the quantity $\Omega_{xx} - \Omega_{xz} \Omega_{zz}^{-1} \Omega_{zx}$ is the variance of $x_i^* - \bar{E}[x_i^* | z_i]$, the difference between the true x_i^* and its best linear predictor using z_i , while $\Omega_{xx} - \Omega_{xz} \Omega_{zz}^{-1} \Omega_{zx} + \Upsilon_i$ is the variance of $x_i - \bar{E}[x_i^* | z_i]$, the difference between the measured x_i and the best linear predictor given z_i .

Thus R_{xxi} has the form of a reliability matrix for the variable $x_i - \bar{E}[x_i^* | z_i]$, the portion of x_i not accounted for by z_i . Using (29) to (32), the overall best linear predictor of x_i^* given both x_i and z_i is

$$(33) \quad \bar{E}[x_i^* | x_i, z_i] = x_i R_{xxi} + \bar{E}[x_i^* | z_i] (I - R_{xxi})$$

while the best linear predictor of z_i^* is just

$$(34) \quad \bar{E}[x_i^* | x_i, z_i] = z_i .$$

The HEIV estimator is then the regression of y_i on $\bar{E}[x_i^*|x_i, z_i]$ and $\bar{E}[z_i^*|x_i, z_i]$:

$$(35) \quad y_i = (x_i R_{xxi} + \bar{E}[x_i^*|z_i](I - R_{xxi}))\beta + z_i\gamma + [\varepsilon_i + (x_i^* - (x_i R_{xxi} + \bar{E}[x_i^*|z_i](I - R_{xxi})))\beta].$$

The EIV estimator is of the same form, but with R_{xxi} replaced by

$$R_{xx} = (\Omega_{xx} - \Omega_{xz}\Omega_{zz}^{-1}\Omega_{zx} + \Upsilon_{xx})^{-1}(\Omega_{xx} - \Omega_{xz}\Omega_{zz}^{-1}\Omega_{zx}),$$
 which leads to the regression

$$(36) \quad y_i = (x_i R_{xx} + \bar{E}[x_i^*|z_i](I - R_{xx}))\beta + z_i\gamma + [\varepsilon_i + (x_i^* - (x_i R_{xx} + \bar{E}[x_i^*|z_i](I - R_{xx})))\beta].$$

The term $\bar{E}[x_i^*|z_i](I - R_{xx})$ is in the space spanned by the z_i . So one would obtain the same estimate of β from a regression of y_i on $x_i R_{xx}$ and z_i . However, in the case of (35), the term $\bar{E}[x_i^*|z_i](I - R_{xxi})$ cannot be dropped because, though $\bar{E}[x_i^*|z_i]$ is in the space spanned by z_i , the coefficients, $I - R_{xxi}$ vary by observation.

When only a single independent variable is measured with error, x_i and R_{xii} are scalars and (33) implies that the best linear predictor of x_i^* is a convex combination of the observed variable x_i and its prediction, $\bar{E}[x_i^*|z_i] = z_i\Omega_{zx}\Omega_{zz}^{-1}$, based on the other independent variables. The weight on x_i is the fraction of the variance of $x_i - \bar{E}[x_i^*|z_i]$ not attributable to measurement error.

When only a single independent variable is measured with error, the results of sections III and IV appear to give a good guide to the behavior of the EIV and HEIV estimators for β if one associates r and r_i with R_{xx} and R_{xxi} , respectively. That is, the correct multivariable analog of the single variable reliability ratio is the reliability ratio for the quantity, $x_i - \bar{E}[x_i^*|z_i]$. This reflects the frequently made point that what may seem to be a relatively small amount of measurement error in a variable can still be quite significant if, given the other variables in the regression, there is little independent variation in the variable.

Computing the EIV and HEIV estimators

StataCorp (1999) has a built-in procedure (eivreg) that can compute the EIV estimator for the case of a diagonal Υ matrix. One simply supplies the procedure with the reliability ratios corresponding to the mean level of measurement error. Computing the HEIV estimator, or the EIV estimator in the case of a nondiagonal Υ matrix, requires only modestly more work. Given their definitions in terms of OLS regressions, they can be implemented using standard regression software without extensive programming or high computational expenses.

In the case in which there is a single regressor subject to measurement error, the various components of the calculation can be easily obtained once the weighted estimate of Ω is constructed. If the variable x_i^* subject to measurement error is ordered first, then the first row and column of $\hat{\Omega}_w$ will need to be computed using the weights described above. However, the rest of the matrix, corresponding to the variables, z_i^* measured without error, is just the standard

$\hat{\Omega}_{wzz} = n^{-1} \sum z_i' z_i$. Standard regression programs then can take the covariance matrix $\hat{\Omega}_w$ as

input to compute the coefficients of the regression of x_i on z_i , $\hat{\Omega}_{wxz} \hat{\Omega}_{wzz}^{-1}$, the associated fitted values, $z_i \hat{\Omega}_{wxz} \hat{\Omega}_{wzz}^{-1}$, and the variance of the residuals, $\hat{\Omega}_{wxx} - \hat{\Omega}_{wxz} \hat{\Omega}_{wzz}^{-1} \hat{\Omega}_{wzx}$, which are the main quantities needed to compute the EIV and HEIV estimators.

VI. Conclusion

Having varying degrees of measurement error in the independent variables of a linear regression model is a common problem in applications when the data are taken from earlier rounds of statistical analysis based on samples of varying sizes. Simply replacing the assumed-common variance of the measurement error in the standard errors-in-variables estimator yields a consistent EIV estimator when measurement errors are heteroskedastic. But the results of this paper suggest that there are significant gains in efficiency from using the alternative, HEIV estimator. It is, moreover, straightforward to compute using standard regression software.

References

- Aaronson, Daniel and Daniel G. Sullivan (1999), "Worker Insecurity and aggregate wage growth," Federal Reserve Bank of Chicago, working paper, no 99-30.
- Blanchard, Olivier and Lawrence F. Katz (1997), "What we know and do not know about the natural rate of unemployment," *Journal of Economic Perspectives*, volume 11, number 1, Winter, pp. 51-72.
- Blanchflower, David and Andrew Oswald (1994), *The Wage Curve*, Cambridge, MA: MIT Press.
- Card, David, and Dean Hyslop (1997), "Does Inflation 'Grease the Wheels of the Labor Market'?", in *Reducing Inflation: Motivation and Strategy*, Christina Romer and David Romer, Chicago: University of Chicago Press, pp. 71-114.
- Dempster, A.P., N.M. Laird, and D.B. Rubin (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 39, 1-38.
- Greene, William H. (1997), *Econometric Analysis*, Third Edition, Upper Saddle River, NJ: Prentice Hall.
- Campbell, Jeffrey R. and Hugo A. Hopenhayn (2001), "Market Size Effects in U.S. Cities' Retail Trade Industries," unpublished working paper, University of Chicago, <http://home.uchicago.edu/~jcampbe>.
- StataCorp (1999), *Stata Statistical Software: Release 6.0*, College Station, TX: Stata Corporation.
- White, Halbert (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, volume 48, number 4, pp. 817-838.

Table 1: Finite-sample and asymptotic variances for state data simulation

Variance Component	Sample Size			Asymptotic Limit
	100	500	1000	
Λ_{eiv}^s : n EIV Structural Variance	2.100 (0.031)	2.035 (0.029)	2.016 (0.028)	2.00
Λ_{eiv}^s : n HEIV Structural Variance	1.758 (0.025)	1.711 (0.025)	1.690 (0.024)	1.68
Λ_{eiv}^e : n EIV Prediction Variance	1.715 (0.025)	1.741 (0.025)	1.756 (0.025)	1.73
Λ_{heiv}^e : n HEIV Prediction Variance	0.563 (0.008)	0.551 (0.008)	0.551 (0.008)	0.55
Λ_{eiv} : n EIV Mean Squared Error	3.749 (0.054)	3.783 (0.049)	3.767 (0.054)	3.73
Λ_{heiv} : n HEIV Mean Squared Error	2.231 (0.034)	2.279 (0.033)	2.238 (0.032)	2.23
EIV Robust Variance Estimate	3.576 (0.012)	3.699 (0.006)	3.719 (0.004)	3.73
HEIV Robust Variance Estimate	2.221 (0.007)	2.223 (0.003)	2.227 (0.002)	2.23

Figure 1: $E[r_i x_i (x_i^* - r_i x_i) | \tau_i^2]$ as a function of τ_i^2 / ω^2

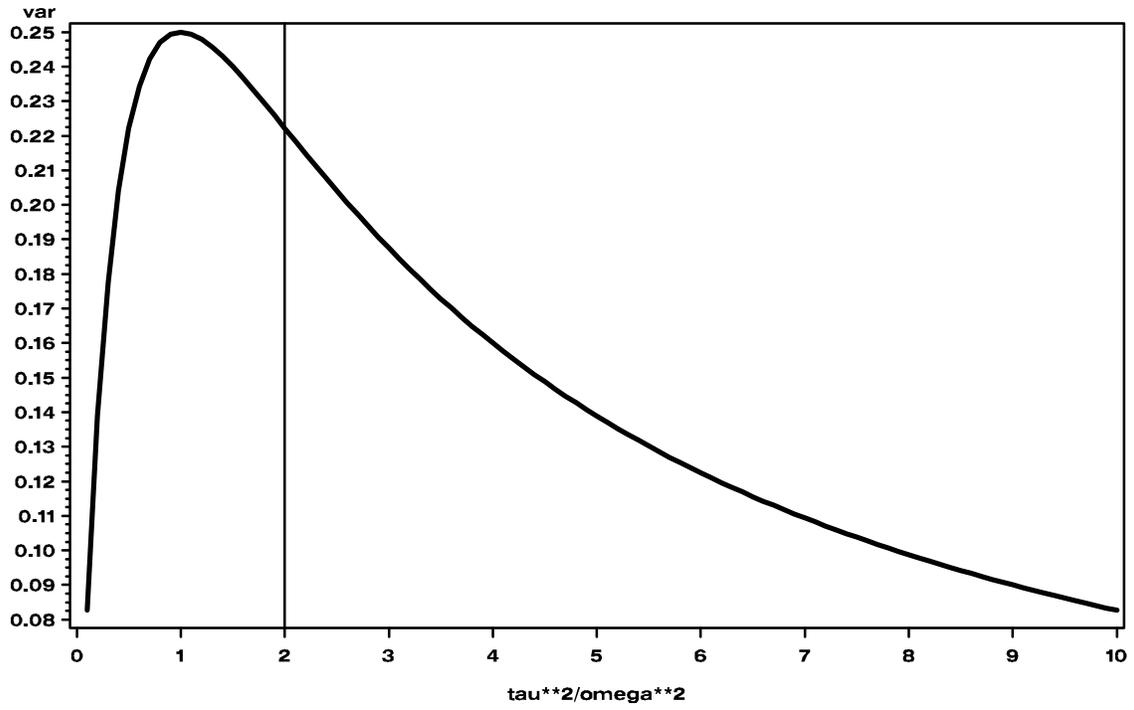


Figure 2: $\Lambda_{heiv}^s/\Lambda_{eiv}^s$ and $\Lambda_{heiv}^s/\Lambda_{eiv}^s$ as a function of τ^2/ω^2 for state-data simulation

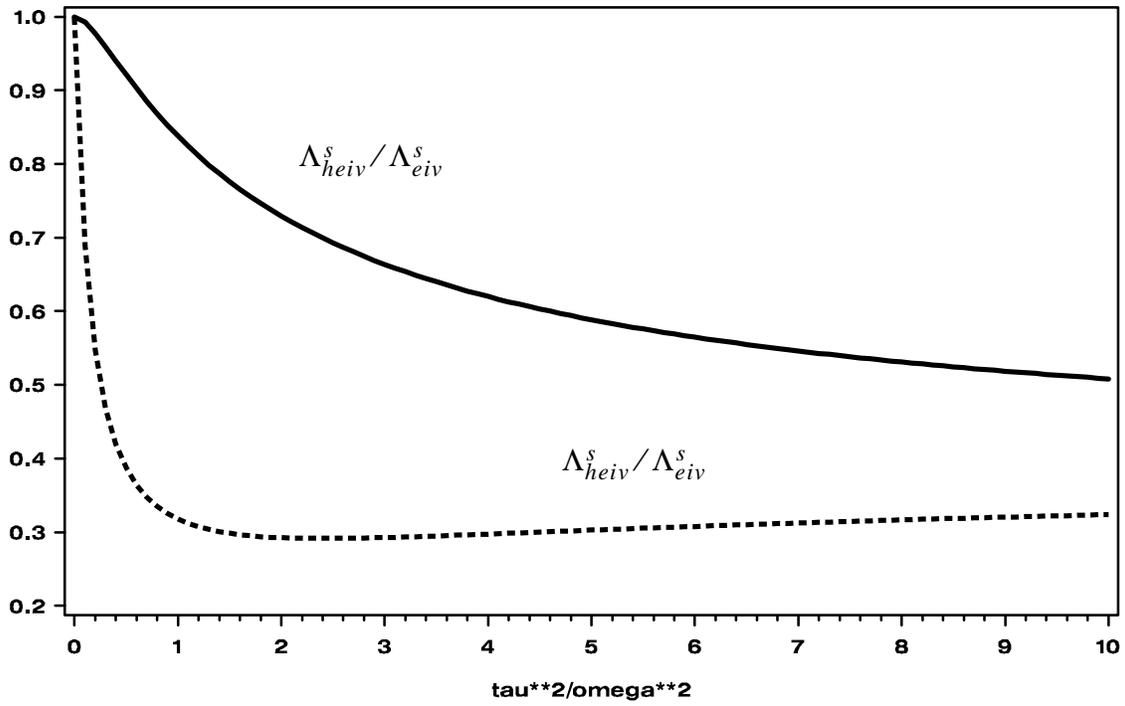


Figure 3: $Var[r_i x_i(x_i^* - r_i x_i)]$ and $Var[rx_i(x_i^* - rx_i)]$ as a function of τ^2/ω^2 for state-data simulation

