NHH

INSTITUTT FOR FORETAKSØKONOMI

DEPARTMENT OF FINANCE AND MANAGEMENT SCIENCE

FOR 10 2009

ISSN: 1500-4066

SEPTEMBER 2009

Discussion paper

A simple improvement of the IV estimator for the classical errors-in-variables problem

BY

JONAS ANDERSSON AND **JARLE MØEN**

Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS AND BUSINESS ADMINISTRATION

A simple improvement of the IV estimator for the classical errors-in-variables problem

Jonas Andersson and Jarle Møen*

September 10, 2009

Abstract

Two measures of an error-ridden explanatory variable make it possible to solve the classical errors-in-variable problem by using one measure as an instrument for the other. It is well known that a second IV estimate can be obtained by reversing the roles of the two measures. We explore a simple estimator that is the linear combination of these two estimates, that minimizes the asymptotic mean squared error. In a Monte Carlo study we show that the gain in precision is significant compared to using only one of the original IV estimates. The proposed estimator also compares well with full information maximum likelihood under normality.

Keywords: Measurement errors, Classical Errors-in-Variables, multiple indicator method, Instrumental variable techniques

JEL-codes: C13, C30, C80

*Department of Finance and Management Science, Norwegian School of Economics and Business Administration, Hellevn. 30, N-5045 Bergen, Norway.

Email: jonas.andersson@nhh.no and jarle.moen@nhh.no.

We have received useful comments from Erik Biørn, Gernot Doppelhofer and Arngrim Hunnes. We are grateful to Anne Liv Scrase for proof-reading the manuscript.

1 Introduction

It is well known that ordinary least squares (OLS) is inconsistent and biased if one or more explanatory variables are measured with error. It is also well known that instrumental variables (IV) can be used to deal with the problem. Graduate text books in econometrics typically present the classical errors-in-variables model where one explanatory variable is measured with error and the measurement error is uncorrelated with all explanatory variables in the model as well as with the unobserved disturbance. A second measurement of the mismeasured variable is introduced, and it is assumed that the measurement error in the second measure is uncorrelated with the measurement error in the first as well as with all other variables including the disturbance. The second measure is then a valid instrument for the first. Papers that have made important contributions using this technique include Ashenfelter and Krueger (1994), Borjas (1995), Barron et al. (1997) and Krueger and Lindahl (2001).[1]

The favourite text book example of instrumental variables used to solve a measurement error problem in economics is the analysis of returns to education by Ashenfelter and Krueger (1994). Ashenfelter and Krueger simultaneously account for ability bias and measurement errors by using a sample of twins. Identical twins are similar with respect to family background and genetic endowment, but measurement errors in education are exacerbated when ability is differenced out. The ingenuity of the Ashenfelter and Krueger study is that they obtain two measures of education by asking each twin both about his or her own education and about the education of the sibling.

The classical errors-in-variables model with two indicators constitutes a three-equation system and can be estimated with full information maximum likelihood using the latent variable framework of Goldberger (1972) and Jöreskog (1978).[2] In the applied econometrics literature, however, IV seems to be the preferred approach when two measures are available. In fact, the only papers we have found in economics journals that present a full information maximum likelihood estimate are Ashenfelter and Krueger (1994) and the

[1]We briefly review some papers that instrument one mismeasured variable with another in Appendix 1.

[2]This estimation approach is usually implemented by using the software packages LISREL, see e.g. Jöreskog et al. (2001).

follow up study by Rouse (1999).[3]

When two indicators are available and allow for an IV-solution, it is not obvious which measure should be used as explanatory variable, and which measure should be used as instrument. Whichever is chosen, a second estimate can be produced by reversing the role of the variable and the instrument. Several studies present both estimates, but no discussion of the choice between them appears to be available in the econometrics literature. In a comprehensive chapter on measurement errors in the *Handbook of Econometrics*, Bound et al. (2001) note that the availability of two estimates gives "some capacity to test the underlying assumptions of the model". Otherwise, the issue is left untouched.

The preference for IV among applied econometricians is probably explained by the fact that this method is intuitive and computationally easy to implement. IV estimates often have low precision, however. In the present paper we explore a simple improvement of the classical IV solution. The proposed estimator is a linear combination of the two IV estimates that is obtained by using a pair of indicators both ways. The improved estimate is based solely on the two original estimates and by-products obtained when these are calculated. It is optimal in the sense that it minimizes the variance among linear combinations of the two IV estimators, and without co-variates it is a special case of the Chamberlain (1982) Π-matrix approach.[4] In a Monte Carlo study we show that the gain in precision is significant compared to using only one of the two original IV estimates. Both the asymptotic and the small sample efficiency are in the range of 70-85 percent. Moreover, the proposed estimator compares very well with full information maximum likelihood under normality. This holds even for small sample sizes, and, unlike maximum likelihood, it does not require any numerical optimization nor any distributional assumption. Somewhat counter-intuitively, our analysis reveals that those who present only a single ordinary IV estimate should use the indicator suspected to be most contaminated by measurement errors as variable and the other as instrument.

[3]Given that Ashenfelter and Krueger (1994) is the leading text book example of IV as a solution to measurement errors bias, this is somewhat ironical. We have not found any graduate text book that mentions the full information approach.

[4]See Chamberlain (1982, p.24). Ashenfelter and Krueger (1992) and Behrman and Rosenzweig (1999) apply Chamberlain (1982) as an alternative to IV in a setting with classical measurement error and two indicators. Ashenfelter and Krueger use the simple regression framework while Behrman and Rosenzweig include control variables. Ashenfelter and Krueger (1992) is a preprint of their famous 1994-paper.

Section 2 reviews the classical errors-in-variables model and Section 3 presents the improved IV estimator. Section 4 contains the Monte Carlo study exploring the small sample properties of the various estimators and Section 5 concludes.

2 The model

The problem at hand is a linear regression where one of the explanatory variables is measured twice, both times with measurement errors. We consider the case of classical errors-in-variables, i.e. we assume that the measurement errors are independent of each other and of the underlying variable it is supposed to measure. Our main interest is to estimate the parameter β in the model

$$y_i = x_i^* \beta + \boldsymbol{w}_i' \boldsymbol{\gamma} + \varepsilon_i \tag{1}$$

$\boldsymbol{w}$ is a k-dimensional exogenous variable, i.e. all elements have the property $Cov(w_{ij}, \varepsilon_i) = 0$ for all $i = 1, ..., n$ and $j = 1, ..., k$. Furthermore, $Cov(\varepsilon_i, \varepsilon_j) = \sigma^2$ for $i = j$ and zero otherwise.

The explanatory variable x^* is observed with measurement error through the variables x_1 and x_2 given by

$$x_{1i} = x_i^* + \delta_{1i} \tag{2}$$

and

$$x_{2i} = x_i^* + \delta_{2i} \tag{3}$$

where δ_1 and δ_2 are independent measurement errors with variances τ_1^2 and τ_2^2. $Cov(\delta_{ki}, \varepsilon_i) = 0$ for $k = 1, 2$.

Regressing y on e.g. x_1 using OLS is problematic since x_1 is correlated with the error term. This can be seen by considering the regression equation

$$y_i = x_i^* \beta + \boldsymbol{w}_i' \boldsymbol{\gamma} + \varepsilon_i = x_{1i} \beta + \boldsymbol{w}_i' \boldsymbol{\gamma} + \varepsilon_{1i} \tag{4}$$

where $\varepsilon_{1i} = \varepsilon_i - \beta \delta_{1i}$. The correlation between ε and x_1 is $-\beta \tau_1^2$ created by the common term δ_1 in the equations (2) and (4). An analogous result is true when using x_2 as the regressor with the corresponding error term $\varepsilon_{2i} = \varepsilon_i - \beta \delta_{2i}$.

3 An improved IV estimator

There are two possible instrumental variable estimators. x_2 is correlated with x_1, but uncorrelated with ε_1. This means that x_2 is a valid instrument when x_1 is used as regressor. [5] Likewise, x_2 is uncorrelated with ε_1 and is a valid instrument when x_2 is used as regressor. For the sake of exposition we first show how to form the estimator, $\hat{\beta}_{1,IV}$, where x_1 is used as an instrumental variable for x_2. For this purpose we form the $n \times (k+1)$-quantities $X_1 = [\boldsymbol{x}_2, W]$ and $Z_1 = [\boldsymbol{x}_1, W]$ where $\boldsymbol{x}_i$, $i = 1, 2$ are $n \times 1$-vectors containing the explanatory variables and the instrumental variable, respectively. Furthermore, the $n \times 1$-vector $\boldsymbol{y}$ containing the observations of the dependent variable, the corresponding vector of error-terms $\boldsymbol{e}_1$ and $\boldsymbol{\pi} = [\beta, \boldsymbol{\gamma}']'$ enables us to rewrite (4) as

$$\boldsymbol{y} = X_1 \boldsymbol{\pi} + \boldsymbol{e}_1 \tag{5}$$

The IV estimator of $\boldsymbol{\pi}$ can now be written:

$$\hat{\boldsymbol{\pi}}_{1,IV} = \begin{bmatrix} \boldsymbol{x}_2'\boldsymbol{x}_1 & \boldsymbol{x}_2'W \\ W'\boldsymbol{x}_1 & W'W \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{x}_1'\boldsymbol{y} \\ W'\boldsymbol{y} \end{bmatrix} \tag{6}$$

The matrix inversion can be beneficially accommodated by use of the particular partition of the matrix used in (6). This enables us to obtain a direct expression for the IV estimator of β, $\hat{\beta}_{1,IV}$, which will be useful in what follows. The expression for $\hat{\beta}_{1,IV}$ is

$$\hat{\beta}_{1,IV} = K_1 \boldsymbol{y} \tag{7}$$

where

$$K_1 = (\boldsymbol{x}_1'\boldsymbol{x}_2 - \boldsymbol{x}_1'W(W'^{-1}W'\boldsymbol{x}_2)^{-1}(\boldsymbol{x}_1' - \boldsymbol{x}_1'W(W'^{-1}W') \tag{8}$$

Similarly, when x_2 is used as an instrumental variable for x_1 we obtain

$$\hat{\beta}_{2,IV} = K_2 \boldsymbol{y} \tag{9}$$

where

[5] Note that both x_1 and x_2 are endogenous variables in our model, only the latent x_i^* is truly exogenous. Nevertheless, the orthogonality conditions for valid instruments are satisfied. Thus, as pointed out by Biørn (2009) p. 348, endogenous variables can be useful as instruments in models with measurement errors.

$$K_2 = (\boldsymbol{x}_2'\boldsymbol{x}_1 - \boldsymbol{x}_2'W(W'^{-1}W'\boldsymbol{x}_1)^{-1}(\boldsymbol{x}_2' - \boldsymbol{x}_2'W(W'^{-1}W') \tag{10}$$

Our aim in this section is to find the linear combination of these two estimators which has the smallest variance.[6] Since the estimators are consistent and thereby asymptotically unbiased, for large samples, this can also be seen as finding the linear combination which minimizes the asymptotic mean square error.

The asymptotic variances of the IV estimators $\hat{\beta}_{k,IV}$, $k = 1, 2$ are

$$v_1 = Var(\hat{\beta}_{1,IV}) = \sigma_1^2 K_1 K_1' \tag{11}$$

and

$$v_2 = Var(\hat{\beta}_{2,IV}) = \sigma_2^2 K_2 K_2' \tag{12}$$

respectively.[7] The variances σ_1^2 and σ_2^2 are the error term variances in each of the regressions.

Finally, an the asymptotic covariance between $\hat{\beta}_{1,IV}$ and $\hat{\beta}_{2,IV}$, is given by

$$c_{12} = Cov(\hat{\beta}_{1,IV}, \hat{\beta}_{1,IV}) = \sigma_{12} K_1 K_2' \tag{13}$$

where σ_{12} is the covariance between the error terms in the two regressions. The quantities σ_1^2, σ_2^2 and σ_{12} can be estimated by the corresponding sample moments of the residuals from the two IV regressions through

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum \hat{\varepsilon_1}^2, \tag{14}$$

$$\hat{\sigma}_2^2 = \frac{1}{n}\sum \hat{\varepsilon_2}^2, \tag{15}$$

and

$$\hat{\sigma}_{12} = \frac{1}{n}\sum \hat{\varepsilon_1}\hat{\varepsilon_2} \tag{16}$$

[6]Note that our approach cannot be generalized to a situation where x_1 is a proxy of the type $x_1 = \alpha x^* + \delta_1$. The second IV estimator is then needed to solve for α. See Lubotsky and Wittenberg (2006) for a recent discussion and extension of this model. Our approach can, however, be generalized to a case where one indicator is systematically smaller than the other if the difference can be modelled in the form of an intercept in one of the measurement equations. Such an intercept can be transferred to the y-equation and included in W.

[7]These variances are just sample versions of the asymptotic variance of IV estimators (e.g. Mardia et al., 1994, p. 188).

where $\hat{\varepsilon_1}$ and $\hat{\varepsilon_2}$ are the residuals from the two IV regressions. Our new estimator is

$$\hat{\beta}_{\lambda,IV} = \lambda\hat{\beta}_{1,IV} + (1-\lambda)\hat{\beta}_{2,IV} \tag{17}$$

and has variance

$$Var(\hat{\beta}_{\lambda,IV}) = \lambda^2 v_1 + (1-\lambda)^2 v_2 + 2\lambda(1-\lambda)c_{12} \tag{18}$$

Minimizing this with respect to λ gives us an estimator which is optimal in the sense that it is the linear combination of the two IV estimators which has the smallest variance. The optimal λ is given by

$$\lambda_{opt} = \frac{v_2 - c_{12}}{v_1 + v_2 - 2c_{12}}. \tag{19}$$

4 Small sample properties

We have performed various simulation studies in order to investigate the performance of the estimator $\hat{\beta}_{\lambda_{opt},IV}$. In section 4.1 we use a simple regression framework and vary the ratio between the variances of the measurement errors of the two indicators. In section 4.2 we include a correctly measured covariate w and vary the correlation between this covariate and the latent variable x^*.

4.1 Simple regression

We start out exploring the small sample properties of the proposed estimator using a simple regression model without intercept. The simulated data generating process (DGP) is

$$\begin{cases} y = 0.5x^* + \varepsilon \\ x_1 = x^* + \delta_1 \\ x_2 = x^* + \delta_2 \end{cases} \tag{20}$$

where $Var(x^*) = 1$, $\sigma^2 = Var(\varepsilon) = 0.5$, $\tau_1^2 = Var(\delta_1) = 0.25$ and $\tau_2^2 = Var(\delta_2)$ is varied in between 0.25 and 1. In addition, ε, δ_1 and δ_2 are assumed to be independent and normally distributed. The results are given in table 1 and table 2 and show significant improvement compared to using one single IV estimator. Furthermore, for all of the investigated sample sizes, the cases studied indicate that the MSE of the improved IV estimator is very close to the full information maximum likelihood estimator.

n	OLS	IV1	IV2	ML	SIMP	non-conv
100	14.766	7.418	7.510	6.444	6.459	11
1000	10.455	0.735	0.710	0.631	0.631	21
5000	10.064	0.144	0.138	0.121	0.121	16

Table 1: MSE of estimators when the true value of $\beta = 0.5$, $\sigma^2 = 0.5$, $\tau_1^2 = 0.25$ and $\tau_2^2 = 0.25$. 1000 simulation replicates. The last column shows the number of replicates where the ML-estimator did not converge. Those replicates were removed for all estimators. The number of observations is given by n, and SIMP is our improved IV estimator.

n	OLS	IV1	IV2	ML	SIMP	non-conv
100	14.427	10.621	11.662	8.659	8.549	9
1000	10.389	0.923	1.133	0.788	0.785	0
5000	10.057	0.208	0.231	0.172	0.172	0

Table 2: MSE of estimators when the true value of $\beta = 0.5$. $\sigma^2 = 0.5$, $\tau_1^2 = 0.25$, $\tau_2^2 = 1$. 1000 simulation replicates. The last column shows the number of replicates where the ML-estimator did not converge. Those replicates were removed for all estimators. The number of observations is given by n, and SIMP is our improved IV estimator.

It is also interesting to know which of the two original IV estimators that dominates the other, and how this depends on the two measurement errors. This is investigated by means of figure 1 where λ_{opt} is plotted against the ratio of the two measurement errors for the parameter values in the simulation study. For the simple DGP studied above, λ_{opt}

can be written as a function of this ratio

$$\lambda_{opt} = f(\kappa) = \frac{\beta^2 + (\sigma^2 + \beta^2\tau_1^2)\kappa}{\beta^2 + \sigma^2 + (\sigma^2 + \beta^2 + 2\beta^2\tau_1^2)\kappa} \tag{21}$$

where we use the notation $\kappa = \tau_2^2/\tau_1^2$ and have assumed that $Var(x^*) = 1$. Furthermore, in order to avoid dependence on data we have substituted the K-matrices, which contain data, with population moments, e.g. we have used $Var(x_1) = 1 + \tau_1^2$. Figure 1 shows a plot of this function for the parameter values used in the simulation study above.

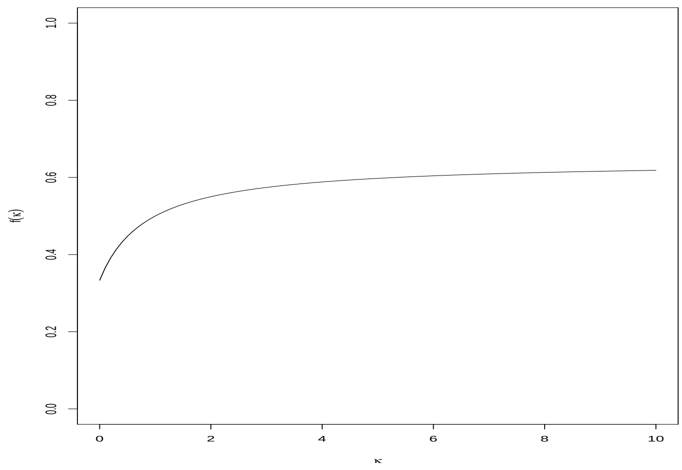

Figure 1: The optimal weight λ_{opt} for different choices of κ, the ratio of the measurement errors of the two indicators.

At least two interesting observations can be made from this graph. First, if only a standard IV estimator is used and the econometrician has an opinion regarding which of the measurements is least prone to measurement errors, the measurement thought to have the smallest errors should be used as the instrumental variable. This can be seen by observing the fact that for values of κ larger than one, (relatively small measurement error in x_1) the optimal λ_{opt} is large, implying a large weight on the IV estimator where x_1 is the instrumental variable. Secondly, even for cases where the measurement error in one variable is huge relative to the other, a gain is to be made from weighting them together. This is seen by the asymptote of the function. Even for $\kappa = 100$, i.e. when one x-variable has a measurement error variance that is 100 times larger than the measurement error variance of the other, a significant weight should be given to both estimators. However, it should be noted that we do not consider other alternatives than the IV estimators. If the ratio is large simply because one of the variables is measured almost without error, OLS would be better than any of the two IV estimators. If the ratio is large because one of the indicators is extremely noisy, OLS may also be preferable. In this case, however, a trade-off between bias and precision has to be made.[8]

We can also see that when $\kappa = 1$, then $\lambda_{opt} = \frac{1}{2}$ which means that if the measurement error is of the same magnitude for both measurements and one insists on using just one instrumental variable estimator, then the choice of estimator is irrelevant. However, an improved estimate can be obtained by weighting the two together, and the optimal estimator is simply the average of the two original IV estimates in this case.

Figure 2 shows how the variance of the improved estimator relates to the variance of that of the ordinary IV estimators with the smallest variance. The improvement increases with the ratio of the measurement errors of the two indicators.

[8] An early and interesting contribution to the measurement error literature by Feldstein (1974) discusses this trade-off and suggests and evaluates alternative procedures for "balancing the loss of efficiency in IV estimation against the potential gain of reduced bias". He proposes a so-called WAIV estimator which is a weighted average of the OLS and IV estimates. Feldstein finds that the WAIV estimator is consistent and has a "smaller MSE than the IV estimator in a wide class of conditions and otherwise has an equal MSE".

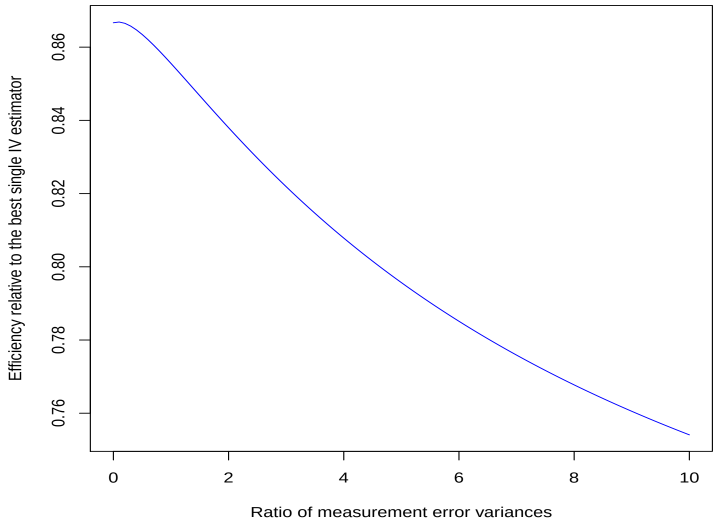

Figure 2: The efficiency of the improved IV estimator relative to the best single IV estimator for different choices of κ, the ratio of the measurement errors of the two indicators.

4.2 Adding a covariate without measurement error

Our proposed estimator allows for an arbitrary number of correctly measured covariates in addition to the mismeasured variable of main interest. Most relationships in applied work contain such covariates. In this section we explore whether the main results from the simulation study above are robust to including a covariate. The DGP that we simulate from is

$$\begin{cases} y = 0.5x^* + 0.5w + \varepsilon \\ x_1 = x^* + \delta_1 \\ x_2 = x^* + \delta_2 \end{cases} \tag{22}$$

where $Var(x^*) = 1$, $Var(w) = 1$, $\sigma^2 = Var(\varepsilon) = 0.5$, $\tau_1^2 = Var(\delta_1) = 0.25$ and $\tau_2^2 = Var(\delta_2) = 0.25$. The covariance between x^* and the extra regressor w, σ_{x^*w}, is varied between -0.5, 0 and 0.5. As before, ε, δ_1 and δ_2 are assumed to be independent and normally distributed.

None of the tables 3, 4 or 5 reveal any fundamental difference from the results in the previous section. For the parameter values studied, the improved IV estimator performs better than both the original IV estimators and it is not significantly outperformed by the ML estimator.

n	OLS	IV1	IV2	ML	SIMP	nonconv
100	21.085	10.847	10.258	9.067	8.931	19
500	16.363	2.034	1.921	1.663	1.654	1
1000	16.191	1.003	1.086	0.877	0.881	1

Table 3: MSE of estimators when the true value of $\beta = 0.5$, $\gamma = 0.5$, $\sigma_{x^*w} = -0.5$, $\sigma^2 = 0.5$, $\tau_1^2 = 0.25$ and $\tau_2^2 = 0.25$. 1000 simulation replicates. The last column shows the number of replicates where the ML-estimator did not converge. Those replicates were removed for all estimators. The number of observations is given by n, and SIMP is our improved IV estimator.

n	OLS	IV1	IV2	ML	SIMP	nonconv
100	14.850	7.968	7.697	6.976	6.939	5
500	11.052	1.321	1.380	1.152	1.155	2
1000	10.382	0.699	0.739	0.627	0.624	3

Table 4: MSE of estimators when the true value of $\beta = 0.5$, $\gamma = 0.5$, $\sigma_{x^*w} = 0$, $\sigma^2 = 0.5$, $\tau_1^2 = 0.25$ and $\tau_2^2 = 0.25$. 1000 simulation replicates. The last column shows the number of replicates where the ML-estimator did not converge. Those replicates were removed for all estimators. The number of observations is given by n, and SIMP is our improved IV estimator.

n	OLS	IV1	IV2	ML	SIMP	nonconv
100	20.796	10.427	10.330	8.912	8.709	4
500	16.697	1.971	2.016	1.669	1.666	7
1000	16.166	1.023	1.068	0.892	0.891	18

Table 5: MSE of estimators when the true value of $\beta = 0.5$, $\gamma = 0.5$, $\sigma_{x^*w} = 0.5$, $\sigma^2 = 0.5$, $\tau_1^2 = 0.25$ and $\tau_2^2 = 0.25$. 1000 simulation replicates. The last column shows the number of replicates where the ML-estimator did not converge. Those replicates were removed for all estimators. The number of observations is given by n, and SIMP is our improved IV estimator.

5 Conclusion

An easy-to-implement improvement of the IV estimator of the classical error-in-variables model has been proposed and investigated with a Monte Carlo study. In terms of MSE, the estimator significantly outperforms the standard IV estimator, and, more surprisingly, performs well compared to a full (Gaussian) maximum likelihood estimator even under normally distributed errors.

References

Ashenfelter, O. and A. Krueger (1992): "Estimates of the Economic Return to Schooling From a New Sample of Twins," NBER Working Paper No. 4143.

——— (1994): "Estimates of the Economic Return of Schooling from a New Sample of Twins," *American Economic Review*, 84, 1157–1173.

Ashenfelter, O. and C. Rouse (1998): "Income, Schooling, and Ability: Evidence From a New Sample of Twins," *Quarterly Journal of Economics*, 253–284.

Barron, J. M., M. C. Berger, and D. A. Black (1997): "How Well Do We Measure Training?" *Journal of Labor Economics*, 15, 507–528.

Behrman, J. R. and M. R. Rosenzweig (1999): ""Ability" Biases in Schooling Returns and Twins: A Test and New Estimates," *Economics of Education Review*, 18, 159–167.

Benhabib, J. and M. M. Spiegel (1994): "The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data," *Journal of Monetary Economics*, 34, 143–174.

Biørn, E. (2009): *Økonometriske emner*, Unipub, Oslo.

Bjerk, D. (2007): "Measuring the Relationship Between Youth Criminal Participation and Household Economic Resources," *Journal of Quantitative Criminology*, 23, 23–39.

Bonjour, D., L. F. Cherkas, J. E. Haskel, D. D. Hawkes, and T. D. Spector (2003): "Returns to Education: Evidence from U.K. Twins," *American Economic Review*, 93, 1799–1812.

Borjas, G. J. (1995): "Etnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review*, 85, 365–390.

Bound, J., C. Brown, and N. Mathiowetz (2001): "Measurement Error in Survey Data," in *Handbook of Econometrics*, ed. by J. Heckman and E. Leamer, Elsevier Science, chap. 59, 3705–3843.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–42.

Drago, F. (2008): "Self-Esteem and Earnings," IZA Discussion Paper No. 3577.

Feldstein, M. (1974): "Errors in Variables: A Consistent Estimator with Smaller MSE in Finite Samples," *Journal of the American Statistical Association*, 69, 990–996.

Goldberger, A. (1972): "Maximum-Likelihood Estimation of Regressions Containing Unobservable Independent Variables," *International Economic Review*, 13, 1–15.

Jöreskog, K. (1978): "Structural-analysis of covariance and correlation matrices," *Psykometrica*, 43, 443–477.

Jöreskog, K., D. Sörbom, S. Du Toit, and M. Du Toit (2001): *LISREL 8: New Statistical Features*, Scientific Software.

Krueger, A. and M. Lindahl (2001): "Education for Growth: Why and For Whom?" *Journal of Economic Literature*, 39, 1101–1136.

Lubotsky, D. and M. Wittenberg (2006): "Interpretation of Regressions with Multiple Proxies," *Review of Economics and Statistics*, 88, 549–562.

Mardia, K., J. Kent, and J. Bibby (1994): *Multivariate Analysis*, Academic Press.

Rouse, C. E. (1999): "Further Estimates of the Economic Return to Schooling from a New Sample of Twins," *Economics of Education Review*, 18, 149–157.

The appendixes below are not intended for journal publication.

Appendix 1: Examples of papers that have two indicators and use IV to correct for measurement error bias

While severe measurement error is often a problem in economic data, it is not often the case that two measures of an error ridden variable is available to correct the bias. With an increasing availability of micro data, this is likely to change, however. Researchers who are able to find a second measure of an important, mismeasured variable, often make important contributions. Below are some papers that succeeded Ashenfelter and Krueger (1994) which we discussed in the introduction.

- Borjas (1995) in the *American Economic Review* shows that residential segregation gives rise to ethnic externalities in the human capital accumulation of the young generation. Parental skills are imprecisely measured, but a large number of siblings in the data makes it possible to instrument each individual's own report with the average of the siblings' report in the intergenerational transmission equation.

- Barron et al. (1997) in the *Journal of Labor Economics* use a survey where both employers and their employees have provided an estimate of on-the-job training. Their analysis suggests that previous estimates of the effect of training on wages and productivity growth have been underestimated by a factor of nearly three.

- Ashenfelter and Rouse (1998) in the *Quarterly Journal of Economics* study the correlation between ability and schooling and the extent to which the return to schooling varies with ability level. They use data for identical twins and use one twin's report of both twin 1 and twin 2's education as an instrument for the other twin's report of the same measures. They find that more able individuals attain more schooling because they face lower marginal costs of schooling, not because of higher marginal benefits.

- Krueger and Lindahl (2001) in the *Journal of Economic Literature* establish that the lack of a significant effect of changes in education on economic growth in the famous paper by Benhabib and Spiegel (1994) is due to measurement error in education. They use an additional data set with educational information to instrument for the education variable used by Benhabib and Spiegel (1994).

- Bonjour et al. (2003) in the *American Economic Review* estimate the returns to education using data on UK twins and follow the approach of Ashenfelter and Krueger (1994) by asking each twin to report both his or her own education and that of the other twin.

- Bjerk (2007) in the *Journal of Quantitative Criminology* use a household's percentiles in the income and wealth distributions as two indicators of economic resources when studying the effect of a household's economic resources on youth criminal participation.

- Drago (2008) in a recent IZA Working Paper analyses the effect of self-esteem on earnings and has measures of self-esteem from two surveys conducted seven years apart.

Appendix 2: Programming code to implement the improved IV estimator in R

```
optimal.iv=function(y,x1,x2,W)
{
n=length(y)
y=matrix(y,n,1)
x1=matrix(x1,n,1)
x2=matrix(x2,n,1)
W=as.matrix(W)
X1=cbind(x2,W)
Z1=cbind(x1,W)
pi1=solve(t(Z1)%*%X1)%*%t(Z1)%*%y
pi2=solve(t(X1)%*%Z1)%*%t(X1)%*%y
e1=y-X1%*%pi1
e2=y-Z1%*%pi2
What=W%*%solve(t(W)%*%W)%*%t(W)
K1=solve(t(x1)%*%x2-t(x1)%*%What%*%x2)%*%(t(x1)-t(x1)%*%What)
K2=solve(t(x2)%*%x1-t(x2)%*%What%*%x1)%*%(t(x2)-t(x2)%*%What)
b1=K1%*%y
b2=K2%*%y
v1=var(e1)*K1%*%t(K1)
v2=var(e2)*K2%*%t(K2)
c12=cov(e1,e2)*K1%*%t(K2)
lambda=(v2-c12)/(v1+v2-2*c12)
bopt=lambda*b1+(1-lambda)*b2
sb=sqrt(lambda^2*v1+(1-lambda)^2*v2+2*lambda*(1-lambda)*c12)
cl=bopt-1.96*sb
cu=bopt+1.96*sb
return(list(b1=b1,b2=b2,bopt=bopt,pi1=pi1,pi2=pi2,lambda=lambda,cl95=cl,cu95=cu))
}
```