

## American Economic Association

---

Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left

Author(s): Jerry Hausman

Source: *The Journal of Economic Perspectives*, Vol. 15, No. 4 (Autumn, 2001), pp. 57-67

Published by: American Economic Association

Stable URL: <http://www.jstor.org/stable/2696516>

Accessed: 14/12/2009 12:08

---

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

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

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

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).



American Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Economic Perspectives*.

<http://www.jstor.org>

## **Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left**

Jerry Hausman

**T**he effect of mismeasured variables in statistical and econometric analysis is one of the oldest known problems, dating from the 1870s in Adcock (1878). In the most straightforward regression analysis with a single regressor variable, the least squares estimate is downward biased in magnitude toward zero.<sup>1</sup> While a mismeasured right-hand side variable creates this problem, a mismeasured left-hand side variable under classical assumptions does not lead to bias. The only result is less precision in the estimated coefficient and a lower *t*-statistic.

I will begin by reviewing these classical issues involving mismeasured variables. By the “classical mismeasurement assumptions,” I mean that the error of measurement is uncorrelated with the true variables and is uncorrelated with the stochastic disturbance in the regression specification. Then I will consider three recent developments for the effect of mismeasurement on econometric models. The first issue involves difficulties in using instrumental variables as a method of addressing the problem of mismeasured variables. A second involves the consistent estimators that have recently been developed for mismeasured nonlinear regression models. Finally, I return to mismeasured left-hand side variables, where I will focus on issues in binary choice models and duration models but also will point out some open issues in other models.

<sup>1</sup> Surveys of these classical results are found in Aigner et al. (1984), Fuller (1987), and Hausman, Newey and Powell (1995). For related issues, see Griliches (1986). Henceforth, I will assume that the “true” parameter is positive so that I will not repeat the “in magnitude” qualifier.

■ *Jerry Hausman is John and Jennie S. MacDonald Professor of Econometrics and Public Finance, Massachusetts Institute of Technology, Cambridge, Massachusetts. His e-mail address is [jhausman@mit.edu](mailto:jhausman@mit.edu).*

## Linear Models with Mismeasured Variables

I begin with the classic linear regression mismeasurement model. I assume a linear specification in which means of all the variables have been subtracted off so that we need not worry about an intercept term.<sup>2</sup>

$$y_i = \beta z_i + \varepsilon_i = \beta x_i + \varepsilon_i - \beta \eta_i = \beta x_i + \tau_i \quad i = 1, \dots, n$$

$$x_i = z_i + \eta_i$$

Begin with  $y_i$  as the dependent variable, with the true regressor  $z_i$  and an error term  $\varepsilon_i$ . The problem is that  $z_i$  cannot be observed accurately—it is mismeasured. Instead, the observed variable is  $x_i$ , which, as the second equation shows, is equal to  $z_i$  plus an error in observation  $\eta_i$ , which is taken to have mean zero and to be uncorrelated with  $z_i$  and with  $\varepsilon_i$ . (I will assume, here and throughout the paper, that the sample is independent and identically distributed.) Thus, the equation actually estimated is a regression of  $y_i$  on  $x_i$ , with an error term  $\tau_i$  that mixes together the effects of the “true” error  $\varepsilon_i$  and the measurement error  $\eta_i$ .

The classic result is usually that the estimated coefficient calculated by the ordinary least squares regression based on the observable variable  $x_i$ —call it  $b$ —is biased downward from the coefficient that would have been calculated based on the true variable  $z$ . Thus, the classic result is that usually the least squares estimator  $b < \beta$ . This result follows because  $x_i$  in the above equation is negatively correlated with the stochastic disturbance term  $\tau_i$ .<sup>3</sup>

These results do not hold in general if the measurement error is correlated with  $z$  or with  $\varepsilon$ , although the downward bias still often occurs.<sup>4</sup> For instance, Aigner (1973) showed that with misclassification error, if  $z_i$  is a binary indicator variable, the ordinary least squares estimator  $b$  is again biased downward. At MIT I have called this the “Iron Law of Econometrics”—the magnitude of the estimate is usually smaller than expected. It is also called “attenuation” in the statistics literature. In the large sample result, the extent of the bias depends on the ratio of the variance of the “signal” (true variable) to the sum of the variance of the signal and the variance of the “noise” (error in measurement). As the variance of the noise becomes relatively larger compared to the variance of the signal, the bias increases. One situation where the bias increases in this manner is when the regression specification is panel data with fixed effects. Estimation of the fixed effects typically

<sup>2</sup> Other right-hand side variables, assumed to be measured without error, have been “partialled out” of the model. For the effect on the estimated coefficients of the other variables in the regression specification that have been “partialled out,” see Meijer and Wansbeck (2000).

<sup>3</sup> In terms of a large sample result,  $p \lim b = \alpha \beta < \beta$ , where  $\alpha = \sigma_z^2 / \sigma_x^2 = \sigma_z^2 / (\sigma_z^2 + \sigma_\eta^2) < 1$ . The result also holds in expectation in finite samples if the stochastic terms are both normal or in the non-normal case if the model conditional expectation is linear. Then the result is  $E(b) = \alpha \beta$ .

<sup>4</sup> A straightforward calculation demonstrates the effect of correlated measurement error on the estimator  $b$ .

increases the variance of the noise relative to the variance of the signal, as Griliches and Hausman (1986) discuss. Griliches and Hausman also demonstrate how to estimate the magnitude of the observation error and recommend determining if its size is “reasonable” in the context of the model. Card (2000) reviews empirical evidence of downward bias in estimation of the coefficient of the return to schooling from the many papers in the literature. He finds that the downward bias is often in the range of 25 percent to 33 percent, a substantial downward bias.

Another classic result arises if the left-hand side and right-hand side variables are interchanged in the regression specification. The inverse of the ordinary least squares estimator of the coefficient in the reverse regression,  $g$ , gives an upward-biased estimate of the true regression coefficient that would have been derived from a regression on  $z$ . Thus, the downward-biased estimate and the upward-biased estimate can be thought of as bracketing the true value of the coefficient so that the true coefficient  $b < \beta < g$ .<sup>5</sup> Indeed, a less well-known result tells the width of the bounds based on the regression  $R^2$ , which will be the same for both the original regression and the reverse regression. This equation holds that  $b/g = R^2$ . Thus, for cross-section econometrics where  $R^2$  is often about 0.3, then the  $b$  coefficient will be only one-third of the  $g$  coefficient, and the bounds will be quite wide. However, if the  $R^2$  is quite high, as is often (but not always) true in time series applications, mismeasurement will not have a large effect on the least squares estimate. For example, a test of the times series version of the “permanent income hypothesis,” in which the true regressor  $z$  is permanent income but the observed variable  $x$  is measured income, the estimate of the marginal propensity to consume will not be greatly affected by mismeasurement of permanent income, because the  $R^2$  in this time series data is quite high.

A last result is to focus on the consequences of a mismeasured left-hand side variable, where the right-hand side variable is measured without error. In this case, the ordinary least squares estimator would be unbiased under a wide range of assumptions. To understand this point, think of an observed left-hand variable  $q_i$  that is equal to the sum of the unobserved “true” left-hand variable  $y_i$  and a measurement error term  $\omega_i$  assumed to be uncorrelated with the error term  $\varepsilon_i$  in the regression. Then

$$\begin{aligned} y_i &= \beta z_i + \varepsilon_i & \text{and} & & q_i &= y_i + \omega_i, & \text{so} \\ q_i &= \beta z_i + \nu_i, & \text{where} & & \nu_i &= \varepsilon_i + \omega_i. \end{aligned}$$

Thus, when the left-hand side variable is mismeasured, the result is actually an error term on the right-hand side with increased variance, since the new error term  $\nu_i$  includes both a component for measurement error on the right-hand side and a

<sup>5</sup> This result has been known since 1921. If the error in measurement is correlated with the error in the equation, no bounds for the true coefficient exist. For a discussion of using prior information and for references to other generalizations considered in the literature, see Erickson (1993).

component for the original error expected in the regression. Overall, the outcome will be an unbiased estimate of  $\beta$ , but reduced precision in the estimate  $b$ , a lower  $t$ -statistic and a reduced  $R^2$ .

What happens if both left- and right-hand side variables are mismeasured, with errors of measurement uncorrelated with each other, the true variables, and with  $\varepsilon_i$ ? In this case, the estimate of the regression coefficient  $b$  remains biased downward, compared with the coefficient that would have been derived from a measurement of the “true”  $x$  variable. The extent of the bias is not changed by mismeasurement of the left-hand side variable. However, as a result of mismeasurement on the left-hand side, the coefficient  $b$  is measured with less accuracy, and the  $R^2$  of the regression will decrease. If one carries out the reverse regression estimation mentioned, the upward-biased coefficient  $g$  in the reverse regression will be expected to rise. Since  $b/g = R^2$ , when  $b$  remains the same and  $R^2$  declines,  $g$  must rise. Intuitively, when measurement error occurs in both the left- and right-hand side variables, the result is a less statistically accurate estimate of  $b$ , and because of the decline in statistical accuracy, the estimation bounds for the true coefficient also increase.

### **Instrumental Variables as a Solution to Mismeasurement: The Problem of Weak Instruments**

Most solutions to the mismeasurement problem for the linear specification in econometrics depend on the use of instrumental variables.<sup>6</sup> Instrumental variables are assumed to be correlated with the true  $z_i$  but uncorrelated with either the basic error term  $\varepsilon_i$  and/or the measurement error term  $\eta_i$ .<sup>7</sup> A typical approach here is to use two-stage least squares.<sup>8</sup> Let the vector of instrumental variables be  $w_i$ , and the instrumental variable estimator will use a linear combination of the  $w_i$  to achieve a consistent estimator  $b_{IV}$  of the true coefficient  $\beta$ .

The instrumental variables approach can be useful, if handled carefully. However, a significant understanding has emerged over the past few years that instrumental variables estimation of the errors-in-variables model can lead to problems of estimation and inference in the situation of “weak instruments.” Weak instruments problems arise when the instruments do not have a high degree of explanatory power for the mismeasured variable(s), when the size of the mismeasurement is large or when the number of instruments becomes large. In a situation

<sup>6</sup> Another solution to the measurement error problem is to find an independent measure of the reliability of the variable, usually from a resampling approach. This solution can be given an instrumental variables interpretation.

<sup>7</sup> Although the focus of the discussion here is on the basic situation that assumes no heteroskedasticity or serial correlation, it is worth noting that in more general situations of conditional heteroskedasticity, the White (1984) instrumental variables estimator should be used.

<sup>8</sup> The instrumental variables approach is discussed in this symposium in the paper by Angrist and Krueger.

of weak instruments, the danger is that the instrumental variables test will fail to discern a problem of mismeasured variables, even when significant (finite sample) bias is present, because the estimated standard errors are not very accurate.

Hahn and Hausman (1999) offer an approach to determining whether instrumental variable estimation is reliable in a particular situation. The new specification test, which takes the general approach of Hausman (1978), begins by carrying out a conventional two-stage least squares estimate of the coefficient of the right-hand side mismeasured variable. Then, it applies a two-stage least squares approach to the reverse regression using the same set of instruments. If the instrumental variable estimate is a reliable way of adjusting for measurement error, then the two estimates should be very similar. If they are not, then problems of inference are likely to be present. When the reverse regression test using two-stage least squares raises a warning flag that mismeasurement may be a problem, Hahn and Hausman then offer further specification tests and possible methods of estimation. This specification test may be quite useful in cross-section situations with a mismeasured right-hand side variable where the instruments do not have high explanatory power for the mismeasured variable.

## **Nonlinear Models with Mismeasured Variables**

In a nonlinear regression framework, with measurement error in the right-hand side variables, two-stage least squares estimates (in either linear or nonlinear form) no longer lead to consistent estimates of coefficients (Amemiya, 1985). The underlying problem is that the instruments will end up being correlated both with the mismeasured right-hand side variable and also with the error term in the econometric specification, as demonstrated by Hausman, Newey and Powell (1995).

One approach to this problem of nonlinearity is to use a polynomial specification. For example, allowing the right-hand side variables to take on powers of two, three, four or more allows the relationship between  $x$  and  $y$  to be a curving line. Hausman et al. (1991) demonstrate how to solve the problem of measurement error and how to achieve consistent estimates in the case of a polynomial specification. Again, begin with a regression where the unobserved independent variable  $z_i$  is replaced by  $x_i$ , which is observed with measurement error  $\eta_i$ . In this framework, Hausman et al. develop a consistent estimator under either of two sets of assumptions. The first is a “repeated measurement” specification, which can be applied to a situation where the same variable is measured more than once and where later measurements can be used to provide information about measurement errors in earlier estimates. The second specification uses instrumental variables.

Hausman, Newey and Powell (1995) use these approaches to estimate Engel curves for family expenditure in the United States, a subject that has long been an area of interest among econometricians. In recent research, it has been common to examine the “Leser-Working” form (Leser, 1963) of Engel curve, in which budget

shares spent on certain areas of household consumption—in this study, food, clothing, recreation, health care and transportation—are regressed on the log of income or expenditure. However, few studies of Engel curves have used estimators other than least squares or nonlinear least squares.

This study sought to estimate the Engel curves in a polynomial specification using a variety of methods to deal with measurement error: specifically, two different kinds of instrumental variable estimation and the “repeated measurement” approach. A first instrument for current expenditure is expenditure in future periods, which follows from a life-cycle model approach to consumption. A second set of instruments was based on determinants of income and expenditure, such as education and age. The study used the 1982 U.S. Consumer Expenditure Survey (CES), which collects data from families over four quarters, so that total expenditures from the second quarter of 1982 could be used to correct for mismeasurement of total expenditures in the first quarter of 1982.

Hausman, Newey and Powell (1995) find that the usual assumption of constant budget share elasticities, which is imposed by the Leser-Working specification, appears inconsistent with the data. They also find that a Hausman-type specification test (Hausman, 1978) of the instrumental variable estimates versus ordinary least squares estimates strongly rejects the ordinary least squares estimates, indicating the importance of the measurement error specification. Thus, there is strong evidence that using current expenditure in estimation of Engel curves on microdata involves an errors-in-variables problem.<sup>9</sup> In estimation of Engel curves, an ordinary least squares approach that uses current expenditure and does not account for measurement error leads to inconsistent estimates where the usual downward bias is present.

Perhaps most intriguing of all, Hausman, Newey and Powell (1995) offer an empirical investigation of a theoretical result from a notable paper by Gorman (1981). Gorman considered Engel curves in which either expenditure or budget shares are specified as polynomials in functions of expenditure. Given a certain functional form, Gorman demonstrates that the rank of the matrix of coefficients for the polynomial terms in income is at most three. More concretely, this finding implies that if one estimates two budget share equations, one based on a quadratic form and the other on a cubic form, the ratio of the coefficients of the cubic terms to the coefficients of the quadratic terms will be constant across budget share equations. Indeed, Hausman, Newey and Powell find a rather remarkable result that the ratios of the coefficients, estimated to take measurement error into

<sup>9</sup> In a recent doctoral thesis at MIT, Schennach (2000) has extended the Hausman et al. (1991) results from the polynomial specification to the consistent estimation of nonlinear models with measurement errors in the explanatory variables when one repeated observation exists. Thus, regression models with nonlinear functions such as  $g(z_i)$ , where  $z_i$  is an unobserved variable, can be estimated consistently. Her approach is considerably more flexible. Schennach also estimates nonlinear Engel curves using Consumer Expenditure Survey data since 1984 and finds that her approach, which allows for mismeasured right-hand side variables, is preferable to a traditional nonlinear least squares estimator that does not take account of mismeasured right-hand side variables.

account, are extremely close in actual values and estimated precisely. Specifically, the ratios of the coefficients for the different budget share equations are  $-25.0$ ,  $-25.2$ ,  $-25.1$ ,  $-3.3$  and  $-25.6$ . Thus, use of consistent estimates of a mismeasurement model for Engel curve analysis provides a situation in which a restriction derived from economic theory applies quite well to the data.

### Measurement Error in the Left-Hand Side Variables: Probit and Logit

In the usual linear regression specification, a mismeasured left-hand side variable does not lead to a biased coefficient, as discussed above, but only to less statistical precision in estimation. However, in certain contexts, misclassification of the left-hand side variable can lead to estimators that are biased and inconsistent. In particular, this situation often arises when the dependent variable is limited in some way: for example, in a probit or logit estimation where the dependent variable takes on only two values, zero or one.<sup>10</sup> For example, consider a case in which the left-hand variable is whether a person has changed jobs or not.

I now consider the consequences of mismeasurement in the left-hand side variable in binary outcome models where the observed left-hand side variable is a function of an unobserved (latent) dependent variable. The usual latent variable specification has an observed latent variable  $y_i^*$  and an observed variable  $y_i = 1$  if  $y_i^* \geq 0$  and  $y_i = 0$  if  $y_i^* < 0$ . The regression specification that allows for misclassification is

$$q_i = \alpha_0 + (1 - \alpha_0 - \alpha_1)F(z_i\beta) + \varepsilon_i,$$

where  $\alpha_0$  is the probability that  $q_i = 1$  although the true  $y_i = 0$ ,  $\alpha_1$  is the probability that  $q_i = 0$  although the true  $y_i = 1$ , and  $F$  is the cumulative probability distribution for the probit model or logit model. With no mismeasurement,  $\alpha_0 = \alpha_1 = 0$ , and this equation becomes a typical probit model or logit model. If one tries to estimate a standard logit or probit model without allowing for misclassification when it exists, the result will be biased and inconsistent estimates. However, Hausman, Abrevaya and Scott-Morton (1998) demonstrate that maximum likelihood estimation of this equation provides straightforward and consistent estimates so long as a monotonicity condition holds:  $\alpha_0 + \alpha_1 < 1$ . (This monotonicity condition is relatively weak, since it says that the combined probability of misclassification is not so high that on average you cannot tell which result actually occurred.) They also show that this equation provides a consistent coefficient if estimated by nonlinear

<sup>10</sup> Probit and logit estimation, and binary choice models more generally, are discussed by Horowitz and Savin in this symposium.



least squares methods. Indeed, the estimated coefficients of  $\alpha_0$  and  $\alpha_1$  provide a specification test for whether mismeasurement is a problem.

Thus, two results arise in the binary limited dependent variable cases that are different from the classical regression specification with measurement error in the left-hand side variable: 1) inconsistent estimation results from the mismeasurement; and 2) consistent estimation does not require instrumental variables although a (strong) distributional assumption is necessary. Hausman, Abrevaya and Scott-Morton (1998) use Monte Carlo experiments to demonstrate that relatively small amounts of misclassification, as little as 2 percent, can lead to significant amounts of bias even in large samples.

Hausman, Abrevaya and Scott-Morton (1998) also extend this analysis in a number of ways. They show that this approach extends readily to discrete response models with more than two categories. They also demonstrate that there is no need to assume a particular cumulative probability distribution, like the normal distribution that is assumed in the probit case, and that semiparametric estimation of the binary choice equation with misclassification allowed can also provide consistent estimation. Thus, the strong distributional assumption required for maximum likelihood or nonlinear least squares is no longer necessary. Consistent estimation follows without the need for using instrumental variables.

As an empirical example, Hausman, Abrevaya and Scott-Morton (1998) specify and estimate a model of whether people have changed jobs or not using both the Current Population Survey (CPS) and Panel Study of Income Dynamics (PSID) data sets. The results from the CPS data set, discussed here, demonstrate strong evidence of misclassification. Using a probit specification that allows for misclassification and applying maximum likelihood estimation, they find that  $\alpha_0$ , the probability of misclassification for non-job changers, is estimated to be 6 percent, and  $\alpha_1$ , the probability of misclassification for job changers, is estimated to be 31 percent. Both estimates of misclassification are very precisely estimated. Moreover, the estimates of many of the right-hand side coefficients change by large amounts when misclassification is permitted. Thus, misclassification appears to be a potentially serious problem in microdata.

Interestingly, Hausman, Abrevaya and Scott-Morton (1998) also find that semiparametric estimates of the extent of misclassification estimates are quite similar to the maximum likelihood estimates that allow for misclassification. This finding suggests that the maximum likelihood approach to probit or logit allowing for misclassification may give reasonable results in many actual empirical situations.

## **Measurement Error in the Left-Hand Side Variables: More General Models**

The binary choice model of error in left-hand side variables is relatively straightforward because with only two choices, only two sorts of errors are possible:

0 can be confused for 1, or 1 for 0. Abrevaya and Hausman (1999) suggest a more general model in which the observed left-hand side variable  $q_i$  is a stochastic function of the underlying  $y_i^*$ . They demonstrate that this model can be used to analyze a range of situations: binary choice with misclassification, mismeasured discrete dependent variables and mismeasured continuous dependent variables that arise in duration and hazard models.<sup>11</sup> They demonstrate that conventional estimation of these models leads to inconsistent results, except in special circumstances.

Abrevaya and Hausman (1999) offer methods for estimating all of these models, which will provide consistent estimates as long as the process of mismeasurement follows one sufficient condition. This condition is that the effect of the mismeasurement cannot on average permute the ordering of the observed left-hand side variables with respect to the ordering of the unobserved latent variables; that is, on average it cannot make more frequently occurring outcomes look like less frequently occurring outcomes. The effect of the mismeasurement cannot be “too large” on average.<sup>12</sup> The appropriate question to ask to satisfy the sufficient condition is: “Are observational units with larger ‘true’ values for their left-hand side variable more likely to report larger values than observational units with smaller ‘true’ values?” If the answer is “yes,” the Abrevaya and Hausman techniques can be used. Again, instrumental variables are not required for consistent estimation.

Abrevaya and Hausman (1999) focus on the increasingly used duration and hazard models (for example, Han and Hausman, 1990; Meyer, 1990). In a duration model, the left-hand variable might measure, for example, the amount of time that someone has spent at a current job. A hazard model, in the context of the return to employment for an unemployed person, answers the question of what is the probability of becoming employed in the next time period conditional on being unemployed up to the previous time period. Abrevaya and Hausman demonstrate that conventional estimation of commonly used duration models and hazard models is usually inconsistent when the left-hand side variable is mismeasured and provides coefficients that are biased toward zero.<sup>13</sup>

<sup>11</sup> The general approach discussed here does not cover mismeasurement in the case of the quantile regression estimator. Quantile regression, as discussed in this symposium by Koenker and Hallock, is robust to certain data features that create problems for the classical linear regression model, but it is not robust to a mismeasured left-hand side variable, which does not create problems for the classical linear regression model. The presence of the error in measurement on the left-hand side biases the quantile coefficient estimates toward the coefficient for the median regression estimate. Monte Carlo experiments demonstrate that the bias can be quite substantial when conditional heteroskedasticity is present. Finding a solution to this problem is a good topic for future research.

<sup>12</sup> In more precise terms, the sufficient condition is that the distribution for the observed variable  $q_i$  for a higher latent  $y_i^*$  stochastically dominates the distribution of  $q_j$  for a lower latent  $y_j^*$ . The use of the notion of first-order stochastic dominance is familiar from microeconomics, where it is used to order portfolios or risky distributions (for example, Mas-Colell, Whinston and Green, 1995, p. 195).

<sup>13</sup> The single exception is the highly restrictive Weibull duration model. However, the Weibull specification requires a monotonic (baseline) hazard, which makes it too restrictive for most problems. Also, the particular application cannot allow for censoring, which is typically present in applications of duration models.

Mismeasured durations are common to all data sets that have survey responses. Some common findings are that in the Current Population Survey, about 37 percent of unemployed workers overstated unemployment durations, longer spells have a higher proportion of reporting errors, and there are often “focal responses” (for instance, at a number of weeks that corresponds to an integer month amount, like four or eight weeks). In duration models that do not allow for mismeasurement, it is common to find that the size of the unemployment insurance benefit has a statistically significant effect on the duration of unemployment. However, after allowing for mismeasurement, Abrevaya and Hausman (1999) find a much smaller and statistically insignificant effect of the unemployment insurance benefit levels on unemployment duration. It appears that mismeasurement in the left-hand side variable can have an important effect in duration models.

■ *This paper is given in memory of Zvi Griliches. Jason Abrevaya, Jin Hahn and the editors provided helpful comments.*

## References

- Abrevaya, Jason and Jerry Hausman.** 1999. “Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells.” *Annales D’Economie et de Statistique*. 55–56, pp. 243–75.
- Adcock, R. J.** 1878. “A Problem in Least Squares.” *Analyst*. 5, pp. 53–4.
- Aigner, Dennis J.** 1973. “Regression with a Binary Independent Variable Subject to Errors of Observation.” *Journal of Econometrics*. March, 1:1, pp. 49–59.
- Aigner, Dennis J. et al.** 1984. “Latent Variable Models in Econometrics,” in *Handbook of Econometrics, Volume II*. Z. Griliches and M. D. Intriligator, eds. Amsterdam, New York and Oxford: North-Holland, pp. 1323–93.
- Amemiya, Yasuo.** 1985. “Instrumental Variable Estimator for the Nonlinear Errors-in-Variables Model.” *Journal of Econometrics*. June, 28:3, pp. 273–89.
- Buchinsky, Moshe.** 1994. “Changes in the U.S. Wage Structure 1963–1987: Application of Quantile Regression.” *Econometrica*. March, 62:2, pp. 405–58.
- Card, David.** 2000. “Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems.” NBER Working Paper No. W7769. *Econometrica*. Forthcoming.
- Cavanagh Christopher and Robert P. Sherman.** 1998. “Rank Estimators for Monotonic Index Models.” *Journal of Econometrics*. June, 84:2, pp. 351–81.
- Cox, D. R.** 1972. “Regression Models and Life Tables (with discussion).” *Journal of the Royal Statistical Society*. B, 34, pp. 187–220.
- Erickson, Timothy.** 1993. “Restricting Regression Slopes in the Errors-in-Variables Model by Bounding the Error Correlation.” *Econometrica*. July, 61:4, pp. 959–69.
- Fuller, Wayne.** 1987. *Measurement Error Models*. New York: John Wiley & Sons.
- Gorman, W. M.** 1981. “Some Engel Curves,” in *Essays in the Theory and Measurement of Consumer Behaviour in Honor of Sir Richard Stone*. A. Deaton, ed. Cambridge: Cambridge University Press, pp. 7–29.
- Griliches, Zvi.** 1986. “Economic Data Issues,” in *Handbook of Econometrics, Volume III*. Z. Griliches and M. D. Intriligator, eds. Amsterdam, New York and Oxford: North-Holland, pp. 1465–514.
- Griliches, Zvi and Jerry Hausman.** 1986. “Er-

rors in Variables in Panel Data." *Journal of Econometrics*. February, 31:1, pp. 93–118.

**Hahn, Jinyong and Jerry Hausman.** 1999. "A New Specification Test for the Validity of Instrumental Variables." Massachusetts Institute of Technology, Department of Economics, Working Paper, 99/11. *Econometrica*. Forthcoming.

**Han, Aaron K.** 1987. "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator." *Journal of Econometrics*. July, 35:2-3, pp. 303–16.

**Han, Aaron K. and Jerry Hausman.** 1990. "Flexible Parametric Estimation of Duration and Competing Risk Models." *Journal of Applied Econometrics*. January/March, 5:1, pp. 1–28.

**Hausman, Jerry.** 1978. "Specification Tests in Econometrics." *Econometrica*. November, 46:6, pp. 1251–71.

**Hausman, Jerry, Jason Abrevaya and F. M. Scott-Morton.** 1998. "Misclassification of the Dependent Variable in a Discrete-Response Setting." *Journal of Econometrics*. December, 87:2, pp. 239–69.

**Hausman, Jerry, Whitney K. Newey and J. L. Powell.** 1995. "Nonlinear Errors in Variables: Estimation of Some Engel Curves." *Journal of Econometrics*. January, 65:1, pp. 205–33.

**Hausman, Jerry et al.** 1991. "Measurement Er-

rors in Polynomial Regression Models." *Journal of Econometrics*. December, 50:3, pp. 273–95.

**Koenker, Roger W. and Gilbert Bassett, Jr.** 1978. "Regression Quantiles." *Econometrica*. January, 46:1, pp. 33–50.

**Leser, C. E. V.** 1963. "Forms of Engel Functions." *Econometrica*. October, 31:4, pp. 694–703.

**Liviatan, Nissan.** 1961. "Errors in Variables and Engel Curve Analysis." *Econometrica*. July, 29:3, pp. 336–62.

**Mas-Colell, Andreu, Michael D. Whinston and Jerry R. Green.** 1995. *Microeconomic Theory*. Oxford: Oxford University Press.

**Meijer, Erik and Tom Wansbeek.** 2000. "Measurement Error in a Single Regressor." *Economic Letters*. December, 69:3, pp. 277–84.

**Meyer, Bruce D.** 1990. "Unemployment Insurance and Unemployment Spells." *Econometrica*. July, 58:4, pp. 757–82.

**Nagar, A. L.** 1959. "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations." *Econometrica*. October, 27, pp. 575–95.

**Schennach, Susanne.** 2000. "Estimation of Nonlinear Models with Measurement Error." Massachusetts Institute of Technology, Ph.D. Thesis.

**White, Halbert.** 1984. *Asymptotic Theory for Econometricians*. Orlando: Academic Press.