CIRJE-F-658

Non-Classical Measurement Error in Long-Term Retrospective Recall Surveys

John Gibson
University of Waikato

Bonggeun Kim
Seoul National University

September 2009

Non-classical measurement error in long-term retrospective recall surveys

John Gibson[1]
and
Bonggeun Kim[2]

August 29, 2009

Abstract

Applied microeconomic researchers are beginning to use long-term retrospective survey data in settings where conventional longitudinal survey data are unavailable. However, inaccurate long-term recall could induce non-classical measurement error, for which conventional statistical corrections are less effective. In this paper, we use the unique Panel Study of Income Dynamics Validation Study to assess the accuracy of long-term retrospective recall data. We find underreporting of transitory variation which creates a non-classical measurement error problem.

JEL: C33, J64

Keywords: Earnings, mean reversion, measurement error, retrospective surveys

Acknowledgements
We are grateful to comments from Ole Rummel, an anonymous referee and the editor, and seminar audiences at the Universities of Canterbury and Otago, the Labour Econometrics Workshop at the University of New South Wales and the International Conference on Measurement Error at Aston University. Financial support from Marsden Fund grant UOW0504 is gratefully acknowledged.

[1]Department of Economics, University of Waikato, New Zealand.
[2]Department of Economics, Seoul National University, Korea.

I. Introduction

Applied microeconomic researchers are beginning to use long-term retrospective survey data when conventional longitudinal survey data are unavailable. For example, Fleisher and Wang (2005) study the changing returns to schooling in China over 40 years, based on lifetime histories collected by Zhou (2000). These data include retrospective information on each respondent's income in 1955, 1960, 1965, 1975, 1978, 1984, 1987, and 1991-94, as remembered in 1994. Boucher, Smith, Taylor, and Yúnez-Naude (2007) use a dynamic panel estimator to examine how U.S. policy initiatives such as NAFTA and immigration reform affected the supply of Mexican labour to U.S. farms. They estimate their model on a sample of 340 Mexican households who were interviewed in 2003 about their migration histories from 1980 to 2002.

Practitioners who use these long-term retrospective survey data may well prefer to use more conventional longitudinal surveys, where respondents are revisited several times and typically have to recall only over the year prior to the latest interview. But in many settings, particularly developing countries and for hard to survey groups like immigrants, longitudinal surveys either may not exist or have started too recently to provide sufficient waves of data for analysis. Since longitudinal data are needed for transition studies, such as between locations (e.g., migration), and also allow individual fixed effects to be controlled for so as to alleviate estimator biases due to unobservable factors like ability, long-term retrospective survey data may seem a tempting alternative. However, inaccurate long-term recall could induce non-classical measurement error, for which conventional statistical corrections are less effective. Inaccuracy may occur because respondents either completely forget events or mis-date them. For example, many unemployment spells are forgotten in retrospective interviews (Jacobs, 2002) and transitions out of unemployment are often inconsistently dated (Paull, 2002).

The objective of this note is to test the accuracy of long-term retrospective survey data and to examine the nature of the measurement error in such data. We use the unique Panel Study of Income Dynamics Validation Study (PSIDVS) to compare long-term retrospective reports of earnings with more conventional longitudinal survey data gathered by repeatedly interviewing respondents over several years. While several previous studies compare retrospective recall data with standard longitudinal survey data collected more frequently (Peters, 1988; Pierret, 2001; Smith and Thomas, 2003), they do not validate data from either type of survey. Consequently there is continued debate about the accuracy of long-term retrospective recall and whether data gathered in this manner can be a substitute for more conventional longitudinal surveys (Kennickell and Starr-McCluer, 1997; Campbell, 2000; Jacobs, 2002). In contrast, the PSIDVS contains accurate information on labour market outcomes from a company's records (which acts as a "gold standard"). Previous analysis with PSIDVS has compared longitudinal survey data with the gold standard (Pischke, 1995) but has not included the retrospective recall in the comparisons. Moreover, unlike the cross-sectional validation studies reported by Bound et al. (2001), since the PSIDVS was conducted in two waves four years apart, it also provides information on measurement error in retrospectively recalled *changes* in variables.

II. Data

We use three types of PSIDVS data: company records from 1981 to 1986 that provide the validation information; the longitudinal survey data gathered each year, referring to the previous year; and the long-term retrospective recall data that were gathered in 1987 but refer to each year from 1981 to 1986. Comparisons with the validation data allow us not only to identify any long-term recall bias, but also to measure its size relative to the bias (which was shown to exist by

Pischke (1995)) in the contemporaneously surveyed longitudinal survey data. These comparisons can also establish whether the recall errors are non-classical (e.g., mean-reverting), which would make them contrary to the assumptions used in most treatments of measurement error.

Table 1 contains descriptive statistics for the company records, surveyed earnings, and recalled earnings of the 219 sample workers in the PSIDVS sample.[1] Retrospectively recalled earnings initially appear to be a good proxy for true earnings. The ratios of the means of company records of earnings to the sample means of log earnings in the retrospectively recalled data are very close to one, ranging from 0.997 to 1.002.

However, measures of variance for recalled data do not appear accurate, with ratios of company to recalled records from 0.763 to 1.275. This variance ratio, $Var(\ln y_{it}^{true})/Var(\ln y_{it}^{recalled})$ is a reliability ratio under classical measurement error, showing the proportion of true to observed variation. But two of the variance ratios in Table 1 exceed one, so the classical measurement error reliability ratio interpretation does not hold in this case since adding uncorrelated (classical) measurement error would always make the denominator exceed the numerator. Because the recalled 1981 and 1982 earnings show smaller variance than the true earnings, a negative correlation between true earnings and the recall errors is implied.

III. The Measurement Error Model

A hypothesis suggested by the pattern in Table 1 is that people in long-term retrospective surveys under-report transitory variations. Specifically, when asked to report their earnings in earlier years, people tend to report their usual earnings. To examine this hypothesis, note that annual earnings can be written as a sum of two components:

---

[1] Pischke's panel sample size is 234 which is reduced to 219 in our analysis with retrospectively recalled earnings.

$$\ln w_{it} = \ln w_{it}^{P} + \ln w_{it}^{T} = (\alpha_i + \gamma_1 X_{it}) + (\gamma_2 U_t + \varepsilon_{it}) \tag{1}$$

where $\ln w_{it}$ is the i^{th} worker's log real annual earnings in year t, $\ln w_{it}^{P}$ is the permanent or usual component, which consists of the fixed effect α_i representing the combined effect of time-invariant characteristics of worker i and X_{it}, which is worker i's years of work experience as of year t to represent systematic wage growth and $\ln w_{it}^{T}$ is the transitory component, which can be affected by the business-cycle (U_t) or just individual specific transitory events (ε_{it}).

Differencing to either ameliorate omitted variable bias by removing fixed effects or for directly studying earnings growth gives:

$$y_{i,t(s)} \equiv \ln w_{it} - \ln w_{i,t-s} = (\ln w_{it}^{P} - \ln w_{i,t-s}^{P}) + (\ln w_{it}^{T} - \ln w_{i,t-s}^{T}) = s\gamma_1 + y_{i,t(s)}^{T}, \tag{2}$$

where $y_{i,t(s)}^{T}$ represents the true transitory earnings variation. But for a practitioner who is not able to observe true earnings growth $y_{i,t(s)}$ and instead has to work with data from a long-term retrospective survey, possible measurement error has to be added to equation (2):

$$y_{i,t(s)}^{*} = s\gamma_1 + y_{i,t(s)}^{T} + m_{i,t(s)} + v_{i,t(s)} \tag{3}$$

where $m_{i,t(s)}$ is a method effect for *s* recall period and $v_{i,t(s)}$ is a pure random error. If there is underreporting of transitory variation in a long-term retrospective recall survey, $m_{i,t(s)}$ will be negatively correlated with the transitory variation. Hence, the method effect can be expressed as:

$$m_{i,t(s)} = \pi_s y_{i,t(s)}^{T} \tag{4}$$

and combining equations (2)-(4) gives:

$$\begin{aligned} y_{i,t(s)}^{*} &= s\gamma_1 + y_{i,t(s)}^{T} + m_{i,t(s)} + v_{i,t(s)} \\ &= s\gamma_1 + \lambda_s y_{i,t(s)}^{T} + v_{i,t(s)} \\ &= \theta + \lambda_s y_{i,t(s)} + v_{i,t(s)} \end{aligned} \tag{5}$$

where $\theta = (1-\lambda_s)s\gamma_1$ and $\lambda_s = (1+\pi_s)$ represents the potential negative correlation between the true values and the method effect in the measurement error.

Classical measurement error is a special case of equation (5) where $\lambda_s = 1$ and $\theta = 0$. But with correlated errors (e.g. from a long-term retrospective recall survey underreporting the transitory variation), $\pi_s < 0$ and (as long as recalled earnings growth is still positively correlated with true values) the measurement error follows a mean-reverting pattern ($0 < \lambda_s < 1$). The mis-measured earnings growth may be above or below true earnings growth:

$$E(y^*_{i,t(s)}) = \theta + \lambda_s E(y_{i,t(s)}) \gtreqless E(y_{i,t(s)}) \quad iff \quad \theta \gtreqless (1-\lambda_s) E(y_{i,t(s)}). \tag{6}$$

One special case is when people ignore all transitory variation ($\theta = E(y_{i,t(s)}), \lambda_s = 0$).

Measurement error also biases the estimated variance of earnings growth. Unlike the case of classical measurement error, where the variance of the true variable is always less than that of the error-ridden variable,[2] with non-classical error the variance of the error-ridden variable may be less than that of the true variable:

$$Var(y^*_{i,t(s)}) = \lambda_s^2 Var(y_{i,t(s)}) + Var(v_{i,t(s)}) \gtreqless Var(y_{i,t(s)}) \quad iff \quad Var(v_{i,t(s)}) \gtreqless (1-\lambda_s^2) Var(y_{i,t(s)}). \tag{7}$$

Implications

The biases in equations (6) and (7) under the more general framework of non-classical measurement error undermine two tenets of conventional wisdom about the impacts of classical measurement error on a linear regression: (i) error in the dependent variable causes no bias in slope coefficients and (ii) error in the (single) independent variable causes downward (attenuation) bias. Consider some true (bivariate) model:

$$y = \alpha + \beta x + u, \tag{8}$$

[2] This allows the interpretation as a reliability ratio, which cannot exceed 1.0.

where u is a pure random error. If the observed dependent variable is $y^* = \theta + \lambda y + v$ (like equation (5)), a proportional bias (λ) in estimating β results:

$$\beta_{y^*x} = \frac{\text{cov}(y^*, x)}{\text{var}(x)} = \frac{\text{cov}(\lambda\alpha + \lambda\beta x + \lambda u - v, x)}{\text{var}(x)} = \lambda\beta \qquad (9)$$

Mean reverting error in dependent variables tends to make estimated regression coefficients too small in magnitude, which is contrary to the textbook case of no bias in slope coefficients.

For error in the independent variable, with observed $x^* = \theta + \lambda x + v$, the resulting estimate of β in the population regression of equation (8) is

$$\beta_{yx^*} = \frac{\text{cov}(y, x^*)}{\text{var}(x^*)} = \frac{\text{cov}(\alpha + \frac{\beta}{\lambda}x^* - \frac{\beta\theta}{\lambda} - \frac{\beta}{\lambda}v + u, x^*)}{\text{var}(x^*)} = \beta\frac{\lambda\sigma_x^2}{\lambda^2\sigma_x^2 + \sigma_v^2} \qquad (10)$$

The true parameter β is rescaled by $\lambda\sigma_x^2/(\lambda^2\sigma_x^2 + \sigma_v^2)$. Unlike the classical error case this bias parameter could be greater than one, an upward bias, if the 'shrinkage' of the variance in the first term in the denominator due to multiplying by λ^2 (for $0<\lambda<1$) exceeds the effect of adding the variance of the random noise term, *v* to the denominator.

Hypothesis Tests

The above discussion shows that two hypothesis tests about the nature of the measurement error in long-term retrospective recall data are required: first, whether the measurement error is mean-reverting ($0<\lambda<1$), and second, whether the degree of mean-revision (given by λ) relative to the size of the uncorrelated error (σ_v^2) is sufficiently large so as to cause an upward bias when the error-ridden variable is on the right-hand side of a regression. The first test is simply: $H_{0:}$ λ_s=1 versus $H_{1:}$ λ_s<1 for the model in equation (5) and the result directly informs us about the degree of bias in equation (9).

The second test is whether the variance of earnings growth with the mis-measured variable equals the true variance of earnings growth or instead is less than the true variance:

$$H_0: \sigma_{x^*}^2 = \sigma_x^2 \quad vs. \quad H_1: \sigma_{x^*}^2 < \sigma_x^2$$

where we now index the variable by *x* since this test directly informs us about the nature of the bias when the mis-measured variable is on the right-hand side (RHS) of a bivariate model. If the variance of the error-ridden variable is smaller than that of the true variable, the resulting direction of bias when the mis-measured variable is on the RHS is unknown (equation (10)). A sufficient condition to have the traditional attenuation bias (i.e., a reliability ratio less than one) is that the sum of the variance of measurement error and the (negative) covariance between the true variable and its measurement error should be positive. In other words, the bias coefficient multiplying β in the last term of equation (10) needs to be less than the ratio of variances:

$$\frac{\lambda\sigma_x^2}{\lambda^2\sigma_x^2+\sigma_v^2} < \frac{\sigma_x^2}{\lambda^2\sigma_x^2+\sigma_v^2} = \frac{\sigma_x^2}{\sigma_{x^*}^2} \qquad if \;\; 0<\lambda<1 \tag{11}$$

Hence the negative correlation between the true variable and its error (as shown by λ) should not be too strong (i.e., too small relative to σ_v^2). This sufficient condition also plays an important role in whether bounding parameter estimates are feasible, as discussed below.

Results

The PSIDVS data allow us to test both hypotheses since the company records provide measures of true earnings growth. The tests are carried out for all years available (i.e., all values of $y^*_{i,t(s)}$ as the recall period *s* changes). If the length of the recall period affects the magnitude of recall bias, mean-reversion would be greater for longer recall, so λ_s would decline as *s* increases. The results in the first two columns of Table 2 show that the hypothesis that $\lambda_s = 1$ is strongly rejected in each year in favour of the alternative hypothesis of mean-reverting errors. There is considerable mean reversion, with all $\lambda_s < 0.5$ except $y^*_{i,86(2)}$. The estimates of $\hat{\lambda}_s$ appear to decline somewhat as the recall period lengthens although the trend is complicated, perhaps due to other factors such as business cycles.

In addition to examining the degree of mean reversion relative to the company records it is interesting to compare with the longitudinally surveyed earnings. The last row of Table 2 reports these results, for the only pair of years (1982 and 1986) having both longitudinally and retrospectively reported information. The degree of mean-reversion in the long-term retrospective recall is almost as apparent as in the comparison with the company records.

The results in the last columns of Table 2 show that the null hypothesis of equal variances is strongly rejected in all years but one, in favour of the alternative hypothesis of $\sigma^2_y > \sigma^2_{y^*}$. The negative correlation between the true variable and its measurement error is either very strong or the size of uncorrelated errors (σ^2_v) is small relative to the signal (σ^2_y) except for $y^*_{i,86(2)}$.

An Example

An example where the bias described may occur is if one used long-term retrospective recall data to test whether earnings vary counter-cyclically, non-cyclically, or pro-cyclically with the business cycle (Kim and Solon, 2005). As shown in equation (9), when the dependent variable has mean-reverting error the coefficient on the right-hand side variable (a business cycle indicator in this example) is not the original wage cyclicality parameter β, but rather β rescaled by the measurement error parameter λ. Relative to using conventional longitudinal data, the measured degree of mean-reversion with the retrospective survey data may lead to a further 60% underestimation of the pro-cyclicality of real wages, (as seen from $\hat{\lambda}_s$ in the last row of Table 2).

IV. Possible Statistical Corrections

A conventional solution for classical errors-in-variable bias is IV estimation but this does not work properly for correlated (and mean-reverting) errors, as shown in Black et al. (2000). Reverse regression to form bounding estimators of the unknown true effect (for the simple bivariate linear regression) is another approach. Specifically, when the error in the dependent variable is mean-reverting, as in the recall bias with $0 < \lambda < 1$, one could use the conventional OLS estimate as a lower bound and the inverse slope of the reverse regression coefficient estimate as an upper bound. As shown in equation (9) above, the conventional OLS estimate in the population regression of y^* on x is a lower bound as $\beta_{y^*x} = \lambda\beta < \beta$ when we normalize the data so that $\beta > 0$ **.** But the conventional upper bounding property is not always satisfied, in contrast to the argument of Black et al. (2000) since $\frac{1}{\beta_{xy^*}} = \beta\frac{\lambda^2\sigma_y^2 + \sigma_v^2}{\lambda\sigma_y^2} \gtrless \beta$ as shown in equation (10). Therefore the condition for when the inverse of the slope coefficient in the reverse regression is an upper bound may not hold when there are strong mean-reversions, as in Table 2.

Similarly, in the case of an independent variable with negatively correlated error, the inverse estimate in the population regression of x^* on y can be used as an upper bound as in

$$\frac{1}{\beta_{x^*y}} = \frac{V(y)}{Cov(y,x^*)} = \frac{Var(y)}{Cov(y, \frac{\lambda}{\beta}y - \frac{\lambda\alpha}{\beta} - \frac{\lambda}{\beta}u + \theta + v)} = \frac{\beta}{\lambda} > \beta \quad . \tag{12}$$

However, the use of the conventional OLS estimate as a lower bound is not guaranteed, since equation (10) shows that the absolute value of the OLS estimate of β could exceed the true value. Thus, as a statistical correction, bounding parameter estimates for errors-in-variable bias often may not be feasible for mean-reverting errors-in-variable bias in either an error-ridden dependent variable or error-ridden independent variable.

V. Conclusion

In this note, we assess the accuracy of long-term retrospective survey data. Our results from the Panel Study of Income Dynamics Validation Study suggest that long-term retrospective recall is a poor substitute for genuine longitudinal data. We find underreporting of transitory variation. The resulting error is non-classical, which is unlikely to be properly handled by conventional correction methods such as IV estimation and bounding parameter estimates.

References

Black, D., Berger, M., and Scott, F. (2000) "Bounding parameter estimates with nonclassical measurement error," *Journal of the American Statistical Association* 95(451): 739-748.

Boucher, S., Smith, A., Taylor, J.E., and Yúnez-Naude, A. (2007) "Impacts of policy reforms on the supply of Mexican labor to U.S. farms: New evidence from Mexico" *Review of Agricultural Economics* 29(1): 4-16.

Bound, J., Brown, C., and Mathiowetz, N. (2001) "Measurement error in survey data," In *Handbook of Econometrics* Vol. 5. J. Heckman and E. Leamer (eds), Elsevier, pp. 3705-3843.

Campbell, R. T. (2000) "Longitudinal research" in E. F. Borgatta & R. J. V. Montgomery (Eds.), *Encyclopedia of Sociology* (2nd ed., Vol. 3, pp. 1683-1696). Detroit: Macmillan.

Fleisher, B., and Wang, X. (2005) "Returns to schooling in China under planning and reform" *Journal of Comparative Economics* 33(2): 265-277.

Jacobs, S. (2002) "Reliability and recall of unemployment events using retrospective data." *Work, Employment and Society, 16*(3): 537-548.

Kennickell, A., and Starr-McCluer, M. (1997) "Retrospective reporting of household wealth: evidence from the 1983-89 Survey of Consumer Finances." *Journal of Business and Economic Statistics* 15(4): 452-463.

Kim, B., and Solon, G. (2005) "Implications of mean-reverting measurement error for longitudinal studies of wages and employment" *Review of Economics and Statistics* 87(1), pp.193-96.

Paull, G. (2002) "Biases in the reporting of labour market dynamics" *Institute for Fiscal Studies Working Paper* No. 02/10.

Peters, E. (1988). "Retrospective versus panel data in analyzing lifecycle events." *Journal of Human Resources, 23*(4): 488-513.

Pierret, C. R. (2001) "Event history data survey recall: An analysis of the National Longitudinal Survey of Youth 1979 recall experiment," *Journal of Human Resources,* 36(3): 439-466.

Pischke, J-S. (1995) "Measurement error and earnings dynamics: some estimates from the PSID Validation Study," *Journal of Business and Economics Statistics* 13(3): 305-314.

Smith, J., and Thomas, D. (2003) "Remembrances of things past: Test–retest reliability of retrospective migration histories." *Journal of the Royal Statistical Society: Series A (Statistics in Society), 166*(1), 23-49.

Zhou, X. (2000) "Economic transformation and income inequality in urban China: evidence from panel data," *American Journal of Sociology* 105(4): 1135-1174.

Table 1. Sample Statistics for Levels of Annual Earnings in the PSIDVS Data, N=219

Variable	*Mean*	*Ratio with company record,* $\frac{E(\ln w_{it})}{E(\ln w_{it}^*)}$	*Var.*	*Ratio with company record ,* $\frac{Var(\ln w_{it})}{Var(\ln w_{it}^*)}$	F-Test for the homogeneous variances $H_0{:}\sigma^2_{\ln w^*}=\sigma^2_{\ln w}$ $H_1{:}\sigma^2_{\ln w^*}<\sigma^2_{\ln w}$
					$p=$ $\Pr(F>\frac{\hat{\sigma}^2_{\ln w}}{\hat{\sigma}^2_{\ln w^*}})$
Company Records					
$\ln y_{i81}$ (log1981 real annual earnings, company record)	10.292		.0673		
$\ln y_{i82}$	10.408		.0820		
$\ln y_{i83}$	10.344		.0619		
$\ln y_{i84}$	10.410		.0723		
$\ln y_{i85}$	10.468		.0431		
$\ln y_{i86}$	10.476		.0505		
Surveyed Earnings					
$\ln y^m_{i82}$ (log1982 real annual earnings, surveyed in 1983)	10.414	.999	.0861	0.952	$p=.640$
$\ln y^m_{i86}$ (log1986 real annual earnings, surveyed in 1987)	10.485	.999	.0585	0.863	$p=.860$
Recalled Earnings					
$\ln y^r_{i81}$ (log1981 real annual earnings, recalled in 1987)	10.286	1.001	.0602	1.118	$p=.204$
$\ln y^r_{i82}$	10.430	.997	.0643	1.275	$p=.036$
$\ln y^r_{i83}$	10.352	.999	.0668	0.926	$p=.712$
$\ln y^r_{i84}$	10.388	1.002	.0917	0.787	$p=.960$
$\ln y^r_{i85}$	10.451	1.001	.0564	0.763	$p=.976$
$\ln y^r_{i86}(\equiv \ln y^m_{i86})$	10.485	.999	.0585	0.863	$p=.860$

Note: Variables with superscript *m* are from the longitudinal survey carried out each year, those with superscript *r* are from the retrospective survey carried out in 1986 and those without superscripts are from the company records.

Table 2: Tests for Correlated Errors of Recalled Earnings Differentials

	Mean-revision parameter		Variance Ratio	
Independent variable	$\hat{\lambda}$ (S.E.)	t-Test for Correlated Errors $H_0 : \lambda = 1$ $H_1 : \lambda < 1$	$\sigma_y^2 / \sigma_{y^*}^2$	F-Test for the homogeneity of variances $H_0 : \sigma_{y^*}^2 = \sigma_y^2$ $H_1 : \sigma_{y^*}^2 < \sigma_y^2$
Company Records as Gold-Standard		$p = \Pr(t < \hat{t})$		$p = \Pr(F > \hat{\sigma}_y^2 / \hat{\sigma}_{y^*}^2)$
$y_{i,86(1)} \equiv \ln w_{i86} - \ln w_{i85}$	.419 (.046) [a]	p=.000	1.569	p=.001
$y_{i,86(2)}$	.727 (.040)	p=.000	1.118	p=.204
$y_{i,86(3)}$	.292 (.051)	p=.000	1.526	p=.001
$y_{i,86(4)}$	.410 (.036)	p=.000	2.255	p=.000
$y_{i,86(5)}$	.304 (.041)	p=.000	2.117	p=.000
Surveyed Earnings as Gold-Standard				
$y_{i,86(4)} \equiv \ln w_{i86}^m - \ln w_{i82}^m$	.450 (.033)	p=.000	2.310	p=.000

[a] Standard Errors in the parenthesis.