

UNIVERSITY OF WAIKATO

**Hamilton
New Zealand**

**Measurement Error in Long-term Retrospective Recall Surveys
Of Earnings**

John Gibson
University of Waikato

Bonggeun Kim
University of Waikato and Sungkyunkwan University, Korea

Department of Economics

Working Paper in Economics 3/07

March 2007

Corresponding Author

John Gibson

Department of Economics
University of Waikato,
Private Bag 3105,
Hamilton, New Zealand

Fax: +64 (7) 838 4331

Email: jkgibson@waikato.ac.nz

Abstract

Several recent studies in labour and population economics use retrospective surveys to substitute for the high cost and limited availability of longitudinal survey data. Although a single interview can obtain a lifetime history, inaccurate long-term recall could make such retrospective surveys a poor substitute for longitudinal surveys, especially if it induces non-classical error that makes conventional statistical corrections less effective. In this paper, we use the unique Panel Study of Income Dynamics Validation Study to assess the accuracy of long-term recall data. We find underreporting of transitory events. This recall error creates a non-classical measurement error problem. A limited cost-benefit analysis is also conducted, showing how savings from using a cheaper retrospective recall survey might be compared with the cost of applying the less accurate recall data to a specific policy objective such as designing transfers to reduce chronic poverty.

Keywords

longitudinal data

measurement error

retrospective surveys

JEL Classification

C33, J64

Acknowledgements

We are grateful to comments from seminar audiences at the Universities of Canterbury and Otago and the Labour Econometrics Workshop at the University of New South Wales and to financial support from Marsden Fund grant UOW0504.

I. Introduction

Research in labour and population economics increasingly requires longitudinal data. Such data allow studies of transitions, such as into and out of poverty and unemployment, and also allow biases due to unobservable factors like ability to be alleviated in econometric models that include fixed individual effects. However, longitudinal surveys are costly and are often restricted to small, nationally unrepresentative, samples (Deaton, 1997). One response to this high cost is to use retrospective surveys where a single interview obtains a long-term or even lifetime history (Freedman et al, 1988). For example, the Health and Retirement Study asks about lifetime fertility and the Malaysian Family Life Surveys collect recalled information over a decade (Beckett et al. 2001). In addition to lower cost, other advantages of long-term retrospective recall are that more than one cohort can be studied at a time and sample attrition is less of a problem (Kosloski et al, 1994).

But long-term recall data are possibly inaccurate, although the literature continues to debate the issue. For example, Jacobs (2002: 545) claims that analyses based on such data will be “quite meaningless” and Kennickell and Starr-McCluer (1997: 462) claim that it is a “poor substitute for panel data”. On the other hand, others suggest that when retrospective questions are asked carefully and interviewers are well trained, respondents can provide “accurate and detailed information” (Campbell, 2000, p. 1685). Errors may occur in retrospective surveys because respondents either completely forget events or mis-date them (Dex, 1991). For example, many unemployment spells are forgotten in retrospective interviews, especially for women (Jacobs, 2002) and transitions out of unemployment are often inconsistently dated (Paull, 2002). Moreover, because errors may be systematically correlated with factors such as education (Smith

and Thomas, 2003; Peters, 1988) that might be used as explanatory variables when using survey data, they will tend to bias the coefficients in regression models of respondent behaviour.

Nevertheless, retrospective surveys have some advantages which could outweigh the problem of recall bias. What matters is overall error, rather than error from a particular source such as forgetting. It is possible that reduced bias due to less attrition in retrospective surveys offsets the bias due to respondents forgetting or mis-dating events. Hence a more comprehensive study of the properties of retrospective survey data is required.

The objective of this paper is to test the accuracy of retrospective surveys of earnings. Results from such a survey are compared with a longitudinal survey that gathers data by repeatedly interviewing respondents over several years. While there have been several previous studies comparing retrospective recall data with standard longitudinal survey data collected more frequently (Peters, 1988; Pierret 2001), they have never been able to validate data from either type of survey. In contrast, we use a unique survey, the Panel Study of Income Dynamics Validation Study (PSIDVS), which contains accurate information on labour market outcomes from a company's records (which acts as a "gold standard"). This survey also has retrospectively recalled and contemporaneously surveyed information from the company's workers. 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.

We also consider the possibility of errors in long-term retrospective surveys deviating from the classical assumptions of uncorrelated error. This extends the literature begun by Bound et al. (2001) and Kim and Solon (2005) which shows how realistic departures from the textbook errors-in-variables model can either reverse or strengthen stylized facts that emerge from empirical research that does not allow for non-classical measurement error. For example,

conclusions about the cyclical behavior of real wages may not hold in the light of (potentially correlated) measurement errors in retrospective survey data. Another implication of non-classical measurement error is that conventional correction methods like Instrumental Variables (IV) estimation may not work properly (Black et al, 2000). Possible solutions in this case are discussed towards the end of this paper.

The final feature of the paper is that it reports on initial attempts to carry out a cost-benefit comparison of a retrospective recall survey with a longitudinal survey. Survey agencies will often have a good idea about the costs of a longitudinal survey relative to a single interview that uses retrospective recall. But it is more difficult to put a monetary value on the (potentially) greater accuracy of the longitudinal survey. One use of surveys which gives benefits measurable in monetary terms is for calculating the size and destination of public transfers for poverty reduction. Therefore we use the PSIDVS to consider the cost of a hypothetical poverty reduction using retrospectively recalled data, conventional longitudinal survey data, cross-sectional data with no retrospective component, and the validation data.

In the next section, we describe the data and their measurement errors. Tests of a measurement error model and implications for the literature on the cyclical behaviour of real wages are reported in Section III. In Section IV a cost-benefit comparison of conventional longitudinal data and retrospective recall data is discussed in the context of measuring poverty. Section V discusses statistical corrections for the recall bias and Section VI concludes.

II. Data Description

We use the Panel Study of Income Dynamics Validation Study (PSIDVS), which gives a unique opportunity to assess the accuracy of retrospective labour market data. The PSIDVS

contains accurate information on earnings and hours from a company's records (which acts as a "gold standard") and retrospectively recalled and contemporaneously surveyed information from the company's workers. Comparing retrospectively recalled reports with the company records should provide a wealth of information on the properties of the measurement error in retrospective data. A key difference from other validation studies, such as those reported by Bound et al (2001) is that since PSIDVS was conducted in two waves four years apart, it also provides information on measurement error in retrospectively recalled *changes* in variables.

Specifically, we use three sources of PSIDVS information: the company records from 1981 to 1986 that provide the validation information; the longitudinal survey data gathered each year and 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 retrospective 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. Recalled earnings appear to be a good proxy for true earnings, in terms of sample means for company records that range from 0.997 to 1.002 of the mean of log earnings in the retrospectively recalled data.

However, measures of inequality (or variance) for recalled data do not appear accurate, with ratios of company to recalled records ranging from 0.763 to 1.275. This variance ratio, $Var(\ln y_{it}^{true})/Var(\ln y_{it}^{recalled})$ can also be considered as a reliability ratio under the conditions of

classical measurement error, showing the proportion of true to observed variation. Since two of the variance ratios exceed one, 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 variation than the true earnings a negative correlation between true earnings and the recall errors is implied -- with larger correlations the longer the recall period.

Similarly the ratio of the variance of the recalled change in earnings to the actual change in earnings (2.255 in the last row of Table 1) strongly undermines the assumption of non-correlated errors. Instead, it implies a strong negative correlation between errors in the retrospectively recalled change in earnings and true values of this change. This pattern could occur if the prevalence of underreporting of transitory earnings in recalled data and the degree of underreporting is higher as the length of the recall period increases.

III. The Measurement Error Model

In this section we develop a measurement error model to test the hypothesis that people under-report transitory earnings variations in long-term retrospective surveys. In other words, when asked to report their earnings in previous years people may tend to report their usual earnings by forgetting transitory variations. Annual earnings can be written as a sum of two components:

$$y_{it} = y_{it}^P + y_{it}^T \tag{1}$$

where y_{it} is the i^{th} worker's log real annual earnings in year t , y_{it}^P is the permanent component, and y_{it}^T is the transitory component, which can be affected by a business-cycle or just individual specific transitory events. Survey data on annual earnings have reporting error of the form:

$$y_{it}^* = y_{it} + m_{it} + v_{it}^y \quad (2)$$

where y_{it}^* is the survey response, y_{it} is the true value of annual earnings, m_{it} is a method effect, and v_{it}^y is a pure random error. In other words, different methods of estimating y_{it} , such as using a longitudinal survey versus long-term retrospective recall, may entail different degree of measurement error. This method effect may be (negatively) correlated with the transitory component. Hence, the method effect can be expressed as:

$$m_{it} = \theta + \pi y_{it}^T + v_{it}^m \quad (3)$$

where v_{it}^m is a random deviation for the i^{th} individual from the average method effect. Combining the equations gives:

$$y_{it}^* = \theta + y_{it}^P + \lambda y_{it}^T + v_{it} \quad (4)$$

where $v_{it} (\equiv v_{it}^m + v_{it}^y)$ is a pure random error and $\lambda (\equiv 1 + \pi)$ represents a potential correlation between the true values and the method effect in the measurement error. A variant of equation (4) allows for mean-reversion in the permanent part as well. Classical measurement error is a special case of equation (4) where $\lambda = 1$ and $\theta = 0$. But with correlated errors (e.g. from underreporting the transitory part in a retrospective recall survey), $\pi < 0$ and (as long as measured expenditures are still positively correlated with true values) the measurement error follows a mean-reverting pattern ($0 < \lambda < 1$). In the case of errors negatively correlated with true values, the measurement error type is flexible in terms of mean bias

($E(y_{it}^*) = \theta + E(y_{it}^P) + \lambda E(y_{it}^T) \begin{matrix} > \\ < \end{matrix} E(y_{it})$) and bias in the estimated variance:

$Var(y_{it}^*) = Var(y_{it}^P) + \lambda^2 Var(y_{it}^T) + Var(v_{it}) \begin{matrix} > \\ < \end{matrix} Var(y_{it})$ and the resulting reliability ratios could

be greater than one as in Table 1. More importantly, this mean reversion in dependent variables

will tend to make estimated regression coefficients too small in magnitude, which is contrary to the textbook case where errors in the dependent variable cause no bias in slope coefficients.

In addition, if the length of the recall period affects the magnitude of recall bias, then the degree of mean-reversion will be larger for longer recall. Thus, the measurement error model implies that the under-reporting (of the transitory part) leads to non-classical (mean-reverting) measurement error. The model also implies that the measurement error may affect permanent and transitory components in a different way. We first investigate the effect of the mean-reverting measurement error on earnings transitions. Then we turn to the impacts on the permanent part, using the study of chronic poverty as an example.

A Test of Mean-Reverting Measurement Error

Consider a fixed effects model of earnings transitions like¹

$$y_{it} = y_{it}^P + y_{it}^T = (\alpha_i + \gamma_1 X_{it} + \gamma_2 X_{it}^2) + (\beta U_t + \varepsilon_{it}) \quad (5)$$

where y_{it} is the i^{th} worker's log real annual earnings in year t , the fixed effect α_i represents the combined effect of time-invariant characteristics of worker i , X_{it} is worker i 's years of work experience as of year t , U_t is a business-cycle indicator such as the unemployment rate, and ε_{it} is an individual transitory fluctuation. Instead of true earnings, we use error-ridden variables based on the general (non-classical) errors-in-variables model:

$$y_{it}^* = \theta + (\alpha_i + \gamma_1 X_{it} + \gamma_2 X_{it}^2) + \lambda(\beta U_t + \varepsilon_{it}) + v_{it}. \quad (6)$$

This model assumes that the underreporting (or mean-reversion) does not apply to the systematic part of earnings evolution, but only to the transitory variation, including the part associated with cyclical fluctuations as in the above hypothesis. The textbook errors-in-variable model is the

¹ This fixed effect model follows the specification used in Kim and Solon (2005).

special case that assumes $\lambda = 1$. We can estimate the measurement error parameter λ from the PSIDVS data using a specification like

$$\Delta y_{it}^* = \delta_0 + \delta_1 X_{it} + \lambda \Delta y_{it} + \Delta v_{it}, \quad (7)$$

where $\delta_0 = (1 - \lambda)(s\gamma_1 - s^2\gamma_2)$, $\delta_1 = 2s(1 - \lambda)\gamma_2$.

By replacing the surveyed $y_{i,t}^*$ with the recalled $y_{i,t,R(s)}^*$ with s recall period, we can define the recall bias in terms of $\hat{\lambda}_{R(s)}$ (the degree of mean-reversion) and relative importance of recall bias to the whole error-in-variable bias using the ratio $(\hat{\lambda}_{R(s)} / \lambda)$.

Using the PSIDVS data for the true, surveyed, and recalled earnings of 1982 and 1986, we estimate the mean-reverting measurement error parameter, $\hat{\lambda}_{R(s)}$ and measure its size relative to the (conventional) errors-in-variables bias that may be present in the longitudinal survey data, $(\hat{\lambda}_{R(s)} / \lambda)$. The first and the second columns of Table 2 report the mean-reverting measurement error parameter in the surveyed and recalled earnings respectively. The last column shows the degree of mean-reversion of the recalled variable relative to the (conventional) errors-in-variables bias by regressing the recalled earnings differentials on the surveyed earnings differentials. The measurement errors are clearly mean-reverting for both the surveyed ($\hat{\lambda} = 0.78$) and 4-year recall of earnings ($\hat{\lambda}_{R(4)} = 0.41$). The degree of mean-reversion in the recalled earnings is substantial enough to make it likely that retrospectively recalled earnings will be a poor proxy for true earnings in a regression model focused on transitory variations in earnings.

Furthermore, using $\hat{\delta}_1 = 2s(1 - \lambda)\gamma_2$ with the recalled earnings allows us to test directly whether recall error leads people to under-report transitory variations more than permanent

variations. The calculated t-value of the estimate of $(1-\lambda)$ using the delta method is 1.65, which is significant at five percent under a one-tail test.

Since the PSIDVS provides recalled earnings for each year from 1981 to 1986 along with corresponding validation information it is possible to test the additional hypothesis that the error varies with the length of the recall period. Using equation (7) and by replacing the surveyed $y_{i,t}^*$ with the recalled $y_{i,t,R(s)}^*$ for different s recall periods, the estimated measurement error parameters are reported in Table 3. The estimates of $\hat{\lambda}_{R(s)}$ are such that it appears that the length of the recall period is positively associated with the size of recall bias. However, the time trend in the $\hat{\lambda}_{R(s)}$ estimates is not statistically significant possibly because the degree of error is also affected by other factors such as business cycles rather than just the recall period length.

Implication of Errors in Recalled Earnings for the Cyclicity of Real Wages

As an example of the type of models that can be affected by the measurement errors described above, consider a fixed effects model of earnings transitions, as in equation (5). If we are interested in the transitions of y_{it} due to either the permanent factor or the transitory factor, then we may first-difference equation (5) to get:

$$\Delta y_{it} = \delta_0 + \delta_1 X_{it} + \beta \Delta U_t + \Delta \varepsilon_{it}, \quad (8)$$

where $\delta_0 = (1-\lambda)(s\gamma_1 - s^2\gamma_2)$, $\delta_1 = 2s(1-\lambda)\gamma_2$.

When interested in whether earnings vary counter-cyclically, non-cyclically, or pro-cyclically with the business cycle, one can investigate the sign of β . But instead of true earnings, economists typically have to use dependent variables with non-classical errors-in-variables.

What does this imply for the estimation of cyclicity in real earnings? Substituting equation (7) into equation (6) yields

$$\Delta y_{it}^* = \delta_0 + \delta_1 X_{it} + \lambda \beta \Delta U_t + (\lambda \Delta \varepsilon_{it} + \Delta v_{it}), \quad (9)$$

where $\delta_0 = (1 - \lambda)(s\gamma_1 - s^2\gamma_2)$, $\delta_1 = 2s(1 - \lambda)\gamma_2$.

The coefficient of ΔU_t is not the original wage cyclicity parameter β , but rather β rescaled by the measurement error parameter λ . For instance (mean-reverting) measurement error in standard longitudinal survey data may lead to as much as a 30% underestimation of procyclicality of real wages (Kim and Solon, 2005).

What is the situation with retrospectively recalled data? We can infer β from the estimated measurement error parameters in Table 2. The measured degree of mean-reversion $\hat{\lambda}_{R(s)}$ with the recalled survey data may lead to as much as a 60% underestimation of the pro-cyclicality of real wages. Thus retrospective surveys would not appear to be suitable for estimating how cyclical are wages, due to their mean-reverting errors which exceed those already found to exist with longitudinal survey data.

IV. An Example of Cost-Benefit Analysis of Retrospective Recall Data

Lower cost is an important advantage of retrospective surveys but it is difficult to put a monetary value on the cost of their (potentially) greater inaccuracy. One area where survey measures can lead directly to monetary costs is in setting budgets for income support or poverty alleviation interventions. Therefore in this section we use the PSIDVS data to focus on the permanent part of earnings as part of an analysis valuing data for measuring chronic poverty.

Time-averaged earnings or consumption are often used as long-term welfare indicators since single year measures are noisy indicators of chronic poverty status (Chaudhuri and Ravallion, 1994). We may use either longitudinal survey data or retrospectively recalled data to construct these long-term averages. If retrospective recall data are more error-ridden, as much of the literature suggests, this might be expected to reduce the relative importance of chronic poverty since noisy data is usually indistinguishable from transitory welfare fluctuations (Luttmer, 2001). To see whether this is true, we consider poverty transitions between two periods (1986 and 1982) and we assume that the permanent part of earnings is more strongly auto-correlated than the transitory part. An extreme case will be a unit autocorrelation for the permanent part, and zero for the transitory part. When we divide the population into poor and non-poor, the distribution of the four different combinations is as in Table 4.

The fraction of the chronic poor out of the total poor is $p_{P,P}/(p_{N,P} + p_{P,N} + p_{P,P})$. When the size of transitory earnings becomes smaller, non-diagonal terms ($p_{P,N}$ or $p_{N,P}$) will be reduced. Conversely, the larger the transitory part is, the larger the non-diagonal terms and the smaller the fraction of chronic poor out of the total poor. Under the classical measurement error model, the error-ridden variable is $y_{it}^* = y_{it} + v_{it} = y_{it}^P + (y_{it}^T + v_{it})$ and the fraction chronically poor will be spuriously reduced. However, under non-classical measurement error, as shown in Table 1, the implication is reversed. The relative importance of chronically poor will then be spuriously increased.

This prediction is empirically verified in Table 5 with the PSIDVS data. In the example the bottom 10th percentile of earnings is used as the poverty line. The true $p_{j,j}$ where j =poor or non-poor is in the parenthesis. The fraction of chronic poverty is $3/(18+19+3)$ when using the validation data from the company records. More chronic poverty is apparent with the

longitudinal survey data, with 8/(15+18+8) chronic poor. The highest share of chronic poverty is when the retrospective recall of earnings is used, with 13/(10+10+13). In other words, with a retrospective recall survey the autocorrelation of annual earnings is incorrectly overestimated which tends to exaggerate chronic poverty.

Who much does this non-classical measurement error in retrospective data matter in monetary terms? To answer this we follow Chaudhuri and Ravallion (1994) in calculating the cost of achieving a target welfare level by allocating transfers according to the situation of each person as reported in the company records, longitudinal survey data, or retrospective survey data.

The aggregate poverty measure is:

$$P = \sum_{i=1}^n (\max\{0, 1 - y_i / z\})^2 \quad (10)$$

where z is the given poverty line and the true welfare indicator is y_i .

Step-wise targeted transfers are used, with transfers given to the poorest person until they are raised to the level of the second poorest and so on. We define $P(T)$ as the reduced poverty level achieved with transfer budget T where transfers are based on the true welfare indicator (that is, the gold standard provided by the company records). This sets a frontier which we can use to compare the budget cost, T^* of reaching the same level of aggregate poverty when transfers are based on imperfectly measured data. In other words, $P(T^*)$ with imperfect measures will take a larger transfer budget T^* to achieve the same level of poverty reduction. The difference in the two budgets can be considered as the cost of using an empirical welfare indicator with measurement error. On the other hand the benefit of using retrospective recall would be the saved cost of resurveying respondents.

Figure 1 summarizes the relationship between the transfer budget (T) and the corresponding welfare level (P) with four income indicators: (i) the average of 1982 and 1986

real earnings coming from the company records, (ii) the average for the same years coming from the longitudinal survey data, (iii) the average for the same years coming from the retrospective recall that was asked in 1987 about 1982 and 1986, and (iv) the single year income from 1986 as obtained from the longitudinal survey in that year. To reduce the aggregate poverty level by 50%, the required transfer budget (T) (noting that this is just for the sample rather than for the population that it represents), is approximately \$9,000 when the company records are used as the income indicator. In contrast, using the longitudinal survey data, the required cost of transfers to get the same poverty reduction would be \$14,000 and it would be \$62,000 using the retrospectively recalled data. Since company records are rarely available, the relevant comparison is between two feasible ways of obtaining longitudinal data, using either a longitudinal survey or a retrospective recall. The \$48,000 lower transfer budget when using longitudinal survey data ($T_{surveyed}^* - T_{recalled}^*$) is a benefit that can be compared with the cost of mounting a genuine longitudinal survey. Of course, there are many other benefits to using longitudinal data than just for setting transfers for reducing poverty, so these estimated benefits are very much a lower bound.

Another relevant comparison is between retrospective recall from a single interview and using that same interview to gather just a cross-section of data. It is clear from Figure 1 that the retrospective recall adds value to a single interview, since the transfers budget is less when using retrospective recall than when using the single cross-section. Moreover, the feature of underreported transitory variations in retrospectively recalled earnings could be more efficiently utilized since the size of error in the recalled welfare indicator relative to the precise permanent indicator would be reduced as the time horizon increases. For example, compared to using two years of recalled earnings, we could reduce aggregate poverty by 27% more with the same size

of the transfer budget if recalled earnings over six years were used. Thanks to the underreporting of transitory variations in retrospectively recalled earnings, we have a more reliable proxy relative to the precise long-term welfare indicator over the longer time horizon. Thus, at least in terms of hybrid use of surveys, a retrospectively recalled panel could replace a single year cross-sectional indicator to get better data for a specific policy objective such as setting transfers to reduce chronic poverty.

V. Possible Statistical Corrections for Measurement Errors in Retrospective Surveys

In this paper we have described the nature and magnitude of measurement error biases in retrospective data. Some discussion of statistical treatments for reducing these systematic biases is in order. One solution is to use auxiliary data, such as using PSIDVS to correct estimates based on PSID. As shown above, we can independently measure $\hat{\lambda}_R$ in PSIDVS and rescale $\hat{\lambda}_R \beta$ from PSID by the measurement error parameter $\hat{\lambda}_R$. Unfortunately, this solution is not often feasible because there are very few sources of validation data for retrospective surveys.

In the absence of validation data statistical correction methods such as IV estimation are needed. However IV estimation does not work properly for correlated (and mean-reverting) errors, as shown in Black et al. (2000). In this case one could use bounding estimators of the unknown true effect. Specifically, OLS estimates and their inverse are used to construct the bounding estimates for the true β in the case of an error-correlated dependent variable. When the error is mean-reverting, as in the recall bias with $0 < \lambda_R < 1$, we use the conventional OLS estimate as a lower bound and the inverse of slope coefficient estimate in the population regression of ΔU_t on Δy_{it}^* as an upper bound. It is straightforward to show that the conventional

OLS estimate in the population regression of Δy_{it}^* on ΔU_t is a lower bound since the estimate is biased toward to zero as $|\lambda_R \beta| < |\beta|$.

Unfortunately, the bounding property is not always satisfied, in contrast to the argument of Black et al. (2000) and we need to consider the sufficient (or necessary) condition for when the inverse of the slope coefficient in the population regression of ΔU_t on Δy_{it}^* is an upper bound. The derived sufficient condition is that the sum of the variance of measurement error and the covariance between the true variable and its measurement error should be positive. That is, the (negative) correlation between the true variable and its measurement error should not be too strong. The sufficient condition appears innocuous and our test result for the proposed sufficient condition with the PSIDVS data has been confirmed. However, this solution still requires the validation data to confirm the sufficient condition, so it is not frequently feasible either. Instead, we here propose the inversed IV estimate in the population regression of ΔU_t on Δy_{it}^* as an upper bound ($|\beta/\lambda_R| > |\beta|$) and this bounding property is held under no conditions. Our proposed bounding properties can be easily extended to the case of an independent variable with correlated error.

VI. Conclusion

In this paper, we assess the accuracy of long-term retrospectively recalled data. Such data are increasingly used in labour and population economics due to their lower cost and greater availability than conventional longitudinal survey data. The results based on the unique Panel Study of Income Dynamics Validation Study suggest that retrospective recall is a poor substitute for genuine longitudinal data in the analyses carried out here. We find underreporting of

transitory events due to recall error. The resulting error is non-classical, which is unlikely to be properly handled by conventional correction methods such as IV estimation. We highlight the implications of this type of measurement error bias on some selected literatures such as the cyclical behaviour of real wages and poverty transitions.

These implications suggest that only selective use of retrospectively recalled data is appropriate. Such data are a poor substitute for a conventional longitudinal data with respect to the issues of transitory aspects of labour market outcomes. But retrospective data may improve on the efficiency of a cross-section for policy objectives related to the permanent part of labour market outcomes such as designing transfers to reduce chronic poverty. Our limited cost-benefit analysis shows way one in which these comparisons can be made more explicit. However, a wider range of validation surveys is needed to fully understand the nature and consequences of measurement error in retrospective surveys and the role which such surveys can play in providing data for analyses in labour and population economics.

References

- Beckett, M., Da Vanzo, J., Sastry, N., Panis, C., and Peterson, C. (2001) "The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country," *Journal of Human Resources* 36(3): 593-625.
- 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.
- Bound, J., Brown, C., Duncan, G., and Rodgers, W. (1994) "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data," *Journal of Labor Economics*, 12, 345-368.
- 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.
- Chaudhuri, S. and Ravallion, M. (1994) "How well do static indicators identify the chronically poor?" *Journal of Public Economics*, 53, 367-394.
- Deaton, A. (1997) *The Analysis of Household Surveys* Johns Hopkins University Press, Washington DC.
- Dex, S. (1991) *The Reliability of Recall Data : A Literature Review* (ESRC working paper, No. 11). Colchester: University of Essex.
- Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988) The Life History Calendar: A Technique for Collecting Retrospective Data. *Sociological Methodology*, 18, 37-68.
- 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.
- Luttmer, E. (2001) "Measuring Poverty Dynamics and Inequality: Disentangling Real Events from Noisy Data," Unpublished manuscript, Harvard University.
- Kosloski, K., Datwyler, M. M., & Montgomery, R. J. V. (1994) Evaluating retrospective measures in gerontological research. *Research on Aging*, 16(4), 389-400.
- 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 Economic 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.

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

<i>Variable</i>	<i>Mean</i>	<i>Ratio to company record,</i>		<i>Ratio to company record,</i>	
		$\frac{E(\ln y_i)}{E(\ln y_{it}^{m(or r)})}$	<i>Vari-</i>	$\frac{Var(\ln y_i)}{Var(\ln y_{it}^{m(or r)})}$	<i>ance</i>
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					
$\ln y_{i82}^m$ (log1982 real annual earnings, surveyed in 1983)	10.414	.999	.0861	0.952	
$\ln y_{i86}^m$ (log1986 real annual earnings, surveyed in 1987)	10.485	.999	.0585	0.863	
Recalled Earnings					
$\ln y_{i81}^r$ (log1981 real annual earnings, recalled in 1987)	10.286	1.001	.06023	1.118	
$\ln y_{i82}^r$	10.430	.997	.0643	1.275	
$\ln y_{i83}^r$	10.352	.999	.0668	0.926	
$\ln y_{i84}^r$	10.388	1.002	.0917	0.787	
$\ln y_{i85}^r$	10.451	1.001	.0564	0.763	
$\ln y_{i86}^r$ ($\equiv \ln y_{i86}^m$)	10.485	.999	.0585	0.863	
Differenced Variables					
$\ln y_{i86} - \ln y_{i81}$.184		.0650		
$\ln y_{i86} - \ln y_{i82}$.068		.0709		
$\ln y_{i86} - \ln y_{i83}$.132		.0406		
$\ln y_{i86} - \ln y_{i84}$.066		.0554		
$\ln y_{i86} - \ln y_{i85}$.008		.0254		
$\ln y_{i86}^m - \ln y_{i82}^m$.070	.971	.0727	0.976	
$\ln y_{i86}^r - \ln y_{i81}^r$.198	.929	.0307	2.211	
$\ln y_{i86}^r - \ln y_{i82}^r$.054	1.259	.0314	2.257	
$\ln y_{i86}^r - \ln y_{i83}^r$.132	1.000	.0266	1.526	
$\ln y_{i86}^r - \ln y_{i84}^r$.096	.678	.0495	1.119	
$\ln y_{i86}^r - \ln y_{i85}^r$.033	.242	.0162	1.567	

Table 2. Measurement Error Parameter Estimates

<i>Measurement error parameters</i>	$\hat{\lambda}$	$\hat{\lambda}_{R(4)}$	$(\hat{\lambda}_{R(4)} / \hat{\lambda})$
OLS	.779 (.044) ^a	.410 (.036)	.450 (.033)

^a Standard Errors in the parenthesis.

Table 3. Measurement Error Parameters with Different Recalling Period

<i>Measurement Error Parameters</i>	$\hat{\lambda}_{R(1)}$	$\hat{\lambda}_{R(2)}$	$\hat{\lambda}_{R(3)}$	$\hat{\lambda}_{R(4)}$	$\hat{\lambda}_{R(5)}$
OLS	.419 (.046) ^a	.727 (.040)	.292 (.051)	.410 (.036)	.304 (.041)

^a Standard Errors in the parenthesis.

Table 4. Transitions of Poverty

		<i>T-s</i>	
		Non-Poor	Poor
T	Non-Poor	$P_{N,N}$	$P_{P,N}$
	Poor	$P_{N,P}$	$P_{P,P}$

Table 5. Overstated Chronic Poverty

		<i>T-s</i>	
		Non-Poor	Poor
T	Non-Poor	(186 ^a ,178 ^b ,179 ^c)	(10,18,19)
	Poor	(10,15,18)	(13,8,3)

^a recalled earnings , ^b surveyed earnings, ^c company records

Figure 1. Impacts on Chronic Poverty of Transfers Based on Imperfect Income Indicators

