Can we trust retrospective recall to replace panel data?

John Gibson*
Bonggeun Kim**
University of Canterbury

Preliminary draft
July 29th, 2005

Abstract
Several recent studies in labour and demographic economics have used retrospective surveys to substitute for the high cost and limited availability of longitudinal data. These studies typically focus on employment transitions, migration and fertility, and often use a single interview to obtain a lifetime history. Inaccurate long-term recall could make retrospective surveys a very poor substitute for panel data. Also, long-term recall may induce non-classical error which makes conventional statistical corrections less effective.

In this paper, we assess the accuracy of long-term recall data and develop relevant correction methods for potential recall bias. The results based on the unique Panel Study of Income Dynamics Validation Study suggest that retrospective recall is a poor substitute for genuine panel data in the longitudinal analyses carried out here. We find underreporting of transitory events due to recall error. The resulting error type creates a non-classical measurement error, which cannot be properly handled by conventional correction methods such as IV estimation. A limited cost-benefit analysis is also carried out, comparing the savings from using a cheaper retrospective recall survey with the potential cost of applying the less precise recall data to a specific policy objective such as reducing chronic poverty.

JEL: C33, J64
I. Introduction

Research in the social sciences often relies on longitudinal data, where samples of individuals, households or families are repeatedly monitored over a period of several years. These longitudinal data allow studies of transitions, such as into and out of marriage, poverty and unemployment, which are not possible with cross-sectional surveys. They also allow more reliable statistical models to be estimated because biases due to unobservable factors, such as ability, can be alleviated when there are repeated observations on the same people (Deaton, 1997).

However, longitudinal surveys are costly and in many countries they are restricted to small, nationally unrepresentative, samples. For example, the only mature longitudinal surveys in New Zealand are based on small samples of birth cohorts from Christchurch and Dunedin (Fergusson et al., 2003) which are quite atypical of the current population because of a marked under-representation of ethnic minorities and the elderly. One response to the high cost and limited availability of longitudinal data is to use retrospective surveys. These surveys use a single interview to obtain a long-term or even lifetime history (Freedman et al., 1988). For example, the New Zealand Women: Family, Employment and Education Survey asked a sample of women to recall details about their first five jobs, which required recalling over a period of up to 40 years (Marsault et al., 1997). 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, some authors report scepticism about analyses based on
such data, with claims that it will be “quite meaningless” (Jacobs, 2002, p. 545) and that it is a “poor substitute for panel data” (Kennickell and Starr-McCluer, 1997, p. 462). On the other hand, other authors suggest that when retrospective questions are asked carefully and interviewers are well trained, respondents can provide “accurate and detailed information” (Campbell, 2000, p. 1685). This debate matters not only for users of long-term retrospective surveys, but also for widely used surveys like the Panel Study of Income Dynamics (PSID), where a certain amount of retrospective recall is needed to construct earnings and hours of work for the previous calendar year. Moreover, even short-term recall over a fortnight in household expenditure surveys can cause errors in comparison with more day-to-day reporting methods such as diaries (Gibson, 2002).

Errors occur in retrospective surveys because respondents may either completely forget events or mis-date them (Dex, 1991). For example, in a panel where respondents reported symptoms of depression, less than one half recalled those symptoms several years later (Wells and Horwood, 2004). Similarly, many spells of unemployment appear to be forgotten in retrospective interviews, especially for women (Jacobs, 2002) and transitions for unemployment spells are often inconsistently dated (Paull, 2002). Moreover, because recall 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 distinct advantages and it is possible that these could outweigh the problems 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 and this should be embedded in an overall evaluation of survey errors.

The objective of this paper is to test whether retrospective surveys that rely on long-term recall are a reliable substitute for longitudinal surveys that gather data by repeatedly interviewing respondents over several years. To achieve this objective we use a unique survey, the Panel Study of Income Dynamics Validation Study (PSIDVS). 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, the PSIDVS contains accurate information on labour market outcomes from a company’s records (which acts as a “gold standard”) and retrospectively recalled and contemporaneously surveyed information from the company’s workers.

Moreover, and in comparison to previous research in this area, we 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 the stylized facts that have emerged 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. One implication of non-classical measurement error is that conventional correction methods
like Instrumental Variables estimation may not work properly (Black et al, 2000). Possible solutions in this case are to use either moment conditions from auxiliary (validation) data sets or reverse regressions to construct bounding estimates for the true effect and these are discussed in the 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 cross-sectional survey that relies on 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 in the final section of the paper we use the PSIDVS to consider the cost of a hypothetical poverty reduction using retrospectively recalled data, conventional longitudinal survey data and cross-sectional data with no retrospective component.

The next section of this paper provides a measurement error model with retrospective recall survey and its implications. In Section III, we describe our data and discuss the specific implications of measurement error of recall data in a regression model on the literature of cyclical behaviour of real wages and proposed statistical corrections for the recall bias. In Section IV, we discuss the implications of measurement error of recall data in terms of the measurement of chronic poverty and a cost-benefit comparison of replacing a conventional longitudinal data with a retrospective panel. Section V concludes with a summary of the main findings.
II. A Measurement Error Model

In the course of the comprehensive evaluation of the accuracy of retrospective recall data, we consider the specific hypothesis.

_Do people under-report labour market outcomes such as employment and earnings (especially transitory variations) due to recall error?_

In other words, people who may report their earnings for the previous calendar year tend to report their usual earnings by forgetting the transitory variations. We turn to formalize the hypothesis. As a longitudinal analysis of transitions of earnings, we consider the annual earnings like

\[ y_{it} = y_{it}^p + y_{it}^T \]  \hfill (1)

where \( y_{it} \) is the \( i \)th worker’s log real annual earnings in year \( t \), which consists of the permanent part (\( y_{it}^p \)) and the transitory part (\( y_{it}^T \)), which can be affected by a business-cycle or just individual specific transitory events.

Suppose that survey data on annual earnings is subject to reporting error of the form:

\[ \bar{y}_{it} = y_{it} + m_{it} + v_{it} \]  \hfill (2)

where \( \bar{y}_{it} \) is the survey response, \( y_{it} \) is the true value of annual earnings of the \( i \)th individual, \( m_{it} \) is a method effect, due perhaps to the use of a long-term retrospective recall questionnaire, and \( v_{it} \) is a pure random error. Errors occur in retrospective surveys because respondents may forget the transitory variations of earnings and under-report it. So, the method effect in the measurement error, \( m_{it} \) may be (negatively) correlated with transitory part. Hence, the method effect can be expressed as:
\[ m_i = \pi y^T_i + v^m_i \]  

(3)

where \( v^m_i \) is a random deviation for the \( i \)th individual from the average method effect.

Combining the two equations gives:

\[ y^*_i = y^p_i + \lambda y^T_i + v_i. \]  

(4)

where \( v_i (\equiv v^m_i + v^r_i) \) 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. We also consider the similar variant of equation (4) that would allow for mean-reversion in the systematic part with a \( \lambda \) greater than the \( \lambda \) for the transitory part. Classical measurement error is a special case of equation (4) where \( \lambda = 1 \). But with correlated errors (underreporting of transitory part under 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 addition, if the length of the recall period matters with respect to the magnitude of recall bias, then the degree of mean-reversion will be larger as the length of the recall period or the measurement error parameter \( \lambda \) will be smaller. Thus, the measurement error model implies that the under-reporting (of transitory part) leads to non-classical (mean-reverting) measurement error. The model also implies that the measurement error of recall data may influence differently on the issues of permanent part and the transitory part. In the next section, we first investigate the effect of the mean-reverting measurement error on the transitions of earnings focused on the transitory part. Then, we turn to the issues of the permanent part such as a chronic poverty.

III. A Test of Measurement Error Model and Its Implication in a Regression Model
Data Description

Our empirical assessment of the properties of recall bias will exploit an unusual data source, the Panel Study of Income Dynamics Validation Study (PSIDVS), which provides a unique opportunity to assess the accuracy of retrospective recall of labour market outcomes. The PSIDVS contains accurate information on labour market outcomes from a company’s records (which acts as a “gold standard”) and retrospectively recalled and contemporaneously surveyed information from the company’s workers. Comparing the recalled reports of earnings and hours with the company records should provide a wealth of information on the properties of the measurement error in retrospectively recalled data. Furthermore, because the validation study was conducted in two waves four years apart, it also provides information on measurement error in retrospectively recalled changes in those variables.

We use three sources of PSIDVS information on labour market outcomes for a group of workers in 1981/82: the company records, which serve as the validation information; the longitudinal survey data that were gathered in the following year; and the long-term retrospective recall data that was gathered in 1987 but refer to the 1981/82 period. Comparisons with the validation data allow us not only to identify any retrospective recall bias, but also to measure its size relative to the (conventional) errors-in-variables bias that may be present in the longitudinal survey data where the recall lag is less than one year. These tests can also establish whether the recall errors are non-classical (mean-reverting), which would make them contrary to the assumptions used in most treatments.
of measurement error. The conventional assumptions about errors being independent of either true values or of other variables usually reflect convenience rather than conviction (Bound, Brown, and Mathiowetz, 2001). These comparisons should allow us to test the implications of the measurement error model in the previous section. Table 1 presents a descriptive statistics of 219 annual earnings observations of the PSIDVS data. The recalled 1982 earnings compared to the true and the surveyed earnings show the smaller variation which could be due to the underreporting of transitory variations.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln (1982 real annual earnings, company record)</td>
<td>10.408</td>
<td>.286</td>
<td>8.833</td>
<td>11.086</td>
</tr>
<tr>
<td>ln (1986 real annual earnings, company record)</td>
<td>10.476</td>
<td>.224</td>
<td>9.312</td>
<td>11.015</td>
</tr>
<tr>
<td>ln (1982 real annual earnings, survey data)</td>
<td>10.414</td>
<td>.293</td>
<td>8.698</td>
<td>11.124</td>
</tr>
<tr>
<td>ln (1982 real annual earnings, recall data)</td>
<td>10.430</td>
<td>.253</td>
<td>9.245</td>
<td>11.037</td>
</tr>
</tbody>
</table>

### A Test of Mean-Reverting Measurement Error

Let us consider a fixed effects model of earnings transitions like

\[
y_t = y_{it}^u + y_{it}^T = (\alpha_i + \gamma_1 X_{it} + \gamma_2 X_{it}^2) + (\beta U_i + \epsilon_i)
\]

where \( y_{it} \) is the \( i^{th} \) worker’s log real annual earnings in year \( t \), the fixed effect \( \alpha_i \) represents the combined effect of time-invariant characteristics of worker \( i \), \( X_{it} \) is worker \( i \)’s years of work experience as of year \( t \), \( U_i \) is a business-cycle indicator such
as the civilian unemployment rate, and $\varepsilon_{it}$ is an individual transitory fluctuation term.

Instead of true earnings, we use error-ridden variables with the general (non-classical) errors-in-variables model like:

$$y_{it}^* = \alpha_i + \gamma_{it}X_{it} + \gamma_{2t}X_{it}^2 + \lambda (\beta U_{it} + \varepsilon_{it}) + \nu_{it}. \quad (6)$$

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

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

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

In equation (6), by replacing the surveyed $y_{it}^*$ with the recalled $y_{it}^{R(s)}$ with $s$ recalling 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 with the ratio $\hat{\lambda}_{R(s)}/\hat{\lambda}$. With (7), we can evaluate the measurement error parameter in longitudinal analysis framework in terms of $\hat{\lambda}_{R(s)}$ and $\hat{\lambda}_{R(s)}/\hat{\lambda}$. With the PSIDVS data for the true, surveyed, and recalled earnings of 1982 and 1986, we estimate a mean-reverting measurement error parameter and measure its size relative to the (conventional) errors-in-variables bias that may be present in the longitudinal survey data where the recall lag is less than one year. The estimates of those measurement error parameters of annual earnings are in Table 2. The first and the second columns represent the mean-reverting measurement error parameter with the surveyed earnings and the 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 with the surveyed and recalled (4 year recall period) earnings. The degree of mean-reversion of the recalled earnings is substantial enough to make a retrospective recalled earnings variable a poor substitute for the true earnings in a regression model focused on transitory variations of earnings.

Table 2. Measurement error parameter estimates

<table>
<thead>
<tr>
<th>Measurement error parameters</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{\lambda}_{R(4)}$</th>
<th>$\hat{\lambda}_{R(4)} / \hat{\lambda}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>.779 (0.044)</td>
<td>.410 (.036)</td>
<td>.450 (.033)</td>
</tr>
</tbody>
</table>

a. Standard Errors in the parenthesis.

Furthermore, $\delta_1 = 2s(1-\hat{\lambda})\gamma_2$ with the recalled earnings allows us to test directly whether people more under-report transitory variations than permanent variations due to recall error. The calculated t-value of the estimate of $(1-\hat{\lambda})$ by a delta method is 1.65 and it is significant at 5 percent under one-tail test.

In addition, the second wave of PSIDVS provides the recalled survey from 1981 to 1986 and their corresponding validation information. This unique feature, based on the stable reporting error assumption, allows us to test sub-hypothesis about whether the length of the recall period matters. With (7), by replacing the surveyed $y^*_t$ with the recalled $y^*_t, (1,R,s)$ with different $s$ recalling period, the estimated measurement error parameters are reported in Table 3.
Table 3. Measurement error parameters with different recall period

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

\(^a\) Standard Errors in the parenthesis.

The estimates of \( \hat{\lambda}_{R(i)} \) appear that the length of the recall period is positively associated with the degree of mean-reversion and the size of recall bias, however the trend is not significant and the degree of mean-reversion seemed to be also affected by other factors such as business cycles.

Implication of Recall Bias on Cyclical Behaviors 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 (5). If we are interested in the transitions of \( y_t \) thanks to either the permanent factor or the transitory factor, then we may first-difference equation (5) to get:

\[
\Delta y_t = \delta_0 + \delta_1 X_t + \beta \Delta U_t + \Delta \varepsilon_t,
\]

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 procyclically with the business cycle, one can investigate the sign of \( \beta \). But instead of true earnings, social scientists typically have to use error-ridden dependent variables with the general (non-classical) errors-in-variables model like (6). What does this model of non-
classical measurement error imply for the estimation of cyclicality in real earnings?

Substituting equation (7) into equation (6) yields

$$\Delta y^*_u = \delta_0 + \delta_1 X_u + \lambda \Delta U_y + (\lambda \Delta \epsilon_y + \Delta v_y),$$

(9)

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

The coefficient of $$\Delta U_y$$ is not the original wage cyclicality parameter $$\beta$$, but rather is $$\beta$$ rescaled by the measurement error parameter $$\lambda$$. For instance (mean-reverting) measurement error may lead to as much as a 30% underestimation of procyclicality of real wages (Kim and Solon, 2005). With equation (9), the true coefficient of $$\Delta U_y$$, $$\beta$$, cannot be separately identified. Instead, we can infer $$\beta$$ with the estimated the measurement error parameter in separate estimations as in Table 2. The measured degree of mean-reversion $$\hat{\lambda}_{R(i)}$$ with the recalled survey data (mean-reverting) measurement error may lead to as much as 60% more underestimation of procyclicality of real wages.

Our example of empirical specification is associated with recall bias of recalled dependent variable in a linear regression model, but it can be easily extended to the case of recalled independent variable.

**Statistical Corrections**

In addition to describing the nature and magnitude of measurement error biases in retrospective data, we need to consider more reliable statistical treatments for reducing these systematic biases in the regression context. One solution is the use of auxiliary data sets for correcting recall bias, such as using PSIDVS to correct estimates based on PSID.
As shown above, we can independently measure $\hat{\lambda}_r$ and rescale $\hat{\lambda}_r \hat{\beta}$ 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 longitudinal and retrospective surveys.

In the absence of validation data, we may need some statistical correction methods such as IV estimation for conventional error-in-variable bias. However, even though we have a proper instrumental variable, the IV estimation does not work properly for the correlated (and mean-reverting) error when the error-ridden recalled variable is used as an independent variable as in Black et al. (2000). In this case one could use the bounding estimators for the unknown true effect. Specifically, use OLS estimates in a conventional regression and its inversed regression to construct the bounding estimates for the true $\beta$ in the case of 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 $\Delta U_i$ on $\Delta y^{*}_{it}$ as an upper bound. It is straightforward to show that the conventional OLS estimate in the population regression of $\Delta y^{*}_{it}$ on $\Delta U_i$ is a lower bound since the estimate is biased toward to zero as $|\lambda_p \beta| < |\beta|$. Unfortunately, the other bounding property is not always satisfied and we need to consider the sufficient (or necessary) condition for that the inverse of slope coefficient LS estimate in the population regression of $\Delta U_i$ on $\Delta y^{*}_{it}$ is an upper bound. The inversed OLS estimate in the population regression of $\Delta U_i$ on $\Delta y^{*}_{it}$ can be an upper bound under the specific conditions. 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 detailed discussion of those bounding properties is in Gibson and Kim (2005). The test result for the proposed sufficient condition with the PSIDVS data confirms the upper-bounding property of inversed OLS estimate of the inversed regression. Our proposed bounding properties can be easily extended to the case of independent variable with correlated error. In addition when the error is positively correlated, the bounding properties are always satisfied.

IV. Implications of Recall Error on Measured Chronic Poverty and A Cost-Benefit Analysis of Retrospectively Recalled Panel

In this section, we attempt to carry out a hypothetical cost-benefit comparison of a retrospective recall survey with a longitudinal survey under specific policy objective such as reducing chronic poverty. To do so, we first characterize the implication of measurement error of a retrospective recall data focused on the permanent part of earnings. Then, we turn to discuss a potential way to quantify the potential monetary cost of applying the less precise recall data to the specific policy objective.

We often use a time-averaged earning or consumption as for a long-term welfare indicator such as a chronic poverty status. The single year welfare measure has been considered as a very poor proxy with substantial measurement error (Chaudhuri and Ravallion, 1994). To construct the more reliable measure for a chronic poverty status, we may use a conventional panel data. Alternatively, a retrospectively recalled panel
could be used. Considering a welfare indicator based on a retrospective recall data as just an error-ridden measure, the literature could simply expect that the relative importance of chronic poverty would be spuriously reduced when we measure it as the share of the chronic poverty of the whole poor including the transitory and chronic poverty. To see this, we consider two periods for the transition of poverty status and we assume that the permanent part of earnings is more strongly auto-correlated than the transitory part does. 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 four different combinations is as follows.

<table>
<thead>
<tr>
<th>T</th>
<th>Non-Poor</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T-N-N</td>
<td>P</td>
</tr>
<tr>
<td>T</td>
<td>N-N-N</td>
<td>P</td>
</tr>
<tr>
<td>T</td>
<td>P-P</td>
<td>P</td>
</tr>
<tr>
<td>T</td>
<td>N-N-N</td>
<td>P</td>
</tr>
</tbody>
</table>

The fraction of the chronic poor out of the 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 non-diagonal terms will be observed and the fraction of chronic poor out of the poor will be reduced. Under classical measurement error model, the error-ridden variable is \( y' = y + v = y' + (y' + v) \) and the fraction of chronic poor will be spuriously reduced. However, under the non-classical measurement error assumption as in equation (4) with \( 0 < \lambda < 1 \) and when its reduction dominates the additional pure random error (\( v \)), the implication is reversed. The relative importance of chronically poor will be spuriously reduced.
increased. It is empirically verified as in Table 5 with the PSIDVS data for 1982 and 1986.

Table 5. Overstated chronic poverty

<table>
<thead>
<tr>
<th>T-s</th>
<th>Non-Poor</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>166 (158)</td>
<td>12 (20)</td>
</tr>
<tr>
<td>Poor</td>
<td>16 (21)</td>
<td>25 (20)</td>
</tr>
</tbody>
</table>

We assign the bottom 20 percentile as the imaginary poverty line. The true $p_{j,t}$, where $j=\text{poor or non-poor}$ is in the parenthesis and the recall data increases the fraction of chronic poverty from $20/(21+20+20)$ to $25/(16+12+25)$. Similarly, the fraction of chronic non-poor is also increased. The chronic poverty would be exaggerated. In other words, the autocorrelation of annual earnings would be incorrectly overestimated.

**A Cost-Benefit Analysis**

Thus a time-averaged income based on retrospectively recalled panel could exaggerate the permanent part, so it is also an error-ridden measure to the more reliable (or true) welfare indicator. Here, we try to quantify the potential monetary cost (and benefit) of applying the error-ridden or the less precise indicator to a specific policy objective. In doing so, we may come up with a way to do a direct cost-benefit analysis of replacing a conventional panel data with a retrospectively recalled panel with respect to a long-term welfare indicator.
To be specific, we calculate the additional cost of using recall data in place of conventional data to achieve a target welfare level measured by a normative welfare measure of Chaudhuri and Ravallion (1994) like

$$P_2 = \sum_{i=1}^{n} (\max\{0,1 - y_i / z\})^2$$  \hspace{1cm} (10)

where $z$ is given poverty line.

When we have true welfare indicator $y_i$, $P_2$ is an aggregate poverty measure with a step-wise targeted transfer scheme where transfers will be given to the poorest to be raised to the second poorest household until the given transfer budget is exhausted. With the rankings of household based on $y_i$ distribution, the step-wise transfer scheme will assign $t_i$ to the $i^{th}$ household under the poverty line. We define $P_2(T)$ as the reduced poverty level with the available transfer budget $T$ based on the true welfare indicator (of a conventional longitudinal data) of a household. With the imputed rankings of household of the same $y_i$ distribution based on imperfect measures $y^*_i$, the transfer scheme would assign $t_i$ inefficiently. Thus, $P_2^*(T)$ with imperfect measures will be larger or equal to $P_2(T)$.

To achieve the same level of $P_2(T_R)$, the required transfer budget $T_R$ based on the imperfect welfare indicator (of a recall data) can be calculated. By construction, $T_R \geq T$ and the difference of two budgets will be considered as the cost of using a less-precise welfare indicator to the true indicator. The benefit of using recall data to the conventional longitudinal data would be the saved resurvey (and reassigning transfers) cost.
We use the PSIDVS data for our thought experimental cost-benefit analysis. We assign the bottom 20 percentile as the poverty line and the resulting aggregate poverty level \( P_2 \) is calculated. We consider a time-averaged income of the conventional longitudinal data (1986 and 1982 real annual earnings) as the precise measure and we consider two less precise measures: a time-averaged income of retrospectively recalled earnings and single year earnings (1986) with no retrospective component. The time-averaged annual earnings of the precise measure are ranged from $19,363 to $59,392 and the proposed poverty line is at $29,157. A time-averaged indicator of retrospectively recalled panel assigns the true poorest household to the 5th poorest and the single year or cross-sectional indicator assigns the same household to the 15th poorest, so the imputed rankings of household earnings will be substantially affected by the measurement errors of two less precise measures.

Figure 1 summarizes the relationship between the transfer budget \( T \) as percentage of aggregate poverty gap and the corresponding welfare level \( P_2 \) with three income indicators. To reduce the aggregate poverty level by 50 percent, the required transfer budget \( T \) are 20.3% (of conventional panel as the true indicator), 47.4% (based on retrospectively recalled panel), and 76% (based on cross-sectional indicator) of the aggregate poverty gap ($146,718). The difference \( T_{R} - T \) for the 50 percent poverty reduction is $39,760 and 23.1% of aggregate poverty gap.
To be more realistic, let us use single year measure at the survey year and adjust the rankings of $y_i$ based on the future survey. Alternatively, we use a single retrospective recalled panel for the survey year and for the subsequent future years as well. For example, with two periods only, the monetary cost of applying less precise retrospectively recalled panel to the chronic poverty reduction is

$$
(T_h(\hat{P}_2) - T(\hat{P}_2)) + (T_h(\hat{P}_2) - T(\hat{P}_2))/(1+r)
$$

and its benefit is

$$
(T_s(\hat{P}_2) - T_h(\hat{P}_2)) + C/(1+r)
$$

where $\hat{P}_2$ is the targeted aggregate poverty level, $(1+r)$ is the discount rate, $T_s$ is the required transfer budget with the single year indicator, and $C$ is the resurvey cost. In light of our experiment in the above, we expect that there might be a net gain to use retrospectively recalled panel when $C$ is large enough and we have the longer time...
horizon for the policy. The first point is apparent, but the second point needs some explanations. Even though the PSIDVS data only has two waves of surveyed longitudinal data, but the retrospectively recalled panel and its validation data are available up to 6 years. For the specific policy objective such as chronic poverty reduction, the feature of underreported transitory variations in retrospectively recalled earnings could be more efficiently utilized. The size of error of recalled indicator to the precise permanent indicator would be reduced as the time horizon increases. The simple calculation confirms this prediction. The estimated relative monetary cost of applying less-precise retrospectively recalled panel for the specific policy objective is reduced when we consider relatively longer time horizon. With two years recalled earnings, we only reduce 58 percent of aggregate poverty level with the full transfer budget as in Figure 1. When we do the same calculation for the retrospectively recalled earnings over 6 years, we reduce 74 percent of the aggregate poverty. That is, we reduce the monetary cost \( T_h(\bar{P}_2) - T(\bar{P}_2) \) of the retrospectively recalled panel and we increase the benefit \( (T_s(\bar{P}_2) - T_h(\bar{P}_2)) \) of the retrospectively recalled panel. Thanks to the underreporting of transitory variations in retrospectively recalled earnings, we have a more reliable proxy to the precise long-term welfare indicator over the longer time horizon. At least, in terms of hybrid use of the retrospectively recalled panel and surveyed panel, a retrospectively recalled panel could replace a single year cross-sectional indicator at the survey year to the specific policy objective such as reducing chronic poverty.

In addition we also use the validation earnings as the true indicator to calculate the monetary cost (additional transfer budget based on less precise indicators) of surveyed panel, retrospectively recalled panel, and cross-sectional indicators. Since the
conventional longitudinal survey data is also associated with the short term (less than a year) recall process, the similar pattern of the monetary cost of less precise surveyed longitudinal indicator is also expected. We have found that the mean-reverting measurement error is also present in the surveyed panel data where the recall lag is less than one year, even though the magnitude of mean-reverting is much smaller. Thus, the calculated cost of surveyed panel is smaller than that of retrospectively recalled panel. The calculated monetary cost of retrospectively recalled panel is still between that of surveyed panel and cross-sectional indicator.

V. Conclusion

In this paper, we assess the accuracy of long-term recall data. The results based on the unique Panel Study of Income Dynamics Validation Study suggest that retrospective recall is a poor substitute for genuine panel data in the longitudinal analyses carried out here. We find underreporting of transitory events due to recall error. The resulting error type creates a non-classical measurement error, which cannot be properly handled by conventional correction methods such as IV estimation. We propose some relevant correction methods for potential recall bias in a linear regression model. We highlight the implications of recall bias on some selected literatures such as a cyclical behaviour of real wages and the transitions of poverty status. These implications recommend us a selective use of a retrospectively recalled panel. It is a poor substitute for a conventional longitudinal data with respect to the issues of transitory aspects of labour market outcomes, but it can improve the efficiency of the specific policy objective related to the
permanent part of labour market outcomes such as reducing chronic poverty. Our limited cost-benefit analysis of a retrospectively recalled panel implies that we may improve the efficiency of the specific policy objective such as reducing chronic poverty with a hybrid use of the retrospectively recalled panel and conventional panel.
References


