

Estimating individual vulnerability to poverty with pseudo-panel data

François Bourguignon and Chor-ching Goh, the World Bank,
and Dae Il Kim, Seoul National University

Abstract

This paper presents an original method to study individual earning dynamics using repeated cross-sectional data. Because panel data of individuals are seldom available in developing countries, it is difficult to study individual earning dynamics and related issues such as the propensity of earners to fall into poverty or vulnerability to poverty because of changes in earning. This paper shows that under the assumption that individual earning dynamics obey some basic properties and follow a simple stochastic process, the main parameters of this process can be recovered from repeated cross-sectional data. The knowledge of these parameters then permits simulation of the earning dynamics of an individual, and estimate other measures of interest, such as an individual's vulnerability to poverty. Our results show that model parameters recovered from pseudo-panels approximate reasonably well those estimated directly from a true panel. Moreover, implications of the model, in this case pseudo-panel measures of vulnerability to poverty, reflect closely those based on actual panel data.

World Bank Policy Research Working Paper 3375, August 2004

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.

Acknowledgment

We wish to thank Gary Fields for his helpful comments on an earlier draft and suggestion that we make a comparison of results between panel and cross-sectional data which forms the gist of this paper.

Introduction

Studying individual earning dynamics requires panel data of individuals that are seldom available in developing countries. Hence, it is difficult to study such issues as the propensity of earners to fall into poverty or vulnerability to poverty due to changes in earnings. Because of the absence of suitable panel data in most developing countries, there is no direct way to examine individual earning dynamics or vulnerability to poverty. It may seem a priori that repeated cross-sectional data are of no use to identify individual earning dynamics because, by definition, such data do not refer to the same individuals at various points in time. However, this paper explores a methodology that permits recovering some parameters of individual earning dynamics from cross-sectional data under a set of simplifying assumptions. The methodology is based on pseudo-panel techniques focusing on second-order moments, as pioneered by Deaton and Paxson (1994). Based on these parameters, it is then possible to derive estimates on the vulnerability to poverty making use of all the cross-sectional information available at a point in time.

Our motivation for studying vulnerability to poverty, defined as the probability of earning below a poverty threshold conditional on initial earnings, stems from concerns expressed by opponents of globalization that integration exposes individuals to the vagaries of international markets, and such shocks may be transmitted to greater volatility and uncertainty in earnings of individual workers. The East Asian financial crisis rekindled this anxiety. There has been little empirical work to investigate the linkage between shocks at the macro level and vulnerability at the level of individual workers. Within the large literature on wage inequality and wage differentials in relation to globalization, only a handful of studies — mostly on Latin American economies perhaps because macroeconomic volatility appears to be structurally higher there — examine this relationship, taking changes in employment as the indicator of vulnerability.¹ De Ferranti and others (2000) summarize issues of worker insecurity and economic openness in Latin America: they find that wage volatility is affected more by inflation than by openness, and that many countries experienced more stable wages during the more open 1990s. Bourguignon and Goh (2004) make a first attempt to investigate this topic in an East Asian context. They find that there was no correlation between trade liberalization and vulnerability to poverty in that region.

The objectives of this paper are to present an original method to study individual earning dynamics using repeated cross-sectional or pseudo-panel data, and to compare the accuracy of these estimates with those produced from true panel data. In our case, a pseudo-panel is formed by following cohorts of randomly selected individuals born in a 5-year interval over time in successive cross-sectional surveys, that is, we are tracking

¹ For instance, Revenga (1997) finds that Mexico's trade reform of 1985-88 reduced employment modestly, but did not reduce wages. Cox Edwards and Edwards (1996) find that Chile's trade liberalization of the 1970s affected workers' duration of unemployment, but its effect was small relative to those of other variables, and declined over time. Arango and Maloney (2002) find some evidence of higher incidence of involuntary separation, mostly among skilled workers, in sectors that are opening to trade in Mexico and Argentina, but the impact is transitory.

over time male workers born in 1946-50 as one homogeneous group; male workers born in 1951-55 as another cohort of group; male workers born in 1956-60 as yet another group and so on. We discuss in section I the method that recovers features of individual earning dynamics from pseudo-panel data. The idea is as follows: if it may be assumed that all individuals within a cohort face a stochastic earning process that has common characteristics, these characteristics may be recovered at the aggregate level, without observing actual earning paths. Observing the evolution of the mean and the variance of earnings within a cohort is sufficient to estimate the common characteristics of individual earning processes. On this basis, simple estimates of the probability of a worker observed in year t to fall into poverty in year $t + 1$ can be worked out. In section II, we apply this method to repeated cross-sectional data in the Republic of Korea, using them as a pseudo panel. We then check the relevance of this approach by applying it to a pseudo panel constructed from true panel data in Korea. Korea was selected because few other developing countries have reasonably long and representative panel data on earnings. The panel data sets that are suitably long enough for us to check the quality of earning dynamics estimates based on pseudo-panel are the Korea Labor Institute Panel Study data and the Korean's Urban Worker Household Income and Expenditure Surveys. In section III, we evaluate the quality of the approximation of pseudo-panel estimates vis-à-vis direct individual panel estimates.

Our results show that the basic earning dynamics parameter--- i.e. the persistence of earnings shocks from one period to the next --- recovered from repeated cross-sectional data, or a pseudo-panel, are not significantly different from those estimated from a true panel. Another parameter of the model, the variance of the earning innovations, recovered from a pseudo-panel also approximate those estimated from a true panel. With regard to our variable of interest, the vulnerability to poverty, estimates simulated from a pseudo-panel track very closely those from a true panel.

I. A model for recovering earning dynamics features from repeated cross-sectional data

Assume that the earnings, w_{it}^j , of individual i belonging to cohort-group j at time t may be represented by the following equation:

$$(1) \quad \ln w_{it}^j = X_{it}^j \beta_t^j + \xi_{it}^j$$

where X_{it} is a set of individual characteristics like age or educational attainment and ξ_{it}^j stands for unobserved permanent earning determinants as well as the transitory component of earnings. Accordingly, assume that this residual term ξ_{it}^j follows an autoregressive process $AR(1)$:

$$(2) \quad \xi_{it}^j = \rho^j \xi_{it-1}^j + \varepsilon_{it}^j$$

where ε_{it}^j is the innovation in earnings and is supposed to have a variance $\sigma_{\varepsilon_{it}^j}^2$.

Suppose now that repeated cross-sectional data are available for periods $t = 1, 2, \dots, T$. If the sample is representative of the whole population at each period, a sample of individuals belonging to each cohort j is observed in each period. It is thus possible to follow cohort j over time. But, because individuals in two successive cross-sections are not identical, it is not possible to observe ξ_{it}^j and ξ_{it-1}^j for the same person i . Thus, model (1)-(2) cannot be readily estimated. Nevertheless, it is possible to extract from these cross-sections some information on the basic dynamic parameters ρ^j and $\sigma_{\varepsilon_{it}^j}^2$. Under the assumption that individuals enter and exit randomly the labor force between two successive periods, it is the case from (2) that the variance $\sigma_{\xi_{it}^j}^2$ of the residual ξ_{it}^j behaves according to the following process:

$$(3) \quad \sigma_{\xi_{it}^j}^2 = \rho^{j2} \sigma_{\xi_{it-1}^j}^2 + \sigma_{\varepsilon_{it}^j}^2$$

The preceding equation may be used to recover the dynamic parameters ρ^j and $\sigma_{\varepsilon_{it}^j}^2$. After having estimated equation (1) on each cohort j separately for each period t , it is a simple matter to get estimates of the residual variance $\sigma_{\xi_{it}^j}^2$. We will need at least three periods to be able to estimate ρ^j by OLS² from equation (3); then, the residuals provide estimates of the variance of the innovation term $\sigma_{\varepsilon_{it}^j}^2$.

While technically three cross-sections will allow us to estimate equation (3), most likely ρ^j will be very imprecisely estimated with such few time observations. This might be remedied by imposing some restriction on the parameter ρ^j across cohorts j . For instance, one could impose this coefficient to be the same across a number of cohorts, or among members of the same cohort belonging to various socio-demographic groups.

If the model is well specified and enough time observations are available, then the estimated $\hat{\rho}^j$ and $\hat{\sigma}_{\varepsilon_{it}^j}^2$ will have the expected signs and magnitude, that is, $0 < \hat{\rho}^j < 1$ and $\hat{\sigma}_{\varepsilon_{it}^j}^2 > 0$ for all t . If estimates are not well-behaved, the hypotheses behind equation (3) –

² We need estimates in equation (3) to behave in a certain way, and must exercise caution when using OLS. First, OLS estimation of equation (3) must be done without an intercept. Second, we must take into account that residuals in equation (3) must be non-negative. Third, the estimated coefficient in equation (3) must be between zero and one. OLS estimation does not automatically satisfy these restrictions. For example, we can use more rigorous ways to impose the second restriction of non-zero residuals by having a half-normal distribution truncating to zero for the residual term in equation (3) (Battese and Coelli (1988)). However, we didn't have to impose such restrictions in the paper because OLS estimates always yield non-zero residuals and a coefficient between zero and one.

i.e. the first-order autoregressive process on earnings or the randomness of entries/exits - have to be rejected. The preceding method has been applied to cross-sectional data from Indonesia, Korea, and Thailand (Bourguignon and Goh (2004)). Reasonable estimates of the parameters of the model were obtained for all countries.

Before discussing the results, two remarks are in order. The first remark concerns how the preceding assumption about individual earning dynamics leads to the mean vulnerability of individuals, observed in cross-section t , to poverty in period $t+1$, conditional on their initial earnings and characteristics. Some additional assumptions are necessary for this last step. The first assumption is that the innovation term is distributed as a normal with mean 0 and variance $\hat{\sigma}_{\epsilon t}^2$, so that earnings are distributed as a lognormal variable, conditional on individual characteristics, X . The second assumption is that some prediction of future individual characteristics \hat{X}_{it+1}^j is available – this is easy for variables like age or educational attainment; other variables might have to be assumed stationary. The same applies to future earning coefficients $\hat{\beta}_{it+1}^j$ and the variance of the innovation, $\hat{\sigma}_{\epsilon t+1}^2$. In both cases, the simplest assumption is that the parameters are stationary. Yet, the intercept coefficient in $\hat{\beta}_{it+1}^j$ may be modified so as to capture the expected growth rate in earnings, whereas $\hat{\sigma}_{\epsilon t+1}^2$ may in some cases reflect the effect of macro-economic shock or on the contrary a stabilization.

Under the preceding assumptions, and denoting $\hat{\xi}_{it}^j$ the estimated residual of the earning equation (1) in period t , the probability of earning less than a poverty threshold, \bar{w} , at time $t+1$, conditional on characteristics of period t is given by:

$$(4) \hat{v}_{it}^j = pr(\ln w_{it+1}^j < \ln \bar{w} | X_{it}^j, \hat{X}_{it+1}^j, \hat{\beta}_{it+1}^j, \hat{\sigma}_{\epsilon t+1}^2) = \Phi\left(\frac{\ln \bar{w} - \hat{X}_{it+1}^j \hat{\beta}_{it+1}^j - \hat{\rho}^j \hat{\xi}_{it}^j}{\hat{\sigma}_{\epsilon t+1}^2}\right)$$

where $\Phi(\cdot)$ denotes the cumulative density of the standard normal. Thus, \hat{v}_{it}^j is the vulnerability of individual j , belonging to cohort j and observed at time t , to falling into poverty at time $t+1$.

The second remark is about the possibility of checking the relevance of the approximation of earning dynamics by the preceding method. Doing so requires true individual panel data. If such data are available, one can compare the indirect estimates of the dynamic parameters $\hat{\rho}^j$ and $\hat{\sigma}_{\epsilon t}^2$ obtained through equation (3) using the cross-sectional nature of the data to the direct estimates of model (1)-(2) obtained using the full panel dimension of the data. It can be seen that the latter is equivalent to estimating the model:

$$(5) \ln w_{it}^j = \rho^j \ln w_{it-1}^j + X_{it}^j \beta_t^j + \gamma_{it-1}^j X_{it-1}^j + \epsilon_{it}^j \quad \text{with} \quad E(\epsilon_{it}^j) = 0 \quad \text{and} \quad V(\epsilon_{it}^j) = \sigma_{\epsilon t}^2$$

In this expression, γ_{it-1}^j actually stands for $-\rho^j \beta_{t-1}^j$ but this is not a restriction as long as the coefficients β_t^j are allowed to change with time. It may also be noted that estimating the preceding model through OLS may be done even when the individual characteristics X_{it}^j do not change over time.

Of course, checking whether the pseudo-panel estimates of earning dynamics are satisfactory can also be done by looking at the implications of the model rather than the estimated parameters. In the present case, this means comparing the estimates of vulnerability to poverty obtained through expression (4) with the actual frequency of falling into (or remaining in) poverty in the panel data.

II. Application to Korea

Repeated cross-sectional data on individual earnings are available in a large number of developing countries, whereas panel data are not easily available. Korea is among the few countries where suitable, albeit very short, panel data are available for evaluating the relevance of the preceding methodology.

The largest cross-sectional data set on individual earnings in Korea is the Wage Structure Survey (WSS), formerly the Occupational Wage Survey, 1991-2000. This is an establishment survey and only wage earners in non-agricultural private firms with 10 or more workers are in the sample. The survey collects information on firms' activity and workers' education, age, job tenure, occupation and monthly wages. The sample size ranges from 450,000 to 500,000 each year. As the survey samples firms with 10 or more workers, sectors with larger firms tend to be over-sampled. In particular, manufacturing is overly represented while retail trade and service sectors are under-represented.

Panel data sets on individual earnings in Korea are much smaller in size and much shorter in the time dimension. Two data sets are available: the Korea Labor Institute Panel Study Data (KLIP), 1998-2001, and the Urban Worker Household Income and Expenditure Survey (UWH), 1994-2000. The KLIP first sampled 5,000 households in urban areas in 1998, approximately 70 percent of which remained in the sample by 2001. The households that left the sample were not replaced. As a result, the survey included 13,738 persons in 1998, but this number fell to 10,179 in 2001. The survey contains information on working status, earnings, and job characteristics such as industry and occupation.

The UWH is a household panel survey covering urban areas. It provides earnings information only for those households headed by a wage/salary worker. It samples 35,000 to 40,000 households each year and provides information on total household earnings and heads' earnings and job characteristics. Although data are available for 1994-2000, the entire sample is replaced every 5 years. Actually, only two short panels are available: 1994-97, and 1998-2000.

The pseudo-panel methodology discussed above requires the maximum number of time observations to yield more precise estimates of earning dynamic parameters. In their true panel dimension, the two panel data sets available in Korea actually permit no more than three observation periods, since the use of lagged values in the equations to be estimated eliminates the first period. Yet, because two data sets are available in the UWH data source, it is possible to use slightly more observation periods in that case. This is the reason our results discussed in this section are based only on the cross-sections of data available in the WSS and the UWH; the KLIP has too short a series to construct a pseudo-panel. For brevity, results are presented and discussed only for male earners – and male household heads in the case of UWH.

Table 1 presents the estimated persistence in the residuals of earnings equation (1), $\hat{\rho}^j$, for the two pseudo-panels. Explanatory variables in that regression include years of age, age squared, educational attainment, marital status and a dummy variable denoting self-employment (for UWH). Since the persistence parameter in equation (3) comes as the square of $\hat{\rho}^j$, a simple transformation was used to obtain an estimate of the standard error of $\hat{\rho}^j$. It can be seen that the estimates of $\hat{\rho}^j$ for both pseudo-panels are reasonably between zero and one, with $\hat{\rho}^j$'s significantly different from zero. The $\hat{\rho}^j$'s are not very precisely measured due to very few observation periods. As an F-test indicates that the $\hat{\rho}^{2j}$'s are not statistically different among cohorts, one can hope to increase precision by pooling the cohorts together and assuming a common $\hat{\rho}$. The last row of Table 1 presents the cohort-combined $\hat{\rho}$'s for the two pseudo-panels, which are 0.63 and 0.85, respectively. Contrary to what we hope, the precision of these estimates is not better than that of cohort-specific estimates because of too much cohort heterogeneity.

That the estimate of persistence is higher with UWH than with WSS is not surprising given that the samples are different. UWH data cover household heads for whom earnings are less volatile and more predictable from one year to the next. In addition, there are likely to be fewer entries and exits from the labor force among household heads, which may reinforce the stability of earnings in UWH.

Based on the estimated persistence in shocks, $\hat{\rho}$, we plot the cohort-specific variance of innovation terms, $\hat{\sigma}_{\epsilon_{it}}^2$, for both pseudo-panels in Figure 1. The repeated cross-sections drawn from the UWH (right graph) show a sharp spike of variance in 1998, reflecting the shock of the financial crisis. Interestingly enough, the WSS data (left graph) show a gradual rise in the variance of the earning innovation that started with the crisis in 1998 and continued an upward trend into 2000. It is tempting to relate these differences again to the definition of the two samples. The story suggested by the two charts in Figure 1 is that the destabilization of the labor market due to the 1998 crisis was limited to the crisis year for household heads, people who generally have steadier career paths and earning profiles. It went beyond the crisis years for secondary, or marginal workers, who are traditionally more mobile across jobs than household heads. This interpretation is reinforced by the fact that, except for the oldest cohort, the variance of

earning innovation for household heads fell back after the crisis to a level higher than that observed before the crisis.

Figure 2 presents the vulnerability measures based on the pseudo-panels and computed according to equation (4). Unsurprisingly, the time evolution of vulnerability to poverty reflects closely the trend of the $\hat{\sigma}_{ijt}^2$. Both data sets show that workers with less education experience greater vulnerability to falling into poverty. They also confirm that the labor market in Korea is fluid, and workers are mobile between sectors.³ Whether a worker is in the tradable manufacturing sector or the non-tradable sector, there is no difference in vulnerability to poverty between sectors.

III. Results from True Panel Data

In this section, we estimate individual earning dynamics based on true panel data and compare the true panel estimates with the cross-sectional (pseudo-panel) estimates to check the precision of the latter. Two sets of panel data are used: the KLIP (1998-2001), and the UWH (1994-2000).

Table 2 presents the persistence in earnings shocks, $\hat{\rho}^j$, for the two panels. As was the case when comparing pseudo-panel estimates obtained with WSS and UWH, true panel estimates of persistence parameters differ between the two panel data sets, KLIP and UWH. They are higher for the sample of male household heads in UWH than for the sample of all male wage/salary workers in KLIP. In both cases, one also observes that the persistence parameter declines when moving from an older cohort to a younger cohort, a fact well documented in the literature on earnings mobility.⁴ In effect, pooling together all cohorts and allowing the persistence parameter to depend linearly on the middle birth year of each cohort (i.e., $\hat{\rho} = \hat{R} + \hat{\gamma} * cohort$) does not reduce significantly the information compared with cohort-specific parameters. In contrast with what was observed with pseudo-panels, however, imposing a constant persistence parameter across cohorts is restrictive.

We now compare the estimates obtained with the pseudo-panel made up of the WSS cross-sections in the previous section with estimates obtained from true panel estimates. The best comparison is with KLIP which does not restrict the sample to household heads. The respective estimates of persistence parameters are shown in Table 3 under alternative restrictions for KLIP. It turns out that the $\hat{\rho}$ based on the repeated WSS cross-section is not significantly different from the $\hat{\rho}$ based on the true panel KLIP when the latter is restricted to be identical across cohorts. The former is .625 whereas the latter is .614. This seems extremely satisfactory. But it should not hide the fact that going back to cohort-specific estimates in Tables 1 and 2, WSS cross-sectional estimates do not pick up at all the age or cohort profile of persistence parameters that is apparent in true panel estimates. This is possibly because of a lack of precision of the

³ See Fields (2000) for a discussion of Korean labor market problems.

⁴ See for instance Atkinson, Bourguignon and Morrison (1992).

cross-sectional estimates. Indeed, comparing the first columns in Tables 1 and 2 shows no significant difference.

Figure 3 presents the cohort-specific $\hat{\sigma}_{ijt}^2$'s obtained from cross-sectional WSS estimates and those estimated on the basis of the true panel KLIP for years 1999 and 2000. The cohort specificity of $\hat{\sigma}_{ijt}^2$ based on the repeated cross-sections approximate very closely those of $\hat{\sigma}_{ijt}^2$ estimated from the panel. Overall, however, KLIP estimates are slightly higher than WSS estimates. From 1999 to 2000, there is slight increase in the variance of innovation with both estimation techniques. While the change is uniform with WSS, it is more cohort-specific with the true panel estimates obtained with KLIP.

Instead of comparing pseudo-panel and true panel estimates obtained from different data sources, it is also informative to compare the two estimates using the same panel data set. In one case, the panel dimension of the data is ignored and only the repeated cross-sections are used to estimate equation (3). In the second case, the panel dimension is used to estimate model (5). The KLIP panel is not very interesting from that point of view because the time dimension of the data is simply too short. This is the reason we now switch to the UWH data set.

Table 4 presents the $\hat{\rho}$ based on the pseudo and true panels obtained from UWH. When the persistence parameter is constrained to be constant across cohorts, the pseudo-panel estimate, at 0.85, is close to, and certainly not significantly different from the panel estimate, at 0.80. As in the preceding comparison, however, the pseudo-panel estimate misses the cohort specificity of the persistence parameters apparent in the true panel estimates.

Figure 4 presents the trend of variance of earnings innovation, $\hat{\sigma}_{ijt}^2$, for all cohort groups combined, based on pseudo and true panels. Note that there are only 4 overlapping years (i.e., 1995-97, and 1999) for the pseudo- and true panel because the UWH survey renewed its sample in year 1998, and we have a first-order autoregressive model. The comparison of pseudo and true panel estimates for each cohort during the overlapping years (not presented here) shows very close approximation, similar to that in Figure 3.

Note that the variance estimated on the basis of the true panel is on average larger than that estimated on the basis of the pseudo panel. These discrepancies can easily be explained. Note that the estimated persistence parameter from the pseudo-panel is above the corresponding estimate from the true panel --- 0.85 versus 0.80 on average. It follows from equation (3) that the variance of earning innovation, σ_{ijt}^2 , is smaller with the pseudo-panel. The gap between the two estimates depends on the variance of earnings residuals of the previous period, σ_{ijt-1}^2 . The variance of earnings is higher during the financial crisis years 1998-99. Accordingly, the gap between pseudo and true panel

estimates of the $\hat{\sigma}_{ijt}^2$ is larger on average in 1999 and 2000. Also, the levels of $\hat{\sigma}_{ijt}^2$ for both pseudo and true panel in 1999 and 2000 are higher than levels in pre-crisis years.

According to the preceding argument, the time evolution of the innovation variance with the pseudo and true panel estimates should be approximately parallel for all cohorts. That this is not the case is due to the fact that the pseudo-panel estimates are not defined on a balanced panel whereas the true panel estimates are. That this makes a difference suggests that exits from the panel cannot always be considered as randomly distributed in the population.

Table 5 presents our variable of interest, vulnerability to poverty, for the first of the two preceding comparisons --- that is, the cross-sectional WSS and the KLIP panel --- for years 1999 and 2000. The poverty threshold is defined as 50 percent of the median. While the point estimates are not identical, cross-sectional and panel vulnerability measures are very close to each other. In both cases, we find that vulnerability does not differ by sector, but depends on educational attainment.

Figure 5 presents the evolution of vulnerability to poverty for our second set of comparisons (that is, the pseudo and true panel of the Urban Worker Family Income and Expenditure Surveys) between 1995 and 2000. We present the comparison between pseudo and true panel estimates for the tradable and non-tradable sectors. Both graphs are close images of each other, reflecting the similar trends of vulnerability in tradable and non-tradable sectors. The trends of pseudo-panel estimates of vulnerability approximate closely the trends of true panel estimates within each sector of employment.

It may be tempting to make a comparison of point estimates of vulnerability measures between pseudo and true panel data, and to make a statement about the precision of the pseudo-panel estimates. However, it's not very meaningful to make an assessment of the point estimates especially in our first set of comparisons between cross-sectional WSS and panel KLIP (Table 5). In this case, we are looking at two different samples. In addition to the $\hat{\sigma}_{ijt}^2$ and $\hat{\rho}$, other parameters such as $\hat{\beta}_{t+1}^j$, average earnings, and the poverty threshold also differ across the data sets. In our second set of comparisons, both the pseudo and true panels come from one data set, but since the pseudo panel is not created from a balanced panel, we are again looking at two different samples. In this case, we have more overlapping years, which allow us to compare the trend of vulnerability based on the pseudo panel with that based on the true panel.

Conclusion

This paper explores a methodology that permits recovering some parameters of individual earning dynamics from cross-sectional data under a set of simplifying assumptions that individual earning dynamics obey some basic properties and follow a simple stochastic process. The knowledge of these parameters then permits one to

simulate individual earning dynamics and estimate vulnerability to poverty, making use of all the cross-sectional information available at a point in time.

The application of this methodology to Korean data yields rather satisfactory results. Two sets of comparisons were undertaken in order to check its relevance in comparison with standard panel data analysis. In the first comparison, estimates of a simple AR(1) earning dynamics model are obtained from a pseudo-panel derived from repeated cross-sectional surveys and a true panel of earnings data. The second comparison is between a panel data set and the pseudo-panel constructed from it. In both cases, it is shown that the estimated parameters of individual earning dynamics processes based on the pseudo-panel approximate very closely the direct panel estimates. The point estimates of the measure for persistence of shocks, based on the pseudo-panel of cohorts, are very close to those based on the true panel. Both estimates are not significantly different from each other. The other key parameter of individual earning dynamics, the variance of the innovation in earnings, estimated from a pseudo-panel of cohorts, tracks closely the direct panel estimates in the overlapping years.

Given that the pseudo and true panel estimates of the earning dynamics are not exactly identical, the vulnerability measures derived from the earning dynamics are similar in trends but not identical in average point estimates.

The methodology developed in this paper has some obvious weaknesses. First, by relying on aggregate data, the degrees of freedom of the estimation depends on the available number of cross-sections. As this number is necessarily limited, not very much precision may be expected. Second, and more importantly, this technique is valid only under the assumption that entries and exits from employment are random with respect to the distribution of individual earnings. Moreover, it focuses on the earning dynamics of those individuals who are employed on a continuous basis. Practically, however, we know that the first assumption is unlikely to be satisfied and also that the main source of vulnerability to poverty may not be in variations in earnings but in the employment status of individuals. Losing one's job and therefore leaving employment may be the most important event behind fluctuations in economic welfare and poverty dynamics. This is a dimension that was not considered in the present paper. Yet, it is likely that the same kind of pseudo-panel techniques used for earnings may be used for employment status, and possibly simultaneously for both. This important dimension of vulnerability to poverty and the way to approach it with cross-sections is left for further work.

Table 1. Estimates of $\hat{\rho}^j$'s based on pseudo-panels constructed from the cross-sectional Wage Structure Survey (WSS), 1990-2000, and the Urban Worker Household Income and Expenditure Survey (UWH), 1994-2000

	$\hat{\rho}^j$ (std error)	
Cohort, by birth year:	WSS	UWH
1941-45	0.686 (.189)	0.769 (.199)
1946-50	0.617 (.182)	0.935 (.179)
1951-55	0.478 (.153)	0.947 (.221)
1956-60	0.421 (.138)	0.866 (.186)
1961-65	0.688 (.201)	0.957 (.181)
1966-70	0.874 (.150)	0.444 (.146)
1971-75	0.762 (.198)	0.756 (.221)
All cohorts combined	0.625 (.189)	0.850 (.202)

Figure 1. Estimates of cohort-specific $\hat{\sigma}_{ijt}^2$ for the WSS and UWH pseudo-panels

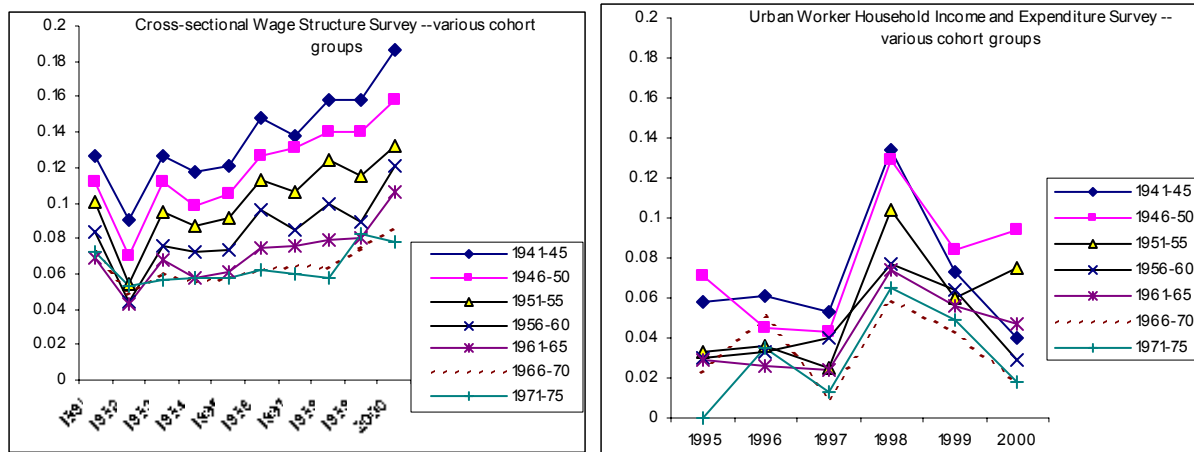
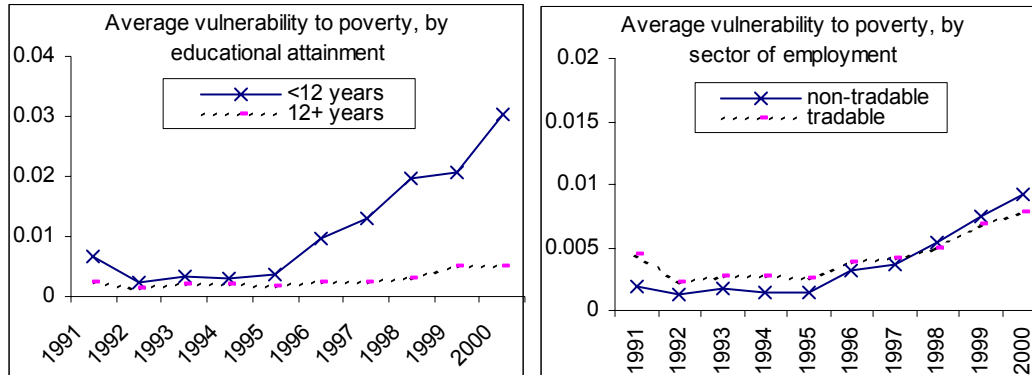


Figure 2. Measures of vulnerability to poverty, by educational attainment and by sectors of employment, for the WSS and UWH pseudo-panels

WSS:



UWH:

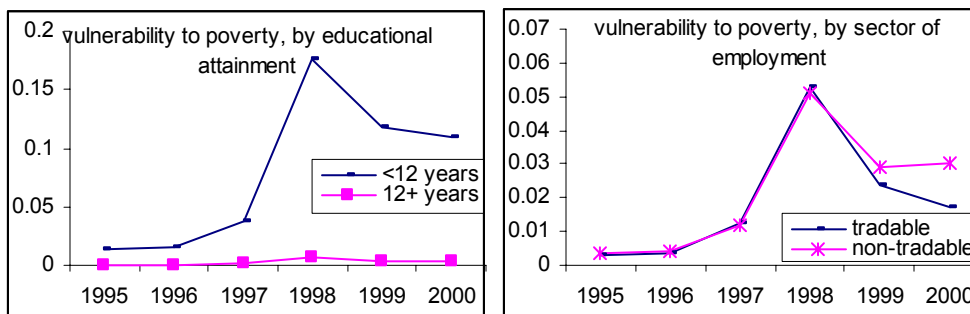


Table 2. Estimates of persistence parameter, $\hat{\rho}^j$, for two panel data sets, the Korea Labor Institute Panel Study data (KLIP), and Korea's Urban Worker Household Income and Expenditure Survey (UWH)

	Persistence parameter (std error)			
Cohort, by birth year:	KLIP		UWH	
1941-45	0.722 (.035)		0.828 (.018)	
1946-50	0.670 (.025)		0.888 (.015)	
1951-55	0.675 (.026)		0.754 (.014)	
1956-60	0.548 (.023)		0.763 (.013)	
1961-65	0.588 (.022)		0.776 (.012)	
1966-70	0.570 (.023)		0.816 (.017)	
1971-75	0.439 (.032)		0.640 (.041)	
	$\hat{\rho} = \hat{R} + \hat{\gamma} * cohort$			
	\hat{R} (std error)	$\hat{\gamma}$ (std error)	\hat{R} (std error)	$\hat{\gamma}$ (std error)
Cohorts combined:	0.748 (.022)	-0.036 (.006)	0.868 (.014)	-0.019 (.004)

Table 3. Estimates of persistence parameter: comparison between the cross-sectional WSS and panel KLIP

<u>Pseudo-panel</u> WSS		<u>Panel:</u> KLIP			
$\hat{\rho}$ (std error)		$\hat{\rho}$ (std error)		$\hat{\rho} = \hat{R} + \hat{\gamma} * cohort$	
				\hat{R} (std error)	$\hat{\gamma}$ (std error)
0.625 (.189)		0.614 (.010)		0.748 (.022)	-0.036 (.006)

Figure 3. Estimates of the variance of earnings innovation $\hat{\sigma}_{ijt}^2$ from cross-sectional WSS and panel KLIPS, 1999 and 2000

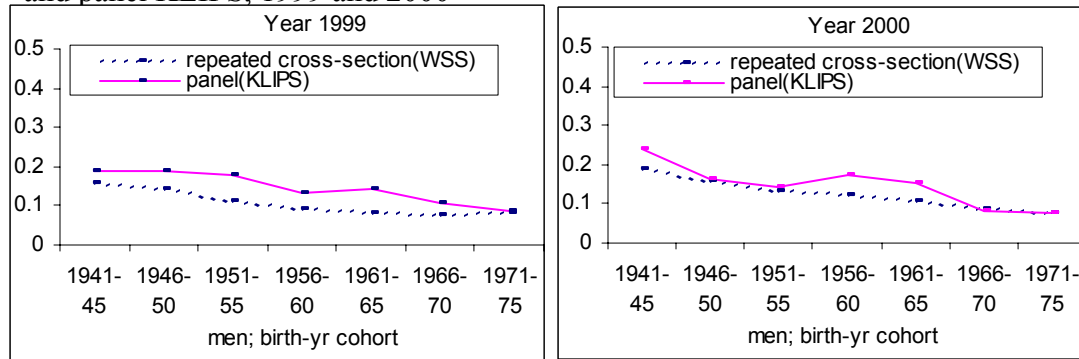


Table 4. Estimates of the persistence parameter based on UWH: comparison between pseudo and true panel estimates

<u>Pseudo panel</u>		<u>True panel</u>		
			$\hat{\rho} = \hat{R} + \hat{\gamma} * cohort$	
$\hat{\rho}$ (std error)		$\hat{\rho}$ (std error)	\hat{R} (std error)	$\hat{\gamma}$ (std error)
0.850 (.202)		0.801 (.006)	0.868 (.014)	-0.019 (.004)

Figure 4. Estimates of the variance of earnings innovation $\hat{\sigma}_{ijt}^2$ based on UWH : pseudo-versus true panel estimates

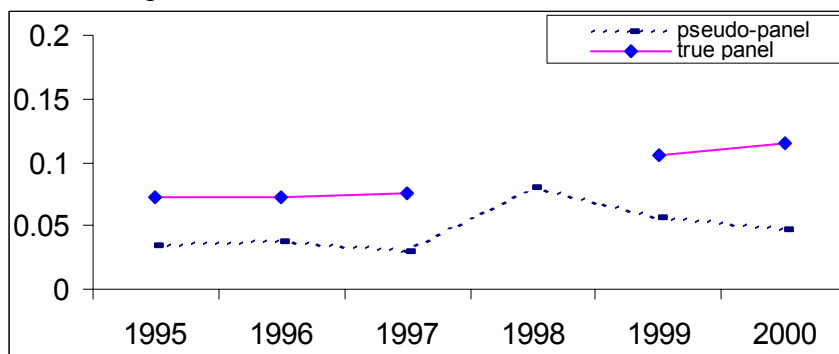
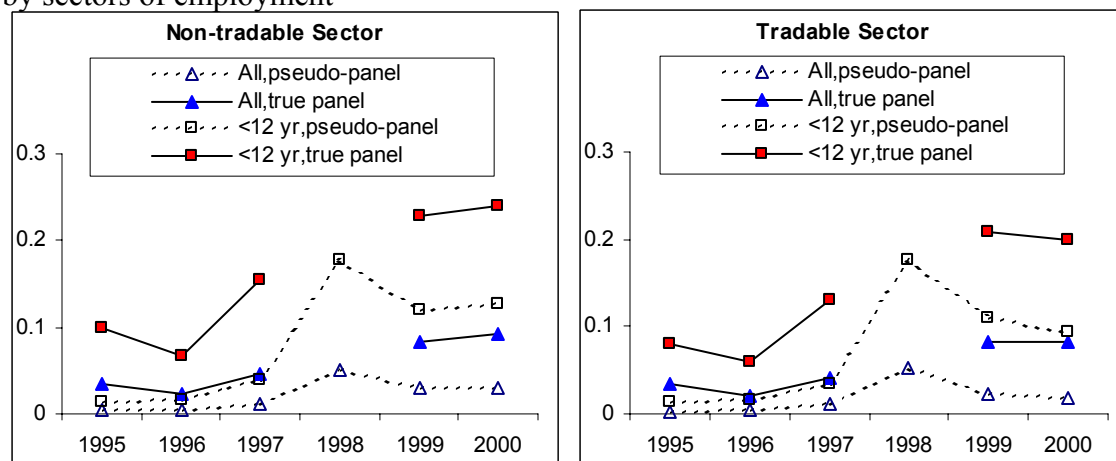


Table 5. Vulnerability to poverty based on the cross-sectional WSS and the panel KLIP, 1999 and 2000

Year		Vulnerability to poverty: $\Phi\left(\frac{\ln \bar{w} - \hat{X}_{it+1}^j \hat{\beta}_{t+1}^j - \hat{\rho}^j \hat{\xi}_{it}^j}{\sigma_{\hat{\xi}_{it+1}^j}^2}\right)$	
		Pseudo-panel from repeated cross-section	True panel
1999	All	.043	.057
	Tradable Sector	.045	.056
	Non-tradable Sector	.041	.057
2000	All	.045	.078
	Tradable Sector	.046	.074
	Non-tradable Sector	.044	.080
1999	with less than 12 years of schooling	.090	.14
	with 12 or more years of schooling	.036	.031
2000	with less than 12 years of schooling	.11	.19
	with 12 or more years of schooling	.035	.043

Figure 5. Vulnerability to poverty based on UWH: pseudo- versus true panel estimates, by sectors of employment



References

- Atkinson, A. B., François Bourguignon and C. Morrison (1992). *Empirical Studies of Earnings Mobility*. Harwood Academic Publishers, Philadelphia, PA.
- Arango, Carlos, and William Maloney (2002). "Unemployment Dynamics in Latin America: Estimates of Continuous Time Markov Models for Mexico and Argentina." World Bank, mimeo.
- Battese, G., and T. Coelli (1988). "Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data," *Journal of Econometrics* vol 38, pp. 387-99
- Bourguignon, François, and Chor-ching Goh (2004). "Trade and labor vulnerability in Indonesia, Republic of Korea, and Thailand," in Kharas, H and Krumm K (eds), *East Asia integrates: a trade policy agenda for shared growth*. World Bank and Oxford University Press, Washington, DC.
- Cox Edwards, A., and Sebastian Edwards (1996). "Trade liberalization and unemployment: policy issues and evidence from Chile." *Cuadernos de Economia*, Año 33, No. 99, pp. 227-50.
- De Ferranti, D., et al. (2000). *Securing our future in a global economy*. World Bank, Washington, D.C.
- Deaton, Angus, and Christina Paxson (1994). "Intertemporal choice and inequality." *Journal of Political Economy*, vol. 102, 437-67.
- Fields, Gary (2000). "The Employment Problems in Korea." *Journal of the Korean Economy*, vol. 1:2, pp.207-27.
- Revenge, Ana (1997). "Employment and wage effects of trade liberalization : the case of Mexican manufacturing", *Journal of Labor Economics* vol. 15 n3(2), pp. S20-43.