

# Estimating Vietnam's income mobility in 2004 – 2008: a test – retest correlation approach

- Paul Glewwe
- Khoa Vu  
University of Minnesota

(Manuscript Received on October 3<sup>rd</sup> 2013, Manuscript Revised on October 7<sup>th</sup> 2013)

## ABSTRACT:

*The economic growth of Vietnam in the 1990s has been a popular topic among the economists because there are many aspects of it are subjected to development studies. This paper attempts to explore one of these aspects, the income mobility of the economy, during the period 2004-2008 by estimating expenditure mobility, using Vietnam Household Living Standard Survey (VHLSS)*

*This is done by applying a methodology that Heise (1969) developed in his work on test-retest correlations, to reduce the classical upward bias due to measurement errors. We estimate the mobility to be 0.035 to 0.092 which indicate a low mobility in Vietnam. This estimation allows us to draw out some implications about income inequality in Vietnam.*

**Keywords:** *income mobility in Vietnam, test-retest reliability and stability coefficients, VHLSS, measurement error.*

## 1. INTRODUCTION

Income mobility is a useful measure for understanding the movements of different income groups in an economy; high mobility implies significant movement of individuals between income groups, i.e. the poor climbing up the income ladder or the rich moves down compared to their original position, while low mobility indicates there is hardly any movement at all. Income mobility is often estimated as a relationship between the income data in different periods. Hence, one simple approach to measure the degree of mobility in a country is to estimate the Pearson correlation coefficient of income in

two successive periods. Another simple way to estimate income dynamics over time is to estimate the slope coefficient on the log of income obtained by a simple linear regression [1, 2]. However, these estimations of mobility suffer an upward bias, because observed income often comes with large errors [3, 4]. As true income is almost impossible, many scholars have developed different ways to reduce the impact of measurement error of the observed income on mobility estimates, one of which is to utilize different variables such as expenditure, education, or body weight as instruments to estimate the correlation coefficient of the regression [3, 5, 6]. Antman and McKenzie prove

that this method would require certain conditions of the instruments in order for the result to be consistent. Specifically, the following three conditions need to hold:

The instrument must be uncorrelated with the measurement error in two periods,

The instrument must be uncorrelated with the error term in the regression, and

The degree of autocorrelation in the instrument must be equal to the degree of autocorrelation in income.

Expenditure, often used as an instrumental variable for income, is likely to be correlated with the measurement error of the income variable as it is most likely to display the same degree of autocorrelation as true earnings. Therefore, expenditure is likely to violate the first two conditions; the third condition makes it even less likely for most possible instruments to be valid except those which do not vary over time. The authors conclude that the instrumental variable method is unlikely to yield a consistent estimate of mobility because it is almost impossible to find an instrument that satisfies all three conditions. Antman and McKenzie also suggest using a pseudo-panel method which “tracks cohorts of individuals, such as birth-cohorts or birth-education cohorts, over repeated cross-sectional survey”. However, this method does not apply for the individual who migrates in or out of the survey area [4], which is common in Vietnam. Therefore, a different method is needed to estimate the income mobility of Vietnam while reducing the impact of measurement error.

## 2. METHODOLOGY

### 2.1 Test-Retest Correlations

This paper takes a different approach by using a methodology developed by Heise [7], which allows us to estimate the true degree of mobility in the absence of true income data. Using panel data of three waves of time, this

“test-retest correlations” method is based on a set of observations about the relations of the true income ( $Y_{i,t}$ ) and the observed income ( $y_{i,t}$ ); these observations are obtained from path analysis developed by Wright [7]. Firstly, one can see that observed incomes are determined by both the true incomes and measurement errors. Secondly, true income in one period is determined by that in the previous periods. These observations are expressed in five equations below:

$$- y_{i,0} = \lambda_0 \cdot Y_{i,0} + \delta_0 \cdot \varepsilon_{i,0}$$

$$- y_{i,1} = \lambda_1 \cdot Y_{i,1} + \delta_1 \cdot \varepsilon_{i,1}$$

$$- y_{i,2} = \lambda_2 \cdot Y_{i,2} + \delta_2 \cdot \varepsilon_{i,2}$$

$$- Y_{i,1} = \beta_{10} \cdot Y_{i,0} + Z_{i,1}$$

$$- Y_{i,2} = \beta_{21} \cdot Y_{i,1} + \beta_{20} \cdot Y_{i,0} + Z_{i,2}$$

(With  $i = 1, \dots, N$ , the  $\lambda_s$ ,  $\delta_s$ , and  $\beta_s$  are unknown, and the error terms are  $\varepsilon_{i,t}$  and  $Z_{i,t}$ .)

The first three observations indicate that the observed income is determined by the true income and the measurement error. The coefficients  $\lambda_i$ , the traditional reliability coefficients, indicate the relationship between the true and observed incomes, and the  $\delta_s$  imply the variance of observed income as a function of errors. The last two equations show the relationship between true incomes in different periods, i.e. the variability of true income over time; therefore, the coefficients  $\beta_{i,t}$  are called stability coefficients. The goal of this method, then, is to estimate the impacts of true income and measurement errors on observed income separately through a test-retest process. The impact of true income comes from the variability of true income over time which is indicated by the stability coefficients. On the other hand, the impact of measurement error is just the residual part because the error terms are random. In order to solve this system, there are three important assumptions to simplify the model:

The unobserved error terms  $\varepsilon_s$  are not correlated with each other. The unobserved error

terms  $Z_s$  are also not correlated with each other. Furthermore, we also assume that  $\varepsilon_s$  are not correlated with the true incomes.

There is no effect of  $Y_0$  on  $Y_2$ ; hence,  $\beta_{20}=0$

The parameters  $\lambda_t$  and  $\delta_t$  are related as followed:

- $\lambda_t = \rho t$
- $\delta_t = 1 - \rho t$
- $\rho_0 = \rho_1 = \rho_2 = \rho$

This constrains the variance of  $Y$  and  $\varepsilon$ . Furthermore, the model assumes that  $\rho_t$  does not change over time; in other word, the ratio between the observed and true variances remains unchanged over the three periods.

These three assumptions allow one to estimate  $\rho^2$  as a function of the correlations among the three observed  $y_{i,t}$ . They also allow one to express the  $\beta_s$  in terms of  $\rho$  and the correlations of the  $y$  variables:

$$\begin{aligned} - \rho^2 &= \frac{r_{01}r_{12}}{r_{02}} \\ - \beta_{10} &= \frac{r_{01}}{\rho^2} \\ - \beta_{21} &= \frac{r_{12}}{\rho^2} \end{aligned}$$

In the first expression,  $\rho^2$  is the expression of true reliability coefficient; the meaning of true reliability coefficient is similar to that of the traditional reliability coefficient: it shows how true income determines observed income separately from measurement error. After obtaining the reliability coefficient, one can estimate the stability coefficients  $\beta_s$  which, then, can be used to calculate the true income mobility as  $(1-\beta)$  [7, 8]. As being indicated by the expression of  $\beta_s$  and  $\rho$ , true income mobility can be estimated in the absence of true income. Furthermore, this method allows us to estimate the income mobility separately from the impact of measurement error on the observed income. Although the reliability coefficient does not directly provide us any information, it implies the extent of the impact of measurement errors on the

observed data because it is the residual of the true income's impact.

## 2.2. Data Description

Similarly to many other mobility studies, in this paper, the observed income ( $y_{i,t}$ ) is represented by the observed expenditure, because standard economic theory measures household welfare as a function of consumption, not income [9]. Hence, the household expenditure will be used as the measure of the observed income ( $y_{i,t}$ ) in our model from this point. This study will apply the test-retest correlation method on the expenditure data from the VHLSS in 2004-2008 to estimate the extent of income mobility in Vietnam. The VHLSS is a longitudinal survey conducted every two years in Vietnam; the questionnaires cover many social/economic aspects of households and communes, such as expenditure, income, education, health, etc. and the sample has more than 30,000 individuals or 9,000 households each year for the household expenditure data.

One problem that many previous longitudinal studies using the household data have to face is the difficulty in matching the panel data for different years [9, 10, 11]. In our case, there are 9,188 households were interviewed in 2004, 9,189 households were interviewed in 2006, and 9,189 households were interviewed in 2008. In an ideal situation, one can expect half of the household from the first survey to be interviewed for the second survey, and half of that number to be interviewed in the third survey. Indeed, if the matching process happens perfectly, we may obtain roughly one fourth of the initial 9,188 households in the first survey in 2004. In reality, there are 3,830 households from VHLSS 2004 that get re-interviewed in 2006, and only 1,611 households get re-interviewed in 2008, i.e. roughly one fifth of the initial number of households in VHLSS 2004. This number is similar to the result of other matching process

completed by other researchers [10, 11]. This is caused by the high rate of migration of household members or the whole household to different areas; the migration makes the chance that a household is interviewed in all three periods lower than in the ideal situation [9]. Wrong identification number assignment during the re-interviewing process also causes some observations invalid. These observations are by comparing the sex and age variables of the same individuals across the three data sets; we are able to pick out more than 200 false observations from the final pool. This leaves us a sample size of 1,540 households in total. This number is not far off from what we expect in an ideal situation. Moreover, the smaller sample itself is quite similar to the original samples because only the standard deviation of the expenditure increases a little more. Therefore, the smaller sample size would have a minimal effect on the result.

### 3. RESULTS

#### 3.1. Mobility Estimates

As summarized in Table 1a, stability coefficients are estimated to be 0.965 in 2004-2006 and 0.908 in 2006-2008. A simple calculation gives us the values of economic mobility in two periods: 0.035 during 2004-2006 and 0.092 during 2006-2008. The true reliability coefficient's estimation is 0.894. In comparison with the correlation coefficient and the regression coefficient methods, the stability coefficients of the test-retest correlations are significantly higher which means true income is more stable than what the other two methods imply. In other word, the variance of true income is lower than the variance of observed income as shown in Table 1b. The major factor for this difference is that Heise's method separates the effect of measurement errors on the variance of observed data and estimates only the effect of true income instability, as indicated previously. As predicted by other mobility studies, the result shows that

the simple mobility estimation overestimates the true mobility due to measurement error; our result in Table 1b also indicates that measurement error's effect accounts for 68-85% of the simple mobility estimations by correlation coefficient and regression coefficient. Different approaches in estimating mobility can detect different ranges of the effect of errors on mobility estimates: 15-20% (Pseudo panel method), 33-42% (IV method), and 55-85% (transitory shock measurement errors) [4, 6, 9, 12]. Therefore, we can conclude that the test-retest approach is effective among the methodologies to reduce measurement error when estimating income mobility.

#### 3.2. Result Interpretation

In this section, we will explore the implications of our mobility estimates in Vietnam during 2004-2008. Income mobility is defined as the movement of individuals across income groups in the income distribution. Therefore, one may question the existence of a connection of income mobility and income inequality, because income inequality indicates an unequal distribution across different income groups. Indeed, if the poor is given a chance to move up from their original position in the distribution, or the rich moves down, the degree of inequality will go down. However, income mobility is the measurement of income movement of the economy as a whole: it does not indicate the movement of each income group over time, and thus, it is almost impossible to predict the movement of income inequality given only the estimated value of mobility.

However, one can still analyze the implication about inequality by combining observations about the change of individual income over time and mobility data. The first observation, as shown in Figure 1 is that the relationship between the initial income and the level of income change in one time period is

proportional: the higher the initial income, the higher the increase in income, and this observation holds for both periods 2004-2006 and 2006-2008. Moreover, the slopes of the two fitted lines are roughly the same. On the other hand, our low income mobility in Vietnam in these periods indicates that there is hardly any movement of the income groups in the distribution. These two observations point out that the chance for the poor to catch up with the rich is minimal: not only does their increase in income over time is less than that of the upper income groups, but they also hardly move from their original quintiles in the distribution. Therefore, we can conclude that the degree of short-term inequality will not decrease.

Nevertheless, the movement of income distribution and inequality is so complicated that a few measurements cannot tell the whole story. As a matter of fact, although the degree of short-term inequality may not go down, there is evidence that long-term inequality might be reduced. Plotting the initial income level against the percentage of income change as shown in Figure 2, one can realize that the lower income groups are better at exceeding their original income: the percentage change of their income is higher than that of the higher income groups. This means that although their actual amount of increased income is not as high as that of the rich, their growth rate of income is higher. It might be too soon to predict that higher income

growth for the poor would eventually lead to a decline in inequality in the long run but one can conclude that low income mobility does not exclude this possibility because it shows that the income level of the lower groups is rising faster than that of the higher groups in the distribution.

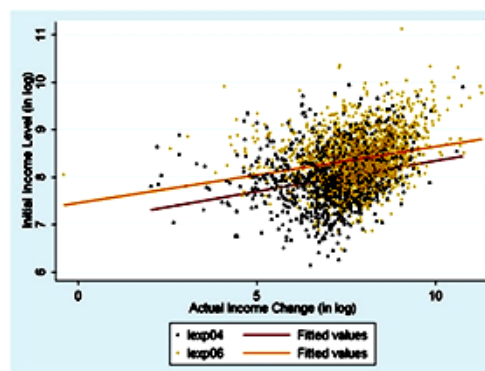


Figure 1. Actual Income Change against Initial Income

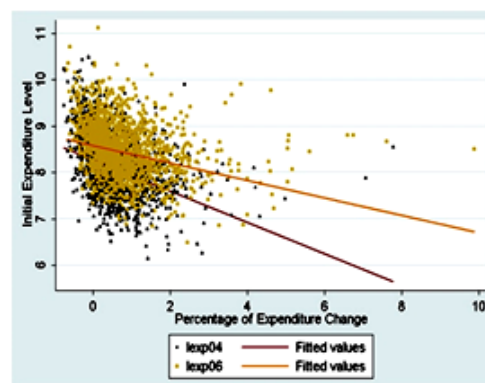


Figure 2. Income Growth Against Initial Income

Table 1a. Test-retest Result Summary

Variables	Mathematical expression	Estimation
True reliability coefficient	$\rho^2$	0.894
Stability coefficient 2004-2006	$\beta_{10}$	0.965
Stability coefficient 2006-2008	$\beta_{21}$	0.908
Mobility 2004-2006	$1-\beta_{10}$	0.035
Mobility 2006-2008	$1-\beta_{21}$	0.092

**Table 1b.** Mobility Estimates Comparison

Stability and Mobility, measured by	2004-2006		2006-2008	
	Stability	Mobility	Stability	Mobility
Correlation Coefficient (with measurement error)	0.755	0.245	0.709	0.291
Regression Coefficient (with measurement error)	0.794	0.206	0.706	0.294
Test-Retest Correlation (separating measurement error)	0.965	0.035	0.908	0.092

#### 4. CONCLUSION

Income mobility is a useful indicator for the movements of different income groups across the distribution. However, it is also subjected large measurement errors which cause the estimates to be biased upward. This paper uses a methodology developed by David Heise to estimate income mobility in Vietnam separately from the impact of measurement errors on the observed data. The Vietnam Household Living Standard Survey (VHLSS) expenditure data in 2004, 2006, and 2008 are used, and the estimated mobility values are 0.035 in 2006 and 0.092 in 2008. These figures imply a very low degree of mobility in Vietnam during these periods.

Although the number indicates that there is hardly any movement in the income distribution and it is observed that the increase in income of the lower income class is lower than that of the higher income class, one needs to be careful when exploring the implication of these observations about income inequality; in fact, we can only predict that short-term inequality may not decrease. The income growth of the lower income class, however, is evidently higher than that of the high income class, which indicates that the poor may still be able to catch up with the high income class in the long term.

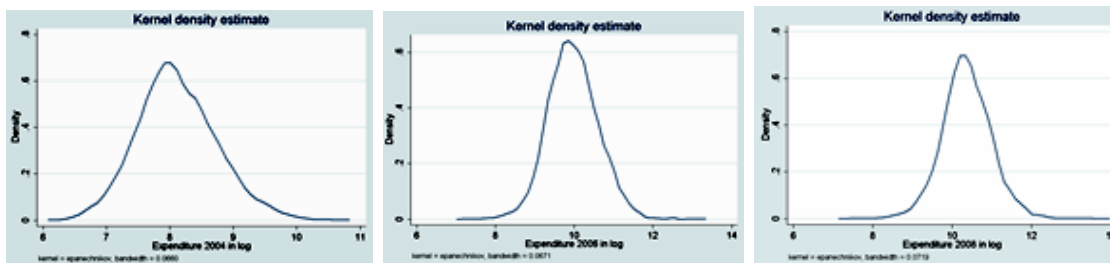
#### APPENDIX

**Table A1.** Observed Expenditure Statistics after Matching Process

Observed expenditure (in log)	Mean	Std. Dev.	Min	Max
$Y_{i, 2004}$	8.093343	0.624063	6.14283	10.48539
$Y_{i, 2006}$	8.442991	0.5701	6.485551	11.11226
$Y_{i, 2008}$	8.899413	0.575923	7.326853	11.48954

**Table A2.** Computed Correlations of Observed Expenditure ( $r_{12}$ ,  $r_{23}$ ,  $r_{13}$ )

Observed expenditure (in log)	$Y_i, 2004$	$Y_i, 2006$	$Y_i, 2008$
$Y_i, 2004$	1		
$Y_i, 2006$	0.758	1	
$Y_i, 2008$	0.6878	0.7131	1



**Figure A1.** Distributions of Observed Expenditure (in log) in 2004, 2006, and 2008

## Ước lượng sự biến đổi thu nhập của Việt Nam giai đoạn 2004-2008: phương pháp kiểm tra-kiểm tra lại

- Paul Glewwe
- Vũ Khoa

Trường Đại học Minnesota, USA

### TÓM TẮT:

Trong bài báo này, chúng tôi tập trung phân tích một khía cạnh quan trọng của nền kinh tế Việt Nam cũng như là một chủ đề được quan tâm trong ngành kinh tế phát triển: đó là tính biến đổi thu nhập của nền kinh tế Việt Nam trong giai đoạn đầu thế kỷ 21. Dựa trên phương pháp kiểm tra-kiểm tra lại và hệ số ổn định và hệ số tin cậy được David Heise phát triển từ năm 1969, chúng tôi xây dựng một mô hình ước lượng tính

biến đổi thu nhập không chịu tác động của sai số đo lường. Ước lượng tính biến đổi thu nhập dựa trên bộ Khảo sát mức sống dân cư ở Việt Nam (VHLSS) trong giai đoạn từ năm 2004 đến năm 2008, mô hình này phát hiện và loại trừ tác động của sai số đo lường trong dữ liệu thu nhập, yếu tố ảnh hưởng đến hơn 80% kết quả mà các phương pháp ước lượng phổ thông mang lại.

**Từ khóa:** Sự biến đổi thu nhập, sai số đo lường trong dữ liệu thu nhập, VHLSS, phương pháp hệ số ổn định và tin cậy

## REFERENCES

- [1]. Atkinson, A. B., Bourguignon, F., and Morrisson, C. , Empirical Studies of Earnings Mobility, *S.I. : Harwood Academic Publishers*, (1992).
- [2]. Jarvis, S. and Jenkins, How Much Income Mobility is There in Britain, *The Economic Journal*, **108**(447), 428-443, (1998).
- [3]. Glewwe. P and Phong Nguyen, *Economic Mobility in Vietnam in the 1990s*, Policy Research Working Paper N° WPS2838, The World Bank, Washington D.C, (2002).
- [4]. Francisca Antman and David McKenzie, Poverty Traps and Nonlinear Income Dynamics with Measurement Error and Individual heterogeneity, *The Journal of Development Studies*, **43**(6), 1057-1083, (2007).
- [5]. Edited by W. S. Fields and G. Pfeffermann, Pathways out of poverty: Private firms and economic mobility in developing countries, *Part I.3: Long-Term Economic Mobility and the Private Sector in Developing Countries: New Evidence*, Kluwer Academic Publishers, (2003).
- [6]. Luttmer, E. F. P, *Measuring Economic Mobility and Inequality: Disentangling Real Events from Noisy Data*, Policy Research Working Paper N° 2549, The World Bank, Washington D.C, (2002).
- [7]. Heise. David R, Separating Reliability and Stability in Test-Retest Correlation, *American Sociological Review*, **34**(1), 93-101, (1969).
- [8]. Michael, Houts and Hastings Orestes, *Reliability and Stability Estimates for the GSS Core Items from the Three-wave Panels, 2006-2010*, GSS Methodological Report # 119, National Opinion Research Center, (2012).
- [9]. Glewwe. P, How Much of Observed Economic Mobility is Measurement Error? A Method to Reduce Measurement Error Bias, with an Application to Vietnam, *World Bank Economic Review*, **26**(2), 236-264 (2012).
- [10]. McCaig, Brian, Summary of 2002-2004 VHLSS Household Panel Construction, Centre for Economic Policy Research-The Australian National University, Discussion paper N° 622, (2009).
- [11]. Le D. Tung and T. Hung. Pham, *Construction of Panel Data for the Vietnam Household Living Standard Survey 2002-2006*, mimeo, Indochina Research and Consulting, Ha Noi, (2009).
- [12]. Glewwe. P. and Dang. H. A. H, Was Vietnam's Economic Growth in the 1990s Pro-Poor? An Analysis of Panel Data from Vietnam, *Economic Development and Cultural Change*, **59** (3), 584-608, (2011).