Was Vietnam's Economic Growth in the 1990's Pro-Poor?
An Analysis of Panel Data from Vietnam

Paul Glewwe

Hai-Anh Hoang Dang

Department of Applied Economics
University of Minnesota

July 2005

Abstract

International aid agencies and almost all economists agree that economic growth is necessary for reducing poverty, yet some economists question whether it is sufficient for poverty reduction. Vietnam enjoyed rapid economic growth in the 1990s, but a modest increase in inequality during that decade raises the possibility that the poor in Vietnam benefited little from that growth. This paper examines the extent to which Vietnam's economic growth has been "pro-poor", with particular attention to two issues. The first is the appropriate comparison group. When comparing the poorest x% of the population at two points in time, should the poorest x% in the first time period be compared to the poorest x% in the second time period (some of whom were not the poorest x% in the first time period) or to the same people in the second time period (some of whom are no longer among the poorest x%)? The second is measurement error. Estimates of growth among the poorest x% of the population are likely to be biased if income or expenditure is measured with error. Household survey data show that Vietnam's growth has been relatively equally shared across poor and non-poor groups. Indeed, comparisons of the same people over time indicate that per capita expenditures of the poor increased much more rapidly than those of the non-poor, although failure to correct for measurement error exaggerates this result.

Keywords: Economic growth, poverty, inequality, measurement error, Vietnam.

I. Introduction

The World Bank, the United Nations and virtually all economists agree that economic growth is essential to reduce poverty in developing countries (World Bank, 2001a; UNDP, 2001). Yet economic growth reduces poverty only of it is broad-based; economic growth that does not change the incomes of the poor cannot reduce poverty. More generally, increased inequality implies that the incomes of the poor grow at a slower rate than the incomes of the rest of the population. Some economists claim that economic growth sometimes leaves the poor behind (see the discussion in Winters, McCulloch and McKay, 2004). Such claims can be resolved only by empirical analysis. This paper investigates this claim for the case of Vietnam

Vietnam's rate of economic growth in the 1990's was one of the highest in the world. The average annual rate of economic growth from 1990 to 2000 was 7.9% (World Bank, 2002). This high rate of economic growth led to a sharp reduction in poverty; estimates based on the 1992-93 Vietnam Living Standards Survey (VLSS) show a poverty rate of 58%, while five years later estimates based on the 1997-98 VLSS indicate that poverty had dropped to 37% (World Bank, 1999). Such a sharp drop in poverty in only five years is an achievement that is rarely seen in any developing country, and Vietnam's continued economic growth since 2000 suggests that the poverty rate will continue to decline in the first decade of the 21st century.

One worrying aspect of Vietnam's recent growth is that it was accompanied by increased inequality. The Gini coefficient for household per capita expenditures rose from 0.33 in 1992-93 to 0.35 in 1997-98. This increase may seem small, but it suggests that the poor in Vietnam benefited less from economic growth than did other households.

This paper uses panel data from the 1992-93 and 1997-98 Vietnam Living Standards Surveys (VLSS) to investigate whether Vietnam's growth has been "pro-poor". It divides the population into five groups – the poorest 20%, the next poorest 20%, etc., up to the wealthiest 20% – and estimates growth rates for each group. A key conceptual issue is whether one wants to compare households who constituted the poorest 20% of the population in 1992-93 with *the poorest 20% of the population* in 1997-98, some of whom may not have been among the poorest 20% in 1992-93, or whether one wants to compare the poorest 20% in 1992-93 with *the same households* in 1997-98, some of whom are no longer among the poorest 20% of the population. A second important issue is more methodological in nature; household expenditures are likely to be measured with error, which can lead to serious biases, especially for the second type of comparison.

The rest of this paper is organized as follows. Section II discusses the data, and Section III explains the methodology. The next section presents estimates of changes in per capita expenditures among different groups of Vietnamese households, using both types of comparisons and presenting estimates that ignore measurement error and estimates that correct for measurement error. A final section summarizes the findings.

II. Data

This paper uses data from the 1992-93 and 1997-98 Vietnam Living Standards Surveys (VLSS), which were conducted by Vietnam's General Statistical Office (GSO). The 1992-93 VLSS covered 4800 households, while the 1997-98 VLSS surveyed 6002 households. Both surveys are nationally representative, and about 4300 households participated in both surveys, which provides a large, nationally representative panel data set. The VLSS household questionnaire covered a wide variety of topics, including education, health, employment, migration, housing, fertility, agricultural activities, small household businesses, income and expenditures, and credit and savings.

Analysis of the welfare of households or individuals in any country requires data on an indicator of household welfare. Income is the most obvious indicator, but most economists would agree that household consumption expenditure is a better indicator because data on expenditures are likely to be more accurate and because consumption expenditures have a stronger link with households' current levels of welfare (see Deaton, 1997). This paper makes extensive use of a variable that measures households' total consumption expenditures that was constructed from the VLSS data by a team of GSO and World Bank staff. The variable includes explicit expenditures on food and non-food items, the value of food produced and consumed at home, and the estimated rental values of durable goods and the household's dwelling. More detailed information on the surveys and the constructed variables are given in World Bank (2001).

The focus of this paper is on the economic growth of the poor, relative to the economic growth of non-poor households. More specifically, for both years the VLSS sample is divided into five groups of equal size (quintiles), ranked by per capita household expenditures. The poorest 20% is referred to as quintile 1, the second poorest 20% as quintile 2, and so forth, up to quintile 5, the wealthiest 20%. As mentioned above, a poverty line defined for use with the VLSS data indicates that 58% of the population was poor in 1992-93, so that quintiles 1, 2 and 3 constituted "the poor" in that year. (This poverty line was set in terms of minimal food requirements plus an allowance for basic non-food items; see World Bank, 1999, for details.) Applying the same poverty line to the 1997-98 VLSS produced a poverty rate of 37%, so that only quintiles 1 and 2 were "the poor" in that year. This paper does not classify particular quintiles as poor or non-poor; for the purpose at hand it is sufficient to examine each quintile over time to see whether expenditure growth is more rapid among the lower or upper quintiles.

III. Methodology

This section presents the methodology used in this paper to investigate the extent to which economic growth in Vietnam in the 1990s was "pro-poor". It begins with a discussion of the appropriate comparisons to make and then addresses in detail the problem of – and a proposed solution to – bias caused by measurement error.

A. The Appropriate Comparison Group. When examining economic growth within quintiles over time, a key issue arises due to the fact that it is likely that some of the households in a given quintile in one year will be in a different quintile in a later year because households' expenditure levels increase (or decrease) at different rates. Thus, for example, some households in quintile 1 in the first year will have moved to quintile 2 (or perhaps an even higher quintile) by the second year, and vice versa.

This movement, known as (relative) economic mobility, implies that there are two distinct ways to measure economic growth for a given quintile. First, one could compare the mean of per capita expenditures of a quintile in the first year with the same mean for the households who are in that quintile in the second year. For example, one can compare the poorest 20% of the population in Vietnam in 1992-93 to the poorest 20% in 1997-98. Second, if one has panel data, for a given quintile one could compare the mean of per capita expenditures in the first year with the mean *for the same households* in the second year, regardless of the quintile they belong to in the second year. As long as there is some economic mobility, the second method will always show higher growth among quintile 1 (the poorest 20% of the population) than the first because the calculation of mean household expenditures under the second method includes some households in higher quintiles in the second year (households whose expenditures in the second year increased enough to place them in the second or a higher quintile) and excludes some households in the poorest quintile in the second year (households in quintile 2 or higher in the first year whose expenditures fell sufficiently to put them in quintile 1 in the second year). More generally, this difference between the two approaches applies to any calculation of expenditure growth for the poorest 15%, 30%, or any other percent, of the population. Similarly, when evaluating the growth rate for the wealthiest 20% (quintile 5), or some other percentage of the wealthiest households, the second method will always produce a growth rate that is lower than that provided by the first method (as long as some households in quintile 5 in the first year are in a lower quintile in the second year).

Both approaches to examining the growth rates of poor and non-poor households provide useful information, and both will be presented in the empirical work in Section IV. The first shows the nature of changes in inequality over time. The second indicates the extent to which there is (relative) mobility; greater mobility implies lower long-run inequality for a given level of inequality at each point in time (see Glewwe, 2005a).

B. A Method to Correct for Measurement Error. The bias caused by measurement error in the expenditure variable when calculating growth rates by expenditure quintiles is relatively small when comparing the poorest x% of the population in one year with the poorest x% of the population in another year (that is, when *not* comparing the same people over time). Indeed, if expenditure follows a lognormal distribution and both inequality and the extent of measurement error are unchanged, there is no bias (see Glewwe, 2005b, for a detailed examination of such comparisons). In Vietnam, where inequality has increased modestly and measurement error is unlikely to change (because the expenditure sections of the two VLSS questionnaires are almost identical), it is likely that the increased inequality leads to a slight overestimate of the expenditure growth of the poor, but the bias is likely to be small. In contrast, measurement error can lead to sizeable biases when calculating growth rates by expenditure quintiles that compare *the same people* over the two years.

Correcting for measurement error is difficult because, by definition, households' true expenditure levels are not observed in either time period. That is, one would like to observe the joint distribution of true values for both years, but one has only the joint distribution of the observed values. The goal is to draw inferences about the joint density of the true values and, ultimately, to simulate that joint distribution. To do so requires some assumptions; the approach taken here will now be explained in detail.

Assume that the relationship between the true values of per capita expenditures in 1992-93, denoted by x*, and the observed values in that year, denoted by x, is determined by a multiplicative random measurement error, ε_x, with $\varepsilon_x > 0$:

$$x = x^*\varepsilon_x, \quad \text{which implies} \quad \ln(x) = \ln(x^*) + \ln(\varepsilon_x) \qquad (1)$$

Assume that $\ln(\varepsilon_x)$ is symmetrically distributed and that $E[\ln(\varepsilon_x)] = 0$. This implies that the medians of $\ln(\varepsilon_x)$ and ε_x are 0 and 1, respectively. Multiplicative measurement error is more reasonable than an additive measurement error, for two reasons: (i) additive errors imply that the dispersion of the errors is unrelated to household expenditures (assuming that the variance of those errors is constant), while it seems more likely that the variance of the errors would be proportional to household expenditure levels, which is the case with multiplicative errors (with constant variance); and (ii) large negative additive errors can result in negative values of observed expenditures, which is nonsensical for consumption expenditures, while the assumption of multiplicative errors rules out this possibility (given the above assumption that $\varepsilon_x > 0$).

Consider next observed per capita expenditures in 1997-98, denoted by y. Assume that an analogous relationship holds between y and its true value, denoted by y*:

$$y = y^*\varepsilon_y, \quad \text{which implies} \quad \ln(y) = \ln(y^*) + \ln(\varepsilon_y) \qquad (2)$$

where $\varepsilon_y > 0$. Assume also that $\ln(\varepsilon_y)$ is symmetrically distributed, with $E[\ln(\varepsilon_y)] = 0$. Lastly, assume that both ε_x and ε_y are uncorrelated with x* and y* and uncorrelated with each other; the implications of relaxing these two assumptions are discussed below.

A key question for equations (1) and (2) is how much of the variance of the observed values, ln(x) and ln(y), is due to variation in the true values and how much is due to variance in their measurement errors. To see how to estimate this decomposition of the variance of x, consider the following equation:

$$\ln(y^*) = \alpha_2^* + \beta_2^*\ln(x^*) + u_2, \qquad (3)$$

where the asterisks on α_2* and β_2* indicate that this relationship is between the true values, x* and y*, and the subscripts indicate that y* is expenditures in the second time period. The residual u_2 has a mean of zero and is, *by definition*, uncorrelated with ln(x*). If ln(x*) were observed one could estimate β_2* using OLS (one need not observe ln(y*) since replacing it with ln(y) in (3) does not cause bias in estimates of β_2*). Equation (3) implies that β_2* equals the covariance of ln(x*) and ln(y*) over the variance of ln(y*):[1]

$$\beta_2^* = \mathrm{Cov}[\ln(x^*),\ln(y^*)]/\mathrm{Var}[\ln(x^*)] \qquad (4)$$

Instrumental variables methods can be used to estimate β_2*. This requires a variable that is correlated with ln(x*) but uncorrelated with the associated measurement error $\ln(\varepsilon_x)$ and uncorrelated with u_2 (uncorrelated with ln(y*) after conditioning on ln(x*)).

If both ln(y*) and ln(x*) in equation (3) were replaced with ln(y) and ln(x), the coefficient on ln(x), call it β_2, would by the same reasoning have a similar property:

$$\beta_2 = \mathrm{Cov}[\ln(x),\ln(y)]/\mathrm{Var}[\ln(x)] = \mathrm{Cov}[\ln(x^*),\ln(y^*)]/\mathrm{Var}[\ln(x)] \qquad (5)$$

where the last expression uses the fact that the covariance between two variables is unaffected by adding (uncorrelated) random measurement errors to each variable. Since ln(x) and ln(y) are observed, OLS provides a consistent estimate of β_2. Combining the estimates of β_2* and β_2 in (4) and (5) yields an estimate of the variance of ln(x*):

[1] To see this, use the standard formula for the covariance of two variables where one or both are weighted sums of 2 or more other variables: $\mathrm{Cov}[\ln(y^*), \ln(x^*)] = \mathrm{Cov}[\beta_2\ln(x^*) + u_2, \ln(x^*)] = \mathrm{Cov}[\beta_2\ln(x^*), \ln(x^*)] + \mathrm{Cov}[u_2, \ln(x^*)] = \beta_2\mathrm{Var}[\ln(x^*)]$, which implies that $\beta_2 = \mathrm{Cov}[\ln(y^*), \ln(x^*)]/\mathrm{Var}[\ln(x^*)]$.

$$\beta_2^*/\beta_2 = \text{Var}[\ln(x)]/\text{Var}[\ln(x^*)], \text{ which implies } \text{Var}[\ln(x^*)] = (\beta_2/\beta_2^*)\text{Var}[\ln(x)] \quad (6)$$

The variance of the measurement error, ε_x, can be obtained by taking the variance of both sides of equation (1):

$$\text{Var}[\ln(\varepsilon_x)] = \text{Var}[\ln(x)] - \text{Var}[\ln(x^*)] = \text{Var}[\ln(x)](1 - (\beta_2/\beta_2^*)) \quad (7)$$

Thus if suitable instrumental variables can be found, one can decompose the variance of ln(x) into the variance of ln(x*) and the variance of $\ln(\varepsilon_x)$. Also, equation (3) implies that:

$$\text{Var}(u_2) = \text{Var}[\ln(y^*)] - (\beta_2^*)^2\text{Var}[\ln(x^*)] = \text{Var}[\ln(y^*)] - \beta_2^*\beta_2\text{Var}[\ln(x)] \quad (8)$$

With a credible instrumental variable to estimate of β_2^*, and a simple OLS estimate of β_2, $\text{Var}[\ln(u_2)]$ can be obtained from (8) if one has an estimate of Var[ln(y*)].

A relatively simple route to obtain an estimate of Var[ln(y*)] is to assume that the (proportional) contribution of measurement error to the variance of ln(x) is the same as the (proportional) contribution of measurement error to the variance of ln(y). That is:

$$\frac{Var[\ln(y)]}{Var[\ln(y^*)]} = \frac{Var[\ln(x)]}{Var[\ln(x^*)]} = \beta_2^*/\beta_2, \text{ which implies } \text{Var}[\ln(y^*)] = (\beta_2/\beta_2^*)\text{Var}[\ln(y)] \quad (9)$$

The assumption that the proportional contributions of measurement error to total variance are the same for both x* and y* is plausible, but it can be avoided by estimating a relationship analogous to that of equation (3), where the roles of x* and y* are reversed. More specifically, consider the following equation:

$$\ln(x^*) = \alpha_1^* + \beta_1^*\ln(y^*) + u_1, \qquad (10)$$

As before, the asterisks on α_1^* and β_1^* indicate that this relationship is between the unobserved true values, x* and y*, the subscripts indicate that x* is expenditures in the first time period, and u_1 is a mean zero random error that is by definition uncorrelated with ln(y*).[2] Similar to equations (4) and (5), one can show:

$$\beta_1^* = \mathrm{Cov}[\ln(x^*),\ln(y^*)]/\mathrm{Var}[\ln(y^*)] \qquad (11)$$

$$\beta_1 = \mathrm{Cov}[\ln(x),\ln(y)]/\mathrm{Var}[\ln(y)] = \mathrm{Cov}[\ln(x^*),\ln(y^*)]/\mathrm{Var}[\ln(y)] \qquad (12)$$

where β_1 is the "slope coefficient" from an OLS regression of ln(x) on ln(y). Unbiased estimates of β_1^* can be obtained if one or more instrumental variables can be found that are correlated with ln(y*) but not correlated with its measurement error, $\ln(\varepsilon_y)$, and not correlated with u_1 (not correlated with ln(x*) after conditioning on ln(y*)). Finally, estimates of β_1^* and β_1 can be combined to estimate the variance of ln(y*):

$$\beta_1^*/\beta_1 = \mathrm{Var}[\ln(y)]/\mathrm{Var}[\ln(y^*)], \text{ which implies } \mathrm{Var}[\ln(y^*)] = (\beta_1/\beta_1^*)\mathrm{Var}[\ln(y)] \quad (13)$$

[2] The relationship in (10) is *not* the same as one that could be obtained algebraically from (3). That is, (3) can be manipulated to obtain $\ln(x^*) = -(\alpha_2^*/\beta_2^*) + (1/\beta_2^*)\ln(y^*) - u_2/\beta_2^*$. Yet in this relationship the residual term u_2/β_2^* is correlated with the regressor ln(y*), as is obvious from inspection of (3), so this is not the same relationship as that given in (10), where u_1 is by definition uncorrelated with ln(y*).

With this estimate of Var[ln(y*)], or the one obtained using equation (9), the variance of u_2 can be obtained using equation (8).

One can also obtain an estimate of the variance of the measurement error $\ln(\varepsilon_y)$:

$$Var[\ln(\varepsilon_y)] = Var[\ln(y) - Var[\ln(y^*)] = Var[\ln(y)](1 - (\beta_1/\beta_1^*)) \quad (14)$$

In addition, taking the variance of both sides of (10), one can estimate $Var(u_1)$:

$$Var(u_1) = Var[\ln(x^*)] - (\beta_1^*)^2 Var[\ln(y^*)] = Var[\ln(x^*)] - \beta_1^*\beta_1 Var[\ln(y)] \quad (15)$$

Finally, an alternative to finding an instrumental variable for ln(y*) to estimate β_1^* is to invoke the "proportional measurement error" assumption in equation (9) to estimate β_1^*:

$$\beta_1^*/\beta_1 = \frac{Var[\ln(y)]}{Var[\ln(y^*)]} = \frac{Var[\ln(x)]}{Var[\ln(x^*)]} = \beta_2^*/\beta_2, \text{ which implies } \beta_1^* = \beta_1(\beta_2^*/\beta_2) \quad (16)$$

Next, consider how to estimate growth in per capita consumption expenditures for each quintile, corrected for measurement error bias. Estimates of α_2^*, β_2^*, Var[ln(x*)] and $Var[u_2]$ can be used, via equation (3), to simulate the joint distribution of ln(x*) and ln(y*) by making four more assumptions. Two key assumptions are that $E[u_2| x^*] = 0$ and that $E[u_2^2| x^*]$ is a constant. The first assumption implies that E[ln(y*)| ln(x*)] is linear in x*, an assumption that can be tested. The other assumptions needed for simulations concern the distributions of (x^*) and u_2. Two convenient assumptions are that both x^* and y^* follow a lognormal distribution, which implies that $\ln(x^*)$ and $\ln(y^*)$ are normally distributed, and that u_2 also follows a normal distribution.

The reasonableness of these normality assumptions can be checked by plotting the distributions of $\ln(x)$ and $\ln(y)$. If the measurement errors $\ln(\varepsilon_x)$ and $\ln(\varepsilon_y)$ are also normally distributed, then $\ln(x)$ and $\ln(y)$ are also normally distributed. This is examined in Figure 1, which compares density estimates of the distributions of $\ln(x)$ and $\ln(y)$ to normal distributions with the same mean and variance, using two different bandwidths. Although the fit is not perfect, and standard specification tests (e.g. the Shapiro-Wilks W test) decisively reject the normality assumption (which is not surprising given the large sample size), that assumption still gives a reasonably close fit for both $\ln(x)$ and $\ln(y)$, which implies that these two normality assumptions will not seriously distort the simulations. Indeed, this claim is further supported below by comparisons of growth rates (by quintile) for observed x and y with the same growth rates calculated from simulations of (observed) x and y that are based on these normality assumptions.

The assumption that $\ln(x^*)$ and $\ln(y^*)$ are related to each other in the simple linear relationships shown in equations (3) and (10) must also be checked. A more general relationship can be expressed by adding $\ln(x^*)^2$ as a regressor in equation (3) and $\ln(y^*)^2$ as a regressor in equation (10). Glewwe (2005a) shows that the assumption that these quadratic terms equal zero can be tested using the observed values, $\ln(x)$ and $\ln(y)$, if $\ln(x^*)$ and $\ln(y^*)$ and their associated measurement errors are symmetric. That is, if one estimates equations (3) and (10) by replacing $\ln(x^*)$ and $\ln(y^*)$ with their observed values and adding quadratic terms (of the observed values) as regressors, insignificant quadratic terms imply that equations (3) and (10) are linear (as opposed to quadratic). The underlying symmetry assumptions can also be checked in Figure 1; some skewness is evident but probably not enough to invalidate this method of checking the linearity of equations (3) and (10). For equation (3), a linear regression based on observed values has an R^2 of 0.493; adding a squared term barely changes the R^2 (0.495), although the associated coefficient is statistically significant at the 1% level (t-statistic of 2.78). Equation (10) shows even less departure from linearity; adding the squared term raises the R^2 by only 0.0007 and the t-statistic on the squared coefficient is only 1.28. Thus the linearity assumption is not strongly rejected by the data, so imposing it is unlikely to influence the simulation results significantly.

Lastly, consider the implications of relaxing two assumptions on the measurement errors. First, suppose that $\ln(\varepsilon_x)$ is correlated with $\ln(x^*)$ and $\ln(\varepsilon_y)$ is correlated with $\ln(y^*)$, perhaps because wealthier households are more likely to underreport expenditure. If this correlation is linear, so that $\ln(\varepsilon_x) = \mu_1 + \lambda_1\ln(x^*) + \ln(\varepsilon_x')$, where $\ln(\varepsilon_x')$ is uncorrelated with $\ln(x^*)$, and $\ln(\varepsilon_y) = \mu_2 + \lambda_2\ln(y^*) + \ln(\varepsilon_y')$, where $\ln(\varepsilon_y')$ is uncorrelated with $\ln(y^*)$, there is no cause for concern. To see why, substitute these expressions into (1) and (2) to derive $\ln(x) = \mu_1 + (1+\lambda_1)\ln(x^*) + \ln(\varepsilon_x')$ and $\ln(y) = \mu_2 + (1+\lambda_2)\ln(y^*) + \ln(\varepsilon_y')$. Inserting these expressions into (3) and (10) yields, after rearranging terms:

$$\ln(y) = \alpha_2^*(1+\lambda_2) + \mu_2 - \mu_1\beta_2^*\left(\frac{1+\lambda_2}{1+\lambda_1}\right) + \beta_2^*\left(\frac{1+\lambda_2}{1+\lambda_1}\right)\ln(x) + (1+\lambda_2)u_2 + \ln(\varepsilon_y') - \beta_2^*\left(\frac{1+\lambda_2}{1+\lambda_1}\right)\ln(\varepsilon_x') \quad (3')$$

$$\ln(x) = \alpha_1^*(1+\lambda_1) + \mu_1 - \mu_2\beta_1^*\left(\frac{1+\lambda_1}{1+\lambda_2}\right) + \beta_1^*\left(\frac{1+\lambda_1}{1+\lambda_2}\right)\ln(y) + (1+\lambda_1)u_1 + \ln(\varepsilon_x') - \beta_1^*\left(\frac{1+\lambda_1}{1+\lambda_2}\right)\ln(\varepsilon_y') \quad (10')$$

In equation (3′) using instrumental variables to overcome the problem that $\ln(\varepsilon_x')$ is positively correlated with ln(x) produces an estimate of $\beta_2^*\left(\frac{1+\lambda_2}{1+\lambda_1}\right)$, not β_2^*, and IV estimates of equation (10′) produce estimates of $\beta_1^*\left(\frac{1+\lambda_1}{1+\lambda_2}\right)$, not β_1^*. Clearly, the IV estimates are consistent if $\lambda_1 = \lambda_2$, that is if the degree of (linear) correlation between the measurement errors and the expenditure variable is the same for both years, i.e. $\text{Cov}[\ln(x^*), \ln(\varepsilon_x)]/\text{Var}[\ln(x^*)] = \text{Cov}[\ln(y^*), \ln(\varepsilon_y)]/\text{Var}[\ln(y^*)]$. This assumption is intuitively plausible and is imposed throughout this paper to extend the methodology to the case where measurement errors are (linearly) correlated with expenditures.[3]

Second, suppose that the measurement errors ε_x and ε_y are correlated. Will this lead to biased estimates of β_1^* and β_2^*? Glewwe (2005a) shows that this depends on the instrumental variables used. If one uses "second measurements" of x* and y* in each period as instrumental variables one will almost certainly underestimate β_1^* and β_2^* if the measurement errors in x* and y* are positively correlated. Yet if the instruments used are variables that are caused by x* and y*, in the sense that two variables, z_1 and z_2, are generated by the equations $z_1 = \kappa_1 + \pi_1\ln(x^*) + w_1$ and $z_2 = \kappa_2 + \pi_2\ln(y^*) + w_2$, and w_1 is independent of ln(x*) and w_2 is independent of ln(y*), then one can use z_1 and z_2 as instruments to obtain unbiased estimates of β_1^* and β_2^*. Yet a final problem remains: now $\text{Cov}[\ln(x), \ln(y)] \neq \text{Cov}[\ln(x^*), \ln(y^*)]$ since $\text{Cov}[\ln(x, \ln(y)] = \text{Cov}[\ln(x^*), \ln(y^*)]$

[3]This assumption is similar, but not identical, to the assumption of proportional measurement error shown in equation (9). Neither assumption implies the other, but if the additional assumption is made that $\text{Var}[\ln(\varepsilon_y')]/\text{Var}[\ln(y^*)] = \text{Var}[\ln(\varepsilon_x')]/\text{Var}[\ln(x^*)]$ then each assumption implies the other.

+ Cov[ln(ε_x), ln(ε_y)]. Thus the second equalities in equations (5) and (12) no longer hold; the expressions in equations (6) and (13) overestimate Var[ln(x*)] and Var[ln(y*)], respectively. This implies that equations (7) and (14) underestimate Var[ln(ε_x)] and Var[ln(ε_y)], respectively. Of course, another set of upper bounds on Var[ln(x*)] and Var[ln(y*)] are Var[ln(x)] and Var[ln(y)], but the expressions for Var[ln(x*)] and Var[ln(y*)] in equations (6) and (13) may well be lower than those given by Var[ln(x)] and Var[ln(y)], and if so they can be used to reduce bias due to measurement error.

In summary, the only possible problem with relaxing the assumptions about the measurement errors is the complications that arise when the measurement errors of x* and y* are correlated with each other. Unfortunately, such correlation is likely; Pischke (1995) has shown this for earnings data in the U.S., and intuitively some reasons for over- or underestimation (e.g. some people have worse memories than others) may persist over time. Since this leads to overestimation of Var[ln(x*)] and Var[ln(y*)], one should make two or more assumptions about Var[ln(x*)] and Var[ln(y*)].[4] Simulations over a range of assumptions can be used to check for the robustness of key simulation results.

IV. Empirical Results

This section applies the methodology presented in Section III to the household survey data from Vietnam. It first shows estimates based on observed data, which are likely to yield biased results for comparisons of the same households over time. It then presents results that correct for measurement error.

---

[4]This amounts to making an assumption about how Cov[ln(x), ln(y)] decomposes into Cov[ln(x*), ln(y*)] and Cov[ln(ε_x), ln(ε_y)]. Note that β_1* = Cov[ln(x*), ln(y*)]/Var[ln(y*)]. Once an assumption is made about Var[ln(y*)] one can solve for Cov[ln(x*), ln(y*)] as β_1*Var[ln(y*)]. This in turn determines the decomposition of Cov[ln(x), ln(y)].

A. Estimates that Are Not Corrected for Measurement Error. Table 1 presents estimates of growth in consumption expenditures per capita from 1992-93 to 1997-98, using the first method of comparison (defining quintiles in 1997-98 in terms of expenditures in 1997-98, not expenditures in 1992-93). All figures are expressed in real terms using January 1998 prices and are calculated at the household level instead of at the individual level (that is, the unit of observation is the household). Calculations at the household and individual levels are almost identical, as will be seen below, and focusing on the household level simplifies the simulations that correct for measurement error. Using the 4799 households in the 1992-93 survey and the 5999 households in the 1997-98 survey with complete expenditure data, mean per capita expenditures among Vietnamese households increased by 48.0% in real terms over five years (from 2,019,000 Dong to 2,998,000 Dong), which implies an annual growth rate of 8.2%. This is slightly higher than the 6.9% growth rate in GDP per capita found in National Accounts data.

The next two lines in Table 1 provide separate estimates for urban and rural areas. Urban households, which account for about one fifth of the population, enjoyed higher economic growth than did rural households; the annual growth for urban households was 9.2%, compared to 6.3% for rural households. Table 1 also shows differences across Vietnam's seven regions. The fastest growing region is the South East, where Ho Chi Minh City (Saigon) is located, which had an annual growth rate of 12.7%. The next fastest are the Red River Delta, which includes Hanoi, and the North Central Coast; their growth rates were 9.5% and 9.6%, respectively. The Mekong Delta had the slowest growth, 4.1%, and the next two slowest regions were the Central Highlands and the South Central Coast, with annual growth rates of slightly more than 5%.

The most important figures in Table 1 are the growth rates for each of the five quintiles. Comparing the poorest 20% of the population in each of the two years (which are not necessarily the same households), expenditure levels increased by 6.5% per year. While this rate of growth is quite high, it is the lowest of the five quintiles. Quintiles 2, 3 and 4 each enjoyed growth rates of about 7% per year, and the mean expenditure level of the wealthiest quintile increased by 8.5% per year. This confirms the finding that inequality in Vietnam is increased in the 1990s.

The last column in Table 1 provides growth rates from the VLSS data that give equal weight to individuals, instead of equal weight to households. The patterns are very similar, especially the growth rates by quintiles. Henceforth, all figures, including those generated from simulations, will use the household as the unit of observation.

Table 2 presents figures when the second method of comparison is used, that is when panel data are used to ensure that the same households are compared over time, based on their characteristics in 1992-93. The panel data set was constructed as follows. Of the 4800 households interviewed in 1992-93, 96 were deliberately excluded from the sample in 1997-98 (because they were from a region that was intentionally slightly "undersampled" in the 1997-98 survey), so under ideal circumstances the panel data set should consist of 4704 households. Yet 399 (8.5%) of these households were not interviewed five years later; most had moved out of their communities during the intervening five years and no attempt was made to follow them. Of the remaining 4305, 21 had experienced so much turnover in membership that the head in 1992-93 was not a member in 1997-98 *and* the head in 1997-98 was not a member in 1992-93. These were deemed to have changed too much to be considered the same household, leaving 4284 households – 91.1% of the original 4704 – for analysis (of which 4281 have complete expenditure data in both years). Thus households are considered to be the same if the head in one of the two years was a member of the household, although not necessarily the head, in the other year. Analyses using a stricter definition of a panel household, that 50% or more of the members over the two years are members in both years, give similar results but are not reported here; they are available from the authors.

Turning to the panel data results in Table 2, for Vietnam as a whole the annual growth rate in per capita expenditures was 7.5%, which is slightly lower than the rate of 8.2% in Table 1. This difference suggests that the 423 households that were dropped from the original 4704 targeted to be interviewed in 1997-98 had higher growth than the 4281 with complete interviews. This is not surprising because re-interview rates were higher in rural areas (93%) than in urban areas (84%), and rural areas had lower growth than urban areas. Examining urban-rural differences, the growth rate in urban areas, 8.5%, is somewhat lower than that in Table 1 (9.2%), yet in rural areas the growth rate is somewhat higher, 7.0% compared to 6.3% in Table 1. Thus it appears that households in rural areas who enjoyed better than average growth in expenditures were slightly more likely to be reinterviewed in 1997-98, while the opposite is true for urban households.

The regional growth rates in Table 2 have a pattern similar to that in Table 1. The Southeast, the North Central Coast and the Red River Delta are still the first, second and third fastest growing regions, respectively, and the Mekong Delta is still the slowest. More generally, for five of the seven regions the annual growth rates in the two tables differ by about 1.5 percentage points or less. However, the growth rate in the Central Highlands is much higher in Table 2 (7.8%) than in Table 1 (5.3%). This may reflect the small sample size of the Central Highlands in the panel data used in Table 2, which contains only 175 households in 1997-98, compared to the cross-sectional sample of 368 households in 1997-98. The other exception is the Southeast, for which the growth rate in Table 1 (12.7%) is much higher than the rate in Table 2 (9.9%). Part of the reason for this appears to be that recent migrants in that region, who are not included in the panel data, are better off. About 3% of the households in the Southeast in the 1997-98 survey are recent migrants, and their mean per capita expenditure (7.82 million Dong) is much higher than that of the 97% who are not recent migrants (4.97 million Dong).

The most important results in Table 2 are the growth rates for different quintiles, which are dramatically different from those in Table 1. Table 2 shows that, when households are classifies by their 1992-93 quintiles, poorer quintiles had much higher growth rates than better off quintiles, while Table 1 shows the opposite, albeit weaker, relationship. Specifically, the growth rate of quintile 1 in Table 2 is 13.9%, which is more than double the growth rate for that quintile shown in Table 1 (6.5%). Conversely, the growth rate for quintile 5 in Table 2 is only 4.6%, which is only about one half of the rate seen for quintile 5 in Table 1 (8.5%), and only one third of the growth rate for the poorest quintile in Table 2.

There are three possible reasons for the dramatic differences in the pattern of growth rates by quintile seen in Tables 1 and 2. First, the results may differ simply because the panel households used in Table 2 are not a representative sample of the entire population, since only about 91% of the households in 1992-93 were reinterviewed in 1997-98. To check this, the last five rows of Table 2 show "cross-section" results for the panel data, that is figures for 1997-98 that classify households into quintiles according to their per capita expenditures in 1997-98 (instead of according to their quintile in 1992-93). These figures show a pattern quite similar to that in Table 1, which implies that the main difference between Tables 1 and 2 is not due to attrition bias in the panel data.

The second possibility is that, as explained above, when the same households are compared over time the growth rates will increase for quintile 1 (and perhaps for quintile 2) as long as some economic mobility is present, and by the same line of reasoning the growth rate will decrease for quintile 5 (and perhaps quintile 4). This simply reflects a change in how households are assigned to quintiles in 1997-98.

The third possible reason is that there is almost certainly measurement error in the household expenditures per capita variable that misclassifies households by quintile in both years. As shown in Glewwe (2005a), random measurement error exaggerates economic mobility, and thus overestimates growth rates in lower quintiles and underestimates them in higher quintiles. To see the intuition, suppose that random measurement error in 1992-93 causes 10% of quintile 1 households to be mistakenly classified as quintile 2 households and 10% of quintile 2 households to be mistakenly classified as quintile 1 households. When the households that are classified as quintile 1 households in 1992-93 are re-interviewed in 1997-98, they include the 10% that were really quintile 2 households in 1992-93 and thus have, on average, higher expenditures in 1997-98 than the households they replaced who really belonged to this quintile, namely the 10% of the households in quintile 1 in 1992-93 that were mistakenly classified as quintile 2 households. This implies that the pattern of much higher growth rates for lower quintile households seen in Table 2 is exaggerated, and the same reasoning implies that the growth rates in the upper quintiles are underestimated. The following subsection presents simulations that correct for the bias caused by measurement error.

B. Estimates that Correct for Measurement Error. The estimates presented in Tables 1 and 2 ignore the possibility that the expenditure variable is measured with error. For most classifications of households this is not a major problem because measurement errors will have little effect when means are taken over household groups. However, as explained above, this is not the case when households in both years are classified by their expenditure quintiles in the first year. This subsection applies to the VLSS data the method developed in Section III that corrects for measurement error.

To obtain estimates of growth rates for each quintile that correct for measurement error, the joint distribution of x* and y* is simulated using equation (10). Simulations could also have been done using equation (3), but the linearity assumption in equations (3) and (10) was rejected for equation (3), but not for equation (10), as explained above.

Simulating the joint distribution of x* and y* in equation (10) requires estimates of β_1^*, α_1^*, the mean and variance of ln(y*) and the variance of u_1. Consider first β_1^*. Recall that using a second measurement of y* as an instrument produces an inconsistent estimate of β_1^* when the measurement errors of x* and y* are correlated, but using a variable caused by y* as an instrument yields a consistent estimate. The best candidate for such a variable in the VLSS data is the average body mass index (BMI) of adult household members. BMI is defined as weight (in kilograms) divided by the square of height (in meters). The intuition for BMI as an instrument caused by household expenditures is simple; households that spend more on food have heavier members. A potential shortcoming of BMI in a developing country setting is that BMI could cause household expenditures because severely undernourished people have a lower capacity to work. Yet this reverse causality is unlikely to be strong in Vietnam because only 3.9% of the adults in the 1992-93 VLSS sample are classified as being "severely underweight" (BMI ≤ 16), and this figure drops to 1.8% for males aged 18 to 60.[5]

Unfortunately, there is another potential problem with using BMI as an instrument to estimate equation (10), which is that adult weight is a stock, so past expenditure levels may determine current weight. This implies that u_1, which indicates idiosyncratic differences in 1992-93 expenditures after conditioning on 1997-98 expeditures, may be positively correlated with weight in 1997-98 and thus BMI in 1997-98 would not be a valid instrument for y* in (10). Furthermore, the first two lines in Table 3 show that Var[ln(x)] is approximately equal to Var[ln(y)], and if the proportion of those variances that is due to measurement error in those two variables is about the same, then $\beta_1^* = \beta_1(\beta_2^*/\beta_2)$ by equation (16). Thus one can use an estimate of β_2^* in equation (3) to estimate of β_1^* in equation (10). Using BMI in 1992-93 as an instrument for x* in equation (3) does not cause problems even though BMI is a stock because u_2 in equation (3) represents idiosyncratic differences in 1997-98 expenditures after conditioning on 1992-93 expenditures, and it is unlikely that household members' weights in 1992-93 would be correlated with this idiosyncrasies (recall that severe malnutrition is rare among Vietnamese adults).

Turn next to estimation of α_1^* and the mean of ln(x*). Since the measurement errors ε_x on ln(x*) have a zero mean, the mean of ln(x) in the first line of Table 3 is an unbiased estimate of the mean of ln(x*). Since u_1 also has a zero mean, α_1^* can be

[5] It is still possible that individuals' *heights* can have a causal impact on their labor productivity and thus on their income, but height is determined during childhood and is insensitive to current expenditures.

estimated as $\overline{\ln(x)}$ - β_1* $\overline{\ln(y)}$. Estimates of β_2, β_2*, β_1, β_1* and α_1* are given in lines 3-7 of Table 3; β_1 and β_2 are from OLS estimates using observed values, i.e. ln(x) and ln(y), while β_2* is estimated using BMI in 1992-93 as an instrumental variable for ln(x*) and β_1* is estimated based on the proportional measurement error assumption.

Estimating the variance of ln(y*) is less straightforward. As explained in Section IV, it is possible to estimate only an upper bound of Var[ln(y*)]. The upper bounds reported in the last two lines of Table 3 indicate that Var[ln(y)] overestimates Var[ln(y*)] by at least 12%. Thus a conservative estimate (in the sense of staying close to the upper bound) of Var[ln(y*)] would be 0.85×Var[ln(y)], which is 0.281. A more speculative, but perhaps more accurate, estimate would be 0.70×Var[ln(y)], which is 0.232. Both assumptions are used in the simulations to check for robustness of results.

The last parameter needed for the simulations based on equation (10) is the variance of u_1. This can be obtained using equation (15): Var(u_1) = Var[ln(x*)] – β_1Var[ln(y*)]. This will vary according to the assumptions made regarding the variances of ln(x*) and ln(y*).

The results in Table 3, along with assumptions on the distances from Var[ln(x*)] and Var[ln(y*)] and their estimated upper bounds, were used to simulate the joint distribution of ln(x*) and ln(y*). Before calculating quintile specific growth rates based on these simulations, it is instructive to use equations (7) and (14) to obtain simulated measurement errors for both ln(x*) and ln(y*) and then simulate the joint distribution of observed expenditures; comparing these to the results in Table 2 shows whether the simulations do a reasonably good job of replicating patterns in the observed data, which indirectly checks the credibility of the simulations of the unobserved data.

Table 4 shows what the simulated joint distribution of ln(x*) and ln(y*) implies for the growth rates of the observed expenditures for the five quintiles. Note that these simulations are completely insensitive to assumptions regarding the extent to shich Var[ln(x*)] and Var[ln(y*)] lie below their estimated upper bounds (if they were sensitive to these assumptions this information could be used to estimate these variances more precisely). Thus there is no need to show different simulations based on different assumptions about those variances.

For Vietnam as a whole, the simulations of the joint distribution of ln(x) and ln(y) indicate a growth rate of 7.3% for the panel households, which is almost identical to the growth rate of 7.5% in Table 2. When panel households are classified by their current quintile in both 1992-93 and 1997-98, the poorest quintile has a growth rate of 7.1% while the wealthiest quintile has a growth rate of 7.5%. The growth rates for quintile 2 is 7.2%, and those for quintiles 3 and 4 are both 7.3%. These results are very close to those seen in Table 2, where the growth rates for the five quintiles, from lowest to highest, are 6.6, 6.7, 6.7, 7.0 and 7.6; the main difference is that the simulated growth rates by current quintile are slightly less dispersed than the actual rates.

The simulated growth rates for quintiles when households in both years are classified according to their quintile in 1992-93 are shown in the bottom half of Table 4. The same pattern is found that was seen in Table 2. Specifically, quintile 1 has a growth rate of 14.0% in Table 4, which is virtually identical to the rate of 13.9% seen in Table 2. The growth rate for the wealthiest quintile is 3.8% in Table 4, which is slightly lower than the rate of 4.6% in Table 2. The growth rates for the middle three quintiles are very close in both tables. The divergence in growth rates is slightly higher in the simulated distribution of observed expenditures, but on the whole the simulated results are remarkably close to the observed results in Table 2.

The similarity of the simulated growth rates for x and y in Table 4 with the actual growth rates in Table 2 suggests that simulations that correct for measurement error should be taken seriously. Such simulations are shown in Table 5; the only difference with those in Table 4 is that the simulated random measurement errors added to ln(x*) and ln(y*) in Table 4 was not added in Table 5. These results are based on the assumption that the upper bounds on Var[ln(x*)] and Var[ln(y*)] are far above their actual values. More specifically, they assume that Var[ln(x*)] = 0.7 Var[ln(x)] and Var[ln(y*)] = 0.7 Var[ln(y)]. The top half of Table 5 shows growth rates by quintile when households are classified by their current quintile in both years. These rates are virtually identical to those in Table 4, which is not surprising since measurement errors have little effect on such estimates. Thus expenditure growth in Vietnam in the 1990s appears to have been almost the same at each point in the income distribution. The only qualification to this inference is that the simulated results in Table 4 for observed expenditure slightly underestimated the dispersion in growth rates seen in Table 2. Since measurement error has very little effect on this type of growth rate, the results when households are classified by their current quintile in each year are probably more accurately portrayed in Table 2 than in Table 5. Even so, economic growth in Vietnam is quite pro-poor in that these growth rates do not differ very much between poor and better off households.

The bottom half of Table 5 shows growth rates by quintile when households are classified by their 1992-93 quintiles in both years. Two results stand out. First, as expected, correcting for measurement error substantially reduces the dispersion in growth rates, reducing growth rates for poor quintiles and raising them for wealthy quintiles. In particular, when measurement errors are removed the growth rate of the poorest 20% falls from 13.9% in Table 2 to 11.1% in Table 5, while the growth rate for the wealthiest 20% increases from 4.6% to 5.1%.

The second result is that, when following the same households over time, the poorest 20% still experienced the highest growth rate and the wealthiest experienced the lowest growth rate even after removing bias due to measurement error. The differences are quite large; for example, the growth rate of 11.1% for the poorest quintile is more than twice the rate of 5.1% for the wealthiest quintile. Overall, these results indicate that the poor in Vietnam are not falling behind the rich, no matter how the comparisons are made. Indeed, in one sense they are doing much better than the rich.

The assumption that the true variances of ln(x*) and ln(y*) are well below their estimated upper bounds could underestimate the true variances. Indeed, the assumption implies that 30% of the variance in ln(x) and ln(y) is due to measurement error, which is a much higher percentage than estimates of the contribution of measurement error to the variance of observed earnings in U.S. data (Pischke). To check the robustness of the results in Table 5, Table 6 assumes that the true variances of ln(x*) and ln(y*) are much closer to their observed upper bounds: $Var[ln(x^*)] = 0.85\ Var[ln(x)]$ and $Var[ln(y^*)] = 0.85\ Var[ln(y)]$. The top half of Table 6 shows growth rates by quintile when households are classified by their current quintile in both years; these rates are almost identical to those in Table 5, as one would expect given the discussion above on this type of comparisons across quintiles. The bottom half of Table 6 shows growth rates by quintile when households are classified by their 1992-93 quintiles in both years. The extent to which growth rates of the poorer quintiles are higher than those of the better of quintiles increases, but not by much. Specifically, the growth rate of the poorest 20% increases from 11.1% in Table 5 to 11.5% in Table 6, while the growth rate for the wealthiest 20% decreases from 5.1% to 5.0%. The true growth rates for each quintile probably fall somewhere between the estimates of Tables 5 and 6, which is a fairly tight range.

VI. Summary and Concluding Comments

Broad-based economic growth in developing countries reduces poverty, almost by definition, and this has clearly been the case for Vietnam. However, rising inequality in observed per capita expenditures raises the question of whether the wealthier groups in Vietnam have benefited more from economic growth than have the poor. In principle, the rich household survey from Vietnam should be able to answer this question easily. Yet there are two complications. First, when looking at growth rates for poorer and better off households, the presence of economic mobility raises fundamental questions of who to compare over time. In particular, should the poorest 20% of households in one year be compared to the same households in a later year or to the poorest 20% of households in that later year? This question does not admit a simple answer, and in this paper both comparisons are made.

A second difficulty is that measurement error is present in the data, and this causes bias in estimates of growth rates by quintiles. This bias is particularly large when growth rates are calculated that compare the same households over time.

This paper has presented a method to correct for measurement error. The method entails estimating the joint distribution of per capita expenditures over time. The results suggest that economic growth in Vietnam has been relatively equitable when comparing the poorest 20% or poorest 40% in 1992-93 with the poorest 20% or poorest 40% in 1997-98. More specifically, Table 2 shows that the growth rates by quintile range from 6.1% for the poorest quintile to 7.6% for the wealthiest quintile when comparisons are made that classify households according to their current quintile in both years. A more interesting result is that when households are classified in both years according to their quintile in 1992-93, economic growth in Vietnam has been very pro-poor. In particular, the rate of expenditure growth among the poorest quintile is more than twice as high as the rate among the richest quintile.

These results suggest that the pessimism of some economists that economic growth inevitably favors the rich over the poor is unwarranted. Vietnam's growth in the 1990s did not follow such a pattern. Yet evidence from a single country is insufficient for drawing general conclusions; further analysis is needed for other countries to see whether Vietnam's experience is typical. While comparable cross-sectional data exist for many developing countries, which allows for comparisons of growth rates based on current quintiles, comparisons of the same households over time requires nationally representative data, which are rare for developing countries. Further analysis using both types of comparisons, controlling for measurement errors using methods such as those presented here, will be needed to see the extent to which economic growth in developing countries is, in general, "pro-poor".

References

Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Johns Hopkins University Press.

Glewwe, Paul. 2005a. "How Much of Observed Economic Mobility is Measurement Error? A Method to Remove Measurement Error, with an Application to Vietnam". Department of Applied Economics, University of Minnesota.

Glewwe, Paul. 2005b. "Measurement Error Bias in Estimates of Income and Income Growth among the Poor: Results and a Correction Formula when Income Follows a Lognormal Distribution". Department of Applied Economics, University of Minnesota.

Pischke, Jörn-Steffen. 1995. "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study" *Journal of Business and Economic Statistics.* 13(3):305-314.

United Nations Development Programme. 2001. *Human Development Report 2001: Making Technologies Work for Human Development.* United Nations, New York.

Winters, Alan, Neil McCulloch and Andrew McKay. 2004. "Trade Liberalization and Poverty: The Evidence So Far." *Journal of Economic Literature* 42(1): 72-115.

World Bank. 1999. *Vietnam: Attacking Poverty*. The World Bank: Hanoi.

World Bank. 2001a. *World Development Report 2000/2001: Attacking Poverty*. The World Bank. Washington, DC.

World Bank. 2001b. "Vietnam Living Standards Survey (VLSS), 1997-98: Basic Information". Development Research Group. The World Bank. Washington, DC.

World Bank. 2002. *World Development Report 2002: Building Institutions for Markets.* The World Bank. Washington, DC.

Table 1: Growth Rates Using Two Cross Sectional Surveys, Based on Observed Data

	Population Distribution, 1992-93 (%)	Mean Per Capita Expenditures 1992-93	Population Distribution, 1997-98 (%)	Mean Per Capita Expenditures 1997-98	Growth over 5 Years (%)	Average Annual Growth Rate (%) Household Level	Individual Level
All Vietnam		2019		2988	48.0	8.2	7.6
Urban	19.9	3329	22.4	5160	55.0	9.2	9.3
Rural	80.1	1692	77.6	2300	35.9	6.3	6.0
By region							
Northern Uplands	15.6	1456	17.9	2098	44.1	7.6	6.8
Red River Delta	21.5	1926	19.6	3037	57.7	9.5	10.0
North Central Coast	12.8	1495	13.8	2361	57.9	9.6	8.8
South Central Coast	11.9	2194	10.7	2808	28.0	5.1	5.4
Central Highlands	3.2	1593	3.7	2062	29.4	5.3	5.1
South East	12.6	3023	12.8	5494	81.7	12.7	12.0
Mekong Delta	22.4	2296	21.5	2808	22.3	4.1	3.7
By quintile							
Poorest 20%	20	808	20	1109	37.3	6.5	6.4
Next 20%	20	1176	20	1635	39.0	6.8	6.8
Next 20%	20	1529	20	2131	39.4	6.9	6.8
Next 20%	20	2065	20	2936	42.2	7.3	7.3
Richest 20%	20	4126	20	6207	50.4	8.5	8.6

Notes: Sample size is 4799 households for 1992-1993 and 5999 for 1997-1998. All per capita expenditure figures are in terms of January 1998 Dong.

Table 2: Growth Rates Using Panel Data with Same Head of Household, Based on Observed Data

	Population Distribution, 1992-93 (%)	Mean Per Capita Expenditures 1992-93	Mean Per Capita Expenditures 1997-98	Growth over 5 Years (%)	Average Annual Growth Rate (%)
All Vietnam		1974	2836	43.7	7.5
Urban	18.7	3247	4893	50.7	8.5
Rural	81.3	1680	2360	40.5	7.0
By region					
Northern Uplands	16.2	1438	2103	46.2	7.9
Red River Delta	20.4	1927	2955	53.3	8.9
North Central Coast	13.6	1496	2299	53.7	9.0
South Central Coast	12.1	2067	2846	37.7	6.6
Central Highlands	3.2	1581	2305	45.8	7.8
South East	12.4	2874	4602	60.1	9.9
Mekong Delta	22.1	2293	2749	19.9	3.7
By 1992-1993 quintile					
Poorest 20%	20	803	1542	92.0	13.9
Next 20%	20	1170	1925	64.5	10.5
Next 20%	20	1514	2387	58.3	9.6
Next 20%	20	2032	2919	43.7	7.5
Richest 20%	20	3992	5006	25.4	4.6
By current quintile					
Poorest 20%	20	803	1104	37.5	6.6
Next 20%	20	1170	1621	38.5	6.7
Next 20%	20	1514	2098	38.4	6.7
Next 20%	20	2032	2844	40.0	7.0
Richest 20%	20	3992	5756	44.2	7.6

Note: Sample size is 4281 panel households for 1992-1993 and 1997-1998. All per capita expenditure figures are in terms of January 1998 Dong.

Table 3: Estimates Used in Simulations

Var[ln(x)]	0.324
Var[ln(y)]	0.331
$\overline{\ln(x)}$	7.409
$\overline{\ln(y)}$	7.763
β_2	0.710
β_2* (IV estimates using BMI as instrument)	0.801
β_1	0.694
β_1* (using proportional variance assumption)	0.783
α_1* (computed as $\overline{\ln(x)} - \beta_1^* \overline{\ln(y)}$	1.331
Upper bound for Var[ln(x*)], as given in equation (6)	0.281
Upper bound for Var[ln(y*)], as given in equation (13)	0.293

Note: All estimates are based on the 4281 households with complete expenditure data in both years.

Table 4: Growth Rates in Observed Expenditure Using Simulated Panel Data

	Mean Per Capita Expenditures 1992-93	Mean Per Capita Expenditures 1997-98	Growth over 5 Years (%)	Average Annual Growth Rate (%)
All Vietnam	1952	2780	42.4	7.1
By current quintile				
Poorest 20%	769	1086	41.2	7.1
Next 20%	1229	1737	41.3	7.2
Next 20%	1661	2362	42.2	7.3
Next 20%	2262	3215	42.1	7.3
Richest 20%	3838	5500	43.3	7.5
By 1992-93 quintile				
Poorest 20%	769	1482	92.7	14.0
Next 20%	1229	2068	68.3	11.0
Next 20%	1661	2546	53.3	8.9
Next 20%	2262	3185	40.8	7.1
Richest 20%	3838	4619	20.3	3.8

Note: Sample size is 50,000 simulated households

Table 5: Simulated Growth Rates in Actual (Unobserved) Expenditure, Assuming Var[ln(x*)] = 0.7Var[ln(x)] and Var[ln(y*)] = 0.7Var[ln(y)]

	Mean Per Capita Expenditures 1992-93	Mean Per Capita Expenditures 1997-98	Growth over 5 Years (%)	Average Annual Growth Rate (%)
All Vietnam	1854	2643	42.6	7.3
By current quintile				
Poorest 20%	867	1226	41.4	7.2
Next 20%	1287	1828	42.0	7.3
Next 20%	1659	2363	42.4	7.3
Next 20%	2144	3052	42.4	7.3
Richest 20%	3314	4747	43.2	7.5
By 1992-93 quintile				
Poorest 20%	867	1465	69.0	11.1
Next 20%	1287	2018	56.8	9.4
Next 20%	1659	2449	47.6	8.1
Next 20%	2144	3027	41.2	7.1
Richest 20%	3314	4258	28.5	5.1

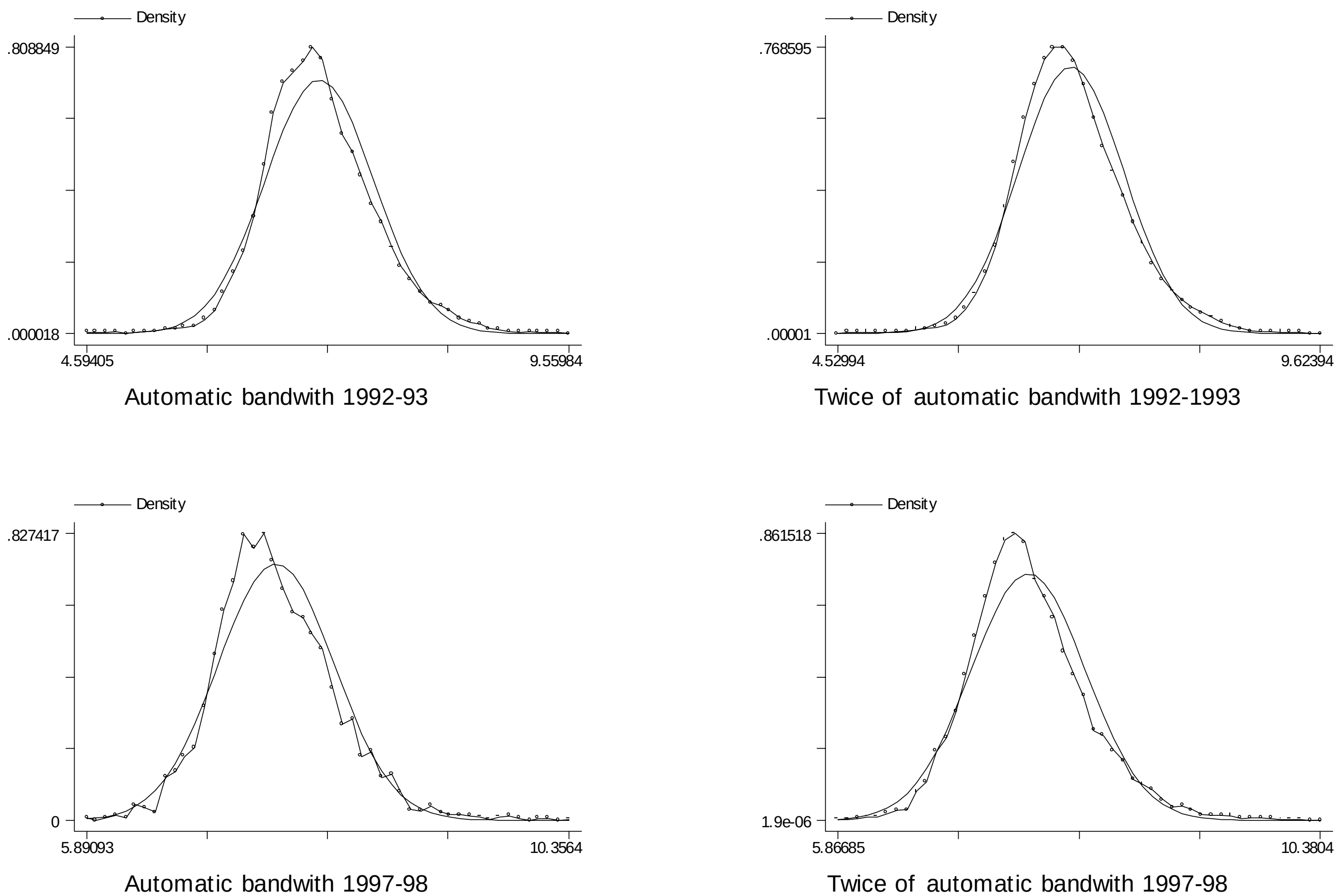
Note: Sample size is 50,000 simulated households

Table 6: Simulated Growth Rates in Actual (Unobserved) Expenditure, Assuming Var[ln(x*)] = 0.85Var[ln(x)] and Var[ln(y*)] = 0.85Var[ln(y)]

	Mean Per Capita Expenditures 1992-93	Mean Per Capita Expenditures 1997-98	Growth over 5 Years (%)	Average Annual Growth Rate (%)
All Vietnam	1901	2710	42.6	7.3
By current quintile				
Poorest 20%	814	1150	41.3	7.2
Next 20%	1256	1783	42.0	7.3
Next 20%	1660	2364	42.4	7.3
Next 20%	2203	3136	42.6	7.3
Richest 20%	3570	5117	43.3	7.5
By 1992-93 quintile				
Poorest 20%	814	1405	72.6	11.5
Next 20%	1256	1997	59.0	9.7
Next 20%	1660	2471	48.9	8.3
Next 20%	2203	3121	41.7	7.2
Richest 20%	3570	4556	27.6	5.0

Note: Sample size is 50,000 simulated households

Figure 1: Density Estimates of Log Per Capita Expenditures, 1992-93 and 1997-98



Note: The solid line is a normal distribution with same mean and variance of the actual distribution.