Consistent Estimation of Income Inequality

John Luke Gallup

Portland State University

jlgallup@pdx.edu

January 31, 2010

Abstract

Measurement error biases inequality statistics because they are measures of
dispersion, and errors spuriously add to the calculated dispersion. Ignoring
measurement error will bias not only the level of income inequality but its change
over time if certain types of workers, such as the self-employed, have incomes
observed with greater error, and the composition of workers changes.

I derive a new consistent estimator of inequality, using survey data which include
both household income and expenditure, which are available in many countries.
The new estimator is decomposable by group and by income source. Using data
from Vietnam, I show that the bias due to uncorrected measurement error is
large and, uncorrected, gives a highly deceptive picture of recent trends in income
distribution.

1 Introduction

The statistics of interest in economics are usually measures of central tendency, such as means or regression coefficients. When the statistics are linear averages, even when the data are subject to random measurement error, the statistics are still unbiased.

Inequality statistics, in contrast, are measures of dispersion. They are directly biased by measurement error because the errors spuriously add to the variability of the data. For example, if all individuals earned exactly the same income, but the income was measured with error, the calculated statistic would show positive inequality.

Even if annual income were measured with perfect accuracy, there is an additional problem. We are usually interested in the inequality of permanent income, not the annual fluctuations of transitory income. The annual income of farmers, for example, varies from year to year due to weather and market price changes. Even if all farmers had the same expected income each year, transitory differences between the locations that had good weather or bad weather would appear as income inequality, even if they balance out from year to year so that there was no inequality in permanent income. The income inequality of sectors with a lot of fluctuation over time, like agriculture and household businesses, would appear more unequal than sectors where income is more stable, like wage labor.

A simple solution to the bias caused by measurement error and by transitory income in current income data is to calculate the inequality of consumption expenditure instead. This only a partial solution, though. Consumption expenditure is also estimated with error, though typically less than with income, so that the measurement error problem has not been solved. Just as important, it is not possible to explore the contribution of different sources of income to inequality using consumption rather than income data, which makes it hard to explore the causes of changing income distribution.

The approach proposed in this paper is to use the covariance of income and consumption expenditure to form a consistent estimate of the variance of permanent income. The two "errors" of concern in income data are measurement error and transitory income. Both of these errors are arguably uncorrelated with consumption. The error in consumption data is only measurement error. There is usually no reason to think that the measurement error in consumption data is correlated with measurement error in income. The covariance of income and consumption is not biased if the errors in each are uncorrelated, even though the own variance of both income and consumption would be

biased.

The covariance of income and consumption provides an immediate estimate of the inequality of permanent income for inequality statistics based on income variance. I am also able to derive approximations to consistent estimates for most other common inequality statistics using the results of Chesher and Schluter (2002).

A previous literature has established the impact of measurement error on inequality statistics. (Chakravarty and Eichhorn 1994) proved that random measurement error biases estimated inequality for a wide class of inequality statistics when measurement errors are additive. Arnold (1980, cited in van Praag, Hagenaars, and van Eck 1983) proved that measurement error biases estimated inequality when the errors are multiplicative.

Van Praag, Hagenaars, and van Eck (1983) estimate the impact of measurement error on inequality using a special Dutch survey where income was independently measured twice, using the assumption that income is distributed log normally. Gottschalk and Huynh (2006) use panel data to estimate the impact of measurement error on inequality and mobility. Both studies find that measurement error has a large impact on estimated inequality. (Cowell and Victoria-Feser 1996) assess the robustness of inequality statistics to idiosyncratic outliers and propose parametric models to correct for the bias, but do not address ubiquitous measurement error. Several papers (Israelsen, McDonald, and Newey, 1984; Chesher and Schluter 2002) explore the effects of different kinds of measurement errors when the magnitude and distribution of the errors are known *a priori*. This can be useful for sensitivity analysis, but does not provide a method for correcting for unknown measurement error bias.

The existing literature contains several methods for correcting measurement error when repeated independent measurements of income are available, or in contexts where the size and distribution of measurement error is already known, but there has not been a practical solution for cross-sectional data when the size of the errors are unknown.

Section 2 demonstrates that measurement error biases inequality statistics. In Section 3, I argue that income and consumption are both noisy estimates of smoothed, or "permanent", income. Section 4 derives unbiased income inequality statistics that combine income and consumption data. Some of these statistics can provide consistent estimates of inequality decomposed by group and by source of income. Section 5 uses data from Vietnam to compare biased and consistent estimates of inequality. Section 6 concludes.

2 Measurement error biases inequality statistics

Suppose that the random variables y and y^* represent respectively the observed and true income in the population under consideration. Since y is observed with error,

$$y = y^* + e \tag{1}$$

where e is an error term. Assume that given a level of true income, y^*, the observed income is an unbiased estimate of the true income, that is,

$$E(y \mid y^*) = y^{*.1} \tag{2}$$

This implies that $E(e \mid y^*) = 0$, and by iterated expectations, $E(e) = 0$.

The variance of observed income, a measure of inequality, is biased upward when income is measured with error. The sample variance of observed income in a random sample of size n is

$$s_y^2 = \left(\frac{1}{n-1}\right)\sum_{i=1}^{n}(y_i - \bar{y})^2 .$$

From (1), $s_y^2 = s_{y*}^2 + s_e^2 + 2s_{y*,e}$ where $s_{y*,e}$ is the sample covariance of y^* and e. Since the sample covariances of a random sample are unbiased estimates of the population covariances, $E(s_y^2) = var(y^*) + var(e) + 2\,cov(y^*, e)$, where $var(\cdot)$ is the population variance and $cov(\cdot)$ is the population covariance. $cov(y^*, e) = E((y^* - Ey^*)e)$ since $E(e) = 0$. Taking iterated expectations with respect to y^*, $cov(y^*, e) = 0$.

$$E(s_y^2) = var(y^*) + var(e) > var(y^*)$$

so the greater the measurement error (the greater is $var(e)$), the more the sample variance of observed income exaggerates the variance of true income.

---

1   This assumption, sometimes called "classical" measurement error, is important. It asserts that neither low nor high incomes are systematically under- or over-reported, which I will argue is plausible for the data used in this paper. If observed low incomes were biased upwards or high income biased downwards, it would counteract the tendency for measurement error to inflate inequality measures. Gottschalk and Huynh (2006) consider the impact of "non-classical" measurement error on inequality measures.

The variance of income is a familiar measure for economists, but what can be said about the impact of measurement errors on inequality statistics in general? We follow Chakravarty and Eichhorn's (1994) proof that measurement error causes an upward bias in a wide class of inequality statistics.

Dalton (1920) first pointed out that underlying each inequality measure is an implicit social welfare function. If inequality is considered bad, then other things being equal, social welfare is lower when inequality is higher and *vice versa*. The impact of each individual in the distribution of incomes on the inequality measure depends on the weight given to a person of that income level in the social welfare function. Let U be a general von Neumann-Morgenstern social welfare function of individuals in the community, which is strictly concave in income. By Jensen's Inequality,

$$E(U(y) \mid y^*) < U(E(y \mid y^*)). \tag{3}$$

Applying assumption (2), $U(E(y \mid y^*)) = U(y^*)$, so that (3) becomes

$$E(U(y) \mid y^*) < U(y^*). \tag{4}$$

Taking expectations of both sides of (4) with respect to y^*, we get

$$E(U(y)) < E(U(y^*)).^2 \tag{5}$$

From assumption (2), we know that y and y^* have the same means. Since U is an arbitrary strictly concave utility function, Kolm (1969) and Atkinson(1970) demonstrate that (5) is equivalent to the statement that y^* Lorenz dominates y. For any inequality index I satisfying symmetry and the Pigou-Dalton transfer principle, Foster (1985) and Chakravarty (1990) demonstrate that y^* Lorenz dominating y is equivalent to $I(y) > I(y^*)$.

The assumption that the social welfare function U is strictly concave ensures that less inequality is preferred to more inequality over all income ranges. Since observed income y has more dispersion than actual income y^*, the expected value of U is lower for observed income, which given equal means, implies that inequality index I is biased upwards due to measurement error. It should not be surprising that measurement error increases estimated inequality. Inequality statistics measure dispersion, and measurement error spuriously increases dispersion.

2　There is a typo in this equation in Chakravarty and Eichorn (1994, their equation 4), confirmed in personal correspondence with Chakravarty.

The condition that *I* satisfies symmetry means that any two individuals in the income distribution can trade incomes without it affecting the index, i.e. the index depends on the income distribution, not on who the individuals are. The Pigou-Dalton principle (given the more evocative name "the Robin-Hood principle" by Atkinson (1970)) ensures that whenever income is taken from a richer individual and given to a poorer individual (as long as it is not so much money the richer individual now becomes ever poorer than the poor individual was) then the inequality index decreases. This makes the inequality index sensitive to equalizing transfers throughout the income range.

The Pigou-Dalton principle rules out quantile-based inequality indices, such as the 90/10 ratio, the ratio of income share of the top decile to the bottom decile. Quantile-based indices are insensitive to equalizing transfers within the quantiles. Other than that, all the common inequality indices are included in Chakravarty and Eichorn's (1994) result, such as the generalized entropy family (Shorrocks, 1980), the Atkinson (1970) index, and the Gini coefficient.

3   The relationship between income and consumption

Access to two estimates of income, both subject to measurement error, but with uncorrelated errors, makes possible a consistent estimate of income inequality. The central idea of this paper is that both current income and current consumption are noisy estimates of permanent income and can be combined to consistently estimate income inequality.

Lifetime consumption is constrained by the lifetime flow of income for the great majority of people who don't receive significant bequests and don't declare bankruptcy. Regardless of the specifics of how individuals or families make decisions, the lifetime budget constraint binds: you cannot spend more in your lifetime than you earn.

When income is variable, households smooth out consumption to iron out the ups and downs of income. When income is higher than usual, households tend to save, and when income is lower than usual, they often dissave. Most prefer a steady standard of living to feast and famine.

The desire to smooth consumption can be modeled as the consequence of expected utility maximization of present and future consumption where utility is concave in consumption. A closed-form solution for the level of current consumption as a function

of income requires a whole handful of not very realistic assumptions, but can still serve as a heuristic motivation for the relationship of consumption and income.

Hall (1978) presents the expected utility version of the life-cycle model of consumption. He assumes preferences are intertemporally separable, that each subperiod's utility function is quadratic and identical across periods except for discounting by the rate of time preference. Consumers are able to borrow or lend unlimited amounts at the same interest rate. The rate of time preference and interest rates are constant and equal to each other. Then expected utility maximization implies that consumption in time t, c_t^*, follows a martingale:

$$c_t^* = c_{t-1}^* + \epsilon_t.$$

c_t^* has an asterisk to indicate that it is true consumption, free of measurement error, not observed consumption. Suppose further that there is a single real asset A_t that evolves according to

$$A_t = (1+r)(A_{t-1} + y_{t-1}^* - c_{t-1}^*) \tag{6}$$

where r is the constant interest rate and y_{t-1}^* is (measurement error-free) income in period t-1. We assume there are no bequests, so $A_0 = A_T = 0$, where T is the last period. The lifetime budget constraint is then

$$\sum_{k=0}^{T} (1+r)^{-k} c_k^* = \sum_{k=0}^{T} (1+r)^{-k} y_k^*. \tag{7}$$

We divide income y_t^* into permanent income y_t^p and transitory income u_t, where $E(u_t)=0$. Suppose that permanent income grows at a constant rate g. Taking expectations of consumption and income conditional on the information available at time t (denoted E_t), we find

$$
\begin{aligned}
E_t c_k^* &= c_t^* && \text{if} \quad k > t \\
E_t c_k^* &= c_t^* + \epsilon_k && \text{if} \quad k \leq t \\
E_t y_k^* &= (1+g)^{k-t} y_t^p && \text{if} \quad k > t \\
E_t y_k^* &= (1+g)^{k-t} y_t^p + u_k && \text{if} \quad k \leq t
\end{aligned}
$$

Applying these results to equation 7,

$$c_t^* \sum_{k=0}^{T} (1+r)^{-k} \;=\; y_t^p \sum_{k=0}^{T} \frac{(1+g)^{k-t}}{(1+r)^k} \;+\; \sum_{k=0}^{t-1} (1+r)^{-k}(u_k - \epsilon_k).$$

Simplifying,

$$c_t^* \;=\; m(1+g)^{-t} y_t^p \;+\; \alpha \sum_{k=0}^{t-1} (1+r)^{-k}(u_k - \epsilon_k) \tag{8}$$

where $\alpha = r/[1+r - (1+r)^{-T}]$, $m = \alpha \alpha'$, where $\alpha' = \dfrac{(1+g)/(1+r) - [(1+g)/(1+r)]^{-T}}{(1+g)/(1+r) - 1}$.

Equation 8 provides a linear relationship between current consumption and current permanent income. Given the number of more or less arbitrary assumptions in the life cycle model, it may be just as satisfying to assume that consumption is proportional to permanent income, as done by Friedman (1957). The life cycle model provides us with an internally consistent model of how expected utility maximization implies consumption smoothing, even if most of the predictions of the model have been rejected by empirical testing (Deaton, 1992, especially chapter 2). Many of the arbitrary assumptions used here can be avoided in a life cycle model if we can give up a closed form relationship between current consumption and permanent income, but we need this to evaluate income inequality using cross-sectional data.

In the end the most unrealistic assumption of the life cycle model may be the hyper-rationality of life-time intertemporal maximization. The whole field of behavioral finance is based upon (and makes money on) the identification of contexts in which people, even professional investors, fail to make optimal intertemporal decisions. It is reassuring that the relationship between consumption and permanent income derived from the life cycle model is essentially the same as the simple rule of thumb of Friedman, which has been robustly confirmed in empirical work (Romer, 1996, p. 313).

4 Consistent estimation of income inequality

The relationship in Equation 8 makes it possible to relate the covariance of observed consumption expenditure and observed income. Observed consumption, c_t, is the sum of actual consumption, c_t^*, and a measurement error, e_t^c, uncorrelated with other disturbances:

$$c_t = c_t^* + e_t^c.$$

As before, observed income, y_t, is the sum of true income, y_t^*, and a measurement error, e_t. Breaking out true income into permanent income, y_t^p, and transitory income, u_t,

$$y_t = y_t^p + u_t + e_t.$$

The consumption and income measurement errors are uncorrelated with each other and all other variables, since they are generated by sample survey data collection that is unrelated to household consumption and income generation. Transitory income at time t, u_t, is uncorrelated with permanent income and past realizations of transitory income by construction, and uncorrelated with past consumption innovations, ϵ_k. Hence,

$$cov(y_t, c_t) = cov(y_t^p, c_t^*).$$

Applying Equation 8,

$$cov(y_t, [m(1+g)^{-t}]^{-1} c_t) = var(y_t^p). \qquad (9)$$

This brings us closer to our goal of an unbiased estimate of the variance of permanent income, but we also need a measure of $m(1+g)^{-t}$. This can be obtained from estimating Equation 8 as a regression. The estimation will be applied to cross-sectional household data, so we drop the subscript t to simplify the notation and include the subscript i for households. Since $c_i^* = c_i - e_i^c$ and $y_i^p = y_i - u_i - e_i$, Equation 8 can be rewritten as an estimating equation in terms of observables c_i and y_i

$$c_i = m(1+g)^{-t_i} y_i + \eta_i \qquad (10)$$

where $\eta_i = e_i^c - m(1 - g t_i)(u_i + e_i) + \alpha \sum_{k=0}^{t_i-1} (1+r)^{-k}(u_{ik} - \epsilon_{ik})$, and t_i, the time since the household began earning income, is also observable (imperfectly) as the age of the household head.

If we estimated m and g by nonlinear least squares or any method that required the independent variables to be uncorrelated with the error term, the estimates would be biased. The unobserved error term η_i is correlated with y_i through u_i, transitory income, and e_i, measurement error. Luckily, we have an abundance of good instruments for y_i: traditional determinants of income like the education and experience of working household members as well as demographic composition of the household, both to account for the number of workers contributing to total household income and to capture Chayanov-type (1925) effects where adults with young children work harder to provide for them. These variables are correlated with observed income, but are unlikely to be correlated with transitory income or measurement error in the residual. These variables are correlated with the dependent variable, consumption, but through income's causal effect on consumption, not with consumption's measurement error in the residual.

With instruments in hand, we can estimate m and g consistently with nonlinear instrumental variables estimation. Due to the presence of t_i in the residual, which varies across households, both in the $1 - g\,t_i$ term and $t_i - 1$ in the limit of summation, the residual is heteroskedastic, which is addressed with robust standard errors.

The estimation of m and g can be simplified by approximating $(1+g)^{-t}$ linearly as $1\text{-}gt$, which will be reasonably accurate for small income growth rates. The estimating equation becomes

$$c_i \;=\; m\,y_i \;-\; mg\,t_i y_i \;+\; \eta_i \tag{11}$$

where mg can be identified in the second right hand side term from an interaction of y_i and t_i.

We are able to estimate consistently the variance of permanent income using equation 10 or 11. Let $\tilde{c}_i = [\widehat{m}\,(1-\widehat{g})^{t_i}]^{-1} c_i$ (or $\tilde{c}_i = (\widehat{m} - \widehat{mg}\,t_i)^{-1} c_i$ in the linearized case), where $\widehat{m}$, $\widehat{g}$, and $\widehat{mg}$ are the estimates of the coefficients in Equation 10 or 11 and t_i is the age of the household head. Then a consistent estimate of the variance of permanent income is $s_{y_i,\tilde{c}_i}$, the sample covariance of y_i and $\tilde{c}_i$.

The variance of income is not an ideal inequality measure because its level depends on the choice of units for income. A better income inequality measure which is invariant to

scale is the Generalized Entropy Index of order 2, $I_2 = \dfrac{s_y^2}{2\,\bar{y}}$. Generalized Entropy indices are decomposable by population group and by income source (Shorrocks, 1980). I_2, like all other inequality indices in the broad class described in Section 2, is biased upwards by random measurement error. So we replace the biased index with a consistent estimate:

$$\widehat{I}_2 = \frac{s_{y_i,\tilde{c}_i}}{2\,\bar{y}} \qquad (11)$$

since $\bar{y}$ is a consistent estimate of the mean of permanent income.

5 Income inequality in Vietnam

Vietnam is a country that has had tremendous structural change since moving from pervasive central planning to a market economy around 1990. The result has been the third fastest economic growth in the world in the period since then, with average incomes now three times the level twenty years ago, and higher in the major cities. From 1993 to 2006, the rural population fell from 80% to 73%, and the movement out of agriculture into home enterprises and wage labor has been even more rapid. These changes have big implications for income distribution, since agriculture has both the lowest average incomes and the most equally distributed.

The data used to calculate income inequality come from a series of household living standards surveys, initially conducted under the direction of the World Bank and now continued by the Vietnamese General Statistics Office. These surveys were conducted in 1993, 1998, 2002, 2004, and 2006, although the 2002 data are not used here because it was the first time the Vietnamese government oversaw the data collection, they increased the sample size by five times over the 1998 survey to 30,000 households, and the interview quality control broke down. The 2004 and 2006 surveys have more modest sample sizes (just over 9000 households) and high quality, internally consistent data on income as well as household consumption expenditures.

The calculation of household income is straightforward for wage employment. Monthly compensation was annualized according to the number of months worked in the past year. The calculation for the self-employed, both farmers and those with other household enterprises, was more involved. Net revenues were calculated from detailed

questions on the purchase price of inputs and the sale price of outputs. Investment costs were excluded because of their irregular timing, and the data were not suffi ciently detailed to attribute annual depreciation. The data used for self-employment income is unusually rich, but net revenue still surely includes substantial measurement errors and omissions.

For each survey year, I estimated the relationship between total household income and total household expenditure using equation 10. Instruments for household income were education of the household head, and the average level of education, experience, and experience squared for the other adult workers over age 16. Experience is equal to age minus years of education, so it is collinear with the included variables for education and age for the household head. Addition instruments were numbers of children 0-5 years old, 6-16 years old, and adult men and women in the household, as well an indicator for ethnic minority status of the household head.

Figure 1 shows the uncorrected I_2 inequality index for income per equivalent adult, the I_2 index using consumption expenditure per equivalent adult (instead of income), and the consistent $\widehat{I}_2$ index from Equation 10. The weights for equivalent adults come from the OECD schedule, equal to 0.67 for the fi rst adult and 0.33 for subsequent adults and children over age 14. Children under age 14 have a weight of 0.20 (Cowell, 2008, p. 103). The difference between the uncorrected income inequality and the consistent estimates is striking. For all but 2004, the uncorrected inequality levels are more than twice the consistent estimates. Just as striking is the contrast in the pattern over time. The consistent estimates are concave, rising between the fi rst two surveys and then gradually declining. The uncorrected estimates tell almost the exact opposite story, showing declining inequality that accelerates up to 2004 before rising sharply. The diff erence between the two estimates is the consequence of both pure measurement error and transitory income. This method does not allow us to tell which contributes, but from the point of view of the distribution of welfare in Vietnam, they are both spurious variation that does not sustained diff erences in well-being.

The inequality of household consumption is much closer to the consistent estimates of income inequality, both in level and shape over time, although the estimated measurement error of consumption inequality is bigger in the fi rst and last years than the middle years. The estimates show that consumption inequality is a much better measure of permanent income inequality than uncorrected income inequality, especially in a context like Vietnam where self-employment is more common than wage employment. The drawback of consumption inequality alone is that it is not possible to

decompose it into sources of income.

Figures 2 and 3 compare the contribution of different income sources with and without measurement error correction. The two figures tell dramatically different stories. In the uncorrected estimates, farming contributes almost as much to total inequality as wage income. Wage income ends up having the lower contribution to inequality of any income source in 2006. The corrected estimates, in contrast, show that the highly equal farming income, when adjusted for its high volatility, actually reduces overall inequality until 2006, and wage income makes a large, steadily rising contribution to inequality, becoming the most disequalizing income source in the final year.

The qualitative patterns of the corrected and uncorrected income inequality estimates could not be more different. The uncorrected estimates of inequality are off by about a factor of two, and the evolution over time is close to the opposite of that shown by the corrected estimates, increasing where the consistent estimates decrease and vice versa. In the context of Vietnam, addressing measurement error in income inequality changes the picture entirely.

6   Conclusion

Random measurement error and transitory income directly bias income inequality upwards for a wide class of inequality measures that includes all the common indicators except for quantile-based measures, which are less sensitive to income changes.

Applying the life cycle model of consumption, I derive a relationship between current consumption and permanent income which relies on restrictive assumptions, but is similar to the empirically robust relationship postulated by Friedman (1957). This relationship is the basis of a new estimator of inequality that is consistent in the presence of measurement error. A bonus is that one can estimate the size of the measurement error.

Previous methods for correcting the impact of measurement error on income inequality statistics relied either on repeated independent measures of income, which are rare, or prior knowledge of the size and distribution of the errors, which are usually unknown. The new estimator proposed in this paper can be applied to cross-sectional household survey data as long as it contains measures of both income and consumption. Many high and low income countries regularly conduct such surveys, making it possible to use them to study the evolution of inequality over time.

Applying the new estimator to a series of household surveys in Vietnam, I find that the impact of measurement error (including transitory income) is very large, and correcting it dramatically changes the apparent time path and source of income inequality. The consistent estimates show an pattern of inequality that is almost the opposite of that shown by uncorrected estimates.

The importance of correcting for measurement error is greater in countries where self-employment is substantial and where the prevalence of self-employment is changing, because the self-employed have more volatile income, and it is measured with more error. The self-employed are the majority of workers in most low income countries.

References

Arnold, B. C. 1980. Misreporting Increases Apparent Inequality. Riverside, California: University of California, Riverside.

Arnold, Bany C. 1990. "The Lorenz order and the effects of taxation policies.." *Bulletin of Economic Research* 42(4): 249-264.

Atkinson, A. B. 1970. "On the measurement of inequality." *Journal of Economic Theory* 2(3): 244–263.

Chakravarty, Satya R., and Wolfgang Eichhorn. 1994. "Measurement of income inequality: Observed versus true data." In *Models and measurement of welfare and inequality*, Berlin: Springer-Verlag, p. 28-32.

Chayanov, A.V. 1925[1966]. *The Theory of Peasant Economy*. Homewood, Ill.: R. D. Irwin.

Chesher, Andrew, and Christian Schluter. 2002. "Welfare Measurement and Measurement Error." *The Review of Economic Studies* 69(2): 357-378.

Cowell, Frank A. 2008. *Measuring Inequality.* http://darp.lse.ac.uk/papersDB/Cowell_measuringinequality3.pdf, accessed September 12, 2009.

Israelsen, L.D., J.B. McDonald, and W.K. Newey. 1984. "The Impact of Measurement Error on the Distribution of Income." In *Economic inequality: measurement and policy*, Jai Press.

Kolm, Serge-Christophe. 1969. "The optimal production of social justice," in Margolis, J., Guitton, H., eds. *Public Economics: An Analysis of Public Production and Consumption and their Relations to the Private Sectors*, pp. 145-200. London: Macmillan.

Friedman, Milton. 1957. *A theory of the consumption function*. Cambridge, Massachusetts: National Bureau of Economic Research Books.

Modigliani, Franco, and Brumberg, Richard. 1954. "Utility analysis and the consumption function: an interpretation of cross-section data." In Kenneth K. Kurihara, ed., *Post-Keynesian Economics*, pp 388-436. New Brunswick, N.J.: Rutgers University Press.

Paraje, Guillermo, and Melvyn Weeks. 2002. "How does Income Underreporting Affect Inequality Measures? A Simulation Approach," unpublished paper, Faculty of Economics and Politics, University of Cambridge.

Romer, David A. 1996. *Advanced Macroeconomics*. Columbus, Ohio: McGraw Hill.

Shorrocks, A. F. 1980. "The class of additively decomposable inequality measures." *Econometrica: Journal of the Econometric Society*: 613–625.

Van Praag, Bernard, Aldi Hagenaars, and Wim van Eck. 1983. "The Infl uence of Classification and Observation Errors on the Measurement of Income Inequality." *Econometrica* 51(4): 1093-1108.

Figure 1: Income Distribution, Expenditure Distribution, and Consistent Income
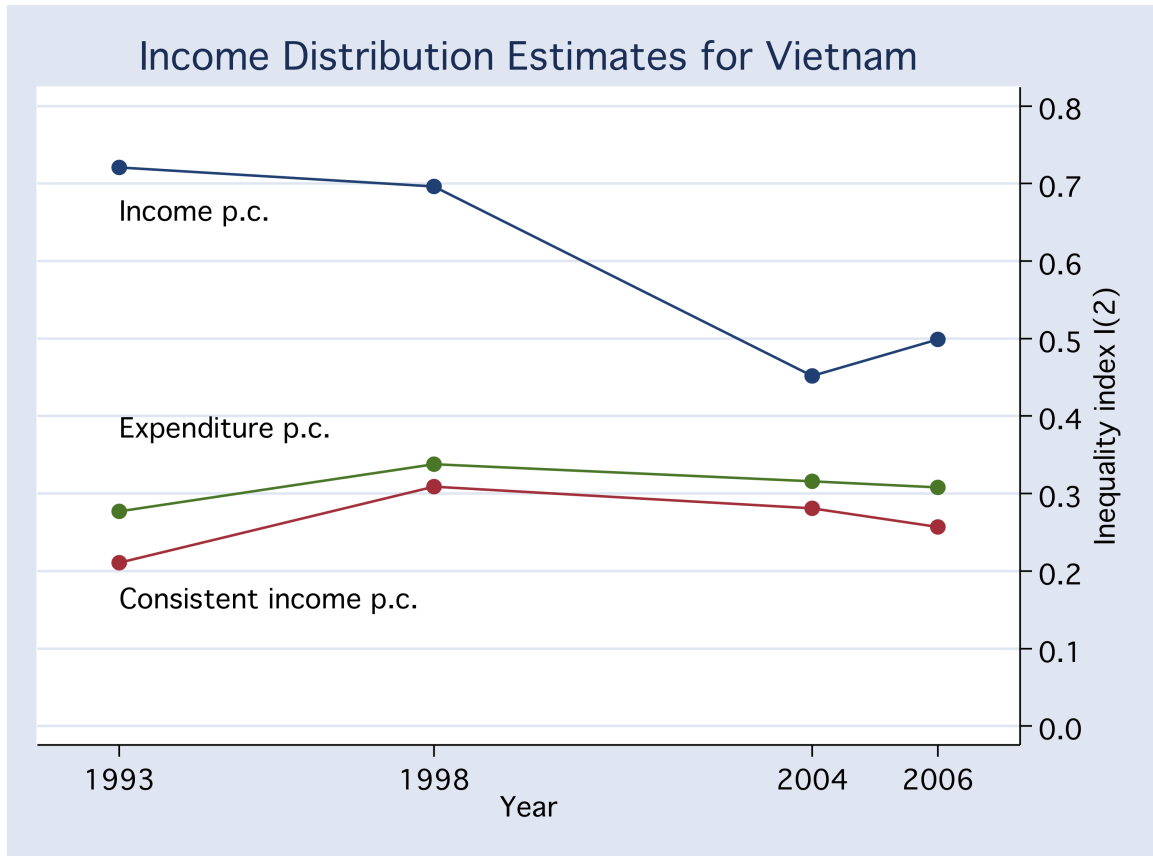Distribution, 1993-2006.

Simple estimates

Household enterprise

Other income

Wage income
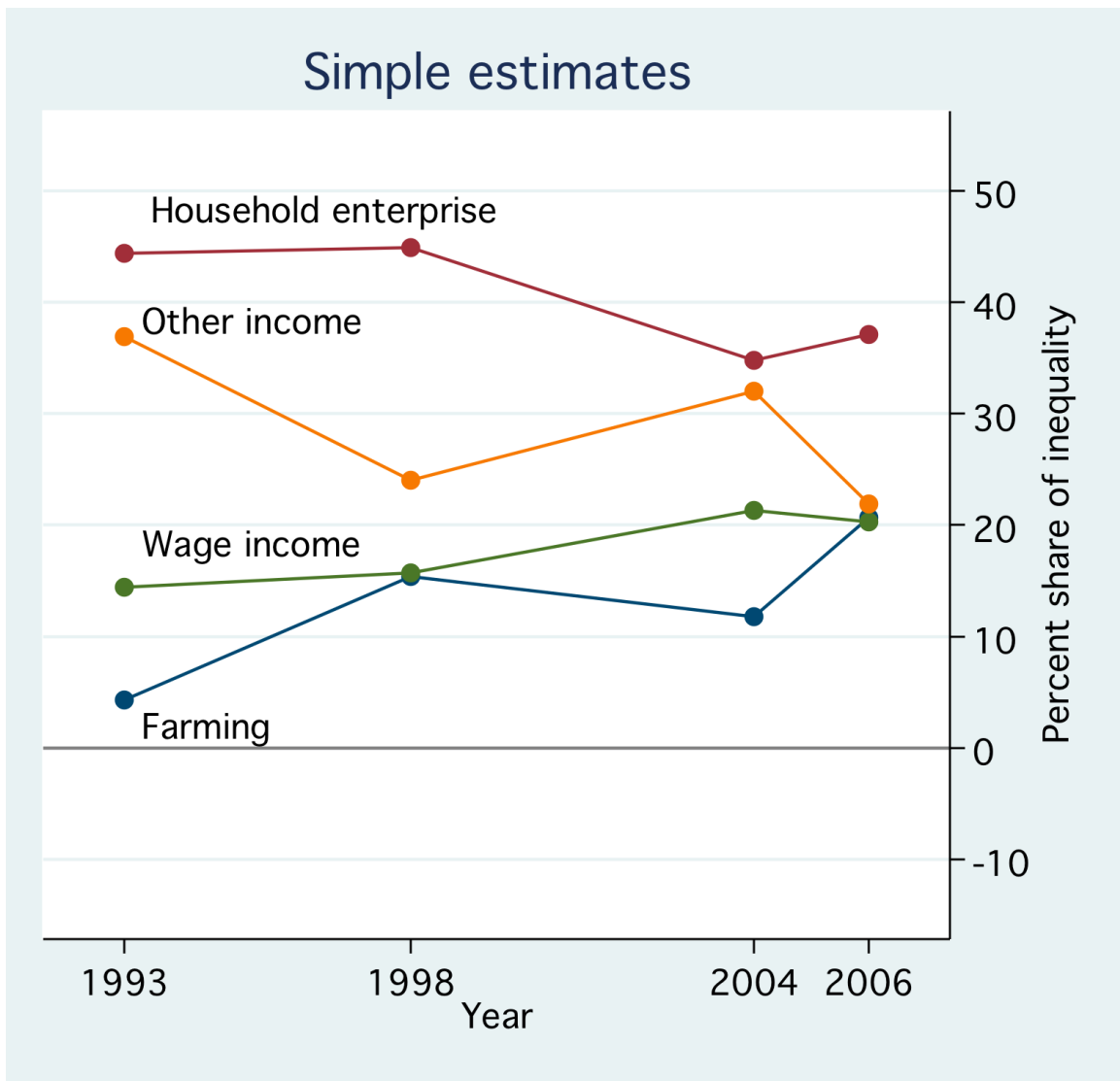
Farming

Year

Percent share of inequality

Figure 3: Inequality by Source of Income, Correcting for Measurement Error, 1993-2006.