Columbia University

Department of Economics
Discussion Paper Series

Assessing Household Vulnerability to Poverty from Cross-sectional Data: A Methodology and Estimates from Indonesia

Shubham Chaudhuri
Jyotsna Jalan
Asep Suryahadi

Discussion Paper #:0102-52

Department of Economics
Columbia University
New York, NY 10027

April 2002

Assessing household vulnerability to poverty from cross-sectional data: a methodology and estimates from Indonesia*

Shubham Chaudhuri
Columbia University, New York, U.S.A.
(E-mail: sc301@columbia.edu)

Jyotsna Jalan
Indian Statistical Institute, New Delhi, India
(E-mail: jjalan@isid.ac.in)

Asep Suryahadi
SMERU, Jakarta, Indonesia
(E-mail: suryahadi@smeru.or.id)

March 2002

Abstract

A household's observed poverty status is an ex-post measure of a household's well-being (or lack thereof). But for thinking about forward-looking anti-poverty interventions that aim to prevent rather than alleviate poverty, what really matters is the *vulnerability* of households to poverty, i.e., the ex-ante risk that a household will, if currently non-poor, fall below the poverty line, or if currently poor, will remain in poverty. Ideally, vulnerability at the household level would be estimated with panel data of sufficient length and richness. However, such data are rare, especially in poor, developing economies. We argue in this paper that despite the limitations of purely cross-sectional data, an analysis of these data can potentially be informative. We lay out a simple and fairly flexible methodology for empirically assessing household vulnerability to poverty using cross-sectional survey data, and demonstrate the uses and limitations of the proposed methods through a case study using data from the December 1998 mini-SUSENAS survey from Indonesia.

Keywords: poverty, vulnerability, household consumption
JEL codes: D10, I32, O53

*We thank Benu Bidani, Francois Bourguignon, Gaurav Datt, and Martin Ravallion for extremely helpful comments. We alone are responsible for all remaining errors. This research was funded by the World Bank. Corresponding author: Shubham Chaudhuri, Department of Economics, Columbia University MC-3308, New York, NY 10027. Telephone: (212) 854-4661. Fax: (212) 854-5765. E-mail: sc301@columbia.edu.

1. Introduction

Whether or not a household is poor is widely recognized as an important, albeit crude, indicator of a household's well-being. For more than a decade now, national poverty assessments have been used on a routine basis to inform policy discussions on poverty alleviation in numerous developing economies. These poverty assessments have drawn on cross-sectional household surveys to provide a detailed profile of the poor, and to document the incidence of poverty in various segments of the population.

However, today's poor may or may not be tomorrow's poor. Currently non-poor households who face a high probability of a large adverse shock, may, on experiencing the shock, become poor tomorrow. And the currently poor households may include some who are only transitorily poor as well as other who will continue to be poor (or poorer) in the future. In other words, a household's (or an individual's) observed poverty status—defined in most cases simply by whether or not the household's observed level of consumption expenditure is above or below a pre-selected poverty line—is an ex-post measure of a household's well-being (or lack thereof). But for many policy purposes, what really matters is the ex-ante risk that a household will, if currently non-poor, fall below the poverty line, or if currently poor, will remain in poverty. And the current poverty status of a household, may not necessarily be a good guide to the household's vulnerability to being poor in the future. For thinking about appropriate forward-looking anti-poverty interventions (i.e., interventions that aim to prevent or reduce future poverty rather than alleviate current poverty), the critical need then is to go beyond a cataloging of who is currently poor and who is not, to an assessment of households' vulnerability to poverty.

Vulnerability assessments are likely to differ from the usual poverty assessments on a couple of accounts. First, vulnerability assessments have to be, by definition, explicitly forward-looking. At any point in time, given the data, the vulnerability of households is unobservable to the policy-maker. In contrast, most poverty assessments are couched in atemporal terms and the policy maker, given the right data, does actually observe the current poverty status of the household. But an atemporal approach, if strictly adhered to, is of limited use in thinking about policy interventions that can only occur in the future. In practice, of course, poverty assessments are used in the process of policy formulation, and in doing so, implicit assumptions are being made about the extent to which the situation recorded in the data used to carry out the poverty assessment will be reproduced over time.

A second point of distinction, which follows naturally from the first, is in the treatment of the observed consumption expenditures at a point in time (i.e., from a single cross-section survey) as the outcome (snapshot) of a dynamic process that is occurring in real time. And this means that vulnerability assessments (again, in contrast to poverty assessments which remain largely atheoretical) have to be rooted in explicit models of inter-temporal household behavior.

It should be clear from even the limited discussion thus far that estimation of vulnerability at the household level should ideally be attempted with panel data of sufficient length and richness. However, such data are rare, especially in poor, developing economies. Instead, the best one can usually hope for are cross-sectional household surveys with detailed data on household characteristics, consumption expenditures and in some cases income. Is there then no hope of carrying out vulnerability assessments in these settings?

The primary aim of this paper is to argue that despite the obvious limitations of purely cross-sectional data, a detailed analysis of these data can potentially be informative about the future. The extent to which this exercise will be useful may well vary from setting to setting. But we should not rule out, a priori, that in some settings at least, vulnerability assessments using cross-sectional data may be worthwhile. A simple example demonstrates this point. Imagine that from a cross-sectional survey we observe the cross-sectional dispersion in consumption levels for two separate groups of households, each distinguished by an observable set of characteristics. For instance, group A might consists of rural households with male, elderly married heads of households with a primary school education; and group B might include urban households with male middle-aged married heads with no formal education. Suppose we ask ourselves, how much more likely is it that a household from group A will be poor at some point in the near future? If most of the observed cross-sectional variation in consumption levels across households stems from unobserved (to us) differences across households, say because of unobserved household-specific determinants of consumption levels that are persistent over time, then, clearly, we would not be able to answer this question with any degree of confidence. If, on the other hand, much of the variation can be attributed to the differences in the observable characteristics of households, then even a single cross-section can be quite helpful in answering the question posed above.

In this paper, starting with a definition of vulnerability at the household level as the probability that a household, regardless of whether it is poor today, will be consumption poor tomorrow, we provide a conceptual framework for thinking about the different dimensions of vulnerability to poverty, and then propose a simple method for empirically estimating household-level vulnerability using cross-sectional data. We demonstrate the uses and limitations of the proposed methods through a case study using household-level data from the December 1998 and August 1999 mini-SUSENAS surveys from Indonesia.

The framework we propose is extremely simple in that its empirical implementation entails (if at all) a very modest extension of the standard approach to poverty assessment. That should not be surprising. Poverty and vulnerability (to poverty) are two sides of the same coin. The observed poverty status of a household (defined simply by whether or not the household's observed level of consumption expenditure is above or below a pre-selected poverty line) is the ex-post realization of a state, the ex-ante probability of which can be taken to be the household's level of vulnerability. So if we are able to generate predicted probabilities of poverty for households with different sets of characteristics (which some but not all poverty assessments attempt), we will have, in effect, estimates of the vulnerability of these households. The main distinction between our method and existing approaches, lies therefore, not so much in the empirical strategy and econometric methods we outline, but in the conceptual re-orientation we propose.[1]

Given that we only have cross-sectional household budget survey at our disposal,

[1] Christiaensen and Boisvert (2000) propose an approach to vulnerability assessment that is very similar to the one we do, though there remain differences in orientation and implementation, a key one being their reliance on panel data. Somewhat different, though complementary approaches, again relying on panel data, are suggested by Pritchett et al. (2000) and Kamanou and Morduch (2001).

we face a challenge of not only trying to overcome the lack of the time dimension, but also that of having no information on the risks faced and the options available to the household to mitigate such risks. We thus have to make simplifying assumptions about how shocks evolve over the cross-sectional space. But at the same time we recognize that large common shocks such as economic crises cannot be captured by our method.

The question therefore is how well do our vulnerability estimates match up to the future poverty status of the households? Fortunately, 75% of the 10,000 households in the December 1998 mini-SUSENAS survey were reinterviewed in the August 1999 survey. This provides us with an opportunity to cross validate our vulnerability estimates. We use the December 1998 data to identify the vulnerable households and then use the August 1999 data on the same households to check how well our methods performs in identifying the future "poor" households. Our results indicate that our method predicts the future "poor" households quite accurately over various sub-populations of the data and does so better than other indicators such as a household's current poverty status or consumption level.

The paper is organized as follows. The following section proposes a definition of vulnerability and discusses a number of conceptual issues in applying that definition. Section 3 discusses the estimation strategy and statistical issues. Section 4 introduces the Indonesia data. Section 5 clarifies the uses and interpretation of vulnerability estimates. Results of the implementation for Indonesia are presented in Sections 6 through 9. Section 10 presents the results for some cross-validation exercises, and section 11 concludes.

2. Assessing vulnerability to poverty: a conceptual overview

We define vulnerability, within the framework of poverty eradication, as the ex-ante risk that a household will, if currently non-poor, fall below the poverty line, or if currently poor, will remain in poverty. Certainly this is not the only definition possible. In fact, in much of the recent work on the vulnerability of different segments within a population (see for instance, Glewwe and Hall (1998), Cunningham and Maloney (2000)), vulnerability is defined in terms of exposure to adverse *shocks* to welfare, rather than in terms of exposure to poverty.[2] The difference is substantive. Our definition would include among the vulnerable, households who are currently poor and have a high probability of remaining poor even if they do not experience any large adverse welfare shocks. On the other hand, our definition would exclude those households among the non-poor who face a high probability of a large adverse shock but are currently well-off enough so that even were they to experience the shock, they would still remain non-poor.[3]

[2] In a separate paper, Cunningham and Maloney(2000) take a step towards bridging this gap by considering exposure to adverse shocks, weighted by a household's initial position in the distribution of welfare.

[3] Even within the realm of poverty, our definition of vulnerability is restricted in one important respect. Poverty reflects deprivation on multiple fronts, and hence vulnerability to poverty ought also to be a multidimensional construct. However, given standard data constraints, to be able to empirically assess the extent to which various characteristics of households make them more or less vulnerable, the notion of vulnerability has to be made much more concrete. We therefore, limit ourselves to a focus on vulnerability to poverty defined in terms of a single measure, namely current consumption expenditure.

Formally, the vulnerability level of a household h at time t is defined as the probability that the household will find itself consumption poor at time $t+1$:

$$v_{ht} = \Pr(c_{h,t+1} \leq z) \tag{2.1}$$

where $c_{h,t+1}$ is the household's per-capita consumption level at time $t+1$ and z is the appropriate consumption poverty line. Note that the level of vulnerability at time t is defined in terms of the household's consumption prospects at time $t+1$.

The difference is noteworthy because it reflects an important distinction between the notion of vulnerability and the concept of poverty. Vulnerability is a forward-looking or ex-ante measure of a household's well-being, whereas poverty is an ex-post measure of a household's well-being (or lack thereof). This implies that while the poverty status of a household is concurrently observable-i.e., with the right data we can make statements about whether or not a household is *currently* poor-the level of vulnerability is not. We can *estimate* or *make inferences* about whether a household is *currently* vulnerable to *future* poverty, but we can never directly observe a household's current vulnerability level.

An assessment of vulnerability is, therefore, innately a more difficult task than assessing who is poor and who is not. To assess a household's vulnerability to poverty we need to make inferences about its future consumption prospects. And in order to do that, we need a framework for thinking explicitly about both the inter-temporal aspects and cross-sectional determinants of consumption patterns at the household level.

Over the last two decades, a large literature has developed which addresses precisely these issues (See Deaton(1992) and Browning & Lusardi(1995) for excellent overviews). This literature suggests that a household's consumption in any period will, in general, depend on a number of factors. Among them its wealth, its current income, its expectations of future income (i.e., lifetime prospects), the uncertainty it faces regarding its future income and its ability to smooth consumption in the face of various income shocks. Each of these will in turn depend on a variety of household characteristics, those that are observable and possibly some that are not, as well as a number of features of the aggregate environment (macroeconomic and socio-political) in which the household finds itself. At a general conceptual level, this suggests the following reduced form expression for consumption:

$$c_{ht} = c(X_h, \beta_t, \alpha_h, e_{ht}) \tag{2.2}$$

where X_h represents a bundle of observable household characteristics, β_t is a vector of parameters describing the state of the economy at time t, and α_h and e_{ht} represent, respectively, an unobserved time-invariant household-level effect, and any idiosyncratic factors (shocks) that contribute to differential welfare outcomes for households that are otherwise observationally equivalent.

Substituting from (2.2) into (2.1) we can rewrite the expression for the vulnerability level of a household as:

$$v_{ht} = \Pr\left(c_{h,t+1} = c(X_h, \beta_{t+1}, \alpha_h, e_{h,t+1}) \leq z \mid X_h, \beta_t, \alpha_h, e_{ht}\right) \tag{2.3}$$

The expression above makes clear that a household's vulnerability level derives from the stochastic properties of the inter-temporal consumption stream it faces, and these in turn depend on a number of household characteristics and characteristics of the environment in which it operates. And at a conceptual level, the expression is very general in a number of respects.

First, it allows for the possibility of complicated interactions between the multiple cross-sectional determinants of a household's vulnerability level. For instance, X_h could include variables such as the educational attainment of the head of the household, presence of a government poverty scheme in the community in which the household resides, as well as interactions between the two to capture potential inequities in the level of access to public programs.

Second, because a household's vulnerability is defined in terms of its future consumption prospects *conditional* on its current characteristics, both observed and unobserved, the possibility of poverty traps and other non-linear poverty dynamics is implicitly built in.

And third, the possible contribution of aggregate shocks and unanticipated structural changes in the macro-economy to vulnerability at the household level is also incorporated through inclusion of the time-varying set of parameters, β_t.

In practice, as will be clear in the next section, data constraints will usually not permit estimation of vulnerability at the level of generality embodied in expression (2.3). Nevertheless the formulation is useful in providing a basis for thinking through the possible implications of the various restrictions that will need to be imposed in any attempt to estimate vulnerability with the sorts of data that are usually available.

3. Empirical strategy and statistical issues

The probability that a household will find itself poor depends, not just on its expected (i.e., mean) consumption looking forward, but also on the volatility (i.e., variance, from an inter-temporal perspective) of its consumption stream, and possibly on higher moments of the consumption process as well. A salaried low-level government employee with an expected level of consumption roughly similar to that of a self-employed proprietor of a small business may nevertheless be much less vulnerable to poverty because of the relative stability of the former's consumption stream.[4]

To estimate a household's vulnerability to poverty we need therefore to, at a minimum, estimate both its expected consumption and the variance of its consumption. Ideally, this would be done using longitudinal data (where the same households are tracked over a number of periods) of sufficient length. With such data, one could directly estimate the inter-temporal variance of consumption at the household-level without the need for auxiliary assumptions. Longitudinal data are, however, rare. And even where longitudinal data are available, the cross-sectional coverage of these data tends to be very limited, reducing their usefulness for policy analyses that require nationally representative samples.[5]

Cross-sectional household surveys are much more widely available than are lon-

[4] Of course at times of macroeconomic crises accompanied by rapid inflation, the situations may easily be reversed.

[5] Pritchett et al. (2000) adopt a different approach in outlining how vulnerability to poverty may be quantified using *panel* data.

gitudinal surveys.[6] These cross-sectional surveys provide the raw data for most of the poverty assessments that are now routinely done for numerous developing economies. We therefore propose a method for estimating vulnerability to poverty that can be implemented using cross-sectional data.

Not surprisingly, to estimate vulnerability from a single cross-section, we have to make a number of fairly stringent assumptions regarding the stochastic process generating consumption. Essentially assumptions limiting the degree of unobservable heterogeneity in the future consumption prospects of households that are, at the time of the analysis, observationally identical along a number of dimensions. Chaudhuri(2000) provides a detailed description of the assumptions that are needed to interpret the estimates we obtain in terms of vulnerability to poverty.[7]

We begin by assuming that the stochastic process generating the consumption of a household h is given by:

$$\ln c_h = X_h \beta + e_h \tag{3.1}$$

where c_h is per capita consumption expenditure, X_h represents a bundle of observable household characteristics, characteristics such as household size, location, educational attainment of the household head, etc., β is a vector of parameters, and e_h is a mean-zero disturbance term that captures idiosyncratic factors (shocks) that contribute to different per capita consumption levels for households that are otherwise observationally equivalent.

Implicit in expression (3.1) is the assumption that the idiosyncratic shocks to consumption are identically and independently distributed *over time* for each household. This implies that we are ruling out unobservable sources of persistence (arising for example, from serially correlated shocks or unobserved household-specific effects) over time in the consumption level of an individual household. We also assume that the structure of the economy (captured by the vector β) is relatively stable over time, ruling out the possibility of aggregate shocks (i.e., unanticipated structural changes in the economy). That is, in assuming a fixed β over time, we are assuming that the uncertainty about future consumption stems solely from the uncertainty about the idiosyncratic shock, e_h, that the household will experience in the future. We are ignoring uncertainty about future consumption that arises from uncertainty about the future structure of the economy. However, as we note below, we do *not* assume that they are identically distributed *across households*.

Both these assumptions are forced upon us because we are attempting to estimate vulnerability from a *single* cross-section. Without longitudinal data we cannot identify the parameters driving persistence in individual consumption levels. And without a long enough time-series of repeated cross-sections, we cannot identify the stochastic process generating β.

[6]Examples include the SUSENAS (National Socioeconomic Survey) in Indonesia, the Family Income and Expenditure Survey (FIES) in the Philippines, the National Sample Survey (NSS) in India, and the series of Living Standards Measurement Surveys (LSMS) carried out in a number of developing economies in collaboration with the World Bank.

[7]We however, note two things here: first, the validity of these assumptions, and hence the usefulness of the vulnerability estimates generated by the method we propose is ultimately an empirical matter. And so, below, we devote a section to a number of cross-validation exercises. Second, the assumptions we make are certainly no stronger than, and in some cases, are in fact less stringent than those that are implicitly made when interpreting the findings of a typical poverty assessment.

We do however allow the variance of e_h (and hence of $\ln c_h$) to depend upon observable household characteristics in some parametric way. There are a number of ways in which this can be done. The estimates we report are generated assuming the following extremely simple functional form:

$$\sigma_{e,h}^2 = X_h \theta \tag{3.2}$$

We estimate β and θ using a three-step feasible generalized least squares (FGLS) procedure suggested by Amemiya(1977). A summary of the estimation procedure is provided in the statistical appendix, and details are available in Chaudhuri(2000).

Using the estimates $\widehat{\beta}$ and $\widehat{\theta}$ that we obtain we are able to directly estimate expected log consumption:

$$\widehat{E}\left[\ln c_h \mid X_h\right] = X_h \widehat{\beta} \tag{3.3}$$

and the variance of log consumption:

$$\widehat{V}\left[\ln c_h \mid X_h\right] = \widehat{\sigma}_{e,h}^2 = X_h \widehat{\theta} \tag{3.4}$$

for each household h. By assuming that consumption is log-normally distributed (i.e., that $\ln c_h$ is normally distributed), we are then able to use these estimates to form an estimate of the probability that a household with the characteristics, X_h, will be poor, i.e, of the household's vulnerability level. Letting $\Phi(.)$ denote the cumulative density of the standard normal, this estimated probability will be given by:

$$\widehat{v}_h = \widehat{\Pr}\left(\ln c_h < \ln z \mid X_h\right) = \Phi\left(\frac{\ln z - X_h \widehat{\beta}}{\sqrt{X_h \widehat{\theta}}}\right) \tag{3.5}$$

The method we have outlined is the standard one used in most poverty assessments that rely on regression methods, but with one important difference. In poverty assessments, the disturbance term is implicitly thought of as stemming from measurement error or some unobserved factor that is incidental to the main focus of the analysis. In most cases, therefore, rather than specify a separate equation such as (5), so that the variance of e_h is allowed to also depend upon the particular characteristics of the household, it is assumed that this variance is the same for all households. Thus, an estimate of β and a single common estimate of σ, the standard deviation of e_h (and hence, $\ln c_h$), are obtained from ordinary least squares (OLS) estimation of (3.1). With the same additional assumption that we make, which is that $\ln c_h$ is normally distributed, these estimates are used to derive the probability that a household with characteristics X_h will be poor.

There are two problems with the assumption that the variance of the disturbance term (and of log consumption) is the same for all households.

Within the framework we propose-in which the variance of the disturbance term is interpreted in *economic* terms as the *inter-temporal variance of log consumption*-the assumption that the variance of log consumption is the same for all households seems quite restrictive, regardless of its statistical import. That is because it forces the estimates of the mean and variance of consumption to be monotonically related across households, ruling out the possibility that a household with a lower mean consumption may nevertheless face greater consumption volatility than a household with a higher average level of consumption. Both formal and anecdotal evidence points to high levels of income and consumption volatility for poor households.

Moreover, in purely statistical terms, unlike in other settings where failure to account for heteroskedasticity results in a loss of efficiency but need not bias the estimates of the main parameters of interest, here, the standard deviation of the disturbance term enters directly (see (3.5) above). A biased estimate of this parameter will therefore lead to a biased estimate of the probability that a household is poor. Recognizing this point, some poverty analyses do explicitly model the variance of the disturbance term (see for instance, Elbers et al. (2001)), but this step is seen as just a necessary heteroskedasticity correction with little economic relevance beyond that.

4. Data

The data we use come from two sources. The main data on household characteristics and consumption expenditures come from the Mini-SUSENAS, which is a smaller version of the SUSENAS (National Socio-Economic Survey) that is the primary household expenditure survey in Indonesia.[8] We combine these with data from the 1996 "Village Potential" (PODES) Survey which provides a wide range of information on the characteristics of the villages/communities ("desa") in which these households reside.

The Mini-SUSENAS survey was first conducted in December 1998 and again in August 1999, using the same sample frame, and moreover, with about 75% of the original 10,000 or so households being surveyed on both occasions. The Mini-SUSENAS therefore provides a 2-period panel for roughly 7,500 households. In terms of cross-sectional sample size, this is considerably smaller than the main SUSENAS survey, which covers about 65,000 households in each round. We however chose to limit ourselves to an analysis of the data for the panel households, because we hope to carry out cross-validation exercises checking the usefulness of our cross-sectional vulnerability estimates. We therefore use only the data for December 1998 cross-section to generate the vulnerability estimates, and then bring in the August 1999 data for the same sample of households for the cross-validation exercises.

We estimate equations (3.1) and (3.2) separately for each of 13 geographical domains-the province of Jakarta (which is completely urban) and the rural and urban areas of six clusters of provinces that we define: Sumatra, West Java, Central Java and Yogyakarta, East Java and Bali, Kalimantan and Sulawesi, and the rest of Indonesia. A listing of the provinces that are included in each of these clusters is given in Table 1. We adopt this disaggregated estimation strategy because we wished to allow for some heterogeneity in the structural parameters underlying the consumption processes of households in these different regions. Given the differences in the structures of local economies of different regions, it is likely that key structural parameters-for instance, the returns to education or experience-may differ across regions.

To construct the poverty lines, which we need in order to generate our vulnerability estimates, we started with the set of regional poverty lines for February 1999 calculated by Pradhan et al. (2000). We then deflated these regional poverty lines to December 1998 and August 1999. We use as deflators, a set of re-weighted provincial

[8] Details about the Mini-SUSENAS survey, and the procedure used to construct the consumption aggregates that we use are available in BPS (2000).

CPIs (Pradhan et al. (2000)). The Indonesian CPI has a food share of 0.4, while the food share of the poverty lines is 0.8, reflecting the importance of food to the poor. So for each province we calculate a re-weighted CPI with a food share of 0.8. Another weakness of the CPI is that it is based solely on urban prices. Unfortunately, this weakness is carried over to the re-weighted CPI. We are forced to use the same deflator for urban and rural poverty lines within a province. This amounts to assuming the inflation rates in urban and rural areas in a province, during the periods of interest to us-i.e., December 1998 to February 1999, and February 1999 to August 1999-were the same. Table 1 displays the complete set of poverty lines that we obtained and use in our analysis.

The covariates we included in the regressions were: household size (level and its square), proportion of household members in the age-groups 6-12 years, 13-15 years, 16-18 years, proportion of adults in the household, whether the head of the household is single, married, divorced, age and age-squared of the head of household, and a series of dummies for whether the household head is illiterate, has attended primary school, attended junior-high school, attended senior high school, has some tertiary eduction; whether the head of the household is male, whether the household head is self-employed with no assistance, self-employed with some assistance from family and temporary workers, self-employed with permanent employees, and salaried workers in either the government or private sector.

5. Interpreting and using vulnerability estimates

Before we proceed to a discussion of the results, in this section, we make a few prefatory points regarding the uses and interpretation of vulnerability estimates. Mean vulnerability level within a group should, in the absence of aggregate shocks (and assuming the group is large enough), approximately equal the observed poverty rate for that group. To see this, consider the following stylized example. Suppose that 50% of a particular population has a vulnerability level of 0.40 (i.e., they face a 40% probability of being poor) while the remaining 50% has a vulnerability level of 0.10. The mean vulnerability level in this population is, therefore, 0.25. What fraction of the population would we expect to actually be poor at any given point in time? If the risks of poverty are independently distributed across individuals (i.e., there are no aggregate shocks) the low vulnerability group should contribute 5% (10% of 50%) while the high vulnerability group should add 20% (40% of 50%) for a total of 25% that should end up being poor.

This example illustrates that the observed poverty rate may represent one particular summary statistic (namely, the mean) of the underlying distribution of poverty.[9] And that in turn highlights the wealth of additional information that a vulnerability assessment (in which an attempt is made to estimate the entire distribution of

[9] Note that because we estimate vulnerability from a single cross-section where we only have data on cross-section variation in consumption levels, in purely statistical terms, our estimates indicate how various household characteristcs influence the concurrent probability of poverty). Of course, the basic premise underlying our approach is that the probabilities of current poverty map directly into probabilities of future poverty. But again, on a purely statistical basis, if our model fits the data well, the mean estimate of (what we are interpreting as vulnerability to future poverty) should come reasonably close to the current observed incidence of poverty, both in the aggregate and for various sub-samples.

vulnerability) can, in principle, provide in comparison with a poverty assessment, which, ultimately, focuses only on the mean.

In practice, as the literature on income distributions and inequality amply demonstrates, it will not always be feasibe to directly compare entire distributions of vulnerability. And so, in most instances, we will need to summarize the key properties of the underlying distribution through some well-chosen summary measures. The mean level of vulnerability is, of course, one. A second possibility, one that we use repeatedly below, is the fraction of the population that has a vulnerability level above some threshold, and can therefore be called vulnerable.

The choice of a vulnerability threshold is ultimately quite arbitrary. However, two thresholds stand out as possible focal points. The first, which we term the relative vulnerability threshold, is the observed current poverty rate in the population. The idea is that because the observed poverty rate represents the mean vulnerability level in the population, anyone whose vulnerability level lies above this threshold faces a risk of poverty that is greater than the average risk in the population and hence can legitimately be included among the vulnerable.

An alternative more stringent threshold is 0.50, which we term the high vulnerability threshold. A household whose vulnerability level exceeds 0.50 is more likely than not to end up poor and can be considered, therefore, to be highly vulnerable.

Secondly, two populations may have similar observed poverty rates but very different incidences of vulnerability. Consider two populations. In the first, call it A, 20% of the population has a vulnerability level of 1 whereas 80% has a vulnerability level of 0. In the other, call it B, 100% of the population has a vulnerability level of 0.20. In both populations, the observed poverty rate will be approximately 20%. But the fraction of the population that is vulnerable (with a relative vulnerability threshold) is dramatically different. Only 20% of population A is vulnerable, whereas with the same threshold the entire population of B is vulnerable. This dramatic difference has important implications for policy that we discuss in a later section.

The above example also illustrates that the ratio of the fraction of the population that is vulnerable (given a threshold) to the fraction that is poor-which Pritchett et al. (2000) term the vulnerability to poverty ratio-can provide an useful measure of how dispersed vulnerability is in the population. In general, for any given vulnerability threshold, a higher vulnerability to poverty ratio indicates a more dispersed ("egalitarian") distribution of vulnerability, whereas a lower ratio suggests that vulnerability is concentrated among a few. For population A, where vulnerability is limited to 20%, the vulnerability to poverty ratio is 1, while for population B where the entire population is vulnerable, the corresponding ratio is 5 (100% of the population is vulnerable, but only 20% is poor at any point in time).

6. Aggregate poverty and vulnerability in Indonesia

The estimated aggregate distribution of vulnerability for Indonesia is depicted in Figure 1, which plots the incidence of vulnerability at vulnerability thresholds ranging from 0 to 1, for the population as a whole as well as by observed poverty status. By construction, as the threshold increases, the incidence of vulnerability (the fraction of the population that has an estimated probability of being poor higher than the threshold) declines. Thus, at a threshold of zero, everyone is vulnerable while no one is vulnerable at the threshold of one. Perhaps not surprisingly, for any given threshold, the incidence of vulnerability is higher for the poor than for the population as a whole, which in turn is higher than the incidence of vulnerability amongst the nonpoor. More significantly, Figure 1 suggests that for a wide range of thresholds, poverty and vulnerability are significantly different from each other. Not all the poor are vulnerable while a significant proportion of the nonpoor are vulnerable.

Table 2 makes this notion more precise for our preferred vulnerability threshold, which is the observed incidence of poverty in the population. At the national level, while 23% of the population is observed to be poor, we estimate that 45% of the population is vulnerable to poverty. That is, have an estimated probability of experiencing poverty in the near future which is greater than the average risk of poverty (equal to the observed incidence of poverty) in the population. These estimates appear to support the often-stated (and vaguely defined) claim that the observed incidence of poverty underestimates the fraction of the population that is vulnerable to poverty.

Thus there may be some households whose ex-ante probability of poverty (vulnerability level) may be high who are nevertheless observed to be non-poor; and conversely, there may be some households who are observed to be poor, whose vulnerability level is, nevertheless, low enough for them to be classified as non-vulnerable. Of the 78% of the population that is observed to be non-poor, over 36% are nevertheless estimated to be vulnerable. This implies that 28% of the population, though not currently poor is vulnerable to poverty. Amongst the poor, 78% are estimated to be vulnerable. Of course that implies that we estimate that 22% of the observed poor is non-vulnerable, and while that may seem surprising, it simply reflects the stochastic nature of the relationship between poverty and vulnerability alluded to in the previous paragraph. And this is also apparent in the fact that of those we classify as non-vulnerable, 9% are nevertheless observed to be poor. (Table 2)

Amongst the vulnerable, we distinguish between the relatively vulnerable (i.e., those who have an estimated vulnerability level greater than the observed incidence of poverty but less than 0.5) and the highly vulnerable (i.e., have an estimated vulnerability level greater than 0.5). The relatively vulnerable constitute over four-fifths of the vulnerable and 37% of the overall population while the highly vulnerable make up 8% of the overall population.

Two main messages emerge from these aggregate numbers. First, the fraction of the population that faces a non-negligible risk of poverty (and hence, by definition, is taken to be vulnerable) is considerably higher than the fraction that is observed to be poor. And second, while poverty and vulnerability are closely related concepts, there remain important distinctions between the two and neither is a subset of the other. The characteristics of those who are observed to be poor at any given point in time may differ from the characteristics of those who are estimated to be vulnerable to poverty, whether or not they are currently poor. Interventions and programs that aim to reduce the level of vulnerability in the population may therefore need to be targeted differently from those aimed at poverty alleviation. We return to this point in later sections.

7. Poverty and vulnerability profiles

Table 3 presents the poverty and vulnerability profiles for Indonesia in December 1998. We report both the overall estimates for rural and urban Indonesia and also disaggregated by regions and certain select demographic and community characteristics. Table 3 provides us with some insights on average, about the geographical location of the vulnerable as well as their socio-economic characteristics.

We begin by detailing the spatial distribution of poverty and vulnerability. Poverty and vulnerability in Indonesia are largely rural phenomena. Relative to their share in the population, rural households are over-represented among the poor and the vulnerable. While 61% of Indonesia's population is rural, 80% of the observed poor live in rural areas as do 82% of those we estimate to be vulnerable. The highly vulnerable are even more disproportionately rural, with 91% of this group located in rural areas. The disproportionate contribution of rural households to overall poverty and vulnerability stems from the much higher incidence of poverty and vulnerability in rural areas. About 30% of the rural population is observed to be poor, whereas in urban areas, the observed poverty rate is 12%. Similarly, while we estimate that 20% of the urban population is vulnerable, 60% of the rural population

The imbalances in the contributions of rural and urban areas to overall poverty and vulnerability are reproduced at the regional level. Urban areas, regardless of region, are under-represented among the poor and the vulnerable, relative to their shares in the population. With the exception of rural Sumatra, rural areas tend to be over-represented. In absolute terms, rural areas of Java, Kalimantan and Sulawesi contribute the largest numbers to the populations of the poor and vulnerable. And of the 9% of the population that we estimate to be highly vulnerable, a fifth are found in rural areas of Kalimantan and Sulawesi and another 20% live in rural areas of West Java.

The tremendous variation in the poverty rates across the far-flung regions of Indonesia has been documented elsewhere (see Pradhan et. al (2000)). The fifth column of Table 3 confirms the presence of these regional disparities. The fraction of the population that is observed to be poor ranges from a low of 2% in Jakarta to a high of 56% in rural areas of West and East Nusa Tengarra, Papua and Maluku (which have collectively been labeled "Rest of Indonesia"). Except for Central Java and Yogyakarta, where 22% of the urban population is observed to be poor, urban areas have lower observed poverty rates than rural areas.

Inter-regional differences in the estimated incidence of vulnerability are even more pronounced than the regional disparities in poverty rates. The fraction of the population estimated to be vulnerable ranges from a low of 2% in Jakarta to a high of 77% in rural Central Java and Yogyakarta. Again, while urban areas generally have lower vulnerability rates, Central Java and Yogyakarta are exceptional in that 46% of the urban population in these two provinces is estimated to be vulnerable.

A comparison of the observed poverty rates and the estimated incidences of vulnerability across the 13 geographic domains we have defined reveals two points, both indicative of the ways in which the distribution of vulnerability can differ across regions.

First, in keeping with our findings at the national level, in each of the domains, the estimated incidence of vulnerability is at least as high and in most cases higher, than the observed incidence of poverty. However, there is considerable variation in the ratio of the fraction of the population that is vulnerable to the fraction that is poor. The vulnerability to poverty ratio is 1.00 in Jakarta and 1.27 in urban Sumatra indicating that vulnerability to poverty is quite concentrated in these two regions. In contrast, in several other regions, mostly rural, vulnerability to poverty is dispersed in the population, with the fraction that is vulnerable more than the double the fraction that is poor.

Second, two regions with roughly similar observed poverty rates may have very different incidences of vulnerability. For instance, in both East Java and Bali and what we term the "Rest of Indonesia", about 8% of the urban population is observed to be poor. However, we estimate that only 10% of the population of urban East Java and Bali is vulnerable, whereas in the "Rest of Indonesia," over 21% of the urban population is vulnerable.

Turning next to the other correlates of poverty and vulnerability, the one that stands out is the educational attainment of the household head. Of the 69% of the population that lives in households headed by individuals with at most a primary school education-who comprise 88% of the poor and an overwhelming 95% of the vulnerable-nearly 30% are poor while 63% are vulnerable to poverty.

Within this group, households headed by individuals with no schooling are particularly at risk-28% of the population in such households is estimated to be highly vulnerable.

In sharp contrast, within the populations in the two highest educational attainment categories, which together make up 21% of the overall population, the observed poverty rate is only 5%, the vulnerability rate is 2% and the fraction that is vulnerable is less than 1%. Even among households headed by individuals with at most junior schooling, the poverty rate, at 12%, is less than half that for households just one step down in the educational attainment hierarchy. The drop in the incidence of vulnerability to just 14% from 61% is even more striking.

If we divide up the sample according to the employment status of the household head we do not get such a clear trend though the incidence of vulnerability is understandably lower for salaried workers in the public and private sectors than it is for those in other employment categories. Somewhat surprisingly, the group with the highest rates of poverty and vulnerability is those who are self-employed with some help from family and hired workers. Of the 31% of the population belonging to this group, more than half are vulnerable.

When the population is split along other demographic characteristics, there is, surprisingly, hardly any difference in the poverty and vulnerability rates for different groups. So for instance, households with high dependency ratios are as likely to be poor and vulnerable as households with low dependency ratios, and households headed by females are as likely to be poor and vulnerable as male-headed households. Perhaps the only difference of note is the higher fraction of female headed households that is estimated to be highly vulnerable.

Community characteristics such as the availability of transport facilities, the presence of a bank or cooperative in the community, industrial activity and access to clean water are all associated with lower levels of vulnerability and poverty. Of these, access to clean water is associated with the sharpest drops in poverty and especially vulnerability.

8. Geography, poverty and vulnerability

Regional disparities in poverty are common in many developing economies and are especially pronounced in some such as Indonesia (as Table 3 showed). Geography has therefore often been the basis for poverty targeting. For instance, in China, counties that are classified as national poor or provincial poor (based on assessments of the extent of poverty) selectively receive additional government support. And in India, plan allocations to the states at least partially reflect the degree of need as captured by the level of poverty.

With the increasing number of fiscal decentralization initiatives, under which funds and expenditure authority are being devolved down to sub-national jurisdictions and local government institutions, a better understanding of the geographic aspects of poverty has become even more crucial. If the severity of poverty in a region is to be included in the criteria for determining the allocation of central funds, information on the geographic distribution of poverty is obviously essential. And in that respect, the recent development of poverty mapping techniques should prove extremely useful (see Hentschel et al. (1998)).

The methods we propose here complement these efforts in making possible (though perhaps not at the same level of geographic dis-aggregation) an assessment of the geographic distribution of vulnerability to poverty. But doing so raises the question of whether funds for poverty alleviation efforts should be allocated on the basis of the incidence of poverty or the incidence of vulnerability to poverty. If the rankings of sub-national units in terms of vulnerability and poverty coincided (or largely overlapped), the question could obviously be sidestepped. However, as the earlier comparison of poverty and vulnerability rates at the regional level indicated (see Table 3), this need not be the case.

This point is reinforced in Figure 2, where the estimated incidence of vulnerability (on the vertical axis) is plotted against the observed incidence of poverty (on the horizontal axis) for each of Indonesia's 26 provinces. The figure is meant mainly to be illustrative because the limited overall size of the Mini-SUSENAS implies that the samples at the provincial level are quite small for some of the smaller provinces. The provincial level findings for these provinces therefore need to be treated with caution, and to assist in this, we have used a smaller plotting symbol for those provinces where the samples are particularly small.

In keeping with our findings at the national level, for most provinces, the estimated incidence of vulnerability is higher (often considerably higher) than the observed incidence of poverty. This can be seen from the fact that most of the points lie above the 45-degree line. Even in some of the larger provinces such as East Java and Central Java, where sample size is not a concern, the ratio of the fraction of the population we estimate to be vulnerable to the fraction we observe to be poor is well above 2.

More noteworthy still is the substantial re-ranking that takes place when provinces are ordered in terms of the incidence of vulnerability rather than the observed incidence of poverty. Because the provinces are ordered along the horizontal axis in terms of increasing incidence of poverty, the re-ranking is reflected in the non-monotonicity of the scatterplot. Some of the re-ranking is undoubtedly due to the noise introduced by the small sample sizes, but the extent of re-ranking is striking nevertheless.

The re-rankings are particularly striking in the case of those provinces that appear in the upper left quadrangle defined by the vertical and horizontal lines indicating, respectively, the poverty and vulnerability rates at the national level. The poverty rate in these provinces is below the national rate, and so any poverty-targeting scheme based on poverty rates would allocate relatively fewer funds, on a per-capita basis for these provinces. However, in terms of the incidence of vulnerability to poverty, these provinces are above the national rate, and should, in principle, receive, on a per-capita basis, proportionally more funds for poverty programs.

The key to resolving this apparent dilemma lies in distinguishing ex-ante poverty prevention interventions from ex-post poverty alleviation interventions. An example drawn from public health makes this distinction clearer. Consider a situation where public health interventions are aimed at reducing the incidence of some disease. Suppose information is available on both the incidence of disease in different regions, as well as on the fraction of the population in different regions that is at high risk of contracting the disease. Funds for treatment of those already afflicted should clearly be directed to regions where the incidence of the disease is highest. But funds for preventive measures (such as vaccinations) ought to be directed to regions where the fraction of the population at risk is the largest. And the two sets of regions need not coincide. Regions with a higher incidence of the disease may also be regions where the risk of contracting the disease is concentrated among those afflicted. So the fraction of the population at risk may well be lower than in other regions where the incidence of the disease is lower.

The analogy with our treatment of vulnerability should be clear. The incidence of poverty, like the incidence of the disease, should determine the allocation of funds for treatment, which in the case of poverty means funds for ex-post poverty alleviation programs. The allocation of funds for preventive interventions-ex-ante interventions aimed at poverty prevention-should however be guided by the incidence of vulnerability to poverty. The funds for focused ex-post interventions such as food-for-work schemes or means-tested transfer programs are likely to be disbursed through very different channels than funds for ex-ante interventions. The latter will in general be much more varied in nature, and depending on the context may range from vocational training schemes, agricultural extension programs, social investment funds to major irrigation projects.

9. Sources of vulnerability

Consider Figure 3, which shows the simulated consumption streams (over a 50-period time horizon) for two different households.[10] The consumption streams of the two households look very different. Household A, on average, enjoys a much higher level of consumption, but its consumption is quite volatile. Household B, on the other hand, has a relatively stable inter-temporal consumption profile, but with much lower levels of consumption, on average. What is special about these two households is that despite the obvious differences in their mean levels of consumption and in the volatility of their consumption streams, the simulations have been constructed so that *their vulnerability levels are the same.*

[10] The simulations are based on actual estimates of mean consumption and consumption variance for two households in the December 1998 Mini-SUSENAS data set from Indonesia.

Figure 3 illustrates, rather starkly, the general point that households with similar levels of vulnerability may be vulnerable for very different reasons.[11] For some, vulnerability may stem primarily from low long-term consumption prospects (household B above). For others, consumption volatility may be the main source of vulnerability to poverty (household A above). From a policy perspective it will be important to distinguish between these two possibilities. For instance, vulnerability due to high volatility may call for ex-ante interventions that reduce the risks faced by households or insure them against such risks. On the other hand, to address vulnerability due to low endowments what might be needed are transfer programs. Clearly, a decomposition of the sources of vulnerability at the household level into the two components described above can help inform that choice.

At the same time it should be recognized that the two possibilities represent stylized extremes which are potentially interconnected in subtle ways. For instance, it may be that with inadequate risk management instruments at their disposal, households forego risky but, on average, high return earnings opportunities in favor of lower risk but lower return income streams. And in that case while the vulnerability of the household may appear to be due to low endowments, the true source of vulnerability may lie in an inability to adequately deal with risk.

What does the data tell us? Figure 4 indicates that in the Indonesian data they in fact do differ, and quite markedly. There, we have plotted our estimates of the mean and standard deviation of consumption for households with selected levels of vulnerability thereby constructing, empirically, a number of iso-vulnerability curves-i.e., combinations of mean consumption and standard deviation of consumption that imply the same level of vulnerability.[12]

Consider the cluster of points associated with a vulnerability level of 0.25, which is slightly above the threshold level of vulnerability (0.22) above which we consider households to be vulnerable. All the households represented in this cluster have estimated levels of vulnerability in the range 0.245 and 0.255. Yet the normalized mean consumption levels estimated for these households-the ratio of estimated mean consumption to the poverty line-ranges from a low of about 1.25 (with correspond-

[11] Conversely, two households with the same mean level of consumption may have very different levels of vulnerability if the degree of consumption volatility they are subject to, differs substantially.

[12] The shapes of the iso-vulnerability curves depicted in Figure 4 merit some elaboration. When mean consumption is above the poverty line, increases in variance, not surprisingly, increase the probability of poverty and hence, by definition, increase the level of vulnerability. Starting from a given level of mean consumption, an increase in the variance of consumption has, therefore, to be offset by an increase in mean consumption if the level of vulnerability is to remain the same. Hence, the upward slope of the iso-vulnerability curves to the right of the vertical line corresponding to the poverty line.

However, when mean consumption is below the poverty line, an increase in the variability of consumption (holding mean consumption fixed) may reduce the level of vulnerability because it increases the likelihood of consumption levels above the poverty line. To get a sense of how this might arise, consider the extreme case where a household's consumption is fixed at some level below the poverty line with absolutely no volatility whatsoever. Clearly such a household faces the certainty of poverty in the future implying a vulnerability level of 1. Were some variability in consumption to be now introduced, this household would have at least a small chance of realizing a consumption level above the poverty line, and this by definition would reduce its vulnerability level from 1. So, for a low enough initial level of consumption variability, an increase in variability would have to be offset by a reduction in mean consumption to maintain the same level of vulnerability. And this would imply that when mean consumption is below the poverty line segments of the iso-vulnerability curves would be negatively sloped.

ingly lower levels of normalized volatility) to a high of about 1.75. Within this group, therefore, some households are vulnerable because they have low levels of mean consumption whereas others are vulnerable because their consumptions are more volatile.

Figure 4 also illustrates another important point, which is the mean and standard deviation of consumption need not be monotonically related across households. For instance, amongst households with an estimated vulnerability level of 0.4, a household with highest estimated standard deviation of consumption, has both a higher estimated standard deviation of consumption as well as a lower estimated mean level of consumption than several of the households with lower estimated levels of vulnerability. This possibility for a household with a lower mean level of consumption to face greater consumption volatility is, as we noted earlier, not allowed in the methods used in most poverty assessments. The standard there is to implicitly force the estimated variance of consumption to always be higher for households with higher estimated mean consumptions. Figure 4 therefore highlights the importance of keeping the estimation strategy adequately flexible for the mean and variance of consumption to be separately estimated.

To facilitate the discussion of the sources of vulnerability we adopt a three-way classification of households. The first group are those with an estimated vulnerability level below the threshold level of 0.22, who, by construction, will have estimated levels of mean consumption well above the poverty line. These households are the non-vulnerable. The second group, whom we label the high volatility (HV) vulnerable, are those with an estimated vulnerability level above the threshold, but estimated mean consumption above the poverty line. These households are vulnerable because their consumptions are volatile; were we to eliminate the variability in their consumptions, these households, because their mean consumptions lie above the poverty line, would no longer be vulnerable to poverty. The third and final group consists of those households with mean levels of consumption below the poverty line. We term these the low-mean (LM) vulnerable. By construction, these households have vulnerability levels above 0.5, and their vulnerability stems primarily from their low levels of mean consumption in that a reduction in consumption variability would still leave them highly vulnerable to poverty (and may even increase their vulnerability).[13]

We estimate that 40% of the population is vulnerable due to the high volatility of their consumption while 5% is vulnerable because of low levels of mean consumption (Table 4). Thus, of the 45% of the population that is vulnerable, nearly nine-tenths

[13] There is a crude parallel between the classification we propose above and the more familiar distinction between the non-poor, the transient poor and the chronic poor. Loosely speaking, households who are HV-vulnerable are in a sense more likely to be only transitorily poor, whereas households who are LM-vulnerable are more likely to be chronically poor. But the parallel should not be taken too far because there are important distinctions between the two classification schemes. HV-vulnerable households may have very high levels of vulnerability and may therefore more often be poor than not. For instance, in Figure 4, almost all the households with estimated vulnerability levels close to 0.55 have estimated mean consumption levels above the poverty line and are hence HV-vulnerable. Should these households be included among the transient poor? Ultimately, the two taxonomies differ fundamentally because of the different questions they pose. The distinction between the transient poor and the chronic poor is based on the question: how often is the household poor? On the other hand the distinction between HV-vulnerable and LM-vulnerable households is based on the question: why is the household poor?

are so due to the high volatility of their consumption. Consumption volatility is also the main source of vulnerability for those currently poor. Of the 78% of the poor whom we estimate to be vulnerable, over four-fifths are vulnerable because their consumptions are volatile. Put another way, 64% of the poor would not be poor if ways could be found to stabilize their consumption streams, while maintaining their mean consumption levels.[14]

Figure 5 presents a decomposition of the sources of vulnerability for different segments of the population. The figure reinforces, from a slightly different perspective, the basic conclusion that the vulnerability of different groups stems from different sources. The starting point for the decomposition exercise was the choice of a hypothetical reference household whom we endowed with the median predicted level of mean consumption, the median predicted variance of consumption, and, it turned out, close to the median level of vulnerability. For each household, we then decomposed the difference between its vulnerability level and that of the reference household into two components: a portion due to the difference in the predicted means, and a portion due to the difference in the predicted variances. The average of these components, along with the overall average difference is plotted in Figure 5 for different segments of the population.

Two broad patterns stand out. Households in rural areas, regardless of region tend to have higher vulnerability levels than the reference household, and this difference is due to the lower predicted mean consumption levels in rural areas. In fact, relative to the reference household, rural households in all region face lower predicted levels of consumption volatility. For urban areas, the pattern is generally reversed, with the estimated vulnerability levels (generally lower than the reference household) attributable to consumption volatility.

A similar clear pattern emerges in differences in the sources of vulnerability across educational attainment categories. In both rural and urban areas, the component of vulnerability attributable to low predicted mean consumption levels is largest for the least educated and diminishes uniformly and markedly as the educational attainment of the household head goes up. The contribution of consumption volatility to vulnerability follows just the reverse pattern. However, even here, the differences in the sources of vulnerability for the rural population and the urban population are evident. For instance, though relative to urban households headed by highly educated individuals, urban households headed by individuals with no schooling are more vulnerable because of lower levels of mean consumption, in comparison with their counterparts in rural areas, these households' vulnerability stems more from consumption volatility.

10. Cross-validation exercises

How well do our cross-sectional vulnerability estimates identify who is likely to be poor in the future? To answer this question we need panel data where the same household is tracked for at least two periods. The mini-SUSENAS did track the same households at two points in time: in December 1998 and in August 1999. We

[14] This last qualifier is important because, even without any public intervention there might well have been ways in which these households could have reduced the volatility of their consumption streams. That they "chose" not to do so suggests that the cost incurred in terms of a reduction in mean consumption, in stabilizing consumption may have been too high.

use the August 1999 data, merged in with the December 1998 cross-section to carry out a number of cross-validation exercises aimed at evaluating the usefulness of our vulnerability estimates.

Table 5 displays the mean vulnerability levels (estimated from the December 1998 cross-section) for four groups of households classified by their poverty status in 1998 as well as 1999. Figure 6 plots the histograms of the underlying vulnerability estimates from which these means are calculated. The results indicate that the cross-sectional vulnerability estimates do a reasonably good job of identifying, those among the non-poor who are less vulnerable and are hence likely to remain non-poor, and those among the poor who are more vulnerable and are hence likely to remain poor. This can be seen from Table 5, where the mean vulnerability estimate for the group that is non-poor in both periods is considerably lower than the mean for the group that ends up poor in 1999, despite being non-poor in 1998. This difference is reflected in Figure 6 in the fact that the distribution of estimated vulnerability for those who are non-poor in both periods is visibly more left-skewed than that for the group that ends up poor in 1999. Similarly, the mean vulnerability for those who are poor in both 1998 and 1999 is substantially higher than the mean for those among the poor in 1998 who exit poverty between 1998 and 1999. Keep in mind that the estimates of mean vulnerability are generated exclusively from the December 1998 cross-section.

The predictive power of our vulnerability estimates is even more directly evident in Figure 7. This figure presents a comparison of the predicted poverty rate (i.e., mean estimated vulnerability level from the 1998 cross-section) and actual poverty rate in 1999 (from the 1999 cross-section) for each decile of the vulnerability distribution estimated using the December 1998 cross-section.[15] By and large, our vulnerability estimates reproduce the ordinal properties of the true distribution of vulnerability in the population. Thus, if in December 1998, had we used our vulnerability estimates to order the population into deciles in terms of the probability of ending up poor in 1999, the ordering of the poverty rates that was realized in 1999 would coincide with our ordering.

Figure 8 reinforces this conclusion in a slightly different way. There we have plotted estimated kernel densities of actual consumption levels in August 1999 for households in various intervals of the vulnerability distribution estimated from the December 1998 cross-section. As we move up the vulnerability distribution from groups of households with low levels of vulnerability to groups with higher levels of vulnerability, the distribution of future consumption for the group moves further and further to the left, indicating increasingly adverse consumption prospects and a higher likelihood of poverty.

Finally Table 6 reports a comparison of predicted and observed poverty rates for various segments of the population, grouped along dimensions other than estimated vulnerability. Except along the regional dimension, where there is clear evidence of different regions recovering from the crisis at different rates, our estimates do quite well in predicting the ordering of groups.

[15] To construct Figure 7 we ordered and grouped the households into deciles based on the vulnerability estimates generated using the December 1998 cross-section. For each decile, we calculated the estimated mean vulnerability (which is the predicted poverty rate for that ventile) and the actual observed poverty rate in the August 1999 cross-section (for the same households). Figure 7 presents the comparison of the two.

11. Conclusions

Poverty alleviation is widely acknowledged as the ultimate objective of development policy. And poverty assessments, a cataloging of who is poor, who is not, and the characteristics of those who are, have been the main analytic tools for structuring discussions of poverty policy. However, poverty is a stochastic phenomenon. The poor today may or may not be tomorrow's poor, and some of the non-poor today may well end up poor tomorrow. In this paper, we have argued that in thinking about appropriate forward-looking anti-poverty policy interventions, we need therefore, to look at, not just who is poor today, but also who is likely to be poor in the future. In other words, we need to identify who is vulnerable to poverty.

We define vulnerability at the household level, within the framework of poverty eradication, as the probability that a household, regardless of whether it is poor today, will be consumption poor tomorrow. We provide a conceptual framework for thinking about the different dimensions of vulnerability to poverty, and then propose a simple method for empirically estimating household-level vulnerability using cross-sectional data.

We implement the methods we propose using data from Indonesia. Three main conclusions emerge from this analysis. First, the fraction of the population that faces a non-negligible risk of poverty is considerably greater than the fraction that is observed to be poor. While 22% of the Indonesian population was observed to be poor in December 1998, we estimate that 45% of the population was vulnerable. Second, the distribution of vulnerability across different segments of the population can differ markedly from the distribution of poverty. We argue that this highlights the need for a distinction between poverty prevention programs-i.e., those aimed at reducing vulnerability-and poverty alleviation programs, and for differential targeting of the two. Third we find striking differences in the sources of vulnerability for different segments of the population. For rural households and for less-educated households, the main source of vulnerability appears to be low mean consumption prospects; for urban households and for more highly educated households, on the other hand, vulnerability to poverty stems primarily from consumption volatility. This too has important implications for the types of poverty prevention programs that are needed to address the vulnerability of different groups within the population.

References

[1] Amemiya, T. (1977) "The maximum likelihood estimator and the non-linear three stage least squares estimator in the general nonlinear simultaneous equation model," *Econometrica*, 45, 955-968.

[2] Browning, M. and A. Lusardi (1996) "Household Saving: Micro Theories and Micro Facts", *Journal of Economic Literature*, 34, 1797-1855.

[3] Chaudhuri, S. (2000) "Empirical Methods for Assessing Household Vulnerability to Poverty", *mimeo*, Department of Economics, Columbia University: New York.

[4] Christiaensen, Luc J. and Richard N. Boisvert (2000). "On measuring household food vulnerability: case evidence from northern Mali," *Cornell University Department of Agricultural and Resource Economics Working Paper* #2000-05, March.

[5] Cunningham, W. and W.F. Maloney (2000), "Measuring Vulnerability: Who Suffered in the 1995 Mexican Crisis?", *mimeo*, The World Bank: Washington DC.

[6] Deaton, A. (1992) *Understanding Consumption*, Oxford University Press: New York.

[7] Elbers, C., J. Lanjouw and P. Lanjouw (2001), "Welfare in Villages and Towns: Micro-level estimation of poverty and inequality," *mimeo*, The World Bank.

[8] Glewwe, P. and G. Hall (1998) "Are some groups more vulnerable to macroeconomic shocks than others? Hypothesis tests based on panel data from Peru" *Journal of Development Economics*, 56, 181-206.

[9] Hentschel, J., J. Lanjouw, P. Lanjouw, and J. Poggi (1998) "Combining Census and Survey Data to Study Spatial Dimensions of Poverty" *Policy Research Working Paper* # 1928, The World Bank: Washington DC.

[10] Jalan, J. and M. Ravallion (2000) "Transient poverty in post-reform rural China", *Journal of Comparative Economics*, 26, 337-358.

[11] Kamanou, Gisele and Jonathan Morduch (2001). "Measuring vulnerability to poverty," *mimeo*, Department of Economics, New York University.

[12] Pradhan, M., A. Suryahadi, S. Sumarto, and L. Pritchett (2000) "Measurements of Poverty in Indonesia, 1996, 1999, and Beyond" *Policy Research Working Paper* # 2438, The World Bank: Washington DC.

[13] Pritchett, Lant, Asep Suryahadi and Sudarno Sumarto (2000). "Quantifying vulnerability to poverty: a proposed measure with application to Indonesia," *mimeo*, Social Monitoring and Early Response Unit, Jakarta.

APPENDIX

We begin by assuming that the stochastic process generating the consumption of a household h is given by:

$$\ln c_h = X_h\beta + e_h \tag{A.1}$$

where c_h is per capita consumption expenditure, X_h represents a bundle of observable household characteristics, characteristics such as household size, location, educational attainment of the household head, etc., β is a vector of parameters, and e_h is a mean-zero disturbance term that captures idiosyncratic factors (shocks) that contribute to different per capita consumption levels for households that are otherwise observationally equivalent.

We assume that the variance of e_h is given by:

$$\sigma^2_{e,h} = X_h\theta \tag{A.2}$$

We estimate β and θ using a three-step feasible generalized least squares (FGLS) procedure suggested by Amemiya(1977).

First we estimate equation (11.1) using an ordinary least squares (OLS) procedure. We use the estimated residuals from equation (11.1) to estimate:

$$\widehat{e}^2_{OLS,h} = X_h\theta + \eta_h \tag{A.3}$$

using OLS. The predictions from this equation are used to to transform the equation as follows:

$$\frac{\widehat{e}^2_{OLS,h}}{X_h\widehat{\theta}_{OLS}} = \left(\frac{X_h}{X_h\widehat{\theta}_{OLS}}\right)\theta + \frac{\eta_h}{X_h\widehat{\theta}_{OLS}} \tag{A.4}$$

This transformed equation is estimated using OLS to obtain an asymptotically efficient FGLS estimate, $\widehat{\theta}_{FGLS}$. Note that $X_h\widehat{\theta}_{FGLS}$ is a consistent estimate of $\sigma^2_{e,h}$, the variance of the idiosyncratic component of household consumption.

The estimates:

$$\widehat{\sigma}_{e,h} = \sqrt{X_h\widehat{\theta}_{FGLS}} \tag{A.5}$$

are then used to transform equation (11.1) as follows:

$$\frac{\ln c_h}{\widehat{\sigma}_{e,h}} = \left(\frac{X_h}{\widehat{\sigma}_{e,h}}\right)\beta + \frac{e_h}{\widehat{\sigma}_{e,h}} \tag{A.6}$$

OLS estimation of equation (11.1) yields a consistent and asymptotically efficient estimate of β. The standard error of the estimated coefficient, $\widehat{\beta}_{FGLS}$, can be obtained by dividing the reported standard error by the standard error of the regression.

Using the estimates $\widehat{\beta}$ and $\widehat{\theta}$ that we obtain we are able to directly estimate expected log consumption:

$$\widehat{E}\left[\ln c_h \mid X_h\right] = X_h\widehat{\beta} \tag{A.7}$$

and the variance of log consumption:

$$\widehat{V}\left[\ln c_h \mid X_h\right] = \widehat{\sigma}^2_{e,h} = X_h\widehat{\theta} \tag{A.8}$$

for each household h. By assuming that consumption is log-normally distributed, we are then able to use these estimates to form an estimate of the probability that a household with the characteristics, X_h, will be poor, i.e, to estimate the household's vulnerability level. Letting $\Phi(.)$ denote the cumulative density of the standard normal, this estimated probability will be given by:

$$\widehat{v}_h = \widehat{\Pr}\left(\ln c_h < \ln z \mid X_h\right) = \Phi\left(\frac{\ln z - X_h\widehat{\beta}}{\sqrt{X_h\widehat{\theta}}}\right) \tag{A.9}$$

Two substantive issues arise in the implementation of the procedure outlined above, both having to do with the estimation of the variance of consumption. The first has to do with the possibility of measurement error in the observed data on consumption expenditures. Measurement error is a major concern in most consumption (and income) measures drawn from household surveys. The presence of such errors can lead to significant overestimates of the variance of log consumption from (11.1) and (11.1). Why? Because the mean of the squared residuals from (11.1) will be biased upwards by the variance of the measurement error and that bias will be transmitted to the estimate of the intercept in equations (11.1) and (11.1). And if that were the case, we would overestimate predicted mean consumption *levels* (which, given log-normality of consumption, is an increasing function of the variance of log consumption). To control for this, we make a multiplicative adjustment to the estimated variances such that the predicted mean consumption equals the actual mean consumption for each of the geographic domains for which we estimate a separate set of regressions.

This adjustment also corrects for overestimates of variance because of unobserved, but deterministic components of consumption. For instance, suppose two households look identical in terms of the observables *we include* in the consumption equation (11.1). Nevertheless, they have different consumption prospects because of some unobserved but deterministic factor-e.g. rural cultivating households who live in areas with more fertile soil may have better consumption prospects though they appear to be identical to households in areas with less fertile soil. This will bias upwards the mean of the squared residuals from (11.1).

A second, somewhat different complication stems from the possibility of *unobserved local shocks* that are common to households in particular areas. For instance, suppose a particular area is subject to a localized shock, which is reflected in the consumption data from that area. Households from that area will, depending on whether the shock was positive or negative, have higher or lower consumption levels than otherwise observationally equivalent households from other areas. If we include a set of area dummies in log consumption equation (11.1) to capture the effects of such localized common shocks and include the estimated dummies in estimating the mean of log consumption we would bias (either upwards or downwards) the latter estimate. If we instead include a set of area dummies in the variance-estimating equations, we risk overestimating the variance of log consumption for households in areas that experience large relative shocks. A reason for including area dummies would be to control for unobserved deterministic components of consumption. But since we address that issue through the adjustment we describe above, we chose ultimately not to include any area dummies in any of the regressions we estimated.

A third more minor issue is the fact that, given the simple linear specification we have adopted, there is no guarantee that the estimate of $\sigma^2_{e,h}$, $X_h\widehat{\theta}$, will be positive. In practice we did not find this to be a problem except for a few observations, so we simply dropped them from the sample. An alternative would have been to choose a different specification for the variance-estimating equation (11.1), such as a logistic specification (as in Elbers et al.(2001)). That would force the estimate to always be positive, though the estimate would then have to be constructed from a Taylor approximation.

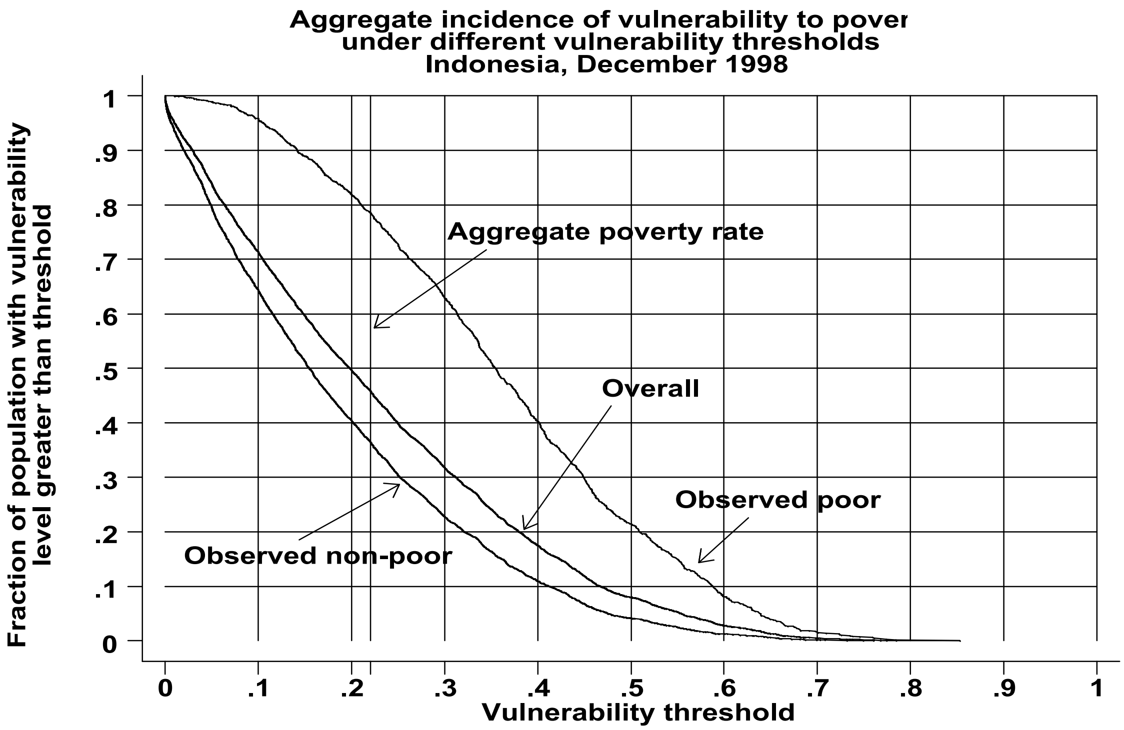

Figure 1

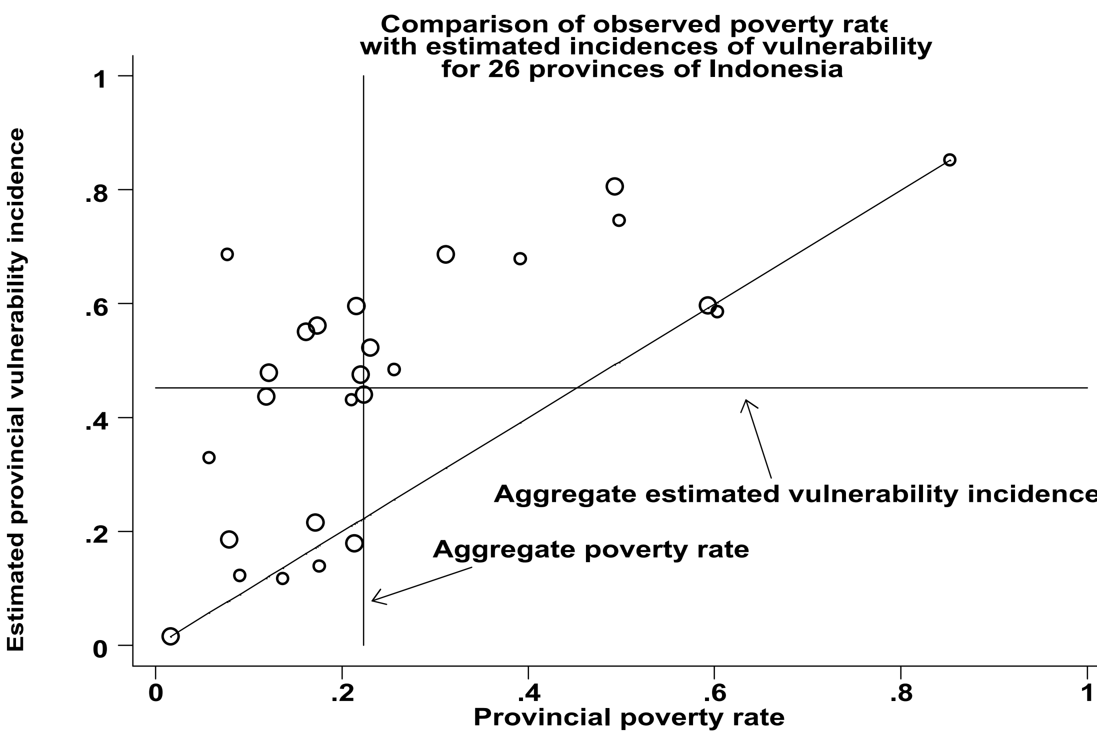

Figure 2

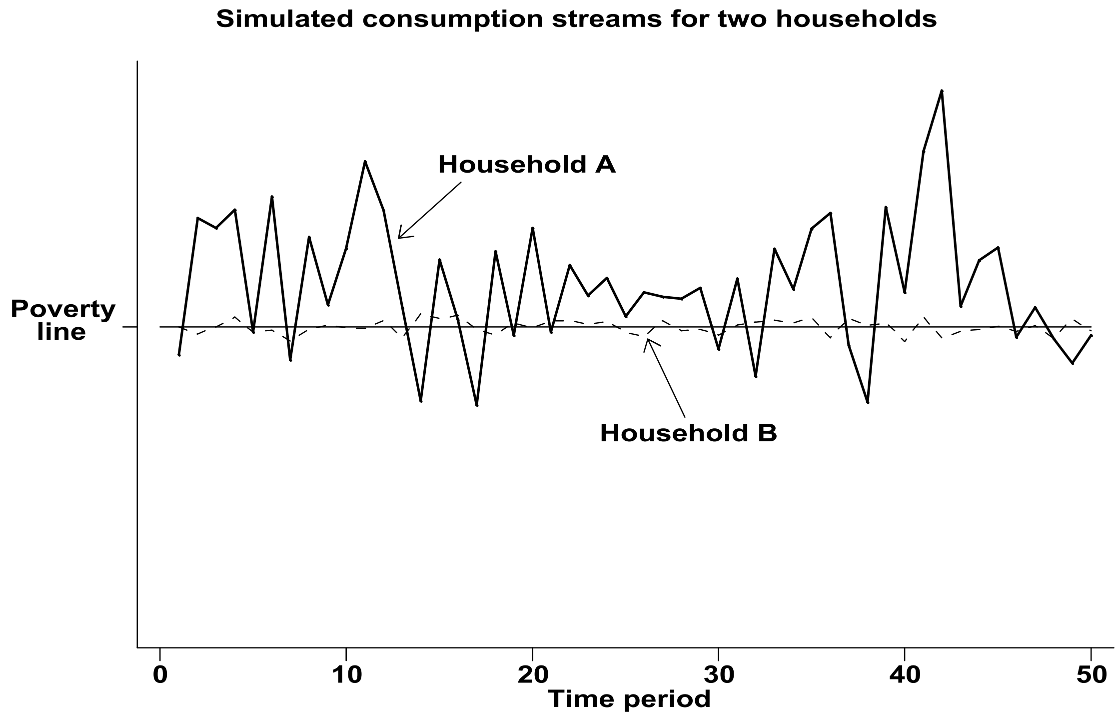

Figure 3

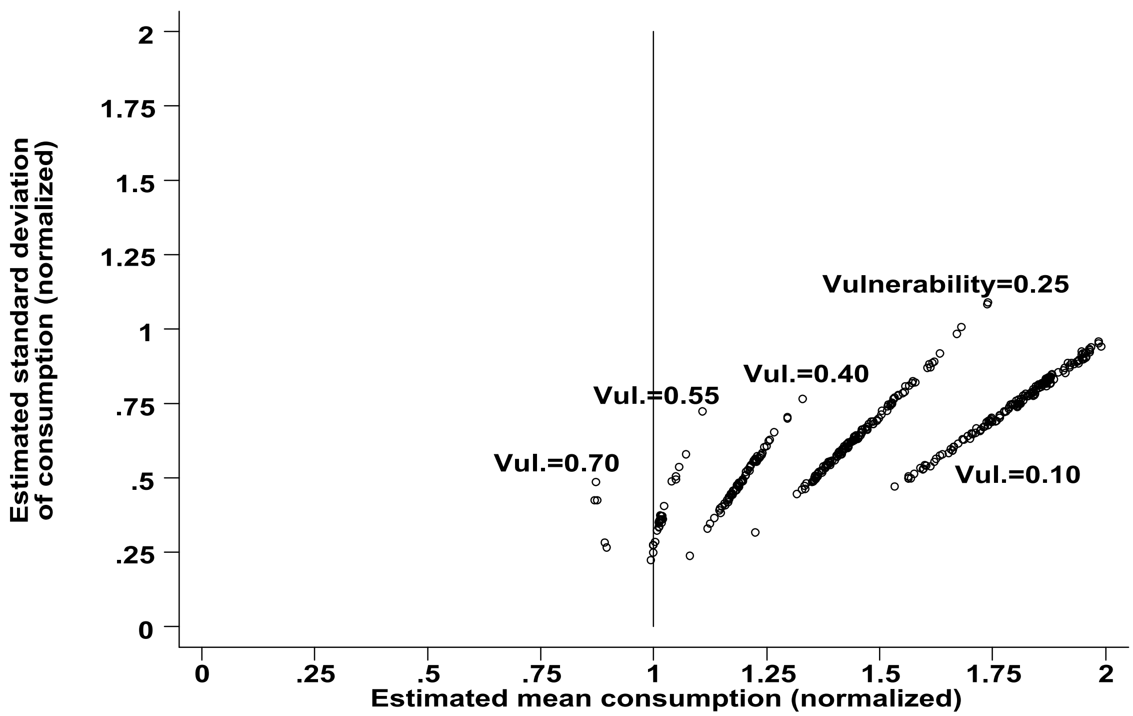

Figure 4

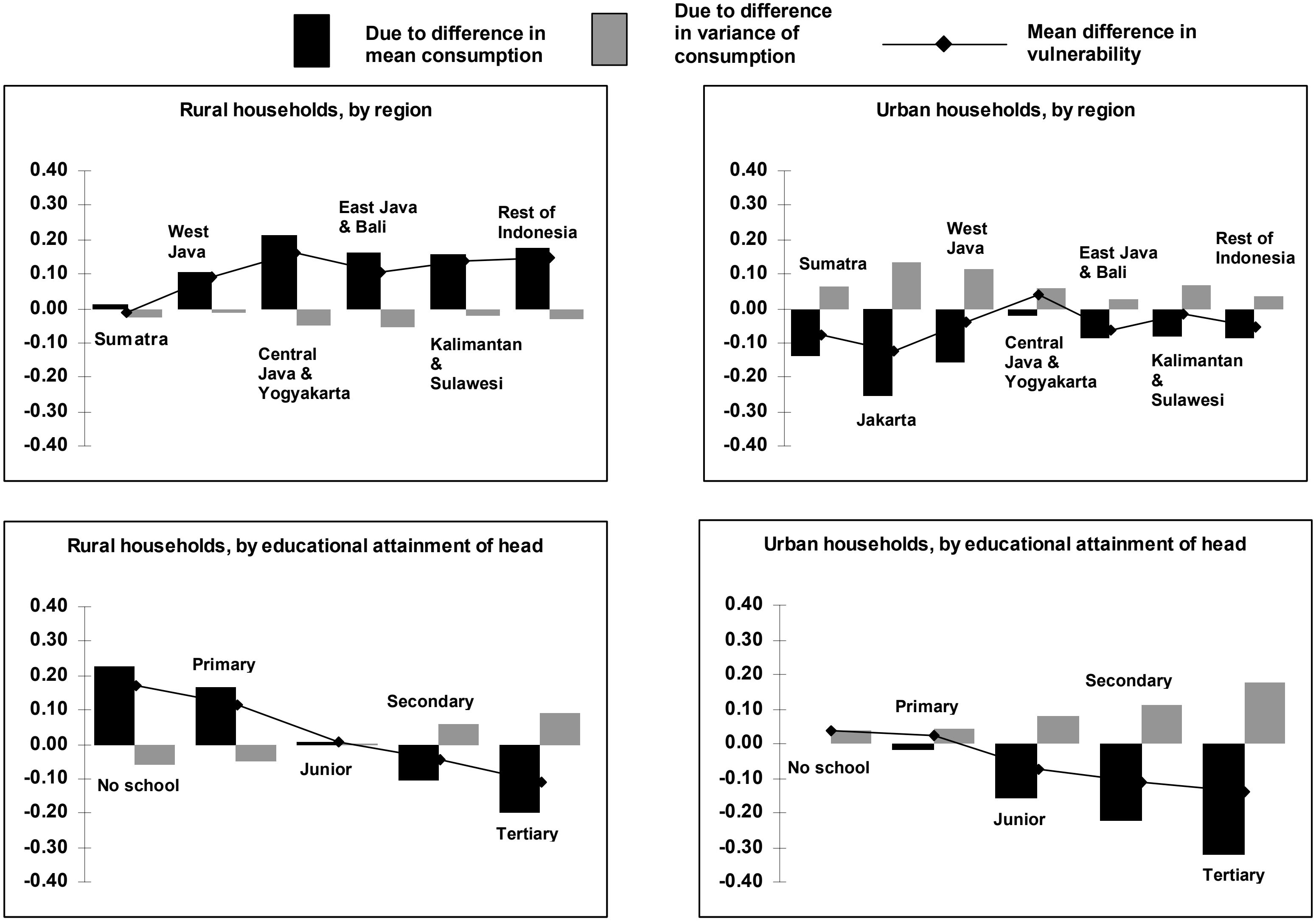

Figure 4
Decomposition of sources of vulnerability relative to reference level for different segments of the population

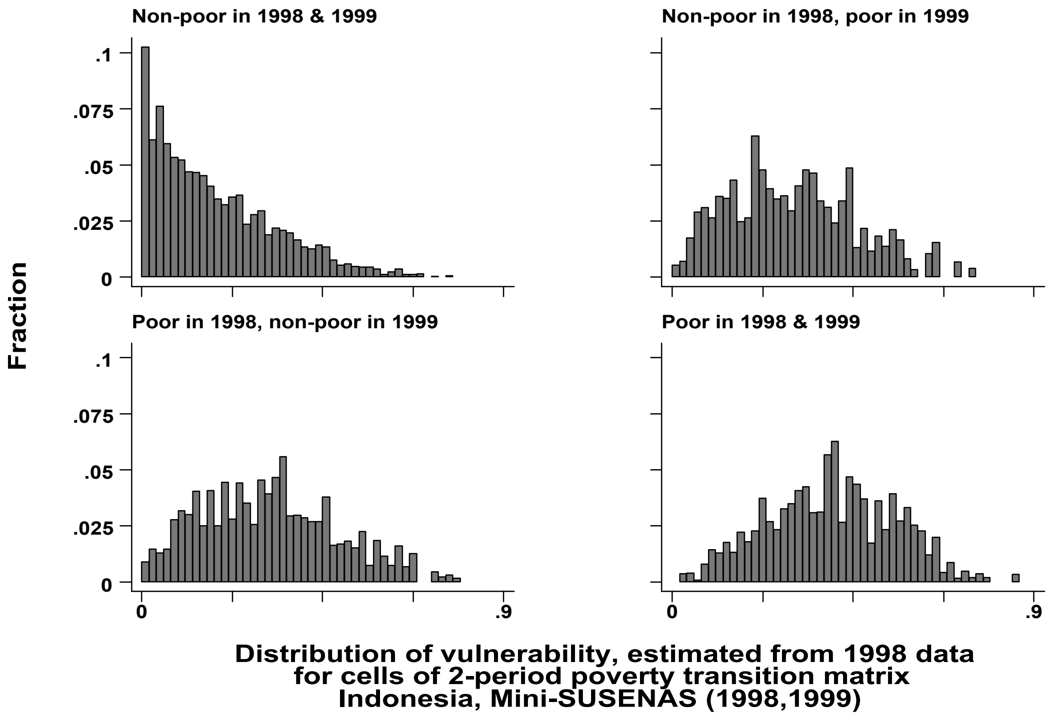

Figure 6

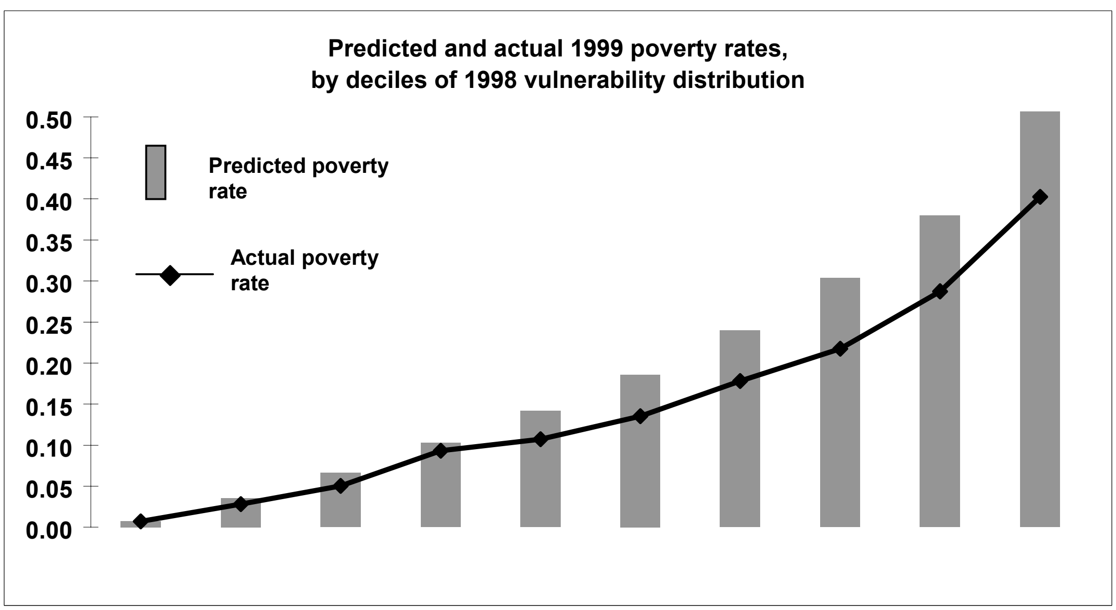

Figure 7

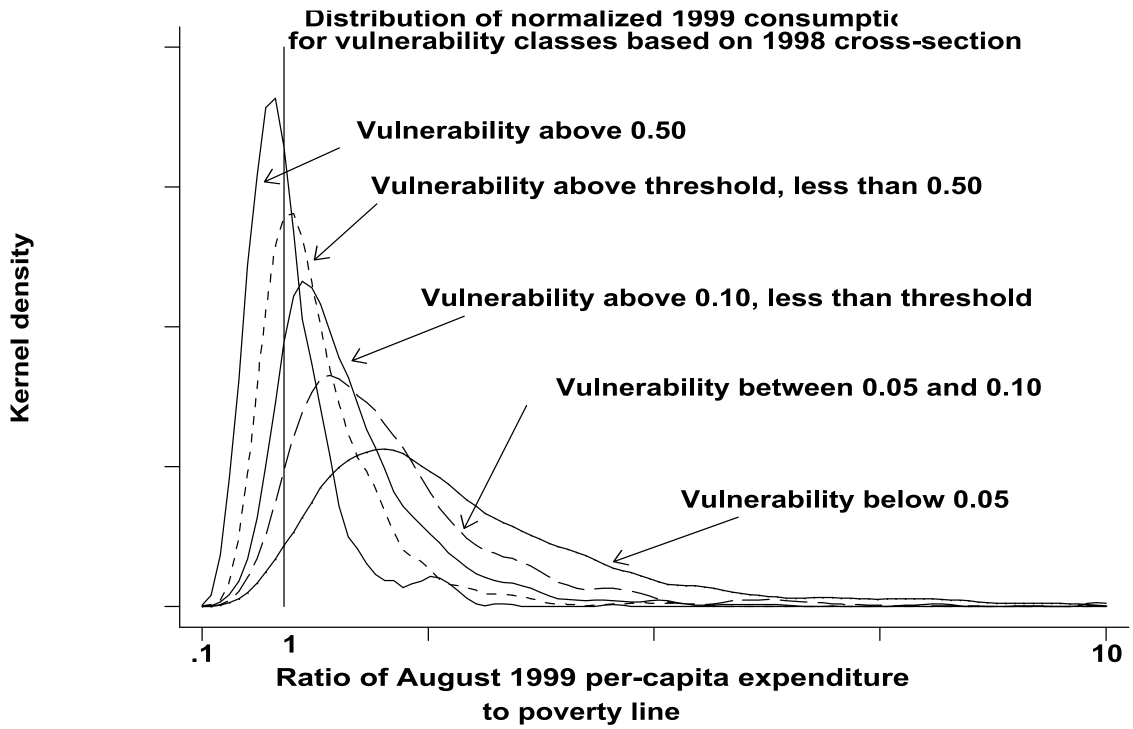
Distribution of normalized 1999 consumptic
for vulnerability classes based on 1998 cross-section
Kernel density
Vulnerability above 0.50
Vulnerability above threshold, less than 0.50
Vulnerability above 0.10, less than threshold
Vulnerability between 0.05 and 0.10
Vulnerability below 0.05
.1
1
10
Ratio of August 1999 per-capita expenditure
to poverty line

Figure 8

Table 1
Geographic domains and poverty lines (Rupiah per capita per month)

Cluster/Province	**Rural**	**Urban**	**Cluster/Province**	**Rural**	**Urban**
Sumatra			**East Java & Bali**		
Aceh	65426	69050	East Java	74256	78899
North Sumatra	71811	80493	Bali	87040	90164
West Sumatra	74091	80568	**Kalimantan & Sulawesi**		
Riau	81915	92510	West Kalimantan	83963	89114
Jambi	70535	78057	Central Kalimantan	81955	91461
South Sumatra	74140	79348	South Kalimantan	77642	81377
Bengkulu	73249	80821	East Kalimantan	86570	89450
Lampung	70034	79153	North Sulawesi	76743	81688
Jakarta			Central Sulawesi	73932	78215
Jakarta	…	96659	South Sulawesi	68200	77539
West Java			Southeast Sulawesi	76600	82660
West Java	81952	89936	**Rest of Indonesia**		
Central Java & Yogyakarta			West Nusa Tenggara	76714	79489
Central Java	73373	79497	East Nusa Tenggara	72262	78098
Yogyakarta	79289	88179	Maluku	85829	88081
			Papua	91002	82905

Table 2
Aggregate poverty and vulnerability profiles for Indonesia, December 1998

	Overall	**Amongst the non-poor**	**Amongst the poor**	**Amongst the non-vulnerable**	**Amongst the vulnerable**	**Amongst the relatively vulnerable**	**Amongst the highly vulnerable**
Mean per-capita expenditure (Rupiah/month)	138897	161061	65209	171797	97851	101693	82718
Fraction poor	0.22	0.00	1.00	0.09	0.40	0.34	0.63
Mean vulnerability	0.23	0.19	0.37	0.10	0.40	0.35	0.61
Fraction vulnerable	0.45	0.36	0.78	0.00	1.00	1.00	1.00
Fraction relatively vulnerable	0.37	0.32	0.57	0.00	0.80	1.00	0.00
Fraction highly vulnerable	0.08	0.04	0.21	0.00	0.20	0.00	1.00

Table 3
Poverty and vulnerability within different segments of the population, Indonesia, December 1998

	Population share	Share of poor	Share of vulnerable	Share of highly vulnerable	Fraction poor	Mean vulnerability	Fraction vulnerable	Vulnerability to poverty ratio	Fraction highly vulnerable
Overall					0.23	0.23	0.44	1.92	0.09
By location:									
Rural	0.61	0.80	0.82	0.91	0.30	0.30	0.60	1.99	0.13
Urban	0.39	0.20	0.18	0.09	0.12	0.13	0.20	1.66	0.02
Sumatra: urban	0.06	0.03	0.02	0.02	0.08	0.10	0.10	1.27	0.00
Jakarta: urban	0.05	0.00	0.00	0.00	0.02	0.03	0.02	1.00	0.00
West Java: urban	0.10	0.05	0.04	0.00	0.12	0.13	0.23	1.95	0.00
Central Java & Yogyakarta: urban	0.07	0.06	0.07	0.06	0.22	0.23	0.48	2.16	0.08
East Java & Bali: urban	0.06	0.02	0.01	0.00	0.08	0.11	0.12	1.52	0.00
Kalimantan & Sulawesi: urban	0.04	0.03	0.03	0.01	0.18	0.16	0.30	1.66	0.02
Rest of Indonesa: urban	0.01	0.00	0.00	0.00	0.08	0.13	0.21	2.59	0.00
Sumatra: rural	0.13	0.12	0.09	0.10	0.16	0.17	0.27	1.74	0.01
West Java: rural	0.12	0.16	0.16	0.20	0.31	0.30	0.62	1.98	0.16
Central Java & Yogyakarta: rural	0.12	0.17	0.20	0.17	0.34	0.35	0.78	2.30	0.14
East Java & Bali: rural	0.14	0.17	0.20	0.17	0.28	0.30	0.65	2.34	0.11
Kalimantan & Sulawesi: rural	0.09	0.12	0.13	0.20	0.31	0.34	0.69	2.22	0.21
Rest of Indonesa: rural	0.03	0.07	0.04	0.06	0.56	0.35	0.74	1.31	0.21
By education of household head									
No schooling	0.12	0.17	0.19	0.36	0.34	0.37	0.74	2.16	0.28
Primary	0.57	0.71	0.76	0.60	0.29	0.28	0.61	2.16	0.10
Junior	0.11	0.06	0.03	0.01	0.12	0.13	0.14	1.35	0.01
Secondary	0.16	0.05	0.02	0.02	0.07	0.08	0.03	0.55	0.01
More than secondary	0.05	0.01	0.00	0.01	0.01	0.03	0.00	0.00	0.00
By employment status of household head									
Unemployed/unpaid	0.13	0.11	0.12	0.12	0.19	0.22	0.43	2.23	0.08
Self-employed: no help	0.24	0.25	0.25	0.13	0.22	0.22	0.46	2.10	0.03
Self-employed: some help	0.31	0.37	0.39	0.43	0.27	0.28	0.57	2.11	0.12
Salaried (private & public)	0.32	0.27	0.24	0.31	0.19	0.19	0.33	1.78	0.08

Table 3 (continued)
Poverty and vulnerability within different segments of the population, Indonesia, December 1998

		Population share	Share of poor	Share of vulnerable	Share of highly vulnerable	Fraction poor	Mean vulnerability	Fraction vulnerable	Vulnerability to poverty ratio	Fraction highly vulnerable
By demographic categories										
Household head less than 60		0.86	0.86	0.85	0.83	0.22	0.22	0.45	2.00	0.08
Household head greater than 60		0.14	0.14	0.15	0.17	0.22	0.25	0.49	2.20	0.10
Female household head		0.08	0.08	0.09	0.13	0.22	0.24	0.46	2.07	0.13
Male household head		0.92	0.92	0.91	0.87	0.22	0.23	0.45	2.03	0.08
Household head not currently married		0.10	0.10	0.10	0.15	0.20	0.23	0.44	2.17	0.12
Married household head		0.90	0.90	0.90	0.85	0.22	0.23	0.45	2.02	0.08
Dependency ratio less than 0.25		0.79	0.81	0.79	0.81	0.23	0.23	0.45	1.99	0.08
Dependency ratio greater than 0.25		0.21	0.19	0.21	0.19	0.21	0.23	0.46	2.22	0.07
By community characteristics										
Transport facilities:	No	0.09	0.15	0.11	0.11	0.41	0.28	0.61	1.48	0.12
	Yes	0.91	0.85	0.89	0.89	0.21	0.22	0.44	2.13	0.08
Industry:	No	0.20	0.25	0.21	0.25	0.29	0.25	0.48	1.63	0.10
	Yes	0.80	0.75	0.79	0.75	0.22	0.23	0.44	2.02	0.07
Bank:	No	0.79	0.83	0.82	0.88	0.24	0.24	0.47	1.90	0.10
	Yes	0.21	0.17	0.18	0.12	0.18	0.20	0.37	2.04	0.05
Cooperative:	No	0.48	0.58	0.57	0.61	0.28	0.27	0.53	1.88	0.11
	Yes	0.52	0.42	0.43	0.39	0.18	0.20	0.37	1.99	0.07
Access to clean water	No	0.74	0.92	0.87	0.91	0.29	0.27	0.52	1.83	0.11
	Yes	0.26	0.08	0.13	0.09	0.07	0.14	0.22	3.01	0.03

Table 4
Sources of vulnerability for Indonesia, December 1998

	Overall	Amongst the non-poor	Amongst the poor	Amongst the non-vulnerable	Amongst the vulnerable	Amongst the high-volatility vulnerable	Amongst the low-mean vulnerable
Mean per-capita expenditure (Rupiah/month)	138897	161061	65209	171797	97851	100925	73168
Fraction poor	0.22	0.00	1.00	0.09	0.40	0.38	0.62
Mean vulnerability	0.23	0.19	0.38	0.10	0.40	0.37	0.60
Fraction vulnerable	0.45	0.36	0.78	0.00	1.00	1.00	1.00
Fraction high-volatility vulnerable	0.40	0.33	0.64	0.00	0.89	1.00	0.00
Fraction low-mean vulnerable	0.05	0.02	0.14	0.00	0.11	0.00	1.00

Table 5
Mean vulnerability level in 1998 by observed poverty status in 1998 and 1999

		Poverty status in 1999		
		Nonpoor	**Poor**	**All**
Poverty status in 1998	**Nonpoor**	.179	.299	.189
	Poor	.330	.398	.361
	All	.201	.359	.228

Note: Based on the Mini-SUSENAS panel of 7220 households

Table 6
Comparison of predicted and observed poverty rates for different segments of the population

	Rural		Urban	
	Predicted poverty rate	Observed poverty rate in 1999	Predicted poverty rate	Observed poverty rate in 1999
Overall	0.29	0.22	0.13	0.09
By geography:				
Sumatra	0.17	0.11	0.10	0.05
West Java	0.30	0.20	0.13	0.12
Central Java & Yogyakarta	0.35	0.24	0.23	0.11
East Java & Bali	0.30	0.24	0.11	0.08
Kalimantan & Sulawesi	0.34	0.28	0.16	0.14
Rest of Indonesia	0.35	0.51	0.13	0.16
Jakarta	-	-	0.03	0.00
	Predicted poverty rate		**Observed poverty rate for 1999**	
Overall	0.23		0.17	
By education of household head:				
No schooling	0.37		0.30	
Primary	0.28		0.21	
Junior	0.13		0.08	
Secondary	0.08		0.04	
More than secondary	0.03		0.01	
By employment status of household head:				
Unemployed/unpaid	0.22		0.13	
Self-employed: no help	0.22		0.18	
Self-employed: some help	0.28		0.22	
Salaried (private & public)	0.19		0.13	
By demographic characteristics of household head:				
Household head less than 60	0.22		0.17	
Household head greater than 60	0.25		0.15	
Female household head	0.24		0.18	
Male household head	0.23		0.17	
Household head not married	0.23		0.18	
Married household head	0.23		0.17	
Dependency ratio less than 0.25	0.23		0.17	
Dependency ratio more than 0.25	0.23		0.18	
By community characteristics:				
Transport facilities: No	0.28		0.28	
Yes	0.22		0.16	
Industry: No	0.24		0.21	
Yes	0.22		0.16	
Bank: No	0.23		0.18	
Yes	0.20		0.12	
Cooperatives: No	0.27		0.22	
Yes	0.19		0.12	
Access to clean water: No	0.26		0.21	
Yes	0.14		0.06	