Ibero-Amerika Institut für Wirtschaftsforschung
Instituto Ibero-Americano de Investigaciones Económicas
Ibero-America Institute for Economic Research
(IAI)

Georg-August-Universität Göttingen
(founded in 1737)

Diskussionsbeiträge · Documentos de Trabajo · Discussion Papers

Nr. 154

Estimating Vulnerability to Covariate and Idiosyncratic Shocks

Kenneth Harttgen, Isabel Günther

November 2006

Platz der Göttinger Sieben 3 · 37073 Goettingen · Germany · Phone: +49-(0)551-398172 · Fax: +49-(0)551-398173
e-mail: uwia@gwdg.de · http://www.iai.wiwi.uni-goettingen.de

Estimating Vulnerability to Covariate and Idiosyncratic Shocks

Isabel Günther* and Kenneth Harttgen

November 14, 2006

Abstract

Households in developing countries are frequently hit by severe idiosyncratic and covariate shocks resulting in high consumption volatility. A household's currently observed poverty status might therefore not be a good indicator of the household's general poverty risk, or in other words its vulnerability to poverty. Although several measurements to analyze vulnerability to poverty have recently been proposed, empirical studies are still rare as the data requirements for these measurements are often not met by the surveys that are available for developing countries. In this paper, we propose a simple method to empirically assess the impact of idiosyncratic and covariate shocks on households' vulnerability, which can be used in a wide context as it relies on commonly available living standard measurement surveys. We apply our approach to data from Madagascar and show, that whereas covariate and idiosyncratic shocks have both a substantial impact on rural households' vulnerability, urban households' vulnerability is largely determined by idiosyncratic shocks.

***JEL* Classification:** I32, D60.

Key words: Vulnerability, idiosyncratic and covariate shocks, multilevel modelling.

*University of Göttingen, Department of Economics, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany, email: isabel.guenther@wiwi.uni-goettingen.de and k.harttgen@wiwi.uni-goettingen.de.

The authors would like to thank Stephan Klasen, Michael Grimm and participants of the Verein for Socialpolitik Annual Meeting and of the Poverty Reduction, Equality and Growth Network Annual Workshop for helpful comments and discussions. Any errors remain our responsibility.

1 Introduction

Households in developing countries are frequently hit by severe idiosyncratic shocks (i.e. household-level shocks, such as death, injury or unemployment) and covariate shocks (i.e. community shocks, such as natural disasters or epidemics), resulting in high income volatility. Although households in risky environments have developed various sophisticated risk-coping strategies to reduce income fluctuations or to insure consumption against these income fluctuations, variance in household consumption remains generally high (see e.g. Townsend, 1994; Udry, 1995). A household's currently observed poverty status is, therefore, in many cases not a very good guide to the household's vulnerability to poverty, i.e. its general poverty risk. Whereas some households might be trapped into chronic poverty, others might only temporarily be poor, whereas other households currently non-poor might still face a high risk to fall into poverty in the future.

Most established poverty measurements, e.g. the FGT poverty measures (Foster et. al, 1984), do, however, only assess the current poverty status of a household without taking into account dynamic consumption fluctuations. Results from these static poverty analysis might therefore be misleading if high consumption volatility persists in a country. Not only might poverty rates fluctuate from one year to another, but even if aggregate poverty rates are constant over time, the share of the population which is vulnerable to poverty might be much higher. Moreover, these poverty measures cannot assess whether high poverty rates are a cause of structural poverty (i.e. low endowments) or a cause of poverty risk (i.e. high uninsured income fluctuations), which is important to know from a policy perspective.

To overcome these shortcoming of traditional poverty assessments, which can only present a static and ex-post picture of households' welfare, vulnerability analysis estimates the ex-ante welfare of households, taking into account the dynamic dimension of poverty. Vulnerability assessments, therefore, try to estimate ex-ante both the expected mean as well as volatility of consumption, with the latter being determined by idiosyncratic and covariate shocks.

Although there has recently been a growing theoretical literature on vulnerability measurement, relevant empirical studies on vulnerability are - largely due to data limitations - still rare. First, to appropriately examine the dynamic aspects of poverty, lengthy panel data would be needed. But for many developing countries, lengthy panel data does not exist and panels with only two or three waves of data or cross-sectional surveys are the only data available. Second, most household surveys were not designed to provide a full accounting of the impact of shocks. Information on idiosyncratic and covariate shocks is, therefore, in most data sets either completely missing or very limited. Hence, existing empirical studies have so far either only examined the aggregate vulnerability of households, ignoring the causes of the observed vulnerability, or have only studied the impact of selected idiosyncratic or covariate selected shocks on households' consumption, leaving out an analysis of the relative importance of different shocks on households' vulnerability. In addition, concentrating on selected shocks might lead to biased and inefficient estimates of the impact of these shocks on households' vulnerability.

The objective of this paper is to assess the relative impact of idiosyncratic and covariate shocks on households' vulnerability to poverty. More precisely, we both analyze how much of households' vulnerability is structural and risk induced, as well as provide an estimate of the share of consumption volatility that is idiosyncratic and covariate respectively. We propose a simple method which can be applied to commonly available standard household surveys without being constrained by the usual data limitations for vulnerability analysis; i.e. the method allows to estimate the impact of idiosyncratic and covariate shocks on households' vulnerability without lengthy panel data and information on a wide range of shocks. The suggested approach is an integration of multilevel analysis (Goldstein 1999) into the widely applied method by Chaudhuri (2002) to estimate vulnerability from cross-sectional or short panel data.

The remaining paper is structured as follows. Section 2 briefly discusses the current theoretical and empirical literature on vulnerability to poverty, including its shortcomings. Section 3 proposes a methodology that allows assessing the relative importance of idiosyncratic and covariate shocks for households' vulnerability with short panel data or cross-sectional data. Section 4 presents an empirical application to Madagascar. Section 5 concludes.

2 Theory and Empirics of Vulnerability

As discussed in the introduction a household's currently observed poverty status might not be a reliable guide to a household's longer-term wellbeing. Policy makers and researchers in development economics have, therefore, long emphasized that it is critical to go beyond a static ex-post assessment of who is currently poor to a dynamic ex-ante assessment of who is vulnerable to poverty. But although there has been an emerging literature on both the theory and empirics of vulnerability, its significance especially for policy makers is still rather low.

The current state of the theoretical literature on vulnerability can be described in the words of Hoddinott and Quisumbing (2003) as a 'let a hundred flowers bloom' phase of research with numerous definitions and measures and seemingly no consensus on how to estimate vulnerability. Several competing measurements have been offered (for an overview see e.g. Hoddinott and Quisumbing, 2003) and the literature has not yet settled on a preferred definition or measure. However, in principal three main definitions have emerged in the literature.

Combining the literature on imperfect insurance with an assessment of prospective risks, the first proposes to measure vulnerability as uninsured exposure to risks, or in other words, the ability of households to insure consumption against income fluctuations (e.g. Glewwe and Hall, 1998). The second approach defines vulnerability as expected poverty or in other words as the probability that an individual's future consumption will lie below a pre-defined poverty line (e.g. Chaudhuri, 2002; Pritchett, Suryahadi and Sumarto, 2000). The third definition associates vulnerability with low expected utility (Ligon and Schechter, 2003). Based on the micro-economic theory that the utility of risk-averse individuals falls if the volatility of consumption rises, vulnerability is measured with reference to the utility derived from some level of certain-equivalent-consumption, i.e. the level of constant consumption that would yield the same utility as the observed volatile consumption. Last, using an axiomatic approach, Calvo and Dercon (2005) have lately combined the latter two measurements and define vulnerability as 1 minus the expected value of the ratio of households' consumption to the poverty line with an exponent between 0 and 1 to reflect risk aversion.

But independent of the applied definition of vulnerability, vulnerability measures are always a function of the expected mean and variance of households' consumption, where the mean of expected consumption is determined by household and community characteristics whereas the variance in household consumption is determined by the frequency and severity of idiosyncratic and covariate shocks as well as the strength of households' coping mechanisms to insure consumption against these shocks.

For a comprehensive understanding of vulnerability to poverty it is, therefore, important to know both the magnitude of consumption volatility (i.e. the level of vulnerability) as well as the causes of volatility in consumption (i.e. the sources of vulnerability). Currently available data does, however, not even allow for a thorough estimation of either the ex-ante vulnerability of households or the ex-post impact of shocks on consumption, let alone measure both the level and sources of vulnerability at the same time. The existing empirical literature is hence divided into two strands of literature; either concentrating on the measurement of aggregate vulnerability within a population or analyzing the ex-post impact of selected shocks on households' consumption.

The first strand of literature, which intends to estimate the aggregate vulnerability of households, has been pioneered by Townsend (1994) and Udry (1995), who were some of the first using panel data to analyze whether households are able to insure their consumption against idiosyncratic income fluctuations over space and time. In this spirit, several studies followed analyzing consumption fluctuations over time (e.g. Dercon and Krishnan, 2000; Jalan and Ravallion, 1999; Morduch, 2005), concluding that households are partly but not fully capable of insuring consumption against income fluctuations. A severe drawback of this literature is that it relies on rather lengthy panel data, which is very limited for developing countries. The existing studies and drawn conclusions are hence often based on very few rounds (often not more than 2 waves) or observations (often not more than 100 households) of rural panel data, where urban households are mostly ignored (see also Morduch, 2005). A major confounding factor is here also the problem of measurement error as it is quite difficult to distinguish real consumption changes from measurement error in these relatively short panels (see e.g. Luttmer, 2001; Woolard and Klasen, 2005). However, in many developing countries even short panel data is completely missing and one has to rely on cross-section surveys to estimate vulnerability.

The second strand of empirical literature on vulnerability, which estimates the impact of selected shocks on households' consumption, has also large (mostly) data-driven limitations. Information on idiosyncratic and covariate shocks is in most households surveys very limited and sometimes even completely missing (see also Günther and Harttgen, 2005). As a consequence, most authors have only been able to focus on the impact of selected shocks on consumption (see e.g. Gertler and Gruber, 2002; Glewwe and Hall, 1998; Kochar, 1995; Paxon, 1992; Woolard and Klasen, 2005). Concentrating on certain shocks does however not allow for an analysis of the relative impact of various shocks on households' consumption to assess which shocks should be given first priority in anti-poverty programs. Moreover, these studies have rarely been able to analyze the impact of these shocks on the vulnerability of households, as households' vulnerability to shocks is not only a function of the impact of shocks on households' consumption but also of the frequency distribution of these shocks.

In addition, there are severe econometric problems related to this work, which usually rely on standard regression analysis to study the impact of shocks on households' consumption. First, focusing on certain shocks introduces a considerable omitted variable bias as various shocks are often highly correlated (Mills et al, 2003; Tesliuc and Lindert, 2004; see also Table A3 in the Appendix). The impact of selected shocks on households' consumption is therefore likely to be overestimated. Second, it is often assumed that the impact of shocks on consumption is the same across all households, which is a rather strong assumption to make. Third, the problem of endogeneity might be severe as households' welfare has presumably also an impact on the occurrence of certain shocks, e.g. poor households normally face higher mortality risks.

Most important, several studies, which have analyzed the impact of covariate community shocks might be biased or miss information by a disregard of the hierarchical data structure underlying these estimates (Goldstein, 1997, 1999). We speak of hierarchical data structure or multilevel data whenever variables, i.e. economic indicators, are collected at different hierarchical levels with lower hierarchical levels (e.g. households) nested within higher hierarchical levels (e.g. communities).

If for example covariate community shocks are simply assigned to each household within a community, blowing up data values from a small number of communities to many more household observations, the assumption of independent observations is violated, leading to estimates that might actually be statistically insignificant (Hox, 2002; Steenbergen and Jones, 2002). A related problem of dependent individual observations, leading to biased standard errors that are too small, also occurs in surveys with cluster sampling. Several methods have been proposed to correct the estimated standard errors in clustered survey design (Deaton, 1997) and in principle these correction procedures could also be applied to hierarchical data structure.

However, first most of the proposed procedures assume intraclass correlations between observations within clusters that are equal for all variables, which is usually not the case for variables of different hierarchical levels (Hox, 2002). Second, multilevel models do not only take account of dependencies between individual observations but also explicitly analyze dependencies at each level and across levels as well as the amount of variation at each level (Bryk and Raudenbush, 1992)[1].

We certainly cannot bridge the data gaps that exist with regard to missing panel data and missing data on shocks in developing countries. What we propose is an estimation method, which allows to study the relative impact of idiosyncratic and covariate shocks on households' vulnerability, without lengthy panel data and without facing the discussed econometric problems that usually occur when estimating the impact of certain shocks on household consumption. Furthermore, we estimate the level and sources of vulnerability simultaneously, which has rarely been done. Although we cannot distinguish between the impact of individual shocks, a disaggregation of the impact of covariate community versus idiosyncratic household specific shocks should already be an interesting undertaking.

[1]See Section 3.2 for a detailed discussion

Since covariate community shocks are correlated across households, microeconomic theory states that mutual insurance mechanism within communities should break down during covariate shocks. On the other hand, mutual insurance across communities, which would mitigate the problem of correlated shocks across households, are hypothesized to break down because of information asymmetries and enforcement problematics. On the contrary, it is claimed that households should be able to insure consumption against idiosyncratic shocks, as they are uncorrelated across households even within communities, where information asymmetries are less severe than across communities. Hence, analyzing the relative impact of covariate and idiosyncratic shocks on households' consumption can first of all test the previous stated hypothesis.

In addition, an assessment of the relative importance of idiosyncratic and covariate shocks might help policy makers to set up insurance priorities. Although higher information and enforcement problems prevail for insurance across communities, shocks that occur on the community level are easier to observe and also easier to target with national safety nets as they are geographically clustered.

Few studies (e.g. Carter, 1997; Dercon and Krishnan, 2000) have attempted to estimate the relative importance of covariate and idiosyncratic shocks on households' consumption. Their estimations generally show, that covariate shocks have a larger and more significant impact on households' consumption than idiosyncratic shocks. However, these studies often only analyzed rural households, relied on panel data, which is rarely available for developing countries and also faced the discussed econometric problems of concentrating on some selected idiosyncratic and covariate shocks, without taking into account the hierarchical data structure. In addition, it is often difficult to define ex-ante idiosyncratic and covariate shocks, as certain shocks often do have both a covariate and idiosyncratic component. Hence we think that our approach could contribute to a better understanding of the relative impact of idiosyncratic and covariate risks on households' vulnerability.

3 Methodology

3.1 Mean and Variance in Consumption

Our proposed method is an extension of the methodology proposed by Chaudhuri (2002) to estimate expected mean and variance in consumption using cross-sectional data or short panel data.[2] As for most developing countries lengthy panel data is not available this method has recently become quite popular. The main hypothesis is that the error term in a cross-sectional consumption regression, or in other words the unexplained part of households' consumption, captures the impact of idiosyncratic and community specific covariate shocks, and that this cross-sectional variance also reflects inter-temporal variance in consumption. It is furthermore assumed that this variance in consumption can be explained by household and community characteristics, i.e. that the impact of shocks on consumption fluctuations is correlated with observable variables.

Suppose that a household's h consumption in period t is determined by a set of variables X_h. We can hence set up the equation

$$lnc_h = X_h\beta + e_h \tag{1}$$

where lnc_h is the log of per capita household consumption, X_h a set of household as well as community characteristics, and e_h the part of a household's consumption that cannot be explained. Chaudhuri (2002) suggest that this error term, or the variance in consumption of otherwise equal households, captures the impact of both idiosyncratic and community specific covariate shocks on households' consumption and that this variance is correlated with

[2] Here we only present the estimation procedure for cross-sectional data. For a discussion of implementing the proposed method using panel data with two periods of data see Chaudhuri (2002) or Ligon and Schechter (2004).

observable household characteristics. In other words, whereas standard ordinary least squares (OLS) regression techniques assume homoscedasticity, i.e. the same variance $V(e_i) = \sigma^2$ across all households i, Chaudhuri (2002) assumes that the variance of the error term is not equal across households, i.e. heteroscedastic, reflecting the impact of shocks on consumption.

To estimate the consumption volatility of households, i.e. the impact of shocks on households' consumption, in a second step, the variance of the error term is therefore regressed on the same and/or other household and community characteristics than lnc_h:

$$\sigma^2_{eh} = X_h\theta. \tag{2}$$

If we assume heteroscedasticity, using OLS for an estimation of β and θ would lead to unbiased but inefficient coefficients. To overcome this problem equation 1 has to be reduced to a model where the residuals e_h have a homogeneous variance (for a detailed discussion see Maddala, 1977). Chaudhuri (2002) hence applies a three-step feasible generalized least squares (FGLS) method to estimate efficient coefficients β and θ. For a detailed discussion of the methodology see Appendix, Chaudhuri (2002) and Chaudhuri et al. (2002).

In a third step, for each household the expected mean (equation 3) as well as variance (equation 4) of consumption can be estimated using the consistent and asymptotically efficient estimators $\hat{\beta}_{FGLS}$ and $\hat{\theta}_{FGLS}$.

$$\hat{E}[lnc_h|X_h] = X_h\hat{\beta} \tag{3}$$

$$\hat{V}[lnc_h|X_h] = \hat{\sigma}^2_{eh} = X_h\hat{\theta}. \tag{4}$$

Obviously, in the absence of any time-variant information on consumption, two rather strong assumptions have to be made when using cross-sectional surveys to estimate consumption variance. First, it is assumed that cross-sectional variance can be used to estimate inter-temporal variance in consumption. Certainly, cross-sectional variance can explain part of inter-temporal variance due to idiosyncratic or covariate community-specific shocks. However, the model will miss the impact of inter-temporal shocks on the national level (for example terms of trade shocks).

Second, the existence of measurement error, when using information on consumption from household survey data, remains a major concern for the estimation of the mean and variance of consumption. If measurement error exists, this can lead to a significant overestimation of the variance in consumption, i.e. an overestimation of the impact of idiosyncratic and covariate shocks on households consumption.[3] Hence, it has to be assumed that measurement error in consumption is rather low.

However, the proposed method has the great advantage that it overcomes both the problem of missing lengthy panel data as well as incomplete information on shocks, which might often lead to biased results with regard to the impact of shocks on households' consumption.

In addition, Chaudhuri (2003) demonstrates the robustness of the proposed methodology comparing the predicted poverty rates from cross-sectional estimates of a two-wave panel with the actual incidence of poverty in the second period of the two-wave panel. Also, conducting Monte Carlo experiments Ligon and Schechter (2004) show that the proposed approach of Chaudhuri (2002) is a good estimator of households' mean and variance in consumption whenever expenditure is measured without error and whenever a two-wave panel is at hand.[4]

We expand the proposed method by Chaudhuri (2002) with multilevel analysis (Goldstein, 1999). This first of all allows to differentiate between the unexplained variance on the household level (i.e. the impact of idiosyncratic

[3] More precisely, if measurement error exists, the mean squared residuals from equation 1 would be overestimated by the variance of the measurement error.

[4] In fact, Ligon and Schechter (2004) do not recommend to estimate the mean and variance of households' consumption from pure cross-sectional data.

household specific shocks) and the unexplained variance on the community level (i.e. the impact of covariate community specific shocks). Second, multilevel analysis corrects for inefficient estimators, which might occur whenever the proposed methodology by Chaudhuri (2002) is applied to hierarchical data structures, i.e. whenever variables from various levels are introduced in the regressions.

3.2 Multilevel Analysis

Multilevel models are designed to analyze the relationship between variables that are measured at different hierarchical levels (for an introduction see e.g., Bryk and Raudenbush, 1992; Goldstein, 1999; Hox, 2002). Again, we speak of hierarchical data structure or multilevel data whenever variables, i.e. economic indicators, are collected at different hierarchical levels with lower hierarchical levels (e.g. households) nested within higher hierarchical levels (e.g. communities).

If this data structure is ignored, i.e. if we simply assign to each household living in the same community the same community characteristic, the assumption of independent observations is ignored and the estimated standard errors tend to be underestimated which may result in misleadingly significant results (see also Section 2). One can also think of aggregating the variables of the individual level to a higher level and conduct an econometric analysis on the higher level, which might however lead to a loss of within-group information, in which we are actually interested in.

Using a multilevel model allows to use both individual observations and groups of observations simultaneously in the same model without violating the assumption of independent observations. Multilevel models use the clustering information and explicitly include the various dependencies between variables at different levels without violating the assumption of independent observations, providing correct standard errors and significance tests (Goldstein 1999). In a multilevel model each level is formally represented by its own sub-model which expresses the relationships among variables within the given level and across different levels. In contrast, to control for sample clustering, i.e. to compute efficient estimators, usual regression techniques assume constant intra-class correlations for all variables, ignoring the relationship of variables at each level and between variables of different hierarchical levels.

In addition, and even more important for our case, estimates of the unexplained variance at each level of the model provide the possibility to decompose the unexplained variance of consumption into a household and community component.

To illustrate the basic idea of multilevel modelling suppose $i = 1, ..., n_i$ level one units (e.g. households) and $j = 1, ..., n_j$ level two units (e.g. communities) and that the household i is nested within the community j. If lnc_{ij} is (in our case) per capita household consumption and X_{ij} a set of household characteristics of household i in community j then we can set up a regression equation as follows:

$$lnc_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{ij} \tag{5}$$

where the error term e_{ij} reflects the unexplained variance in the household's consumption. Note that in contrast to standard regression models, the variables in equation (5) are denoted by two subscripts: one referring to the household i and one to the community j, and that coefficients are denoted by a subscript referring to the community j. This means that it is assumed that β_{0j} and β_{1j} vary across communities. Various community characteristics Z can then be introduced to estimate the variance of coefficients across communities.

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \tag{6}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j}. \quad (7)$$

where the error terms u_{0j} and u_{1j} represent level two residuals, i.e. the unexplained variance in consumption of communities.[5] Equation (6) and (7) hence reflect the impact of community characteristics Z on household consumption which differs across communities but which is the same for all households within the same community j.

Substituting equation (6) and (7) into equation (5) provides the full model, which can be written as

$$lnc_{ij} = \overbrace{\gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} Z_j + \gamma_{11} X_{ij} Z_j}^{determinsistic} + \overbrace{(u_{0j} + u_{1j} X_{ij} + e_{ij})}^{stochastic}. \quad (8)$$

and estimated via maximum likelihood (Mason et al., 1983; Goldstein, 1999; Bryk and Raudenbush, 1992).[6] The first part of equation (8) reflects the deterministic part of the equation, including the interaction term $X_{ij}Z_j$, which analyzes cross-level interactions between variables at the household and variables at the community level. The second part, expressed in brackets, captures the stochastic part of the model. In contrast to standard OLS regression the error term in (8) contains not only an individual or household component e_{ij} but also a group or community component $u_{0j} + u_{1j}X_{ij}$. The error term u_{0j} represents the unexplained variance across communities for the intercept β_{0j}. The error term u_{ij} reflects the unexplained variance across communities for the slopes β_{1j}. The error term e_{ij} captures the remaining unexplained individual or household variance in consumption.

The stochastic part in equation (8) demonstrates the problem of dependent errors in multilevel data structure. Whereas the household error

[5]The residuals u_0j and u_{1j} are assumed to have a mean of zero, $E(u_{oj}) = E(u_{uj}) = 0$. The variance of u_0j and u_{1j} is $var(u_{oj}) = \sigma^2_{u0}$ and $var(u_{1j}) = \sigma^2_{u1}$ respectively, and the covariance is $cov(u_{oj}, u_{1j}) = \sigma_{u01}$.

[6]In a more general form, assuming P explanatory variables X at the lowest level, denoted by the subscript $p(p = 1...P)$ and Q explanatory variables Z at the highest level, denoted by the subscript $q(q = 1...Q)$ the equation is $lnc_{ij} = \gamma_{00} + \gamma_{p0} X_{pij} + \gamma_{0q} Z_{qj} + \gamma_{pq} X_{pij} Z_{qj} + (u_{pj} X_{pij} + u_{0j} + e_{ij})$.

component e_{ij} is independent across all households, the community level errors u_{0j} and u_{1j} are independent between communities but dependent, i.e. equal, for every household i within community j. This already leads to heteroscedastic error terms, as the error term of a household depends on u_{0j} and u_{1j} which vary across communities and on household characteristics X_{ij} which vary across households. For the case that the individual error term e_{ij} is heteroscedastic - an assumption we make - multilevel modelling also allows to specify heteroscedasticity at the individual (or household) level.

3.3 Idiosyncratic and Covariate Variance

To assess households' vulnerability to idiosyncratic and covariate shocks using cross-sectional data we simply incorporate multilevel modelling into the method of Chaudhuri (2002). In a first step, we regress the log of per capita household consumption of household i in community j on a set of household X and community covariates Z using a basic two level model.

$$lnc_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}Z_j + (u_{0j} + e_{ij}). \tag{9}$$

The difference to equation (8) is that in equation (9) no cross-level interactions are included so that the interaction term $X_{ij}Z_j$ and the error part $u_{1j}X_{ij}$ are set to zero[7]. Equation (9) hence estimates two error terms u_{0j} and e_{ij}. Following Chaudhuri (2002) it is supposed that the error term at the household level e_{ij} captures the impact of idiosyncratic shocks whereas the error term at the community level u_{0j} captures the impact of covariate shocks on households' consumption.

In a second step we then estimate the variance at the household level σ^2_{eij} and the community level σ^2_{u0j} using the squared residuals e_{ij} and u_{0j} from

[7]The usual procedure for multilevel modelling is to build up the model in several steps. The outset is a model with only level one (household) variables as a benchmark model. Then higher level (communities) variables are included (Hox, 2002), but without any cross-level interaction effects. In a last step interaction terms are included. Incorporating interaction terms and the set-up of a full multilevel model is left for a later version of the paper.

equation (9), again applying multilevel analysis, which provides us with asymptotically efficient and consistent estimation parameters for each variance component. In a third step we predict the mean as well as the idiosyncratic and covariate variance of households' consumption (see Section 3.1 and Appendix). Last, based on the estimated mean and variance of consumption any definition, i.e. measure, of vulnerability can be applied to asses the impact of idiosyncratic and covariate shocks on households' vulnerability.

4 Empirical Illustration

4.1 Data and Model Specification

We empirically illustrate our proposed approach for Madagascar. Madagascar is one of the poorest countries in Sub-Saharan Africa with a GDP per capita of 744 USD PPP and an estimated headcount poverty rate of about 70 percent. Its poor economic performance is also reflected in very low social indicators of human well-being. Life expectancy at birth is 55 years and high rates of child mortality (7.6 percent) and child undernutrition (41.9 percent) persist (World Bank, 2005).

Moreover, households in Madagascar are frequently hit by idiosyncratic and covariate shocks which have an additional severe down-side impact on households' well-being (see Mills et al., 2003; Table A2 in the Appendix). Mills et al. (2003) report that households are most notably hit by frequently occurring covariate shocks, in particular epidemics like malaria and climatic shocks like flooding, which also show a quite strong spatial and temporal correlation (see Table A2 and A3 in the Appendix).

The data which we use for our analysis is derived from a cross-sectional household survey and a cross-sectional community census. The community census is the 2001 ILO/Cornell Commune Levels census which provides information on community characteristics like social and economic infrastructure as well as data on the occurrence of covariate shocks. It covers 1,385 out of the 1,395 communities in Madagascar. Data on household characteristics is taken from the national representative household survey of 2001 (Enquete Aupres Des Menages, EPM), covering 5,080 households in 180 communities.

[Table 1]

To estimate households' expected mean and variance of consumption we include a set of household and community characteristics in our model (Table 1). In addition to the household characteristics listed in Table 1, we consider an agricultural asset index estimated via principal component analysis (Filmer and Pritchett, 2001). At the community level we include population density and the mean educational level of the community as well as several variables reflecting the infrastructure of the community. Also community infrastructure characteristics do not enter separately into the model but as an infrastructure index based again on a principal component analysis.

Using an aggregate index instead of individual variables has two main reason. First, the two chosen indices provide a proxy of the overall agricultural productivity of households and of the infrastructure within communities, respectively. Second, as the individual characteristics are highly correlated, their coefficients are likely to provide no significant effects if they are included separately into the regression.

4.2 Estimation Results

As described in Section 3, we first estimate the expected mean and variance of log per capita consumption using multilevel modelling. We furthermore decompose unexplained consumption variance into an idiosyncratic (household-level) and covariate (community-level) component.[8] To remind, we assume that the estimated variance in consumption on the household level reflects the impact of idiosyncratic shocks on household consumption whereas the

[8]A Whitetest was applied to verify the existence of heteroscedasticity.

estimated variance in consumption on the community level reflects the impact of covariate shocks on households' consumption. In many studies the village has been used as the 'natural' covariate shock or mutual insurance level, but there is no necessity to do so (Genicot and Ray, 2003; Morduch, 2005), and using communities instead, as we do in this analysis, does not seem less useful.

Estimation results are presented in Table 2 separately for rural and urban households, representing 65 percent and 35 percent of national households respectively.[9] The expected per capita (log) consumption of rural households is considerably below the (log) poverty line, whereas the expected per capita (log) consumption of urban households lies considerably above the poverty line. This already indicates that low mean consumption is the main cause for rural vulnerability, whereas consumption volatility might be relatively more important for urban households.

[Table 2]

With regard to the estimated mean variance in consumption, we show that the estimated variance is slightly higher for rural households than for urban households, with a standard deviation of 0.58 compared to 0.51 (see Table 2). Interesting to note is that idiosyncratic variance is higher than covariate variance both for urban and rural households. However, the relative importance of idiosyncratic variance is much higher for urban than for rural households (see also Figure 1). More precisely, whereas among urban households idiosyncratic standard deviation of consumption is 2.12 as high as covariate standard deviation, the respective rate is only 1.52 for rural households. In addition to Table 2, which presents the mean of variance

[9] The detailed regression results are presented in Table A1. All coefficients show the expected signs. The amount of variance that is explained at each level is shown by R_0^2 and R_1^2, where R_0^2=0.38 refers to the explained variance at the household level and R_1^2=0.60 refers to the explained variance at the community level, respectively. The R^2s did not improve when other than the reported household and community characteristics were added.

in consumption, Figure 1 also shows the distribution of the covariate and idiosyncratic variance in consumption across urban and rural households.

[Figure 1]

Both, Table 2 and Figure 1 denote that idiosyncratic shocks have a relatively higher impact on urban consumption whereas covariate shocks have a relatively higher impact on rural consumption. In Section 3.1 we noted that in Chaudhuri's approach (2002) measurement error might lead to an overestimation of the variance in households' consumption. However, even if that were the case, we can still assess the relative importance of idiosyncratic and covariate shocks for rural and urban households.

To obtain a full assessment of the level and sources of vulnerability, we have to assess expected mean and variance of consumption jointly across the entire consumption distribution. Although all possible vulnerability definitions (or measurements) could be applied at this stage, we opt for the measurement proposed by Chaudhuri et al. (2002), defining vulnerability as the probability of a household to fall below the poverty line in the near future.[10] The focus of this paper clearly lies on the estimation of vulnerability parameters (i.e. the mean and variance in consumption) and the chosen measurement of vulnerability only serves for illustrative purposes. Hence we chose a measurement that has in contrast to most other measurements an easy intuitive interpretation, although it has some undesirable axiomatic properties (see also Calvo and Dercon, 2005).

Assuming that consumption is log-normally distributed, we can estimate the probability of a household to fall below the poverty line using the estimated expected mean and variance of consumption:

$$\hat{v}_h = \hat{P}(lnc_h < lnz|X_h) = \Phi\left(\frac{lnz - \widehat{lnc_h}}{\hat{\sigma}^2_{eh}}\right) \tag{10}$$

[10] No difference between vulnerability to short- and long-term poverty is made.

where $\Phi(.)$ denotes the cumulative density of the standard normal distribution function, z denotes the poverty line, $\hat{lnc_h}$ the expected mean of per capita log consumption and $\hat{\sigma}^2_{eh}$ the estimated variance in consumption. The calculation in is conducted separately for estimated idiosyncratic variance σ^2_{eij} and covariate variance σ^2_{u0j} in consumption.

Last, we have to define a probability or vulnerability threshold above which we consider households as vulnerable to poverty as well as the time horizon which we consider as the 'near future'. In this paper we define vulnerability to poverty as a 50 percent or higher probability to fall below the poverty line.[11] The time horizon we apply is *t+2* years. This means, that we consider those households as vulnerable which have a 50 percent or higher probability to fall below the poverty line at least once in the next two years, which is equivalent to a 29 percent or higher probability to fall below the poverty line in any given year.[12] Utilizing the stated vulnerability threshold and time horizon we estimate that 66 percent of households in Madagascar are vulnerable to poverty within the next two years (Table 3). The respective figures for urban and rural households are 87 and 22 percent respectively, indicating that (as expected) rural households are much more vulnerable to poverty than urban households.

We furthermore decompose vulnerability estimates into sources of vulnerability. In other words we first analyze whether vulnerability is mainly driven by permanent low consumption prospects (i.e. structural poverty) or by high consumption volatility (i.e. high poverty risk).[13] We state that

[11] The 50 percent threshold has become a standard vulnerability threshold in the literature (see e.g. Tesliuc and Lindert, 2004).

[12] To illustrate, the probability to fall below the poverty line only in the first year is 0.2059 (0.29 x 0.71). The probability to fall below the poverty line only in the second year is also 0.2059 (0.29 x 0.71). The probability to fall below the poverty line in both years is 0.0841 (0.29 x 0.29). Hence, the probability to fall at least once below the poverty line in the next two years is o.50, i.e. the sum of the probabilities.

[13] Note, that we implicitly assume that low expected mean consumption only reflects structural poverty and is not risk induced, although this does not necessarily have to be the case. Low consumption prospects can also be caused by risk through behavioral responses to risks of households, engaging in low risk but also low return activities (Morduch, 1994;

rural vulnerability is mainly a cause of low expected mean in consumption whereas urban vulnerability is mainly driven by high consumption volatility (Table 3). More precisely, 69 percent of rural households have an expected per capita consumption that already lies below the poverty line, and 'only' 18 percent of the 87 percent vulnerable rural households are vulnerable because of high consumption volatility. In contrast, 14 percent of urban households face risk induced vulnerability (i.e. high consumption fluctuations) whereas only 8 percent face structural induced vulnerability.

[Table 3]

Last, we analyze the impact of idiosyncratic and covariate shocks on vulnerability to poverty. As already indicated in Table 2 and Figure 1 idiosyncratic shocks have a slightly higher influence than covariate shocks on consumption volatility among rural households and a much higher influence than covariate shocks on households' consumption volatility in urban areas. This is supported by Table 3. 84 percent of rural and 22 percent of urban households are vulnerable to idiosyncratic shocks whereas 'only' 78 percent of rural and 15 percent of urban households are vulnerable to covariate shocks.

As an assessment of vulnerability to poverty depends not only on the poverty line but also highly on the chosen vulnerability (or probability) threshold above which we consider households as being vulnerable to poverty, we also show the cumulative density distribution of vulnerability to poverty in Figure 2. It presents the percentage of households that have a i or higher probability to fall below the poverty line. Again, estimates are provided for Madagascar as a whole and for rural and urban households separately.

In Figure 2, we marked the vulnerability threshold of 29 percent, which we used for our vulnerability analysis, providing us with the same estimates as presented in Table 3. What is now interesting to see is that the relative

Elbers et al., 2003).

importance of covariate and idiosyncratic shocks for rural and urban households' consumption depends on the vulnerability threshold chosen. Moreover, if we regard the whole cumulative density distribution of vulnerability to poverty, we observe that the share of urban households that face an idiosyncratic shock induced vulnerability is larger than the share of households that face a covariate shock induced vulnerability for the major part of vulnerability thresholds (Figure 2(b)), whereas the contrary is true for rural households, where covariate shocks seem to be more important for most vulnerability thresholds (Figure 2(a)).

[Figure 2]

5 Conclusion

We propose a simple method to analyze the level and sources of vulnerability using currently available standard cross-sectional or short panel household surveys without any explicit information about idiosyncratic and covariate shocks. In particular, the suggested method allows to estimate expected mean and variance in consumption of households, decomposing variance in consumption into an idiosyncratic and covariate part.

Using the concept of Chaudhuri (2002), defining vulnerability to poverty as the probability of a household to fall below the poverty line, we stated that both covariate and idiosyncratic shocks have a considerable impact on both urban and rural vulnerability. Furthermore, our results indicate that idiosyncratic shocks seem to have an even higher impact on households' consumption volatility than covariate shocks. However, idiosyncratic shocks seem to have a relatively higher impact on urban households' and covariate shocks seem to have a relatively higher impact on rural households' vulnerability.

It is difficult to say whether a higher impact of certain types of shocks on rural or urban households' consumption is the result of a more severe impact of these shocks on households' income or the result of worse insurance mechanisms of households against these shocks. In other words, with the proposed method we can only assess the net (and not gross) impact of shocks on households' consumption. In the following we still provide some cautious explanations for our results.

The suggested higher impact of idiosyncratic shocks in general implies that either insurance mechanisms within communities do not function any better than insurance mechanisms across communities or that idiosyncratic shocks have a much higher impact on households' income than covariate shocks. An explanation might be that idiosyncratic shocks (often referring to the death or job loss of a household member) cause much higher income drops or that covariate shocks are in many cases more anticipated than idiosyncratic shocks, so that ex-ante coping strategies take place.

The relatively higher impact of covariate shocks on rural households' consumption might be explained by the fact that there are certainly many more covariate shocks (such as climatic shocks) which have a higher impact on rural (agricultural) households than on urban (non-agricultural) households. Also, it is possible that urban households face even higher information and enforcement problems and that therefore community based informal insurance mechanisms against idiosyncratic shocks work better among rural than among urban households.

Last, we noted that the relative importance of consumption fluctuations (versus low mean consumption) seems to be even greater for urban households' welfare than for rural households' welfare. Hence, urban households should - if possible - be included into vulnerability studies, which have so far mostly focused on rural villages and households.

We are aware of the fact that some rather stringent assumptions have to be made to apply the proposed method. However, we argue that as long as lengthy panel data with comprehensive information on idiosyncratic and covariate shocks is missing, the suggested approach can provide quite interesting insights into the relative impact of idiosyncratic and covariate shocks on households' vulnerability. Moreover, we recommend, that any study which analyses the influence of covariate shocks on households' consumption - no matter if cross-sectional or panel-data is used and independent of the extent of shock data available - should apply multilevel modelling as it appropriately takes into account the hierarchical structure of the data that is used for such analysis.

References

Bryk, A.S. and S.W. Raudenbush (1992), *Hierarchical Linear Models: Applications and Data Analysis*, Newbury Park: Sage Publications.

Carter, M. (1997), Environment, Technology, and the Social Articulation of Risk in West African Agriculture. *Economic Development and Cultural Change*, 45 (3): 557-590.

Calvo, C. and S. Dercon (2005), Measuring Individual Vulnerability, *Discussion Paper* No. 229, University of Oxford, Oxford.

Chaudhuri, S. (2002), Empirical methods for assessing household vulnerability to poverty. Mimeo, Department of Economics, Columbia University, New York.

Chaudhuri, S., J. Jalan and A. Suryahadi (2002), Assessing Household Vulnerability to Poverty from Cross-sectional Data: A Methodology and Estimates from Indonesia. *Discussion Paper* No. 0102-52, Columbia University, New York.

Chaudhuri, S. (2003), Assessing vulnerability to poverty: concepts, empirical methods and illustrative examples. Mimeo, Department of Economics, Columbia University, New York.

Deaton, A. (1997) *The Analysis of Household Surveys*, Baltimore: John Hopkins University Press.

Dercon, S. and P. Krishnan (2000), Vulnerability, seasonality and poverty in Ethiopia. *Journal of Development Studies*, 36 (6): 25-53.

Elbers, C., Gunning, J.W. and B. Kinsey (2003), Growth and Risk: Methodology and Micro Evidence. *Tinbergen Institute Discussion Paper* TI 2003-068/2.

Filmer, D. and L.H. Pritchett (2001), Estimating Wealth Effects without Expenditure Data - or Tears: An Application to Educational Enrollments in States of India. *Demography* 38 (1): 115-132.

Foster, J., Greer, J. and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica* 52: 761-5.

Genicot, G. and D. Ray (2003), Endogenous Group Formation in Risk-Sharing Arrangements. *Review of Economic Studies*, 70(1): 87-113.

Gertler, P. and J. Gruber (2002), Insuring Consumption Against Illness. *American Economic Review* 92(1): 55-70.

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 (1): 181-206.

Goldstein, H. (1999), *Multilevel Statistical Models*, London: Arnold.

Gorad, S. (2003), What is Multi-Level for?. *British Journal of Educational Studies*, 51 (1): 46-63.

Günther, I. and K. Harttgen (2005), Vulnerability to Poverty in Madagascar. Background Paper prepared for the World Bank Social Protection Unit.

Hoddinott J. and A. Quisumbing (2003), Methods for Microeconometric Risk and Vulnerability Assessments. *Social Protection Discussion Paper* 0324, World Bank, Washington.

Hox, J. (2002), *Multilevel Analysis - Techniques and Applications*, Mahwah: Lawrence Erlbaum Associates.

Jalan, J. and M. Ravallion (1999), Are the poor less well insured? Evidence on vulnerability to income risk in rural China. *Journal of Development Economics* 58(1): 61-81.

Kochar, A. (1995), Explaining Household Vulnerability to Idiosyncratic Income Shocks. *The American Economic Review* 85(2): 159-164.

Ligon, E. and S. Schechter (2003), Measuring vulnerability, *Economic Journal* 113(486): C95-C102.

Ligon, E. and S. Schechter (2004), Evaluating Different Approaches to Estimating Vulnerability. *Social Protection Discussion Paper Series*, No. 0410, World Bank, Washington.

Luttmer, E.F.P. (2000), Inequality and Poverty Dynamics in Transition Economies: Disentangling Real Events from Noisy Data. Mimeo, World Bank, Washington.

Madalla, G.S. (1977), *Econometrics*, Tokyo: McGraw-Hill

Mason, W.M., Wong, G.M. and B. Entwistle (1983), Contextual analysis through the multilevel linear model. In S. Leinhardt (ed.) *Sociological Methodology*, San Francisco: Jossey-Bass.

Mills, B., Ninno, C. and H. Rajemison (2003), Commune Shocks, Household Assets and Economic Well-Being in Madagascar. Mimeo, World Bank, Washington.

Morduch, J. (1994), Poverty and Vulnerability. *The American Economic Review* 84(2): 221-225.

Morduch, J. (2005), Consumption smoothing across space: Testing theories of risk-sharing in the ICRISAT study region of South India. In S. Dercon (ed.) *Insurance against Poverty*, Oxford: Oxford University Press.

Paxon, C.H. (1992), Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand. *The American Economic Review* 82(1): 15-33.

Pritchett, L., Suryahadi, A. and S. Sumarto (2000), Quantifying Vulnerability to Poverty: A Proposed Measure, with Application to Indonesia. *SMERU Working Paper*, Social Monitoring and Early Response Unit (SMERU), World Bank, Washington.

Sahn, D.E. and D. Stiefel (2003), Exploring Alternative Measures of Welfare in the Absense of Expenditure Data. *Review of Income and Wealth* 49 (4): 463-489.

Steenbergen, M.R. and B.S. Jones (2002), Modelling Multilevel Data Structures. *American Journal of Political Science*, Vol. 46 (1): 218-237.

Tesliuc, E. and K. Lindert (2004), Risk and Vulnerability in Guatemala: A Quantitative and Qualitative Assessment. *Social Protection Discussion Paper* 0404, World Bank, Washington.

Townsend, R.M. (1994), Risk and Insurance in Village India. *Econometrica* 62 (3): 539-591.

Udry, C. (1995), Risk and Saving in Northern Nigeria. *The American Economic Review* 85 (5): 1287-1300.

Woolard, I. and S. Klasen (2005), Income Mobility and Household Poverty Dynamics in South Africa. *Journal of Development Studies* (forthcoming).

World Bank (2005),*World Development Report 2005*, World Bank, Washington.

Tables

Table 1
Summary statistics for household and community characteristics
for Madagascar (2001)

	Urban	Rural	National
Household characteristics			
Age of household head (in years)	42.60	41.71	42.25
Sex of household head (1=male)	76.70	78.07	77.60
Education of household head (in years)	7.80	4.15	6.35
Household size	4.42	4.78	4.56
Total no. children	1.70	2.16	1.88
Number of cattle	0.93	4.88	2.50
Number of chicken	2.63	8.70	5.04
Working in informal sector (%)	22.88	7.04	16.59
Working in formal sector (%)	21.74	5.80	15.41
Working in agricultural sector (%)	41.02	83.00	57.68
Employed (%)	43.86	57.27	49.19
Households having an enterprize in the non-agricultural sector (%)	30.22	20.24	26.26
Community characteristics			
Telephone (%)	83.16	18.75	57.60
Sanitation (%)	75.26	20.54	53.54
Save water (%)	98.43	50.00	79.21
Electricity (%)	98.43	42.00	76.02
Primary education (%)	100	100	100
Secondary education (%)	100	67.86	87.16
Tertiary education (%)	97.89	10.71	63.07
Hospital (%)	93.01	7.14	58.53
National road (%)	93.67	53.75	77.65

Source: Own calculations using the 2001 Enquete Aupres Des Menages (EPM) and 2001 ILO/Cornell Commune Levels census.

Table 2
Estimated mean and variance of consumption
for Madagascar (2001)

	Rural	Urban	National
Households	0.65	0.35	1.00
Consumption			
per capita expenditure	13.54	14.38	13.80
poverty line	13.81	13.81	13.81
Standard Deviation (predicted)			
std total	0.58	0.51	0.56
std idiosyncratic	0.47	0.53	0.49
std covariate	0.31	0.25	0.31
std idiosyncratic / std covariate	1.52	2.12	1.59

Source: Own calculations using the 2001 Enquete Aupres Des Menages (EPM) and 2001 ILO/ Cornell Commune Levels census.
Note: Estimates are household weighted. Mean and standard deviation in consumption refer to log consumption. std=standard deviation.

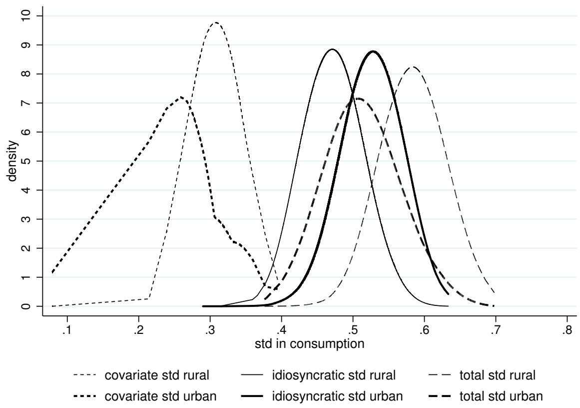
Table 3
Vulnerability decomposition
in Madagascar (2001)

	Rural	Urban	National
Total Vulnerability	0.87	0.22	0.66
Low mean	0.69	0.08	0.50
High volatility	0.18	0.14	0.16
Idiosyncratic Vulnerability	0.84	0.22	0.64
Low mean	0.69	0.08	0.50
High volatility	0.15	0.14	0.14
Covariate Vulnerability	0.78	0.15	0.58
Low mean	0.69	0.08	0.50
High volatility	0.09	0.07	0.08

Source: Own calculations using the 2001 Enquete Aupres Des Menages (EPM) and 2001 ILO/ Cornell Commune Levels census.
Note: Estimates are household weighted.

Figures

Figure 1
Density Distribution of Estimated Standard Deviation of Consumption



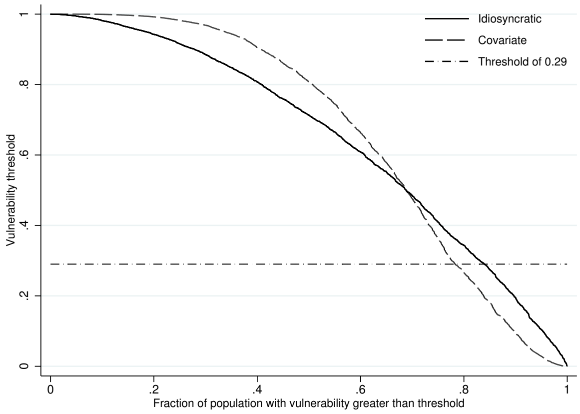
Source: Own calculations.

Figure 2

Cumulative Density Distribution of Vulnerability to Poverty

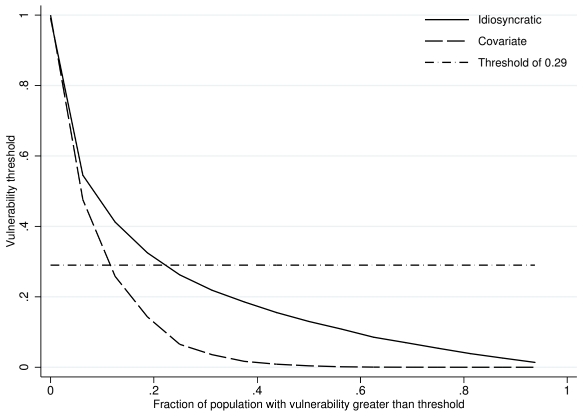
(a)

Rural Level



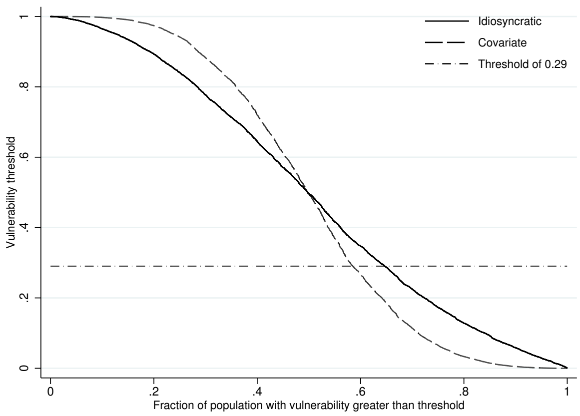
(b)

Urban Level



(c)

National Level



Source: Own calculations.

Appendix

The proposed method by Chaudhuri (2002) to estimate vulnerability to poverty υ_h of a household h at time t is straight forward. First, the assumption is made that a household's per capita consumption expenditure can be approximated by:

$$lnc_h = X_h\beta + e_h, \tag{A1}$$

where c_h is per capita household consumption, X_h a set of observable household characteristics and β a vector of parameters. Usually, the rather stringent assumption is made, that the error term e_h reflects the measurement error of households' consumption. In contrast, Chaudhuri (2002) assumes that the error term e_h, or the variance in consumption of otherwise equal households, reflects the impact of shocks on households' consumption, or in other words, the inter-temporal variance of consumption. He furthermore assumes that this variance of e_h depends on certain household characteristics and can hence be expressed by:

$$\sigma^2_{eh} = X_h\theta. \tag{A2}$$

As it is explicetly assumed that the error term e_h is heteroscedastic and not homoscedastic usual regression techniques would yield inefficient estimates. Thus, Chaudhuri (2002) proposes to use a three-step feasible generalized least squares (FGLS) regression technique. Starting with equation (A1) applying ordinary least squares (OLS), the estimated residuals of equation (A1) are used to estimate

$$e^2_{OLSh} = X_h\theta + \eta_h. \tag{A3}$$

Then, the predicted squared residuals from equation (A3) are used to to transform equation (A3) into:

$$\frac{e^2_{OLSh}}{X_h\hat{\theta}_{OLS}} = \left(\frac{X_h}{X_h\hat{\theta}_{OLS}}\right)\theta + \frac{\eta_h}{X_h\hat{\theta}_{OLS}}. \tag{A4}$$

The estimated coefficients from equation (A4) are the asymptotically efficient FGLS estimators $\hat{\theta}_{FGLS}$ for the variance of households' consumption. We then use the estimates of the standard deviation of consumption

$$\hat{\sigma}_{eh} = \sqrt{X_h\hat{\theta}_{FGLS}} \tag{A5}$$

to transform equation (A1) into

$$\frac{lnc_h}{\hat{\sigma}_{eh}} = \left(\frac{X_h}{\hat{\sigma}_{eh}}\right)\beta + \frac{e_h}{\hat{\sigma}_{eh}}. \tag{A6}$$

OLS estimation of equation (A6) yields a consistent and asymptotically efficient estimate of β. Using the estimator $\hat{\beta}$ and $\hat{\theta}$ we can directly estimate households' expected mean

$$\hat{E}[lnc_h|X_h] = X_h\hat{\beta} \tag{A7}$$

and variance in consumption:

$$\hat{V}[lnc_h|X_h] = \hat{\sigma}^2_{eh} = X_h\hat{\theta}. \tag{A8}$$

Table A1
Regression results of per capita consumption
(two level model)

	Coefficient	Standard error
Household demographic characteristics		
Age	0.008**	(0.003
Age^2/100	0.000	(0.000)
Number of children	-0.074***	(0.009)
Female headed household	0.007	(0.020)
Household Size	-0.087***	(0.006)
Household head socioeconomic characteristics		
Years of schooling	0.053***	(0.002)
Works in informal sector (=1)	0.083**	(0.026)
Works in formal sector (=1)	0.129***	(0.026)
Works in public sector (=1)	0.207***	(0.030)
Employed (=1)	0.221***	(0.032)
Enterprize owner (=1)	0.049**	(0.020)
Land owner (=1)	0.005	(0.005)
Number of cattle	0.004***	(0.001)
Number of chicken	0.001	(0.001)
Agricultural asset index	0.023*	(0.009)
Geographic characteristics		
Infrastructure index	0.035*	(0.020)
Population Density	0.197**	(0.063)
Mean years of schooling per community	0.065***	(0.011)
σ^2_{u0}	0.249	(0.005)
σ^2_{u1}	0.109	(0.011)
R^2_0	0.377	
R^2_1	0.601	
Obs. level 1 (household)	4694	
Obs. level 2 (community)	180	

Source: Own calculations using the 2001 Enquete Aupres Des Menages (EPM) and 2001 ILO/ Cornell Commune Levels census.
Notes: **denotes 10 percent significance, **5 percent significance, ***1 percent significance. Values are household weighted. σ^2_{u0} refer to as the unexplained variance at the household level and σ^2_{u1} to the unexplained variance at the community level. R^2_0 refer to as the explained variance at the household level, R^2_1 refer to as the explained variance at the community level. The agricultural asset index and the infrastructure index are based on factor analysis. For the calculation of the agricultural asset index, various production assets such as tractor, plough, other agricultural equipment, etc. are included. For the calculation of the infrastructure index the following community dummies are included: Bus stop, community road, provincial road, national road, secondary and tertiary school, water, electricity, veterinary, fertilizer, market, bank.

Table A2
Percentage of household with exposure to different shocks
in Madagascar (2000)

Shock	Persons in cummunes with exposure (percent)	Commune correlation across years (1999/2000)*
Human diseases		
Malaria	73.93	0.88
Tuberculosis	54.19	0.81
Typhoid	32.53	0.81
Cholera	33.64	0.44
Agricultural and livestock diseases		
Rice pest	22.72	0.84
Swineflu	39.46	0.63
Newcastle	75.91	0.85
Climate shocks		
Flooding	24.69	0.52
Impassible bridge or road	21.00	0.70
Drought	17.97	0.57
Cyclones	7.37	0.25

Source: Own calculations using the 2001 Enquete Aupres Des Menages (EPM) and 2001 ILO/Cornell Commune Levels census. *Mills, Ninno and Rjemison, 2003.

Table A3
Correlations of different shocks
in Madagascar (2001)

	Malaria	Tuberculosis	Typhoid	Cholera	Rice pest	Swineflu	Newcastle	Flooding	Impassible bridge or road	Drought	Cyclones
Malaria	1.00										
Tuberculosis	0.60	1.00									
Typhoid	0.40	0.44	1.00								
Cholera	0.39	0.36	0.34	1.00							
Rice pest	0.25	0.21	0.18	0.03	1.00						
Swineflu	0.35	0.25	0.07	0.21	0.19	1.00					
Newcastle	0.63	0.49	0.29	0.34	0.15	0.41	1.00				
Flooding	0.15	0.15	0.24	0.17	0.00	-0.05	0.20	1.00			
Impassible bridge or road	0.29	0.18	0.26	0.15	0.14	-0.09	0.19	0.36	1.00		
Drought	0.26	0.24	0.28	0.06	0.20	-0.06	0.18	0.09	-0.02	1.00	
Cyclones	0.07	0.14	0.15	0.11	0.07	0.07	0.03	0.43	0.34	-0.09	1.00

Source: Own calculations using the 2001 Enquete Aupres Des Menages (EPM) and 2001 ILO/Cornell Commune Levels census.