Household Poverty and Vulnerability — A Bootstrap-Approach*

Jesper J. Kühl†

16th May 2003

Abstract:
In the present paper we develop an algorithm to measure vulnerability, defined as the propensity that household consumption will fall below a locally defined poverty line. Based on the distinction between chronic and transient consumption found in the literature, we develop a stochastic process model for household consumption, where we distinguish between the chronic and the transient parts of household consumption. This distinction is utilized to derive an adapted Monte Carlo bootstrap algorithm for the simulation of household consumption, where we take account of household-specific consumption shocks. The outlined method thus extents other approaches to measure vulnerability developed by for example Christiaensen and Boisvert or Chaudhuri et al. by simulating outcomes and obtaining fully household-specific distributions of consumption.
We apply the developed method to broadly representative panel data from the Ethiopian Rural Household Survey, focusing specifically on three villages in the northern highlands, where the households face extreme environmental conditions and poor infrastructure. In a fixed-effect regression, shocks and seasonal variation are found to have a strong impact on household consumption, and on the intra-village distribution of consumption. Vulnerability levels are consistently found to be higher than the poverty levels for various subgroups of the surveyed households, while the overall vulnerability in the village is concentrated among the poor.

1 Introduction

Destitution does not only encompass low current levels of consumption, but also needs to take fluctuations in consumption into account. Households currently observed above the poverty line might face negative shocks pushing their consumption below the poverty line. Poverty and the underlying welfare indicator thus involves not only a chronic component, but also a transient part playing a crucial role for households with low consumption levels and low risk-coping abilities. Where traditional poverty concepts are rather static, the downside risk can be captured by a vulnerability concept, describing the future danger of individuals or households to face adverse shocks. A concept of and measurement method for vulnerability is crucial, especially for poverty prevention policies seeking to help households to avoid destitution in the longer run. The enhanced investigations of the causes and determinants of vulnerability on the household level are thus a crucial ingredient for successful policy designs and implementations to fight and prevent hardship.

*I am very grateful to Stefan Dercon for introducing me to the Ethiopian Rural Household Survey and giving helpful comments on several versions of this paper. Seminar participants in Oxford provided contructive suggestions. All residual errors remain my own.

†Institute of Economics, University of Copenhagen, and UNEP RisøCentre, Denmark; presently visiting the Centre for the Study of African Economies, Oxford; *jesper.kuhl@risoe.dk*.

The present paper is a continuation of the increasing stream of studies on household welfare measures for vulnerability, and we build on empirical measures of vulnerability developed and applied by e.g. Christiaensen & Boisvert [2000], Chaudhuri et al. [2002], Pritchett et al. [2000] and Ligon & Schechter [2002]. We go beyond these studies by developing and applying a useful algorithm to gain a household-specific distribution of consumption as the basis for the computation of vulnerability.

To arrive at an ex-ante distribution of future consumption, we will use a Monte-Carlo bootstrapping approach. Efron [1979] introduced and baptized the bootstrapping technique mainly as a numerical tool to find standard error estimates and other measures of precision, where analytical solutions were impossible or burdensome to reach. A basic idea of the bootstrapping methodology in the context of a regression model is the random reallocation of the residuals. In the present paper we adapt this technique to simulate transient idiosyncratic shocks to household consumption, correcting for household level heteroscedasticty. By this means we arrive at a distribution of consumption for each household.

An application of the bootstrapping technique related to the present paper can be found in Kamanou & Morduch [2002], who use the bootstrap technique to simulate future cross-sectional consumption distributions in the Cote d'Ivoire Living Standards Survey, to arrive at associated population-wide welfare measures. Another, very recent application of the methodology in economics is provided by Elbers et al. [2003], where the parameter estimates and residuals from household surveys are used in a Monte Carlo simulation to gain overall welfare measures based on census data. We extend their contributions by focusing on household-specific distributions of consumption and the attached welfare measures, taking account of the heterogeneity inherent in a household sample.

The paper proceeds as follows. We start with a conceptual discussion of chronic and transient components of household welfare and associated poverty and vulnerability measures. A description of our data and the applied model for household consumption then precedes a stepwise outline of the simulation methodology employing bootstrap techniques to identify the chronic part of consumption and simulate idiosyncratic shocks to find total household consumption. We apply the outlined model to a subsample of the Ethiopian Rural Household Survey, where the simulated samples of household consumption subsequently are used to identify and disaggregate household poverty and vulnerability and derive the response coefficients of key variables. The paper is completed by a conclusion on the main line of analysis and the key findings.

2 Components of Consumption and Poverty

Poor households in developing countries often face a multitude of risks affecting their livelihood. For households with agriculture as their main income source, risk is a pervasive factor and ecological risk such as droughts, crop pests or livestock diseases can reduce income sharply. Price fluctuations in input and output markets affect all households. Sudden illness or death — and even pregnancy and birth — in the household can either reduce the income or necessitate additional consumption expenditures. This pervasiveness of risk is often coupled with an insufficient risk-coping ability due to a low asset base and missing insurance, capital and factor markets. An analysis of their livelihood thus needs to consider the risky and transient aspects of consumption very thoroughly.

2.1 Chronic and transient consumption

In these environments household consumption has to be seen as a stochastic process where the household follows a consumption path, but where transient and unexpected factors determine the final consumption level in each period. Hence, consumption can be modelled similar to the permanent income hypothesis, consisting of a permanent or (for the household) predictable[1] mean consumption aggregated with random shocks(Deaton [1992]).

Denoting the predictable or chronic consumption as c_{ht}^{ch} and the consumption shocks as r_{ht}, total household consumption c_{ht} can be partitioned as

$$c_{ht} = c_{ht}^{ch} + r_{ht} \qquad . \tag{1}$$

Using a structural representation of household consumption, the consumption level c_{ht} for household h in period t is explained by a number of covariates x_{ht},

$$c_{ht} = x_{ht}\eta + \epsilon_{ht} \qquad \text{, with} \quad \epsilon \sim \Psi(0, \Sigma) \quad , \tag{2}$$

allowing for heteroscedasticity across households.

As several studies point out, the clear and precise identification of transient poverty as the fluctuation of consumption around a chronic component is inherently difficult, as not only unexpected consumption shocks but also measurement error are picked up. So at the extreme, it might simply be measurement error that leads us to categorize a household with a fully constant consumption stream as transiently poor.

For the present context, the study by Dercon & Krishnan [2000] is very illuminating and helpful, using the same survey rounds of the Ethiopian Rural Household Survey (ERHS) as drawn upon in the present paper. They find large fluctuations in consumption and transitions between poor and non-poor over the periods, and conclude that measurement error would need to equal a median absolute deviation from the mean level of consumption of 29% to account for these shifts. This seems high and they thus explore other, more systematic explanations for the fluctuations, namely covariate as well as idiosyncratic shocks to consumption and seasonal adjustments in consumption. And indeed, their analysis shows that both aggregate and idiosyncratic shocks as well as seasonal factors have a strong explanatory power for consumption, and account for about half of the observed mobility.

To minimize the blending of measurement error with transient movement in and out of poverty, we will build on their findings, as it will become clearer below.

2.2 Chronic and Transient Poverty

The above distinction between chronic and transient components of consumption certainly applies to most income groups. But it has very decisive consequences for households living near the poverty line.

Traditional poverty measures consider whether a household is poor or non-poor, or what proportion of households is considered poor. Rather, as households face income and consumption fluctuations, poverty requires a dynamic definition. A household may be observed as poor purely due to 'bad luck' such as a failed harvest in the previous season, while having an expected income over time above the poverty line (however defined). This *transient poverty* has to be contrasted with *chronic poverty*, where households have a chronic consumption level below the poverty line. Poverty thus also entails a risk- component, as the vulnerability to risk is intrinsically detrimental to the poor.

The analytical research into the relationship between poverty and risk and the subsequent distinction between chronic and transient poverty dates back to Ravallion [1988], named persistently or transiently poor by him.

Morduch [1994] defines these notions of poverty sample-specific: If a household is poor in every observed period, it is *chronically poor*; otherwise it is *transiently*

[1] We will in the following often use the term 'predictable' to describe the part of consumption which the household with some certainty can anticipate. It is thus not to be equated with the linear part of an econometric regression.

poor, which he also names *stochastic poverty*, as the transient poverty often is caused by stochastic elements. For a formal exposition, describing x as the household's permanent income, c as the household's current consumption, and z as the poverty line, then stochastic poverty can be defined as $c < z < x$, indicating that the current (transient) poverty only arises because they are not able to smooth their income sufficiently. In contrast to the stochastic poor households, the chronic poor suffer from a lack of earning capacity, i.e. both $x < z$ and $c < z$, and mere risk-coping (e.g. borrowing against future earnings) will not keep them above the poverty line.

Jalan & Ravallion [1998] extend this connotation by defining chronic poverty as the poverty that persists in mean consumption over time, while transient poverty can be attributed to fluctuations in the welfare indicator. They define the aggregate inter-temporal poverty measure $P_h = P(c_{h1}, c_{h2}, \ldots, c_{hT})$ for household h, where P_h for example can be defined as the squared poverty gap, $P_i = \frac{1}{T}\sum_1^T(\frac{z-c_{ht}}{z})^2, \text{for} c < z$ (Foster et al. [1984] Foster et al. (1984). The chronic component of poverty can then be characterized as $CH_h = P(\bar{c}_h, \bar{c}_h, \ldots, \bar{c}_h)$, where $\bar{c}_h$ is the (intertemporal) mean consumption for the household. Then transient poverty T_h constitutes the difference between P_h and CH_h, $T_h = P_h - CH_h$.

In a recent summary over thirteen different longitudinal studies of poverty, Baulch & Hoddinott [2000] find that most studies show a higher percentage of transiently poor households than chronically poor households. This should not downplay the unenviable position of the 'always poor' but does stresses the importance of transient poverty and thus risk for the livelihood of poor households.

Before we go on to the formulation of the consumption model and the outline of the bootstrap methodology, we will first discuss and describe the poverty and vulnerability measures to be used. We can use the above disaggregation of consumption into a predictable and a transient part to examine the poverty and vulnerability of the households.

Poverty

As poverty measures we here use the seminal Foster et al. [1984]-measures. They are usually applied as cross-sectional measures for a country, region or village, but the simulation of transient poverty for the single household allows us to use them as a tool to assess household poverty. The general Foster, Greer, Thorbecke (FGT)-poverty measure is

$$P_\alpha = \frac{1}{n}\sum_1^q \left(\frac{z-c}{z}\right)^\alpha \quad . \tag{3}$$

The nature of the measure changes when varying the value of α. The simple headcount ratio is found by $\alpha = 0$, while increasing values of α increase the sensibility of the measure to low outcomes, utilizing the income gap. We will here use $\alpha = 1$, the average income gap, and $\alpha = 2$, a convex weighting of the income gaps.

Vulnerability

The concept of vulnerability has various connotations and deflected measures in social sciences, and even within economics, as e.g. the excellent overview by Alwang et al. [2001] shows. Very generally, vulnerability (in the sphere of economics) can be defined as "the propensity to suffer a significant welfare shock, bringing the household below a socially defined minimum level". To apply this statement, three aspects have to be specified more closely, namely a) the welfare dimension and the nature of the shock, b) the operational implementation of the "propensity ... "-statement, and c) the nature of the "socially defined minimum level".

In the present paper we will focus on vulnerability to poverty. As the EHRS includes consumption data, we define our welfare measure in relation to household consumption. Consumption is preferable to income as a welfare measure when dealing with risk and poverty, as households will use their available means to smooth consumption. Income measures will therefore be excessively volatile compared to the true variability of household welfare. With household consumption as our welfare measure, we then define the socially defined minimum level as a suitable poverty line.

The literature on applied vulnerability measures has turned out a number of different approaches to the operational implementation of the propensity to suffer a significant welfare shock. For informative overviews and discussions see Kamanou & Morduch [2002], Dercon [2001] and Ligon & Schechter [2002]. As we are concerned with the direct consumption poverty and the associated vulnerability, we here use a connotation of vulnerability as "risk of poverty", and formulate vulnerability W as

$$V_{ht} = Pr(c_{h,t+1} \leq z) = \int_{-\infty}^{z} f(c_{h,t+1})dc \qquad , \tag{4}$$

where $f(c_{ht})$ is the pdf of consumption for household h in period t, and z is a consumption poverty line. Vulnerability is thus the ex-ante risk that a household will face poverty in the coming period, or – if currently poor – will remain in poverty. For other applications of this vulnerability measure, see Chaudhuri et al. [2002] or McCulloch & Calandrino [2002].

The restriction to the time period $t+1$ is fairly arbitrary, and with slight modifications equation 4 could be written for a longer horizon. But even the notation $t+1$ signifies the important point that vulnerability is inherently an ex-ante measure concerned with future consumption outcomes.

Vulnerability in the connotation used here is concerned with possible outcomes of consumption in the future. For this reason, the simulations of transient consumption as described above are used as a numerical tool to assess the households' ex-ante probability distribution of their future consumption outcomes. We can use this to implement the more advanced measure of vulnerability based on the same principle developed by Christiaensen & Boisvert [2000]. Here vulnerability is measured as the current probability of falling below the poverty line, multiplied by a conditional probability weighted function of the shortfall below the poverty line:

$$W_{ht,\mu} = F(z) \int_{\check{c}_{t+1}}^{z} (z - c_{t+1})^{\mu} \frac{f(c_{i,t+1})}{F(z)} dc_{t+1} \qquad , \tag{5}$$

where $f(\cdot)$ is an ex-ante distribution of consumption c, and $F(\cdot)$ is the matching cumulative distribution. $\check{c}_{t+1}$ is the lower range of consumption. μ regulates the relative weights shortfalls of different size are assigned, similar to the parameter in the Foster-Greer-Thorbecke poverty gap. The distinction between vulnerability measures 4 and 5 is clearly the relevance of the shortfall's magnitude in Christiaensen & Boisvert [2000]'s measure, wher the depth of poverty is introduced by the income gap-formulation under the integral. In this way the two poverty measures considered here are parallel to the poverty measures above. The vulnerability measure (4) ressembles the poverty ratio, or P_0, in so far as it does not distinguish between the risk to fall just below the poverty line and the risk of a sizable fall below the poverty line. Correspondingly, the second vulnerability measure 5 parallels the weighted poverty gap P_1, as it gives (the risk of) large downfall a higher weight. The empirical implementation of the vulnerability measures also shows an increasing and convex relationship from V_{ht} to $W_{ht,1}$, suggesting a greater weight for low outcomes in $W_{ht,1}$.

The vulnerability measures discussed in this section attribute a vulnerability *level* to all households, but this is obviously not the same as considering all of the households as *vulnerable*. They might exhibit widely ranging vulnerability levels, and it is a question of determining which of the households should be considered vulnerable. This can conveniently be done by introducing a vulnerability threshold, similar to the poverty line, above which households are deemed vulnerable. But the definition of the threshold opens a number of different paths. The first is simply to set a vulnerability level above which households are reckoned to be vulnerable. This approach could for example be based on anekdotal evidence. Using this method, one value stands particularly out, namely 50%. Choosing a threshold of 50% indicates that the household above this threshold has a higher probability to end up poor in the next period than not. Chaudhuri et al. [2002] therefore name this threshold "high vulnerability threshold".

A second approach along a similar line of argument sets the vulnerability threshold at the poverty rate in the population. As Chaudhuri et al. [2002] argue, the poverty rate in a population equals the mean vulnerability level in the community. A household with vulnerability $V_{ht} = v\%$ will end up poor in v out of 100 cases. With N households, it thus contributes $\frac{1}{N} \cdot v$ to the poverty rate in the next period.[2] To add these probabilities we have to assume that vulnerability is independently distributed across households, a condition we will come back to below. Summing the contributions for all households yields the poverty rate for the sample,

$$PR = \sum_{h=1}^{N} \frac{1}{N} v_h \qquad . \tag{6}$$

Simply factoring out the fraction in equation 6 produces the mean vulnerability level,

$$E(V_{ht}) = \frac{1}{N} \sum_{h=1}^{N} v_h \qquad . \tag{7}$$

With the arithmetic thus in place, it can be argued that households with a vulnerability level higher than the poverty rate face a risk of poverty higher than the average household and thus should be considered vulnerable.

In the empirical section below we will in more detail come back to both the welfare measures spelled out in this section as well as to the different methods to define vulnerability thresholds.

3 The Consumption Model

The data for the empirical implementation of the approach put forward come from the three survey rounds collected 1994 and 1995 as part of the Ethiopian Rural Household Survey (ERHS) in collaboration between the Centre for Studies of African Economies, Oxford, and the Economics Department and the Institute for Development Research, both Addis Ababa University. The full ERHS consists of 15 villages representing the main agro-ecological zones, and the resulting sample can be considered broadly representative of the households in the different farming systems in the country.

In the present study we will concentrate on the three villages Shumshaha, Geblen and Haresaw, comprising of 894 households. Round 1 was surveyed in June-July of 1994, followed by a survey in December-January (after the main harvest in October-December) and third round in March-May 1995.

All three villages are situated in the northern highlands of Ethiopia and are part of the agro-ecological zone of grain-plough cultivation, facing harsh climatic conditions,

[2]We have earlier argued that poverty is concerned with the present or past, while vulnerability regards future outcomes. This argument might seem to break down here, but as we are determining poverty and vulnerability based on (the same) simulated consumption values, the temporal distiction between poverty and vulnerability stays rather conceptual.

where rain is increasingly unpredictable and patchy (Bevan & Pankhurst [1996b]). The area is prone to droughts and has been classified as a serious food deficit area for both crops and livestock. Practically all households are engaged in low-scale traditional plough-cultivation of crops, mainly barley, with only limited fertilizer inputs. Soil erosion and exhaustion of soil fertility have rendered land very unproductive. Cattle are crucial as draught power and are used as wealth indicators, while other livestock include goats and sheep. These are sold to overcome liquidity problems, but following draughts and famines the number of livestock in the villages has been decreasing. The villages have very few off-farm income activities, and many households chose to migrate to other parts of the country in the off-season (Febr.-June) to cope with food-shortages.

Using the strength of the panel data, we estimate a fixed effect model to capture the time-invariant part of consumption. This also goes along well with the aim to distinguish chronic consumption from transient consumption, as the latter will be expressed by additional the explanatory variables.

The determinants of household consumption in this environment can be divided into a number of groups. The ERHS allows us to take a detailed account of various shocks and seasonal factors. In the risky environment the households experience a number of negative idiosyncratic shocks, denoted in the vector I_{ht}, affecting crops, livestock, the health of the family members and the availability of labour and oxen. At the same time, also some positive idiosyncratic impacts take place, namely aid (mainly through food-for-work programmes) and public and private gifts. These contributions to household consumption we will collect under the vector G_{ht}. As not only idiosyncratic shocks but also aggregate shocks hitting the village or region as a whole will shape consumption, we moreover define the vector A_{ht}.

Varying consumption levels do not only need to indicate shocks experienced by the household, but can also be an active choice by the household as a response to changes in their external environment. In a time of high prices the household might choose to consume less for a period, as well as increase consumption in times of harvest. As mentioned above, the second round of the survey was done in a post harvest period. To account for these factors, we follow Dercon and Krishnan (2000) and include variables to delineate such trends, denoting them as D_{ht}. Namely, we will use a local price index corrected for inflation and adjusted for village mean prices as well as a discrete variables indicating a period of peak labour demand or post harvest period. To control for household composition, in particular the number of male and female adults, we also add the vector K_{ht}.

Going back to model (2) and using the described determinants of consumption, we can set up a structural model for consumption, which we will use for the following derivation of the simulation of transient consumption and poverty.

We posit that the model

$$c_{ht} = \theta_h + A_{ht}\alpha + I_{ht}\beta + G_{ht}\gamma + D_{ht}\delta + K_{ht}\kappa + \epsilon_{ht} \quad , \tag{8}$$

captures the correlation, where θ_h is a household-specific unobserved effect, A_{ht} is a vector of aggregate shocks such as rainfall/drought hitting the whole village, I_{ht} is a vector of idiosyncratic shocks such as livestock diseases affecting the single household, D_{ht} are variables to control for seasonal variation in consumption, K_{ht} is a vector of controls, and ϵ_{ht} is a vector of residuals. By including shock-variables as well as seasonal factors as advocated by Dercon and Krishnan (2000), we aim to minimize the unexplained part of consumption $\hat{\epsilon_{ht}}$, and — more importantly —calibrate consistent coefficients, as these will be the foundations for the simulation below.

For future reference we will denominate the vector of explanatory variables as $x_{ht} = (A'_{ht}, I'_{ht}, G'_{ht}, D'_{ht}, K'_{ht})$ and the matching full vector of coefficients as $\eta_{ht} = (\alpha'_{ht}, \beta'_{ht}, \gamma'_{ht}, \delta'_{ht}, \kappa'_{ht})$.

Model (8) can thus be abbreviated as

$$c_{ht} = \theta_h + x_{ht}\eta + \epsilon_{ht} \tag{9}$$

4 Simulating consumption

Having thus formulated the model for household consumption, we will now proceed to describe the methodology we will use to simulate total household consumption. Poverty is a static and non-probabilistic concept, stating whether a household is below a predefined poverty line at the present time. Vulnerability on the other hand is a forward-looking measure, incorporating both the level of consumption as well as the variance, or uncertainty of consumption.

While poverty accordingly can be assessed using present data, vulnerability is an ex-ante measure inherently dependent on future outcomes of consumption. And following the broad definition of vulnerability in section 2.2, we moreover need to express the propensity to suffer a significant welfare shock. That is, we would like a distribution of possible future outcomes for the household to express the probability to fall into consumption poverty.

In this section we suggest a methodology for the simulation of total household comsumption as a way to achieve such an ex-ante distribution of future consumption. It is centered around the distintion between the predictable and the transient component of consumption, and employs Monte Carlo techniques adapted from the statistical method of bootstrapping.

For the outline of the methodology below, we first give a structural description of the data. We have a panel dataset with observations on the consumption and the characteristics of households, which are denoted by $h = 1, \dots, H$. For these households there are T observation over periods $t = 1, \dots, T$. With this notation in place, the data points $(c_{ht}, \vec{x}_{ht})$ can be defined, identifying time-specific pairs of consumption and characteristics for the households in the sample. To gain an overview, the data can be specified as in matrix (10).

$$\begin{pmatrix} (c_{11}, \vec{x}_{11}) & (c_{21}, \vec{x}_{21}) & \dots & (c_{H1}, \vec{x}_{H1}) \\ (c_{12}, \vec{x}_{12}) & (c_{22}, \vec{x}_{22}) & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ (c_{1T}, \vec{x}_{1T}) & (c_{2T}, \vec{x}_{2T}) & \dots & (c_{HT}, \vec{x}_{HT}) \end{pmatrix} \tag{10}$$

4.1 Bootstrap methodology

The bootstrap algorithm can be illustrated as in figure 1. The observed data stem from a specific but unknown probability mechanism P, in our case the joint distribution of consumption and the vector of covariates. We observe the random sample (c_{ht}, x_{ht}) 'produced' from the probability model P. As the first step the bootstrap algorithm estimates the probability model P, e.g. as the linear model (8), yielding $\hat{P} = (\hat{\eta}, \hat{\sigma}^2_{ht})$. This model now constitutes the background to resample pseudo-samples of c^*_{ht}, where the random reallocation of error terms plays a crucial role. For a good introduction to bootstrapping techniques and their statistical foundations see e.g. Efron & Tibshirani [1993] or Davison & Hinkley [1997]
We now proceed with a stepwise outline of the bootstrap methodology for the simulation of transient shocks to household consumption, mainly describing the route from the estimated probability model $\hat{P} = (\hat{\eta}, \hat{\sigma}^2_{ht})$ to the bootstrap samples (c^*_{ht}, x_{ht}).

Figure 1: Bootstrap Algorithm

Unknown probability model		Observed random sample		Estimated probability model		Bootstrap samples
$P = (c_{ht}, x_{ht})$	$\rightarrow$	(c_{ht}, x_{ht})	$\Rightarrow$	$\hat{P} = (\hat{\eta}, \hat{\sigma}^2_{ht})$	$\rightarrow$	(c^*_{ht}, x_{ht})

[1]
In our observed sample we observe consumption values c_{ht} and the matching household characteristics x_{ht}, as shown in matrix (10) and model (8). As the first step, the estimation of model (8) yields us parameter estimates $\hat{\eta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}, \hat{\kappa})$. In the standard regression-based bootstrap the full linear part of the regression is taken as the cornerstone, on which the residuals are randomised to find the bootstrap samples (c^*_{ht}, x_{ht}). In line with our focus on predictable consumption and transient shocks we will depart slightly from this. For every household h at time t we find the *predictable component of consumption* as the linear part of model (8), but omit the transient parts A_{ht}, I_{ht}, G_{ht}:

$$\hat{c}^{ch}_{ht} = \hat{\theta}_h + D_{ht}\hat{\delta} + K_{ht}\hat{\kappa} \qquad \forall h, t \quad . \tag{11}$$

We thus include the effect of the control variables as well as the trend variables in our estimate of chronic consumption. To classify these variables as predictable is unquestionably a matter of degree rather than certainty. We nevertheless posit that seasonal variation as well as the composition in the household exhibit a higher degree of predictability than e.g. weather phenomena.

[2]
The contribution of the remaining transient components to consumption can then be found as the effect of the variables in A_{ht}, I_{ht}, G_{ht}, i.e.

$$\hat{r}_{ht} = A_{ht}\alpha + I_{ht}\beta + G_{ht}\gamma \quad \forall h, t \quad , \tag{12}$$

Over the sample population, the estimated $\hat{r}_{ht}$ constitute a distribution of seemingly random terms. In the straight bootstrap methodology, this component would consist merely of the error term and the sample of the error terms would constitute the *empirical distribution function* (EDF) $\hat{F}$, assigning each observation the distributional mass $\frac{1}{M}$. To obtain an unbiased and consistent variance estimate, the data points forming the EDF should be identically distributed (Wu [1986]). This condition will hardly hold for $\hat{r}_{ht}$ in a panel data set with diverse households from different villages and environments. We will come back to this obstacle below, after spelling out the third step in the origianl bootstrap method as a stepping stone for further discussions.

[3]
Using the predictable component of consumption from step [1], $\hat{c}^{ch}_{ht}$ and the empirical distribution function $\hat{F}$, a pseudo- sample of household consumption can be generated as

$$c^*_{ht} = \hat{c}^{ch}_{ht} + r^* \qquad \forall h, t \quad , \tag{13}$$

where r^* are random draws with replacement from $\hat{F}$ (Davison and Hinkley (1997).

Dealing with heteroscedasticity in the sample

All estimated residuals from step [2] were piled up to form the empirical distribution function $\hat{F}$. Drawing a random sample from $\hat{F}$ and reallocating the $\hat{r}_{ht}$ to the linear parts of consumption for the construction of the pseudo-sample in step [3] depends on the exchangeability of the errors and thus implies an assumption that the household observations originate from a single distribution. But it is sensible to assume some degree of household- level heteroscedasticity across the sample, even above the unobservable household effect. The households live across varying environments, with different asset holdings and other differing characteristics, and will therefore encounter diverse shocks. It is therefore relevant to adapt the simulation methodology to take account of not only the household-specific unobservable effect θ_h, capturing time-invariant parts of household consumption, but also a household-specific transient consumption fluctuations.

Econometric studies encountering heteroscedasticity often choose to neglect the inefficiency of the coefficients induced by the unequal disturbances, but correct the standard errors to gain valid parameter tests. In the present case this would be misleading, as the simulation procedure not only depends on the magnitude of the estimated error terms and the consistency of the estimates but also on the precision or efficiency of the estimated coefficients. Therefore we below describe a slight adaptation in step [3] to take account of heteroscedasticity.

Wu [1986] suggested a weighting transformation of the residuals and the use of an auxiliary bootstrapping distribution to avoid a heteroscedasticity bias in regression-based bootstrapping. With the consumption shocks $\hat{r}_{ht}$ and a weighting factor $\nu_{ht} = x_{ht}'(X'X)^{-1}x_{ht}$, pseudo-sample corresponding to prediction (13) in step [3], but purged for heteroscedasticity, can be found as

$$c^*_{ht} = \hat{c}^{ch}_{ht} + \frac{\hat{r}_{ht}}{\sqrt{1-\nu_{ht}}} t^* \qquad \forall h,t \quad , \tag{14}$$

where t^* is a bootstrap sample from an auxiliary distribution A with $E(t) = 0$ and $Var(t) = 1$. A notable change to the assumed homoscedastic case in equation (13) is that the residual $\hat{r}_{ht}$ now stays with the corresponding data point (c^{ch}_{ht}) instead of being reallocated across the sample.

The conditions for the auxiliary distribution are obviously met by a number of distribution, see e.g. Liu [1988] or Wu [1986]. For computational convenience we will here use a simple normal distribution to generate a random sample of t^*,

$$A_1: \qquad t \sim N(0,1) \tag{15}$$

Following Davidson & Flachaire [2001], we can generalize (14) to

$$c^*_{ht} = \hat{c}^{ch}_{ht} + f(\hat{r}_{ht}) \cdot t^* \qquad \forall h,t \quad , \tag{16}$$

thus stressing that the transformation $\frac{\hat{r}_{ht}}{\sqrt{1-\nu_{ht}}}$ stated in (14) is only one of a number of possible heteroscedasticity correcting transformations $f(\hat{r}_{ht})$ of the transitory consumption shock $\hat{r}_{ht}$.

In the present case we will nevertheless argue to omit any transformation $f(\cdot)$ of the transient consumption shock. The estimation of model (8) yields us estimates for the household's transient component of consumption, as shown in step [2] above, equation (12). This estimate of variation reflects the conditions the household is living under. To validate the further analysis of determinants of poverty and vulnerability, both features of the distribution of consumption created by the transient shocks to consumption, it should thus not be altered.

Based on this argument, the prediction equation (13) can then be restated as

$$c^*_{ht} = \hat{c}^{ch}_{ht} + \hat{r}_{ht} \cdot t^* \qquad \forall h,t \qquad . \tag{17}$$

The household's predicted total consumption c^*_{ht} thus consists of its chronic/predictable consumption augmented with a randomisation of the household's transient consumption shock $\hat{r}_{ht}$. But the latter merits some further thoughts on the actual distribution of shocks implied by the random term in equation (17).

We have circumvented heteroscedasticity by pairing the shock $\hat{r}_{ht}$ with its original data point, here represented by $(c^*_{ht}, \hat{c}^{ch}_{ht})$. As a substitute for the random reallocation of residuals in the original bootstrap, we now 'shake' the household-specific residual $\hat{r}_{ht}$ by the draw $t^* \sim N(0,1)$. As the observed shock $\hat{r}_{ht}$ is a constant, the full term $\hat{r}_{ht} \cdot t^*$ is distributed N(0,$\hat{r}^2_{ht}$). The observed $\hat{r}_{ht}$ thus forms the standard deviation of the shock distribution for household h in time period t. This requires strong assumptions on the nature of the shock distribution, and we will utilize the available information to improve the modelling of the shock distribution.

The combined shock term in equation (17), $\hat{r}_{ht} \cdot t^*$, includes on the one hand the normal distribution of t^* and on the other hand the (implicit) distribution of $\hat{r}_{ht}, \forall t$. In the original bootstrap the characteristics of t^*, play a crucial role, as the aim is to preserve the underlying traints of the probability model. The prediction of the dependent vairiable is only of secondary importance, while random samples of e.g. one of the response coefficients are the main goal. In our present application, the intermediate prediction is nevertheless the main aim.

In the employed surveys of the Ethiopian Rural Household Survey we command over three time period observations for every household. Taking the observations $\hat{r}_{ht}, \forall t$ to exhibit three draws from the households distribution of shocks, we can characterize a rough distribution of the household's shock distribution by $E_t(\hat{r}_{ht})$ and $Var_t(\hat{r}_{ht})$. This leads us to the final formulation of the prediction equation, namely

$$c^*_{ht} = \hat{c}^{ch}_{ht} + s^* \qquad \text{where} \quad s \sim N[E_t(\hat{r}_{ht}), Var_t(\hat{r}_{ht})] \qquad . \tag{18}$$

Equation (18) thus generates a series of possible future household consumption as the household's predictable consumption augmented with a draw from the household-specific distribution of transient consumption shocks.

To summarize the derivation of our prediction equation (18), we calibrate a model for household consumption (adult equivalent), calculate the predictable part $\hat{c}^{ch}_{ht}$ for every observation, and augment it with a random draw from an household-specific shock-distribution. In the simulations of possible future consumption outcomes we are thus assuming stationarity over time for the seasonal predictable consumption and for the characteristics of the shock-distribution. Bu if we keep to the definition of vulnerability as the risk of poverty in time period $t+1$ and simply take the simulated consumption outcomes as a possible distribution of consumption in $t+1$, the assumption of stationarity only needs to hold for one time period and should thus not be very misleading.

We will use these simulated consumption outcomes to analyse household poverty and vulnerability for the village of Shumshaha.

4.2 Analytical approaches

Using equation (18), two distinctive approaches to further analysis are possible. An arbitrary number of pseudo-samples of household consumption can be generated for every observation using equation (17) and drawing additional samples of t^* from A_1. Denoting the repeated drawing of t^*-samples by $b = 1, \ldots, B$, the corresponding simulated distributions of consumption $c^{*1}, c^{*2}, \ldots, c^{*B}$ express the true variability of consumption due to consumption shocks, with the predictions for chronic consumption $\hat{c}^{ch}$ as the benchmark.

The further analysis can now be taken along two strands, dependent on the direction of the integration over the distributions c^{*b}. The traditional approach in statistics is to integrate parallel to the dimensions of the simulation, i.e.

$$\begin{pmatrix} \hat{c}^{ch}_{11} \\ \vdots \\ \hat{c}^{ch}_{ht} \\ \vdots \\ \hat{c}^{ch}_{HT} \end{pmatrix} \quad | \quad \begin{pmatrix} \hat{c}^{*1}_{11} \\ \vdots \\ \hat{c}^{*1}_{ht} \\ \vdots \\ \hat{c}^{*1}_{HT} \end{pmatrix} \cdots \begin{pmatrix} \hat{c}^{*b}_{11} \\ \vdots \\ \hat{c}^{*b}_{ht} \\ \vdots \\ \hat{c}^{*b}_{HT} \end{pmatrix} \cdots \begin{pmatrix} \hat{c}^{*B}_{11} \\ \vdots \\ \hat{c}^{*B}_{ht} \\ \vdots \\ \hat{c}^{*B}_{HT} \end{pmatrix}$$

Hereby a statistic ζ (e.g. a mean, median or a coefficient from a regression on the particular distribution) is derived for every pseudo-sample, and the mean or standard deviation for ζ can be found over all distributions. This leads to numerical solutions for analytically difficult characteristics of ζ. Both Kamanou & Morduch [2002] and Elbers et al. [2003] make use of this strength, and find their welfare indicators for every replication of the bootstrap.

An alternative approach is to generate the B pseudo-samples, but rather group them horizontally. To illustrate,

$$\begin{array}{ccc} \left(\hat{c}^{ch}_{11}\right) & | & \left(\hat{c}^{*1}_{11} \quad \ldots \quad \hat{c}^{*b}_{11} \quad \ldots \quad \hat{c}^{*B}_{11}\right) \\ \vdots & & \vdots \\ \left(\hat{c}^{ch}_{ht}\right) & | & \left(\hat{c}^{*1}_{ht} \quad \ldots \quad \hat{c}^{*b}_{ht} \quad \ldots \quad \hat{c}^{*B}_{ht}\right) \\ \vdots & & \vdots \\ \left(\hat{c}^{ch}_{HT}\right) & | & \left(\hat{c}^{*1}_{HT} \quad \ldots \quad \hat{c}^{*b}_{HT} \quad \ldots \quad \hat{c}^{*B}_{HT}\right) \end{array}$$

Thus grouping across pseudo-samples for every observation, yields us B (simulated) observations of consumption for every observation. This approach means that we keep along the same dimensions as our original data, where we have characteristics for every single household in every time period. We can thus find distributional measures based on the simulated consumption values and unambiguously attribute them to the observed characteristics. This lends us a strong analytical tool to assess the welfare of individual households and its determinants. We will come back to these analytical approaches in the empirical section below.

5 Empirical findings on poverty and vulnerability

Now the methodology has been described, we proceed to apply it to the ERHS, and in particular the three villages in the Northern Highlands, Shumshaha, Geblen and Haresaw. As pointed out above, our focus is Shumshaha, for which the simulation and subsequent analysis will be carried through. To obtain a sufficient sample size we nevertheless include Geblen and Haresaw for the calibration of the model parameters for household consumption.

5.1 Model Calibration

To recall, in section 3 we posited the structural model for the subsequent derivation of the simulation methodology, i.e.

$$c_{ht} = \theta_h + A_{ht}\alpha + I_{ht}\beta + G_{ht}\gamma + D_{ht}\delta + K_{ht}\kappa + \epsilon_{ht} \quad , \tag{19}$$

Calibrating the model on the data from the villages of Shumshaha, Geblen and Haresaw, we arrive at the coefficients specified in table 1. Household specific effects such as land holding are subsumed under the household unobserved effect θ_h and the variables in table 1 reflect the time-varying effects, mainly shocks. The erratic environment in the form of variable and patchy rainfall is exposed in the two rainfall coefficients, while non- rain shocks such as pests also have an effect in the mainly farm-based livelihood.

Table 1: Model calibration

Fixed effect regression of log of real consumption per adult equivalent with an unobservable household effect on household characteristics (with robust standard errors)

		coef.	**t**	**sign.**
Aggregate shocks A_{ht}:	Community wide rainfall	0.19	2.15	**
Idiosyncratic shocks I_{ht}:	Rainfall (index)	-0.26	-2.09	**
	Index of crops affected	-0.04	-0.40	
	Non-rain shock	-0.35	-1.96	*
Positive impacts G_{ht}:	Household received food aid	0.11	1.98	**
Trend variables D_{ht}:	Log of price index	0.57	2.22	**
	Post harvest season	0.17	2.25	**
Household controls K_{ht}:	Sex of household head	-0.24	-1.00	
	Number of male adults in household	-0.24	-2.22	***
	Dependency ratio	-0.06	-0.64	

Number of observations: 833; Number of households: 288; Adj. $R^2 = 0.57$
(*: 10%-significance; **: 5%-significance; ***: 1%-significance)

The area is a suffering from serious food-deficit both in crops and livestock, and is as such a large recipient of aid both in the form of direct aid and as food-for-work programmes. As such it can be expected to be a strong determinant of household consumption. At the same time we have to be aware of the possibility of endogeneity, if the distribution of aid is a response to low income or consumption in the household. But Dercon & Krishnan [2003] show that the local targeting of aid to the poor households is weak, both in terms of consumption in the previous round, low rainfalls as well as other agricultural shocks. We therefore assess that the 'return-effect' from consumption to aid is rather ambiguous and endogeneity between aid and consumption can be avoided.

Non-rain shocks cover agricultural damage in the form of frost, pest and diseases, animal damage and weeds. These kind of shocks are quite common in the sampled villages (Dercon & Krishnan [2000]) and have accordingly also a relatively high (negative) impact on household consumption.

The significance of both trend variables show that the households do choose to increase their per capita consumption in postharvest times and respond to price variations, as also found by Dercon & Krishnan [2000] for the full ERHS sample.

In section 4.1 we defined chronic consumption for household h at time t as the predicted consumption level $c_{ht}(\theta_h, D_{ht}, K_{ht})$. This consumption level will thus not be fully constant, but oscillate with changes in the variables in D_{ht}, K_{ht}. Above it was argued that these variables can be seen as sufficiently to be termed predictable, and therefore should not enter the transient shocks to consumption. Based on the model calibration in table 1, the intra-household correlation coefficient of the estimated c^{ch}_{ht} turns out to be 0.90, giving evidence for the claim that this part of consumption is rather stable.

5.2 Empirical approaches

With the calibrated model for household consumption in hand, we can now simulate distributions of household consumption using the bootstrap methodology as discussed in section 4. But as discussed above, there are different ways to use the simulated consumption levels.

As discussed above, the traditional approach in bootstrap applications is to integrate for every pseudo sample, e.g. to characterize the distributions of coefficient estimates. Below, we will briefly use this technique to give an aggregated profile of poverty and vulnerability for Shumshaha.

The second analytical approach rather collects the generated consumption for every observation, as illustrated in section 4.2. This means that we obtain observation-specific distributions of household consumption. This result can now be taken two ways. Firstly, we can pair the found distribution and derived characteristics (e.g. the welfare measures discussed) with the household characteristics in the single observation. We will elaborate on this in section 5.5 below. While this thus treats the panel dataset as if we had a cross-sectional dataset, a second route to take is to assemble the simulated data household-wise. We will follow this route in more detail in section 5.4.

5.3 A poverty profile

Using the calibration in table 1, we run the bootstrap algorithm for Shumshaha. As the poverty line z we use 51 birr, the average local poverty line[3] over the three rounds (Dercon & Krishnan [1998], table A9). Producing 1000 pseudo-samples for transitory consumption and following the first approach outlined gives the simulated expected value for a number of poverty indicators as shown in table 2.

Table 2: Poverty profile for Shumshaha

Measure	*...for predictable consumption*	*...for total consumption*
Mean of consumption	148	120
Standard deviation of consumption	93	74
Poverty headcount ratio	5%	10%
P_1	0.007	0.018
P_2	0.002	0.005
Vulnerability; $V = Pr(c \leq z)$	5%	10%
Vulnerability; W_1	0.16	0.61

The table clearly shows the effect of the transient consumption shocks. While the shape of the consumption distribution stays approximately the same, total mean consumption is lower than predictable mean consumption. This is reflected in the increases of the distribution-based welfare measures. Including idiosyncratic shocks in the consumption levels thus raises the poverty headcount ratio, and the true vulnerability of the households based on total consumption rather than only predictable consumption is higher.

It is important here to keep in mind that we assume homogeneity for the distribution of consumption across the village, and the vulnerability of the single household thus has to be seen in relation to the consumption levels of all their neighbours rather then only their own distribution of consumption. We will come back to more disaggregated analysis below.

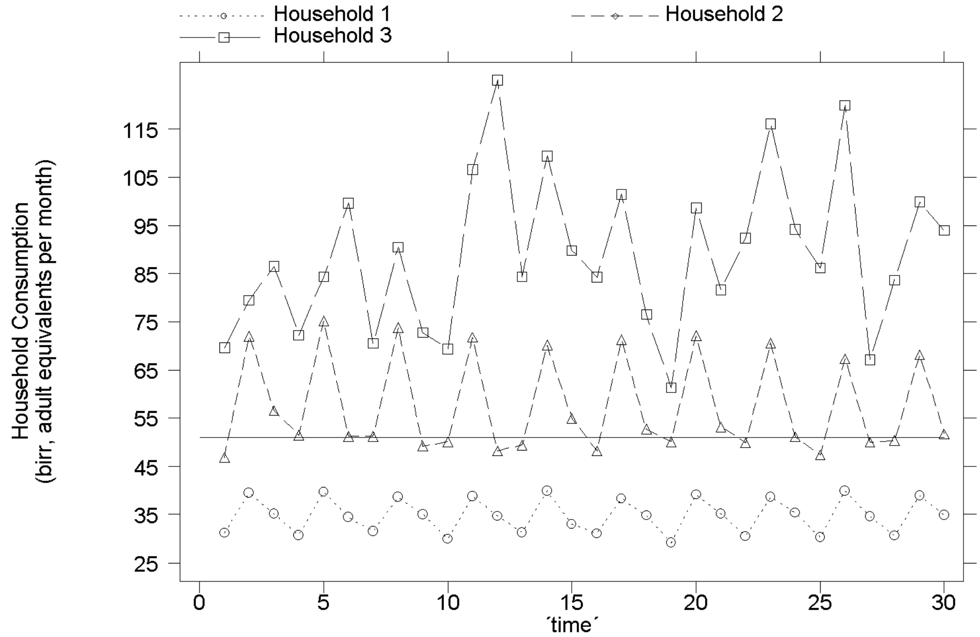
[3]Calculated as a cost-of-basic-needs for 2200Kcal with an additional non-food share

5.4 Household consumption streams

As we are less interested in the outcomes of the pseudo samples *per se* or for the village as a whole, but rather in the single households and the overall conclusions we can draw from the sample of individual households, for the rest of the paper we pursuit the second, household-specific analytical approach in more detail.

In the present section we have stacked the three sets of predictions for every households[4], to form a continuous consumption stream, incorporating for example seasonal variation. This approach can be illustrated as in figure 2, where the simulated consumption paths for three households are depicted. The vertical line shows the local poverty line, see previous footnote.

Figure 2: Simulated consumption streams



The three consumptions paths show som distinct features. First of all, all paths exhibit clear seasonal trends. The households were surveyed trice in 1994/95, and the second round was undertaken in the post-harvest season. This is reflected in the peak consumption levels in the second, fifth, eights,... observation for all households, lending graphical evidence for the significance of the trend variables in the model calibration.

Figure 2 also illustrates the welfare measures discussed in section 2.2, although the figure only depicts 30 out of the 3000 simulated consumption levels[5], see table 3.

Household 1 clearly persists below the poverty line over all time periods, i.e. over all simulated consumption values. It thus exhibits both a poverty ratio P_0 of 1 as well as a vulnerability V_{1t} of 100%. Household 2's consumption fluctuates significantly around the poverty line, with most of the off-harvest observations below the poverty line, and the post-harvest levels above it. This is mirrored in it's poverty ratio of 38% and its vulnerability level V_{2t} of 26%. Household 3 has a highly variable consumption path, but stays above the poverty line in all (depicted) periods. The shocks are even large enough to undo the seasonal trend for example in the 2nd and 3rd period, where

[4]We moreover leave out households, for which we for various reasons have less than three complete sets of observations. This reduces the number of households from 142 to 123.

[5]1000 pseudo samples for every observation, i.e. for every season, are stacked.

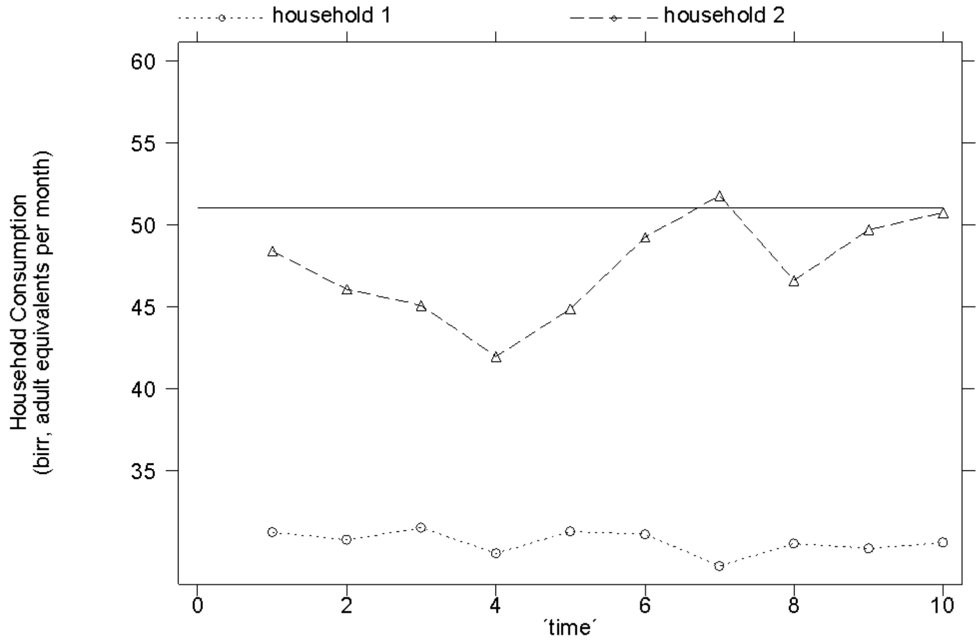
Table 3: Welfare Measures for households in figure 2

Household	*Mean cons.*	*St.dev. of cons.*	*Poverty ratio*	*Vulnerability* V_{ht}
Household 1	34.6	3.6	100%	100%
Household 2	57.6	10.1	38%	26%
Household 3	88.6	16.0	0%	0%

the household's (simulated) consumption does not top in the second period, as for the other two households, but first in the third.

It is also obvious from figure 2 that the household not only incur seasonal variation, but also considerable consumption shocks from other more involuntary sources. Following the outcomes from the same season across the years shows varying outcomes. This becomes even more clear if we depict only the observations for season 1, a pre-harvest season, as in figure 3.

Figure 3: Simulated consumption streams for a single season



Clearly, in this single season over the 'years' the households also experience consumption shocks unrelated to the seasonal trends.

The disaggregation of the households as done for the three households in figure 2 can also be extended and generalized to all households in the sample, see table 4. Using the mean and the variability of household consumption over the simulated consumption stream, we can partition the households in the sample into 7 groups. The poor and non-poor households have expected consumption levels above and below the poverty line, respectively. Similarly, the vulnerable and non-vulnerable households have vulnerability levels below and above the vulnerability threshold. Based on the discussion in section 2.2 we here use the poverty ratio of 7.3% as the threshold.

Among the vulnerable households another distinction can be done, building on the sources of vulnerability. As vulnerability integrates the distribution of consumption for the household, it necessarily depends on all the moments of the distribution, in particular the mean and the variance. Different combinations of the magnitudes of these moments can classify the vulnerable households. Firstly, a household with a expected consumption level above the poverty line can face the risk of falling below the poverty line if it has a high volatility of consumption. Following Chaudhuri et al. [2002] we will label these households high-volatility (HV-)vulnerable. Secondly, a household with a low variability of consumption can equally exhibit high vulnerability if it has an expected consumption level at or below the poverty level. These households will be termed low-mean (LM-)vulnerable.

Table 4: Disaggregation of poverty and vulnerability

	Overall	*The poor*	*Non-poor*	*Vulnerable*	*Non-vuln.*	*HV-vuln.*	*LM-vuln.*
Proportion	100%	7.3%	92.7%	13.8%	85.4%	6.5%	7.3%
Mean hh. cons.	122	44	128	51	134	57	44
Poverty ratio P_0	7.3%	100%	0%	50.0%	0%	0%	100%
$E_h(V_h)$	8.3%	78.3%	2.8%	55.1%	0.3%	32.1%	78.2%
Fraction vuln. hh.	14.6%	100%	7.9%	100%	0%	100%	100%
Fraction HV-vuln.	7.3%	0%	7.9%	50.0%	0%	100%	0%
Fraction LM-vuln.	7.3%	100%	0%	50.0%	0%	0%	100%
Vuln.-Pov. ratio	2	1	$\rightarrow \infty$	2	0	0	1

Table 4 lends itself for some remarks on the occurrence poverty and vulnerability. As conceptual arguments suggest, poverty and vulnerability are close associates. The second column shows that the poor households are clearly highly vulnerable, entirely as a result of their low expected consumption levels. Correspondingly, the vulnerable households in column 4 generally have low expected consumption levels, as their high mean poverty ratio reflects. This column also shows that low expected consumption levels and high volatility are even sources of the vulnerability of the households, but the last two columns show that the high-volatility vulnerable households have an expected consumption above the poverty line, while the low-mean vulnerable households (by definition) have mean consumption below the poverty line.

In the last row of table 4 the vulnerability to poverty ratio is calculated as the ratio of the vulnerable fraction of the group to the poor fraction, where increasing ratios indicate more dispersed ('egalitarian') vulnerability in the specific group (Pritchett et al. [2000]). Vulnerability for the whole village is more dispersed than for the subgroup of poor households, which also follows from the high proportion of poor households among the the vulnerable in column 4. And while the group of poor households has a comparable egalitarian distribution of vulnerability, vulnerability among the non-poor is concentrated on a few households, most likely households with expected consumption near to the poverty line.

Table 4 thus sheds some lights on the overall distribution and the sources of poverty and confirms the findings in for example Chaudhuri et al. [2002]. Maybe the strongest signal to take from table 4 is the relative proportion of vulnerability to poverty. For all groups the vulnerable fraction of the group far exceeds the poor fraction, and thus emphasizes that the volatility of consumption adds a considerable number of households to the group threatened by future frailty. This can help redefine the focus of policy from solely the poor to the larger group of threatened households.

The disaggregation of vulnerability can also be taken into more detail, examining the vulnerability profiles along gender and age lines, see table 5. Female-headed households have a slightly lower mean household consumption than male-headed households, but they show substantially higher poverty and vulnerability figures. The proportion of poor households is higher among these households, and their higher vulnerability is relatively more caused by their lower mean consumption.

Table 5: Poverty and Vulnerability and Characteristics of the Household Head

	Overall	Sex		Age		
		Female	Male	≤ 25	$25-50$	≥ 50
Proportion	100%	24%	76%	4%	58%	38%
Mean hh. cons.	122	120	123	147	132	104
Poverty ratio P_0	7.3%	13.3%	5.4%	0%	4.2%	12.8%
$E_h(V_h)$	8.3%	13.0%	6.9%	0%	4.5%	14.9%
Fraction vuln. hh.	14.6%	20.0%	13.0%	0%	9.9%	23.4%
Fraction HV-vuln.	7.3%	6.7%	7.6%	0%	5.6%	10.6%
Fraction LM-vuln.	7.3%	13.3%	5.4%	0%	4.2%	12.8%
Vuln.-Pov. ratio	2	1.5	2.4	0	2.3	1.8

The age structure shows a similarly clear picture with regard to differences between young, medium-aged and old households. The small proportion of households with young household heads display not only a higher mean consumption, but also no poor or vulnerabile households among them. But their small number would require further studies to confirm these lines. With increasing age, the mean consumption of the households seem to fall, while consequently poverty as well as vulnerability measures increase.

Having thus illustrated the simulated consumption streams and discussed the disaggregation of poverty and vulnerability in Shumshaha, we will now consider possible covariates of the household welfare measures in more detail in the coming sections.

5.5 Regression results

Integrating the bootstrap samples yields us a distribution of possible consumption outcomes for every observation (i.e. for every household in every time period). Using these distributions we can thus assigns vulnerability and poverty estimates to the vector of household and community characteristics in every observation. We can thus undertake an the econometric analysis and regress the household characteristics on the estimates to discuss the determinants of poverty and vulnerability. This is shown in tables 7 to 11 in the appendix.

All the poverty and vulnerability measures given in the tables are distributional properties and hence depend on the moments of the distribution, namely the mean and the variance of household consumption. As we are dealing with fixed effect model with an unobservable household effect, time-invariant factors of household consumption will be subsumed under θ_h and thus not appear in our regression coefficients. Taking the mean of the distributions of household consumption as a proxy for the household's predictable consumption, factors affecting the mean will not show up in the regressions. Nevertheless, in the highly erratic environment the households of Shumsheha are living, the variational dimension of household consumption is surely similarly if not even more important.

A common trait in the regressions of tables 7 to 11 is again the role of seasonal variation. The post-harvest categorical variable is significantly negative in all poverty and vulnerability measures. This not only indicates a lower risk of poverty for the household in the month following the harvest, but also – less comforting – that the households run a significantly higher risk to fall into poverty in the seasons up to the harvest, i.e. the hungry season, as Bevan & Pankhurst [1996b] refer to it. This is well reflected in the described migration of family members and households during these times.

On parts of the shocks, both the danger of illness of family members as well as disease among the livestock contributes to the riskiness of household livelihood. Sickness among household members reduces the earning capacity of the household, while livestock both are productive assets as draught power and transport means as well as important wealth storage.

5.6 Derivation of a gradient vector for welfare

Besides the derivation of the absolute poverty and vulnerability levels for Shumshaha as a whole and the individual households, it is of interest to study the sensibility of these results to changes in the external environment, namely the shocks the households are encountering. Therefore a numerical derivation of the marginal response of the welfare measures is outlined here and applied to the data from Shumshaha.

To examine the marginal impacts of shock variables on vulnerability, we can derive a numerical estimate of the gradient vector $\frac{\delta V}{\delta x_k}$, where x_k indicates a shock variable (see e.g. Pagan & Ullah [1999]).

Adding a slight perturbation $\iota|x^k|$ to the variable x^k simulates an incremental change of the addressed variable x^k.

With the remaining set of variables unchanged, we can now calculate $\hat{c}_{ht}^{ch,+}$ and arrive at pseudo samples of $\hat{c}_{ht}^{*,+}$ by the simulation process outlined in steps [1] to [3] in section 4. This yields us an estimate for a perturbed vulnerability measure $\hat{V}^{k+}$. Correspondingly, subtracting $\iota|x^k|$ from the variable x^k yields a perturbed $\hat{V}^{k-}$. For small values of ι

The simulated central distance estimator of the derivative $\frac{\delta V}{\delta x_k}$ can then be found as

$$\psi_k = \frac{(\hat{V}^{k+} - \hat{V}^{k-})}{(2\iota|x^k|)} \qquad \text{, and} \qquad \frac{\delta V}{\delta x_k} = \lim_{h \to 0} \psi_k \qquad . \tag{20}$$

This numerical derivation of response coefficients follows the lines of the previous simulation of consumption distribution. Moreover, as a nonparametric technique, it does not suppose any specific functional form for the joint distribution of vulnerability and the the household covariates x_{ht}. As it does not give us only a single coefficient for the full sample of households, as e.g. linear regression would do, but an estimate for every household, it also allows for a disaggregation of the responses. We will employ this strength to examine the magnitude of the response coefficients for different groups of households.

In table 6 we depict the simulated response coefficients with respect to the used welfare measures for selected significant variables from the model calibration. For the interpretation of the coefficients we should add that both the poverty ratio and the vulnerability measures V have been upscaled to percentage terms.

Rainfalls and external help both decrease household poverty and vulnerability, as expected. Correspondingly will non-rain shocks and – maybe a little surprisingly – also a higher number of male adults in the household increase the (negative) welfare measures. But the latter is in line with our findings in the model calibration, and can be explainable by the fact that we are modelling adult equivalent consumption, where the per capita consumption decreases if the household with the same means should feed an additional person.

The calculated response coefficients also allow us to interpret the relative effect of the variables with respect to each other. In general, the magnitude of the coefficients suggest that 'natural' shocks – both positive and negative – have higher absolute effects on household welfare than aid and food-for-work. Moreover, to take the analysis a step further, it seems that rainfalls spread the gains of the welfare improvements

Table 6: Numerical Gradients for Welfare

Variable	*Response coefficients*				
	Poverty Ratio	P_1	P_2	V	W_1
Comm.-wide Rainfalls	-7.2	-0.017	-0.006	-7.4	-6.0
Aid and Food-for-work	-4.9	-0.005	-0.001	-4.7	-0.60
Non-rain shock	12.7	0.032	0.012	13.2	3.69
Male adults in hh.	9.8	0.014	0.005	9.2	4.3

more equally across the intra-village consumption distribution. For the poverty ratio and vulnerability V the response coefficients for rainfall are approximately double the size of the aid-coefficients, while their are three to ten times the size for P_1, P_2 and W_1. The distinction between these two groups of welfare indicators is that the latter weight the downfalls below the poverty line while the first treat a household just below the poverty line the same as a household far below the poverty line. The larger relative size of the response coefficients for rain for the weighting indicators therefore indicates that rainfall benefits the poorest households more than aid and food-for-work does. This also follows the results on the effectiveness of aid earlier quoted from Dercon & Krishnan [2003].

5.7 Some Remarks

If households cannot use *ex-post* mechanisms to income smoothing (asset depletion, network assistance, short-term wage work, etc.) households will react to the conceived risks and fluctuations *ex-ante* and adjust their income generation accordingly (using portfolio decisions, technological choice, patronage, etc.). This will mostly entail a trade of lower variability for a lower expected yield from the individual activity, and the overall income and thereby consumption path will be smoother but lower than the 'pure' untouched path. Households thus choose a variety of crops, soil types and field locations to minimize their farming risk (Bevan & Pankhurst [1996a]). Hence, an analysis of transient poverty will understate the true variability the households are facing *ex-ante* their smoothing activities. The analysis is nevertheless informative, as it will yield a lower bound for variability of consumption and the derived welfare measures.

In the simulation of idiosyncratic shocks to household consumption, we do not account for long-term effects of repeated downside shocks and the potential non-reversibilities for the households. Incurring a row of shocks, households will try to smooth their consumption, which can eat into their savings and asset holding, leading to a downward spiral for their livelihood. To consider these potentially very severe consequences, a model taking better account of dynamics in household asset holdings needs to be constructed.

6 Conclusion

The focus here has been on the analysis of traits of household poverty and vulnerability. For poor agricultural households in developing countries, risk is a pervasive factor and ecological risk such as droughts, crop pests or livestock diseases can reduce income sharply. This combination of poverty and risk leads to a dynamic picture of household welfare, as households may enter and leave poverty over time, caused by both seasonal variations and external shocks.

Traditional poverty measures only capture these dynamics incompletely, as they typically only look at the present state of household welfare. This leads us to consider vulnerability as an indicator of welfare. We here use the connotation of vulnerability as "the propensity to suffer a significant welfare shock, bringing the household below a socially defined minimum level". In this light we employ household adult equivalent consumption as our welfare measure, and a local poverty line as a threshold.

Based on the distinction between chronic and transient consumption found in the literature, we develop a stochastic process model for household consumption, where we distinguish between the predictable and the transient parts of household consumption. This distinction is employed in the derivation of an adapted Monte Carlo bootstrap algorithm, where we take account of household-specific consumption shocks. The algorithm now lends us a convenient tool to simulate total household consumption in an erratic environment.

The simulation outcome opens different options to use the obtained data. They can be used to assess poverty and vulnerability in the 'tradtional' way, using the cross-sectional dimension, by integrating across households for every obtained pseudo-sample. On the other hand, and potentially more interesting, the outcome can be integrated for households. This allows us to pair welfare measures obtained from he household-specific consumption distributions with the household's characteristics and analyze traits of household poverty and vulnerability.

We apply the developed method to broadly representative panel data from the Ethiopian Rural Household Survey, focusing specifically on three villages in the northern highlands, where the households face extreme environmental conditions and poor infrastructure. In a fixed-effect regression, shocks and seasonal variation are found to have a strong impact on household consumption, and on the intra-village distribution of consumption. Vulnerability levels are consistently found to be higher than the poverty levels for various subgroups of the surveyed households, while the pverall vulnerability in the village is concentrated among the poor.

The developed simulation algorithm is also utilized to derive numerical gradients for the welfare measures to examine the response coefficients of households welfare to external factors. It is found that 'natural' shocks – both positive and negative – have higher absolute effects on household welfare than aid and food-for-work. Moreover, it seems that rainfall shocks spread the gains of the welfare improvements more equally across the intra-village consumption distribution.

References

Alwang, J., Siegel, P. B., & Jorgensen, S. L. (2001). "Vulnerability: A View from Different Disciplines". *SP Discussion Paper* No.0115; Social Protection Unit; The World Bank.

Baulch, B. & Hoddinott, J. (2000). "Economic mobility and poverty dynamics in developing countries". *Journal of Development Studies*, 36(6), 1–24.

Bevan, P. & Pankhurst, A. (1996a). "Ethiopian Village Studies: Harresaw". *Report*, Dept. of Sociology, Addis Ababa University; and Centre for Studies of African Economies, Oxford.

Bevan, P. & Pankhurst, A. (1996b). "Ethiopian Village Studies: Shumsheha". *Report*, Dept. of Sociology, Addis Ababa University; and Centre for Studies of African Economies, Oxford.

Chaudhuri, S., Jalan, J., & Suryahadi, A. (2002). "Assessing Household Vulnerability to Poverty from Cross-sectional Data: A Methodology and Estimates for Indonesia". *Discussion Paper Series* no. 0102-52; Department of Economics, Columbia University.

Christiaensen, L. J. & Boisvert, R. N. (2000). "On Measuring Household Food Vulnerability: Case Evidence from Northern Mali". *Working Paper* 2000-05; Department of Applied Economics and Management; Cornell University.

Davidson, R. & Flachaire, E. (2001). "The Wild Bootstrap, Tamed at Last". *Mimeo*; Queen's University, Canada, and London School of Economics.

Davison, A. C. & Hinkley, D. (1997). *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge.

Deaton, A. (1992). *Understanding consumption*. Clarendon Press, Oxford University Press. Oxford, England.

Dercon, S. (2001). "Assessing Vulnerability to Poverty". Paper prepared for the Department for International Development (DfID), London.

Dercon, S. & Krishnan, P. (1998). "Changes in Poverty 1989-1995: Measurement, Robustness Tests and Decomposition". Working Paper Series 1998-07; Center for Studies of African Economies; Department of Economics, Oxford University.

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

Dercon, S. & Krishnan, P. (2003). "Risk sharing and public transfers". *Economic Journal*, 113(486), C86–C94.

Efron, B. (1979). "Bootstrap methods: Another look at the jackknife". *Annals of Statistics*, 7(1), 1–26.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall/CRC. New York.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). "Micro-level estimation of poverty and inequality". *Econometrica*, 71(1), 355–364.

Foster, J., Greer, J., & Thorbecke, E. (1984). "A class of decomposable poverty measures". *Econometrica*, 52(3), 761–766.

Jalan, J. & Ravallion, M. (1998). "Transient poverty in post-reform rural China". *Journal of Comparative Economics*, 26(2), 338–357.

Kamanou, G. & Morduch, J. (2002). "Measuring Vulnerability to Poverty". *Discussion Paper* No. 2002/58, World Institute for Development Economics Research (WIDER); United Nations University.

Ligon, E. & Schechter, L. (2002). "Measuring Vulnerability". *Mimeo*; University of California, Berkeley; presented at the IFRPRI-World Bank Conference on Risk and Vulnerability, September 23-24, 2002.

Liu, R. Y. (1988). "Bootstrap Procedures under some non-i.i.d. Models". *Annals of Statistics*, 16(4), 1696–1708.

McCulloch, N. & Calandrino, M. (2002). "Vulnerability and Chronic Poverty in Rural Sichuan". *manuscript*; International Development Studies; University of Sussex.

Morduch, J. (1994). "Poverty and Vulnerability". *American Economic Review*, 84(2), 221–225.

Pagan, A. & Ullah, A. (1999). *Nonparametric Econometrics*. Themes in modern Econometrics. Cambridge University Press.

Pritchett, L., Suryahadi, A., & Sumarto, S. (2000). "Quantifying Vulnerability to Poverty: A Proposed Measure, Applied to Indonesia". *Mimeo*, Social Monitoring and Early Response Unit, Jakarta.

Ravallion, M. (1988). "Expected poverty under risk-induced welfare variability". *Economic Journal*, 98(393), 1171–1182.

Wu, C. (1986). "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis". *Annals of Statistics*, 14(4), 1261–1295.

Appendix

Table 7: Analysis of household poverty rate

Fixed effect regression of simulated household poverty rate with an unobservable household effect (robust standard errors):

	coef.	*t*	*sign.*
Days lost by adults due to sickness	0.004	0.98	
Livestock diseases	0.05	1.99	**
Livestock suffering from lack of water and grazing	0.005	0.19	
Lower harvest due to not having oxen	-0.01	-0.50	
hh. received official gifts during last month	0.05	3.83	***
hh. received private gifts during last month	-0.27	-1.67	*
Post harvest season	0.17	2.25	**
Household size	-0.06	-0.64	

Number of observations: 833
Number of households: 288
Adj. $R^2 = 0.57$
(*: 10%-significance; **: 5%-significance; ***: 1%-significance)

Table 8: Analysis of average poverty gap P_1

Fixed effect regression of simulated household poverty measure P^2 with an unobservable household effect (robust standard errors):

	coef.	*t*	*sign.*
Days lost by adults due to sickness	0.003	1.63	
Livestock diseases	0.006	0.568	
Livestock suffering from lack of water and grazing	0.0001	0.01	
Lower harvest due to not having oxen	0.0002	0.03	
hh. received official gifts during last month	0.01	2.45	**
hh. received private gifts during last month	-0.004	-0.73	
Post harvest season	-0.01	-2.48	**
Household size	0.004	1.35	

Number of observations: 833
Number of households: 288
Adj. $R^2 = 0.57$
(*: 10%-significance; **: 5%-significance; ***: 1%-significance)

Table 9: Analysis of weighted poverty gap P_2

Fixed effect regression of simulated household poverty measure P^2 with an unobservable household effect (robust standard errors):

	coef.	*t*	*sign.*
Days lost by adults due to sickness	0.001	1.56	
Livestock diseases	0.003	0.52	
Livestock suffering from lack of water and grazing	0.001	0.22	
Lower harvest due to not having oxen	0.001	0.38	
hh. received official gifts during last month	0.006	2.04	**
hh. received private gifts during last month	-0.001	-0.38	
Post harvest season	-0.005	-2.12	**
Household size	0.002	1.16	

Number of observations: 833
Number of households: 288
Adj. $R^2 = 0.57$
(*: 10%-significance; **: 5%-significance; ***: 1%-significance)

Table 10: Analysis of vulnerability measure V_{ht}

Fixed effect regression of simulated household vulnerability measure $Pr(c_{ht} \leq z)$ with an unobservable household effect (robust standard errors):

	coef.	*t*	*sign.*
Days lost by adults due to sickness	0.004	0.97	
Livestock diseases	0.05	1.91	*
Livestock suffering from lack of water and grazing	0.005	0.21	
Lower harvest due to not having oxen	-0.01	-0.52	
hh. received official gifts during last month	0.05	3.77	***
hh. received private gifts during last month	-0.02	-1.65	*
Post harvest season	-0.02	-2.29	**
Household size	0.01	1.34	

Number of observations: 833
Number of households: 288
Adj. $R^2 = 0.57$
(*: 10%-significance; **: 5%-significance; ***: 1%-significance)

Table 11: Analysis of vulnerability measure $W_{ht,1}$

Fixed effect regression of simulated household vulnerability measure $W_{ht,1}$ with an unobservable household effect (robust standard errors):

	coef.	*t*	*sign.*
Days lost by adults due to sickness	0.89	1.66	*
Livestock diseases	-0.13	-0.04	
Livestock suffering from lack of water and grazing	0.04	0.01	
Lower harvest due to not having oxen	1.10	0.40	
hh. received official gifts during last month	3.51	2.12	**
hh. received private gifts during last month	-0.26	-0.15	
Post harvest season	-3.36	-2.37	**
Household size	1.21	1.18	

Number of observations: 833
Number of households: 288
Adj. $R^2 = 0.57$
(*: 10%-significance; **: 5%-significance; ***: 1%-significance)