United Nations University

WIDER

World Institute for Development Economics Research

Discussion Paper No. 2003/52

Are Neighbours Equal?

Estimating Local Inequality in Three Developing Countries

Chris Elbers, Peter Lanjouw, Johan Mistiaen, Berk Özler and Ken Simler*

July 2003

Abstract

Based on a statistical procedure that combines household survey data with population census data, this paper presents estimates of inequality for three developing countries at a level of disaggregation far below that allowed by household surveys alone. We show that while the share of *within*-community inequality in overall inequality is high, this does not necessarily imply that all communities in a given country are as unequal as the country as a whole. In fact, in all three countries there is considerable variation in inequality across communities. We also show that economic inequality is strongly correlated with geography, even after controlling for basic demographic and economic conditions.

Keywords: inequality measurement, Ecuador, Madagascar, Mozambique
JEL classification: D63

---

* Elbers is with *Vrije* (Free) University; Lanjouw, Mistiaen and Özler are with the World Bank; Simler is with the International Food Policy Research Institute.

This study has been prepared within the UNU/WIDER project on Spatial Disparities in Human Development, directed by Ravi Kanbur and Tony Venables.

UNU/WIDER gratefully acknowledges the financial contribution to the project by the Government of Sweden (Swedish International Development Cooperation Agency—Sida).

Acknowledgements

We are grateful to the Instituto Nacional de Estadistical y Censo (INEC), Ecuador, Institut National de la Statistique (INSTAT), Madagascar, and the Instituto Nacional de Estatistica, Mozambique, for access to their unit record census data and the Bank Netherlands Partnership Program (BNPP) for financial support. Helpful comments were received from Jenny Lanjouw and participants at a PREM Inequality Thematic Group seminar at the World Bank. The views in this paper are our own and should not be taken to reflect those of the World Bank or any of its affiliates. All errors are our own.

The World Institute for Development Economics Research (WIDER) was established by the United Nations University (UNU) as its first research and training centre and started work in Helsinki, Finland in 1985. The Institute undertakes applied research and policy analysis on structural changes affecting the developing and transitional economies, provides a forum for the advocacy of policies leading to robust, equitable and environmentally sustainable growth, and promotes capacity strengthening and training in the field of economic and social policy making. Work is carried out by staff researchers and visiting scholars in Helsinki and through networks of collaborating scholars and institutions around the world.

www.wider.unu.edu *publications@wider.unu.edu*

UNU World Institute for Development Economics Research (UNU/WIDER)
Katajanokanlaituri 6 B, 00160 Helsinki, Finland

Camera-ready typescript prepared by Lorraine Telfer-Taivainen at UNU/WIDER
Printed at UNU/WIDER, Helsinki

ISSN 1609-5774
ISBN 92-9190-490-2 (printed publication)
ISBN 92-9190-491-0 (internet publication)

1 Introduction

The 1990s witnessed a resurgence in theoretical and empirical attention by economists to the distribution of income and wealth.[1] One important strand of research in the area of political economy and public policy has focused on the appropriate level of government to which can be devolved financial and decision-making power regarding public service provisioning and financing. The advantage of decentralization to make use of better community-level information about priorities and the characteristics of residents may be offset by a greater likelihood that the local governing body is controlled by elites—to the detriment of weaker community members. In a recent paper, Bardhan and Mookherjee (1999) highlight the roles of both the level and heterogeneity of local inequality as a determinant of the relative likelihood of capture at different levels of government. As most of the theoretical predictions are ambiguous, they stress the need for empirical research into the causes of political capture—analysis which to date remains relatively scarce.[2]

Detailed information on local-level inequality has traditionally been available only from case studies which focus on one or two specific localities.[3] Such studies do not provide a basis for generalizations about local-level inequality across large numbers of communities. Construction of comprehensive 'geographic profiles' of inequality across localities has been held back by limitations with conventional distributional data. Detailed household surveys which include reasonable measures of income or consumption are samples, and thus are rarely representative or of sufficient size at low levels of disaggregation to yield statistically reliable estimates. In the three developing countries studied here—Ecuador, Madagascar, and Mozambique—the lowest level of disaggregation possible using sample survey data is to regions that encompass hundreds of thousands of households. At the same time, census (or large sample) data of sufficient size to allow disaggregation either have no information about income or consumption, or measure these variables poorly.

This paper provides, in the next section, a brief description of a recently developed statistical procedure to combine data sources so as to take advantage of the detailed information available in household sample surveys and the comprehensive coverage of a census (Elbers et al. 2003, 2002; Demombynes et al. 2002; Hentschel et al. 2000). Using a household survey to impute per capita expenditures, y, for each household enumerated in

[1] In their introductory chapter to the *Handbook of Income Distribution*, Atkinson and Bourguignon (2000) welcome the marked expansion of research on income distribution during the 1990s, but underscore that much ground remains to be covered.

[2] Although, see Ravallion (1999, 2000) and Tendler (1997).

[3] Lanjouw and Stern (1998) report on a detailed analysis of the evolution of poverty and inequality in a north Indian village over five decades. As their study covered the entire population of the village in all survey years, their measures of income inequality describe the true distribution of income in the village. Such studies are rare. More common are village or community studies which estimate inequality across (often small) samples of households within the village.

the census we estimate inequality at a finely disaggregated level. The idea is straightforward. First a model of y is estimated using the sample survey data, restricting explanatory variables to those either common to both survey and census, or variables in a tertiary dataset that can be linked to both of those datasets. Then, letting W represent an indicator of poverty or inequality, we estimate the expected level of W given the census-based observable characteristics of the population of interest using parameter estimates from the 'first stage' model of y. The same approach could be used with other household measures of wellbeing, such as assets, income, or employment.

Applying this methodology to the three developing countries mentioned above, we examine how well our census-based estimates match estimates from the corresponding household surveys at the level of disaggregation at which the households surveys are representative. Following a description of our data in section 3, and a discussion of implementation of the method in section 4, we find in section 5 that despite the variation in levels of development, geographical context, quality and organization of data, the method seems to work well in all three countries we examine.

In section 6 we turn to a detailed examination of local-level inequality in our three countries under study. We first examine the importance of local-level inequality by decomposing national inequality in all three countries into a within-community and between-community component, where we successively redefine community to correspond to lower levels of disaggregation. We find that in all countries the within-community share of overall inequality remains dominant even after we have disaggregated the country into a very large number of small communities (corresponding to the third administrative level—often representing an average of no more than 1,000-2,000 households). These results might be construed to suggest that there is no basis for expecting communities to exhibit a greater degree of homogeneity than larger units of aggregation. To the extent that local-level inequality is correlated with factors, such as elite-capture, that might threaten the success of local-level policy initiatives such as decentralization and community driven development, this finding sends a cautioning note where initiatives in local-level decision-making are being explored.

However, it is important to carefully probe these decomposition results. Decomposing inequality into a within-group and between-group component effectively produces a summary statistic that can mask important differences. Upon closer examination of the distribution of communities in our datasets, we find that in all three countries considered, a very high percentage share of within-community inequality is perfectly consistent with a large majority of communities having levels of inequality well *below* the national level of inequality. We illustrate how this seemingly paradoxical finding is in fact fully consistent with the decomposition procedure.

Given that in our three countries we observe a significant degree of heterogeneity in inequality levels across communities, we explore in Section 7 some simple correlates. Our aim is not so much to explain local inequality (in a causal sense) but rather to explore the extent to which inequality is correlated with geographic characteristics, and whether this correlation survives the inclusion of some basic economic and demographic controls. In Section 8 we offer some concluding remarks.

2 An overview of the methodology

The survey data are first used to estimate a prediction model for consumption and then the parameter estimates are applied to the census data to derive welfare statistics. Thus, a key assumption is that the models estimated from the survey data apply to census observations. This is most reasonable if the survey and census years coincide. In this case, simple checks can be carried out by comparing the estimates to basic poverty or inequality statistics in the sample data. If different years are used but the assumption is considered reasonable, then the welfare estimates obtained refer to the census year whose explanatory variables form the basis of the predicted expenditure distribution.

An important feature of the approach applied here involves the explicit recognition that the poverty or inequality statistics estimated using a model of income or consumption are statistically imprecise. Standard errors must be calculated. The following subsections briefly summarize the discussion in Elbers et al. (2003, and 2002).

2.1 Definitions

Per capita household expenditure, y_h, is related to a set of observable characteristics, $\mathbf{x}_h$:[4]

$$\ln y_h = \mathrm{E}[\ln y_h \mid \mathbf{x}_h] + u_h \qquad (1)$$

Using a linear approximation, we model the observed log per capita expenditure for household h as:

$$\ln y_h = \boldsymbol{x}_h^T \boldsymbol{\beta} + u_h \qquad (2)$$

where $\boldsymbol{\beta}$ is a vector of parameters and u_h is a disturbance term satisfying $\mathrm{E}[u_h|\mathbf{x}_h] = 0$. In applications we allow for location effects and heteroskedasticity in the distribution of the disturbances.

The model in (2) is estimated using the household survey data. We are interested in using these estimates to calculate the welfare of an area or group for which we do not have any, or insufficient, expenditure information. Although the disaggregation may be along any dimension—not necessarily geographic—we refer to our target population as a 'county'. Household h has m_h family members. While the unit of observation for expenditure is the household, we are more often interested in welfare measures based on individuals. Thus we write W ($\mathbf{m}$, $\mathbf{X}$, $\boldsymbol{\beta}$, $\mathbf{u}$), where $\mathbf{m}$ is a vector of household sizes, $\mathbf{X}$ is a matrix of

[4] The explanatory variables are observed values and need to have the same degree of accuracy in addition to the same definitions across data sources.

observable characteristics and **u** is a vector of disturbances. Because the disturbances for households in the target population are always unknown, we estimate the expected value of the indicator given the census households' observable characteristics and the model of expenditure in (2).[5] We denote this expectation as:

$$\mu = \mathrm{E}[W \mid \mathbf{m}, \mathbf{X}, \xi] \tag{3}$$

where ξ is the vector of all model parameters, i.e., $\boldsymbol{\beta}$ and the parameters describing the distribution of **u**. In constructing an estimator of μ_v we replace the unknown vector ξ with consistent estimators, $\hat{\xi}$, from the first stage expenditure regression. This yields $\hat{\mu} = \mathrm{E}[W \mid \mathbf{m}, \mathbf{X}, \hat{\xi}]$. This expectation is generally analytically intractable so we use Monte Carlo simulation to obtain our estimator, $\tilde{\mu}$.

2.2 Estimating error components

The difference between $\tilde{\mu}$, our estimator of the expected value of W for the county, and the *actual* level of welfare for the county may be written:

$$W - \tilde{\mu} = (W - \mu) + (\mu - \hat{\mu}) + (\hat{\mu} - \tilde{\mu}) \tag{4}$$

Thus the prediction error has three components: the first due to the presence of a disturbance term in the first stage model which implies that households' actual expenditures deviate from their expected values (idiosyncratic error); the second due to variance in the first stage estimates of the parameters of the expenditure model (model error); and the third due to using an inexact method to compute $\hat{\mu}$ (computation error).[6]

Idiosyncratic error

The variance in our estimator due to idiosyncratic error falls approximately proportionately in the number of households in the county. That is, the smaller the target population, the greater is this component of the prediction error, and there is thus a practical limit to the degree of disaggregation possible. At what population size this error becomes unacceptably large depends on the explanatory power of the expenditure model and, correspondingly, the importance of the remaining idiosyncratic component of the expenditure equation (2).

Model error

The part of the variance due to model error is determined by the properties of the first stage estimators. Therefore it does not increase or fall systematically as the size of the target population changes. Its magnitude depends on the precision of the first stage coefficients and the sensitivity of the indicator to deviations in household expenditure. For a given

---

[5] If the target population includes sample survey households then some disturbances are known. As a practical matter we do not use these few pieces of direct information on *y*.

[6] Elbers et al. (2001) use a second survey in place of the census which then also introduces sampling error.

county its magnitude will also depend on the distance of the explanatory variables for households in that county from the levels of those variables in the sample data.

Computation error

The variance in our estimator due to computation error depends on the method of computation used and can be made as small as desired by increasing the number of simulations.

3. Data

In all three of the countries examined here, household survey data were combined with unit record census data. In Ecuador the poverty map is based on census data from 1990, collected by the National Statistical Institute of Ecuador (Instituto Nacional de Estadistica y Census—INEC) combined with household survey data from 1994. The census covered roughly two million households. The sample survey (Encuesta de Condiciones de Vida, ECV) is based on the Living Standards Measurement Surveys approach developed by the World Bank, and covers just under 4,500 households. The survey provides detailed information on a wide range of topics; including food consumption, non-food consumption, labor activities, agricultural practices, entrepreneurial activities, and access to services such as education and health. The survey is clustered and stratified by the country's three main agroclimatic zones and a rural-urban breakdown. It also oversamples Ecuador's two main cities, Quito and Guayaquil. Hentschel and Lanjouw (1996) develop a household consumption aggregate adjusted for spatial price variation using a Laspeyres food price index reflecting the consumption patterns of the poor. The World Bank (1996) consumption poverty line of 45,476 sucres per person per fortnight (approximately $1.50 per person per day) underlies the poverty numbers reported here. Although the 1994 ECV data were collected four years after the census, we maintain the assumption that the model of consumption in 1994 is appropriate for 1990. The period 1990-4 was one of relative stability in Ecuador. Comparative summary statistics on a selection of common variables from the two data sources support the presumption of little change over the period. Additional details on these data are found in Hentschel et al. (2000).

Three data sources were used to produce local-level poverty estimates for Madagascar. First, the 1993 unit record population census data collected by the Direction de la Démographie et Statistique Social (DDSS) of the Institut National de la Statistique (INSTAT). Second, a household survey, the Enquête Permanente Auprès des Ménages (EPM), fielded to over 4,508 households between May 1993 and April 1994, by the Direction des Statistique des Ménages (DSM) of INSTAT. Third, a set of spatial and environmental outcomes at the Fivondrona level (second administrative level or 'districts') were used with the help of GIS.[7] The consumption aggregate underpinning the Madagascar

[7] These data were provided to this project by the non-governmental organization CARE.

poverty map includes components such as an imputed stream of consumption from the ownership of consumer durables. Further details are provided in Mistiaen et al. (2002).

The Mozambique survey data used in this analysis are from the *Inquérito Nacional aos Agregados Familiares sobre as Condições de Vida, 1996-7* (IAF96). The survey is a multi-purpose household and community survey following the World Bank's LSMS format and covering 8,250 households living throughout Mozambique. The sample is designed to be nationally representative, as well as representative of each of the ten provinces, the city of Maputo and along the rural-urban dimension. As the survey was fielded over a period of 14 months, and there is significant temporal variation in food prices corresponding to the agricultural season, nominal consumption values were deflated by a temporal price index. Similarly, spatial differences in the cost of living were addressed by using a spatial deflator based on the cost of region-specific costs of basic needs poverty lines.

In this study, the IAF96 is paired with the *II Recenseamento Geral de População e Habitação* (Second General Population and Housing Census) conducted in Aurgust 1997. In addition to providing the first complete enumeration of the country's population since the initial post-independence census in 1980, the 1997 census collected information on a range of socioeconomic variables. These include educational levels and employment characteristics of those older than six years, dwelling characteristics, and ownership of some consumer durables and productive assets. The 1997 census covers approximately 16 million people living in 3.6 million households. Further details on the Mozambique data can be found in Simler and Nhate (2002).

4 Implementation

The first stage estimation is carried out using the household sample survey. For each of the three countries considered in this paper, the household survey is stratified into a number of regions and is representative at that level. Within each region there are one or more levels of clustering. At the final level, households are randomly selected from a census enumeration area. Such groups we refer to as 'cluster' and denote by a subscript c. Expansion factors allow calculation of regional totals. Our first concern is to develop an accurate empirical model of household consumption. Consider the following model:

$$\ln y_{ch} = E[\ln y_{ch} \mid x_{ch}^T] + u_{ch} = x_{ch}^T \boldsymbol{\beta} + \eta_c + \varepsilon_{ch} \tag{5}$$

where η and ε are independent of each other and uncorrelated with observables. This specification allows for an intracluster correlation in the disturbances. One expects location to be related to household income and consumption, and it is certainly plausible that some of the effect of location might remain unexplained even with a rich set of regressors. For any given disturbance variance, σ_{ch}^2, the greater the fraction due to the common component η_c, the less one benefits from aggregating over more households. Welfare estimates become less precise. Further, failing to account for spatial correlation in the disturbances could bias the inequality estimates.

Thus the first goal is to explain the variation in consumption due to location as much as possible with the choice and construction of explanatory variables. We tackle this in four ways:

1. We estimate different models for different strata in the countries' respective surveys.
2. We include in our specification household level indicators of access to various networked infrastructure services, such as electricity, piped water, networked waste disposal, telephone etc. To the extent that all or most households within a given neighborhood or community are likely to enjoy similar levels of access to such networked infrastructure, these variables might capture unobserved location effects.
3. We calculate means at the enumeration area (EA) level in the census (generally corresponding to the 'cluster' in the household survey) of household level variables, such as the average level of education of household heads. We then merge these EA means into the household survey and consider them for inclusion in the first stage regression specification.[8]
4. Finally, in the case of Madagascar we have merged a Fivondrona level dataset provided by CARE and considered these spatially referenced environmental variables, such as droughts and cyclones, for inclusion in our household expenditure models.

To select variables to reduce location effects, we regress the total residuals, $\hat{u}$, on cluster fixed effects. We then regress the cluster fixed-effect parameter estimates on our location variables and select a limited number that best explain the variation in the cluster fixed-effects estimates. These location variables are then included in the first stage regression model.

A Hausman test described in Deaton (1997) is used to determine whether to estimate with household weights. $\overline{R}^2$'s for our models are generally high, ranging between 0.45 and 0.77 in Ecuador, 0.29 to 0.63 in Madagascar, and 0.27 to 0.55 in Mozambique.[9] We next model the variance of the idiosyncratic part of the disturbance, $\sigma^2_{\varepsilon,ch}$. The total first stage residual can be decomposed into uncorrelated components as follows:

$$\hat{u}_{ch} = \hat{u}_{c.} + (\hat{u}_{ch} - \hat{u}_{c.}) = \hat{\eta}_c + e_{ch} \qquad (6)$$

where a subscript '.' indicates an average over that index. Thus the mean of the total residuals within a cluster serves as an estimate of that cluster's location effect. To model heteroskedasticity in the household-specific part of the residual, we choose somewhere

8 In Madagascar the EA in the household survey is not the same as that in the census. The most detailed spatial level at which we can link the two datasets is the Firaisana ('commune'). Thus, Firaisana-level means were used.

9 Again, see Elbers et al. (2002), Mistiaen et al. (2001) and Simler and Nhate (2002) for details.

between 5 and 20 variables, $\mathbf{z}_{ch}$, that best explain variation in e_{ch}^2 out of all potential explanatory variables, their squares, and interactions.[10]

Finally, we determine the distribution of η and ε using the cluster residuals $\hat{\eta}_c$ and standardized household residuals $e_{ch}^* = \frac{e_{ch}}{\hat{\sigma}_{\varepsilon,ch}} - [\frac{1}{H}\sum_{ch}\frac{e_{ch}}{\hat{\sigma}_{\varepsilon,ch}}]$, respectively where H is the number of households in the survey. We use normal or t distributions with varying degrees of freedom (usually 5), or the actual standardized residual distribution mentioned above when taking a semi-parametric approach. Before proceeding to simulation, the estimated variance-covariance matrix is used to obtain final GLS estimates of the first stage consumption model. At this point we have a full model of consumption that can be used to simulate any expected welfare measures with associated prediction errors. For a description of different approaches to simulation see Elbers et al. (2000).

5 Stratum-level comparisons between survey and census

In this section we examine the degree to which our census-based estimates match estimates from the countries' respective surveys at the level at which those surveys are representative.[11] Table 1 presents estimates for Ecuador of average per capita consumption, the headcount poverty rate and the Gini-coefficient inequality measure from both the household survey and census at the level of the 8 strata at which the household survey is representative. Standard errors are presented for all estimates—reflecting the complex sample design of the household survey for the survey-based estimates, and our imputation procedure for the census based estimates (as described above). In nearly every case, the estimates across the two data sources are within each other's 95 percent confidence interval. In fact, it is striking how closely the point estimates match, particularly for the average consumption and headcount rates.

In the case of the inequality measure, we can see that the census estimates tend to be higher than the survey based estimates, although not generally to such an extent that one can reject that they are the same. The propensity to produce higher estimates of inequality from the imputed census data arises from the fact that inequality measures tend to be sensitive to the tails in the distribution of expenditure. Since the tails are typically not observed in the survey (because of its small size), the survey underestimates inequality.

[10] We limit the number of explanatory variables to be cautious about overfitting and use a bounded logistic functional form.

[11] For a similar analysis, focusing specifically on poverty, see Demombynes et al. (2002).

Table 1: Average expenditure, poverty, and inequality in Ecuador by region (stratum)

Region	Survey Estimate			Census-Based Estimate		
	Mean	FGT(0)	Gini	Mean	FGT(0)	Gini
Quito	126,098	0.25	0.490	125,702	0.23	0.465
	(11344)	(0.033)	(0.023)	(8026)	(0.024)	(0.012)
Urban Sierra	121,797	0.19	0.436	122,415	0.22	0.434
	(8425)	(0.026)	(0.020)	(4642)	(0.017)	(0.011)
Rural Sierra	66,531	0.43	0.393	63,666	0.53	0.457
	(4067)	(0.027)	(0.034)	(2213)	(0.019)	(0.013)
Guayaquil	89,601	0.29	0.378	77,432	0.38	0.416
	(5597)	(0.027)	(0.014)	(2508)	(0.019)	(0.011)
Urban Costa	86,956	0.25	0.359	90,209	0.26	0.382
	(3603)	(0.030)	(0.015)	(2391)	(0.015)	(0.011)
Rural Costa	57,617	0.50	0.346	61,618	0.50	0.400
	(4477)	(0.042)	(0.036)	(2894)	(0.024)	(0.015)
Urban Oriente	110,064	0.20	0.398	174,529	0.19	0.563
	(9078)	(0.050)	(0.035)	(56115)	(0.02)	(0.104)
Rural Oriente	47,072	0.67	0.431	59,549	0.59	0.478
	(4420)	(0.054)	(0.034)	(3051)	(0.025)	(0.014)

Source: See text.

9

Table 2: Average expenditure, poverty, and inequality in Madagascar by province and sector

Province	Survey Estimate			Census-Based Estimate		
	Mean Expenditure	Headcount Index	Gini Coefficient	Mean Expenditure	Headcount Index	Gini Coefficient
URBAN						
Antananarivo	513,818 (48,455)	.544 (.048)	.492 (.027)	576,470 (23,944)	.462 (.015)	.469 (.012)
Fianarantsoa	360,635 (42,613)	.674 (.059)	.430 (.038)	372,438 (21,878)	.646 (.027)	.426 (.015)
Taomasina	445,514 (73,099)	.599 (.086)	.434 (.042)	417,823 (15,406)	.599 (.018)	.402 (.015)
Mahajanga	613,867 (74,092)	.329 (.072)	.371 (.027)	580,775 (31,025)	.378 (.028)	.392 (.016)
Toliara	343,111 (76,621)	.715 (.086)	.514 (.052)	321,602 (32,193)	.713 (.036)	.504 (.030)
Antsiranana	504,841 (46,148)	.473 (.087)	.362 (.025)	693,161 (93,437)	.344 (.031)	.433 (.039)
RURAL						
Antananarivo	312,553 (23,174)	.767 (.037)	.376 (.023)	324,814 (14,378)	.738 (.019)	.404 (.015)
Fianarantsoa	319,870 (45,215)	.769 (.049)	.470 (.050)	251,312 (18,091)	.820 (.025)	.437 (.018)
Taomasina	275,943 (22,832)	.810 (.035)	.352 (.036)	279,239 (15,838)	.786 (.026)	.362 (.017)
Mahajanga	325,872 (30,209)	.681 (.065)	.320 (.026)	321,398 (19,385)	.695 (.039)	.306 (.015)
Toliara	233,801 (22,174)	.817 (.042)	.383 (.029)	259,537 (16,222)	.800 (.027)	.377 (.017)
Antsiranana	486,781 (91,181)	.613 (.073)	.518 (.110)	442,431 (54,869)	.581 (.046)	.453 (.048)

Source: See text.

Note: All figures based on a poverty line of 354,000 Malagasy Francs per capita. Household survey figures are calculated using weights that are the product of household survey weights and household size. Census-based figures are calculated weighting by household size.

10

Table 3: Average expenditure, poverty, and inequality in Mozambique by province

Province	Survey Estimate			Census-Based Estimate		
	Mean Expenditure	Headcount Index	Gini Coefficient	Mean Expenditure	Headcount Index	Gini Coefficient
Niassa	4660 (355)	0.71 (0.038)	0.355 (0.020)	5512 (484)	0.67 (0.042)	0.402 (0.025)
Cabo Delgado	6392 (416)	0.57 (0.042)	0.370 (0.025)	6586 (433)	0.56 (0.036)	0.413 (0.021)
Nampula	5315 (287)	0.69 (0.032)	0.391(0.026)	5547 (279)	0.65 (0.024)	0.400 (0.020)
Zambezia	5090 (208)	0.68 (0.026)	0.324 (0.017)	5316 (274)	0.67 (0.029)	0.366 (0.012)
Tete	3848 (267)	0.82 (0.032)	0.346 (0.019)	4404 (176)	0.77 (0.016)	0.394 (0.018)
Manica	6299 (741)	0.63 (0.059)	0.413 (0.036)	6334 (527)	0.62 (0.044)	0.449 (0.020)
Sofala	3218 (191)	0.88 (0.015)	0.405 (0.031)	4497 (379)	0.78 (0.017)	0.529 (0.032)
Inhambane	4215 (359)	0.83 (0.024)	0.382 (0.037)	4177 (134)	0.81 (0.013)	0.398 (0.012)
Gaza	6024 (356)	0.65 (0.033)	0.380 (0.024)	6521 (355)	0.59 (0.021)	0.421 (0.023)
Maputo Province	5844 (613)	0.66 (0.054)	0.424 (0.029)	8559 (745)	0.55 (0.024)	0.518 (0.029)
Maputo City	8321 (701)	0.48 (0.041)	0.444 (0.033)	11442 (4956)	0.49 (0.047)	0.560 (0.108)

Source: See text.

Note: All figures based on a poverty line of 5433 Meticais daily per capita. Survey figures are calculated using weights that are the product of household survey weights and household size. Census-based figures are calculated weighting by household size.

11

Tables 2 and 3 present results analogous to those presented in Table 1 for Madagascar and Mozambique, respectively. Again, the results indicate that at the stratum level there is little basis for rejecting equality of the survey- and census-based estimates of average per capita consumption, poverty and inequality in the two countries. In Madagascar, standard errors on the survey estimates are quite high, indicating that while the household survey may be representative at the province and sector level, the sample size in these strata is rather smallso that estimates are imprecise. Nonetheless, for our purposes it is encouraging to note that point estimates across all three welfare indicators are often remarkably close.

In Mozambique, as in Ecuador (but less markedly so in Madagascar), inequality estimates tend to be higher than the survey estimates. In some provinces, such as Sofala, Maputo Province and Maputo City, the estimates are not only very high, but are also quite imprecisely estimated in the census. Although these census-level standard errors are large it is due primarily to model error. As a result, and as we shall see below, there is no evidence that estimates become even more noisy at lower levels of aggregation.

6 Decomposing inequality by geographic subgroups

We turn in this section to the important question of how much of overall inequality in a given country is attributable to differences in average consumption across localities as opposed to inequality within localities. It is clear that where national inequality is largely due to differences in mean income across regions, the policy implications are very different from the situation where sub-regions themselves are unequal and national inequality is simply an expression at the country level of a degree of heterogeneity that already exists at the more local level. Decomposing inequality by subgroups enjoys a long tradition in the empirical analysis of inequality, in both developed and developing countries. We decompose inequality using the general entropy class of inequality measures, a class of measures which is particularly well-suited to this exercise. [12] This class of measures takes the following form:

$$I_c = \frac{1}{c(c-1)} \sum_i f_i [(\frac{y_i}{\mu})^c - 1] \qquad \text{for } c \neq 0,1$$

$$= -\sum_i f_i \log(\frac{y_i}{\mu}) \qquad \text{for } c = 0$$

$$= \sum_i f_i \frac{y_i}{\mu} \log(\frac{y_i}{\mu}) \qquad \text{for } c = 1$$

[12] Following Bourguignon (1979), Shorrocks (1980) and Cowell (1980). Cowell (2000) provides a useful recent survey of methods of inequality measurement, including a discussion of the various approaches to subgroup decomposition. Sen and Foster (1997) and Kanbur (2000) discuss some of the difficulties in interpreting results from such decompositions.

where f_i is the population share of household i, y_i is per capita consumption of household i, μ is average per capita consumption, and c is a parameter that is to be selected by the user.[13] This class of inequality measures can be decomposed into a between and within-group component along the following lines:

$$I_c = \frac{1}{c(c-1)}[1 - \sum_j g_j[(\frac{\mu_j}{\mu})^c] + \sum_j I_j g_j (\frac{\mu_j}{\mu})^c \qquad \text{for } c \neq 0,1$$

$$I_c = [g_j \log(\frac{\mu}{\mu_j})] + \sum_j I_j g_j \qquad \text{for } c = 0$$

$$I_c = [\sum_j g_j (\frac{\mu_j}{\mu}) \log(\frac{\mu_j}{\mu})] + \sum_j I_j g_j (\frac{\mu_j}{\mu}) \qquad \text{for } c = 1$$

where j refers to subgroups, g_j refers to the population share of group j and I_j refers to inequality in group j. The between-group component of inequality is captured by the first term to the right of the equality sign. It can be interpreted as measuring what would be the level of inequality in the population if everyone within the group had the same (the group-average) consumption level μ_j. The second term on the right reflects what would be the overall inequality level if there were no differences in mean consumption across groups but each group had its actual within-group inequality I_j. Ratios of the respective components with the overall inequality level provide a measure of the percentage contribution of between-group and within-group inequality to total inequality.

In Table 4 we examine how within-group inequality evolves at progressively lower levels of regional disaggregation in our three countries. At one extreme, when a country-level perspective is taken, all inequality is, by definition, within-group. At the other extreme, when each individual household is taken as a separate group, the within-group contribution to overall inequality is zero. But how rapidly does the within-group share fall? Is it reasonable to suppose that at a sufficiently low level of disaggregation, such as the 3rd administrative level in our three countries (with about 1,000-10,000 households) differences within groups are small, and most of overall inequality is due to differences between groups?

We decompose inequality in our three countries on the basis of the GE(0) measure.[14] In rural Ecuador we see that when we have disaggregated down to the level of 915 'parroquias' (with an average number of households of a little over 1,000) some 86 percent of overall inequality remains within-group. In urban areas of Ecuador, the within-group

[13] Lower values of c are associated with greater sensitivity to inequality amongst the poor, and higher values of c place more weight to inequality among the rich. A c value of 1 yields the well known Theil entropy measure, a value of 0 provides the Theil L or mean log deviation, and a value of 2 is ordinally equivalent to the squared coefficient of variation.

[14] Results remain virtually identical for other values of c.

Table 4: Decomposition of inequality by regional subgroup (GE0)

Level of Decomposition	No. of Subgroups	Within-Group (%)	Between-Group (%)
	Ecuador		
RURAL			
National	1	100	0
Region	3	100	0
Province	21	98.7	1.3
Canton	195	94.1	5.9
Parroquia	915	85.9	14.1
Household	960,529	0	100
URBAN			
National	1	100	0
Region*	5	100	6.6
Province*	19	98.7	7.3
Canton*	87	94.1	8.6
Zonas	664	85.9	23.3
Household	880,001	0	100
	Madagascar		
URBAN	1	100	0
Faritany	6	92.3	7.7
Fivondrona	103	78.3	21.7
Firaisana	131	76.7	23.2
RURAL	1	100	0
Faritany	6	95.2	4.8
Fivondrona	104	84.6	15.4
Firaisana	1117	81.9	18.1
	Mozambique		
National	1	100	0
Province	11	90.7	9.3
District	146	81.6	18.4
Administrative Post	424	78.0	22.0

Source: See text.

Note: Quito and Guayaquil are treated as independent geographic areas.

share, across 664 urban 'zonas' (with 1,300 households on average) is only slightly lower at 77 percent. The same pattern obtains in Madagascar and Mozambique (Table 4). In all three countries no less than three quarters of all inequality is attributable to within-community differences, even after one has disaggregated down to a very low level (corresponding, in our countries, to the lowest level of central government administration). At first glance, one might understand these results as suggesting that even within local communities there exists a considerable heterogeneity of living standards. Such a conclusion might have implications regarding the likelihood of political capture, the feasibility of raising revenues locally, and the extent to which residents in these localities can be viewed as having similar demands and priorities.

However, a blanket statement about the degree of inequality within communities does not follow directly from the above decomposition results. It is important to recognize that the decomposition exercise indicates that *on average* inequality does not fall much with aggregation level. In other words, it is very well possible that at low levels of aggregation the population is characterized by both highly equal and highly unequal communities. A simple example can illustrate this. Consider a population of 8 individuals with consumption values (1,1,2,2,4,4,5,5). This population could be divided into two communities as (1,2,4,5) and (1,2,4,5); or as (1,1,5,5) and (2,2,4,4). In both cases the two communities have the same average consumption. As a result the between-group component from the decomposition exercise is always zero (and thus the within-group share is 100 percent in both cases). However, in the first case inequality in the two communities is exactly equal to national inequality, whereas in the second case one community has higher and the other lower inequality than at the national level. As can be readily seen from the expressions for decomposing the General Entropy class of inequality measures provided above, when average consumption levels are the same for all communities, overall inequality is calculated by taking a population-weighted average of community-level inequality rates. Finding a high within-group share from a decomposition exercise across a large number of communities is thus perfectly consistent with great heterogeneity in inequality levels across communities.

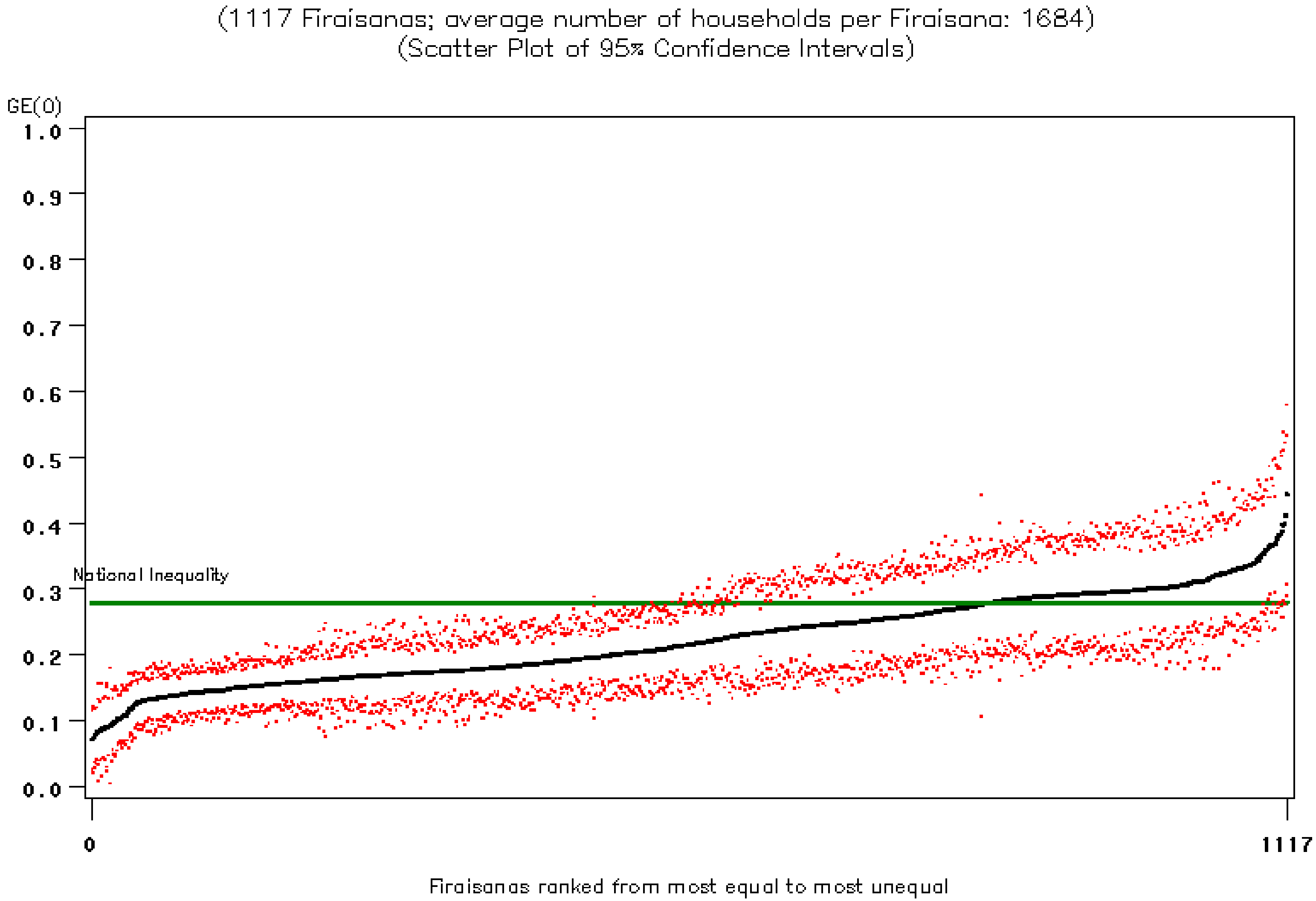
In a situation, such as ours, where the decomposition exercise is carried out across a very large number of communities, it is important to check for variation in the degree of inequality across communities. Are all communities as unequal as the country as a whole? Such a finding would certainly generate a large within-group contribution in a decomposition exercise. Or do communities vary widely in their degree of inequality? That could also yield a high within-group share. In Figures 1-5 we plot community-level inequality estimates and compare these against national-level inequality. Communities are ranked from most equal to most unequal, and 95 percent confidence intervals on each community-level estimate are included as scatter plots.

Figure 1 compares parroquia-level inequality in rural Ecuador against the overall inequality level in rural areas. We see that although the within-group share from the decomposition exercise was as high as 86 percent, this summary statistic masks considerable variation in parroquia inequality levels. A large majority of parroquia-level point estimates are well below the national level in rural Ecuador. Even allowing for the imprecision around the parroquia-level estimates (which are typically 5-15 percent of the point estimate), a sizeable proportion of parroquias are unambiguously more equal than the

Figure 1: Rural Ecuador: distribution across parroquias of parroquia-level inequality

Figure 2: Urban Ecuador: distribution across zonas of zona-leveil inequality

Figure 3: Rural Madagascar: distributino across firaisanas of firaisana-level inequality



18

Figure 4: Urgan Madagascar: distribution across firaisanas of firaisana-level inequality

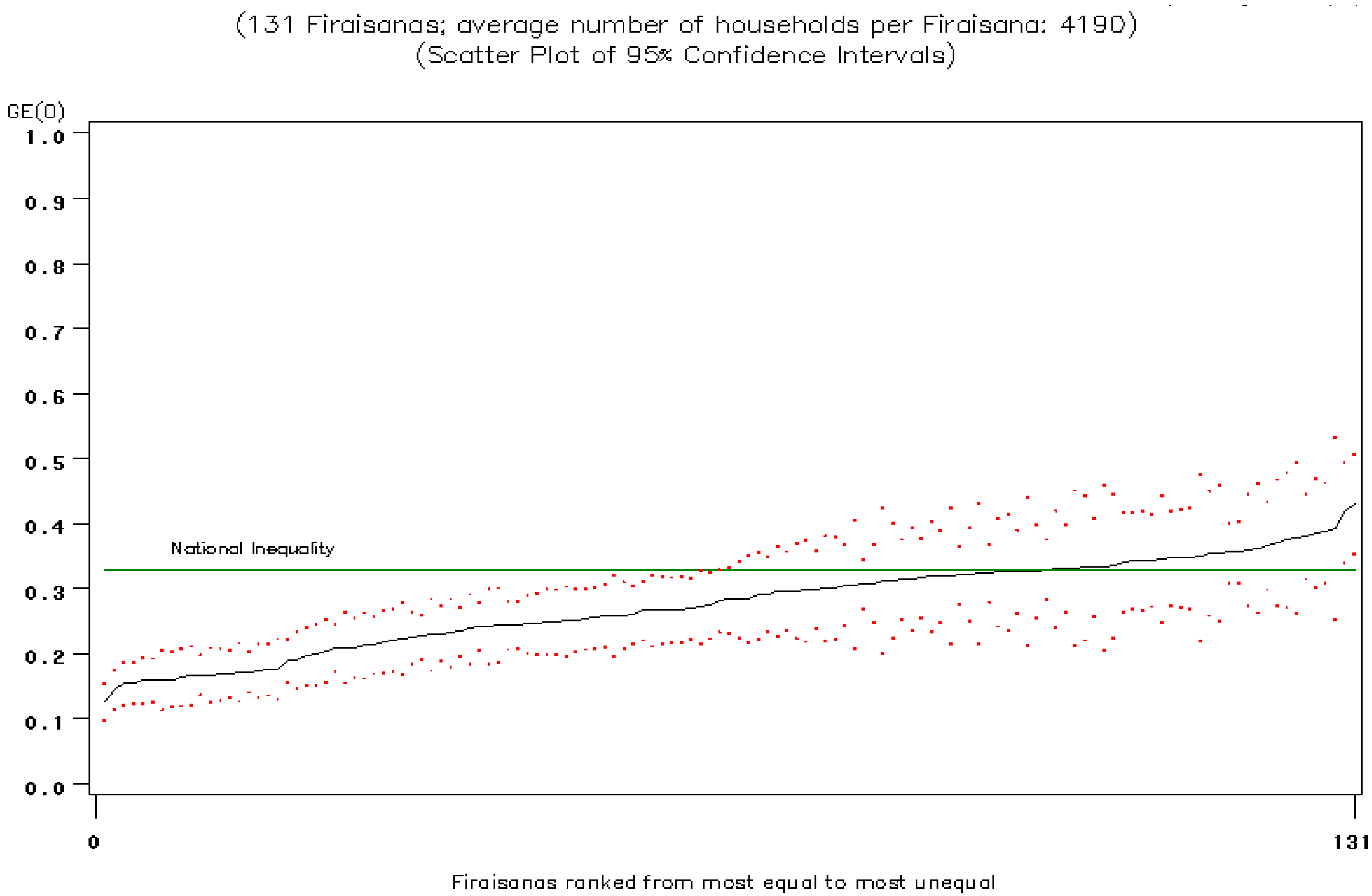
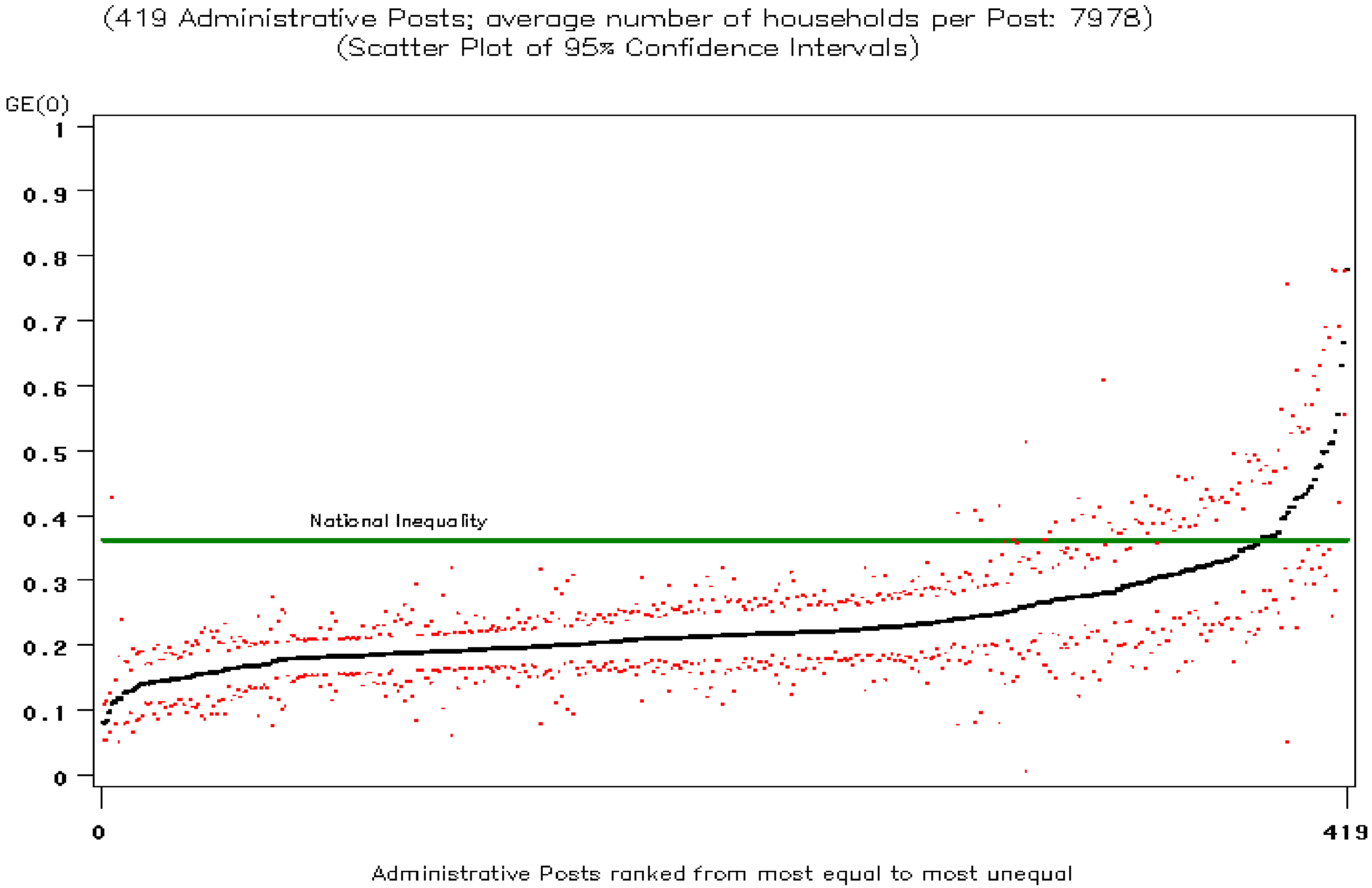


Figure 5: Mozambique: distribution across administrative posts of post-level inequality



picture at the national level. Another sizeable proportion of zonas that have lower inequality than the national-level inequality rate is even higher than in rural areas. The precision of point estimates in urban areas of Ecuador is somewhat higher than in rural areas; accordingly, more zonas lie unambiguously below the national inequality level.

In rural and urban Madagascar (Figures 3 and 4) and in Mozambique (Figure 5) the picture is very similar. In all of the countries considered in this study, there is a clear and sizeable subset of communities with lower inequality than the country as a whole; another large group for which inequality is not significantly different from inequality in the country as a whole; and a small third group of communities with inequality higher than the national level.

7 Correlates of local inequality: does geography matter?

We have found empirical support for both the view that at the local level communities are more homogeneous than society as a whole, and the view that local communities are as heterogeneous as society as whole. The question then arises as to whether it is possible to readily distinguish between communities on the basis of some simple indicators. In particular, we are interested to know whether there are discernable geographic patterns of inequality.

Tables 5a-5e we provide results from OLS regressions of inequality on a set of simple community characteristics. We ask whether inequality levels are correlated with location, controlling for both demographic characteristics of the communities (population size and demographic composition), and mean Per capita consumption. Table 5a for rural Ecuador, finds strong evidence that inequality in the parroquias of the eastern, Oriente, region is significantly higher than province of Pichincha in the central, mountainous, Sierra, region. Communities located in provinces in the western, coastal, Costa, region tend to be more equal, significantly so in the provinces of Manabi, Los Rios, Guayas and El Oro. Relatively few differences are discernable across provinces within the Sierra region.[15] Understanding these geographic patterns of inequality is beyond the scope of this paper, but the evidence is consistent with historical and anecdotal accounts of very a divergent evolution of society and economic structures in the mountainous Sierra vis-à-vis the Costa and Oriente.[16]

In rural Ecuador, there is evidence that larger parroquias tend to be more unequal. An interesting finding is that parroquias with a larger proportion of elderly, relative to the population share of 20-40 year olds, are more unequal. This pattern is consistent with the

[15] We can reject with 95 percent confidence, for both rural and urban Ecuador, the null hypothesis that parameter estimates on province dummies within their respective regions are all equal.

[16] See, for example, 'Under the Volcano', *The Economist*, 27 November 1999 (p.66).

Table 5a: Correlates of mean log deviation (GE0) in rural Ecuador: parroquia-level regression (915 parroquias)

	Basic Regression	+ expenditure
Log population	0.0169 (0.002)***	0.010 (0.002)***
% aged 0-10	-0.139 (0.079)*	0.321 (0.080)***
% aged 10-20	-0.375 (0.104)***	-0.084 (0.096)
% aged 40-60	-0.246 (0.130)*	0.053 (0.120)
% aged 61+	0.269 (0.123)***	0.392 (0.112)***
Log mean per capita expenditure		0.222 (0.085)***
(Log mean per capita expenditure)^2		-0.014 (0.010)
Oriente		
Sucumbios	0.036 (0.013)***	0.036 (0.012)***
Napo	0.051 (0.012)***	0.056 (0.011)***
Pastaza	0.071 (0.015)***	0.077 (0.013)***
Morona_Santiago	0.040 (0.011)***	0.036 (0.010)***
Zamora_Chinchipe	0.034 (0.013)**	0.037 (0.012)***
Costa		
Esmeraldas	-0.012 (0.010)	-0.036 (0.010)***
Manabi	-0.060 (0.010)***	-0.057 (0.009)***
Los Rios	-0.041 (0.013)***	-0.025 (0.012)**
Guayas	-0.050 (0.010)***	-0.035 (0.009)***
El Oro	-0.022 (0.010)**	-0.020 (0.009)**
Galápagos	0.027 (0.023)	-0.000 (0.021)
Sierra		
Carchi	-0.002 (0.012)	0.014 (0.010)
Imbabura	0.024 (0.010)**	0.037 (0.011)***
Cotopaxi	-0.013 (0.011)	-0.001 (0.010)
Tungurahua	-0.025 (0.010)**	-0.010 (0.009)
Bolivar	-0.0002 (0.012)	0.002 (0.011)
Chimborazo	-0.010 (0.010)	0.006 (0.010)
Canar	0.003 (0.012)	0.007 (0.011)
Azuay	0.011 (0.010)	0.014 (0.009)
Loja	0.024 (0.009)**	0.036 (0.008)***
Constant	0.296 (0.060)	-0.571 (0.192)
Observations	915	915
R-squared	0.24	0.38

Source: See text.

Note: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Excluded groups are Pichincha and % population age 20-40.

findings of Deaton and Paxson (1995) regarding the positive association between an aging population and inequality. The quantitative importance and statistical significance of both geographic and demographic characteristics remains broadly unchanged when mean per capita consumption (and its square) are added to the model. In rural Ecuador inequality is positively associated with higher consumption levels. While there is some suggestion of a

Table 5b: Correlates of mean log deviation (ge0) in urban Ecuador: zona-level regression (660 zonas)

	Basic Regression	+ expenditure
Log population	-0.013 (0.015)	-0.003 (0.014)
% aged 0-10	0.231 (0.118)*	0.253 (0.119)**
% aged 10-20	0.283 (0.098)***	0.791 (0.112)***
% aged 40-60	0.001 (0.141)	-0.673 (0.162)***
% aged 61+	0.704 (0.162)***	1.084 (0.161)***
Log mean per capita expenditure		0.025 (0.075)
(Log mean per capita expenditure)^2		0.005 (0.008)
Oriente		
Pastaza	0.052 (0.033)	0.049 (0.031)
Morona_Santiago	0.457 (0.046)***	0.381 (0.045)***
Zamora_Chinchipe	0.031 (0.046)	0.004 (0.044)
Costa		
Esmeraldas	-0.073 (0.013)***	-0.066 (0.012)***
Manabi	-0.084 (0.007)***	-0.069 (0.007)***
Los Rios	-0.077 (0.010)***	-0.049 (0.011)***
Guayas	-0.097 (0.008)***	-0.064 (0.008)***
El Oro	-0.094 (0.009)***	-0.081 (0.009)***
Guayaquil	-0.087 (0.005)***	-0.054 (0.007)***
Sierra		
Carchi	-0.009 (0.017)	0.012 (0.017)
Imbabura	0.022 (0.014)	-0.008 (0.013)
Cotopaxi	0.007 (0.016)	0.006 (0.015)
Tungurahua	-0.008 (0.014)	-0.003 (0.013)
Pichincha	-0.011 (0.010)	-0.000 (0.010)
Chimborazo	-0.025 (0.015)*	-0.026 (0.014)*
Canar	-0.012 (0.024)	-0.018 (0.022)
Azuay	-0.013 (0.010)	-0.018 (0.010)*
Loja	-0.003 (0.013)	-0.010 (0.012)
Constant	0.272 (0.140)	-0.076 (0.242)
Observations	660	660
R-squared	0.52	0.57

Source: See text.

Note: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Excluded groups are Quito and % population age 20-40.

turning point (at around $2,800 per capita per month)—the well known 'inverted U-curve'—the statistical support for this is weak. The correlation between inequality and the population share of young children, relative to 20-40 year olds, switches in sign from negative to positive, depending on whether per capita consumption is included in the specification. It seems clear that the share of young children is likely to be (negatively) correlated with per capita consumption so that the coefficient on this variable is capturing the consumption effect, when average expenditures are excluded from the specification. Once consumption expenditures are controlled for, the correlation between inequality and the share of children in the population becomes positive. Possibly there exists greater heterogeneity in household size in those parroquias with large population shares of young children and that this translates into greater inequality of per capita consumption.

Table 5c: Correlates of mean log deviation (GE0) in rural Madagascar: firaisana-level regression (1,117 firaisanas)

	Basic Regression	+ expenditure
Log population	0.010 (0.002)***	0.012 (0.002)***
% aged 0-5	-0.768 (0.085)***	-0.700 (0.086)***
% aged 6-11	-0.226 (0.127)*	-0.091 (0.126)
% aged 12-14	0.193 (0.241)	0.236 (0.242)
% aged 50-59	-1.757 (0.292)***	-1.747 (0.286)***
% aged 60+	0.462 (0.152)**	0.696 (0.152)***
Log mean per capita expenditure		0.886 (0.118)***
(Log mean per capita expenditure)^2		-0.034 (0.005)***
Provinces		
Antananarivo	-0.068 (0.006)***	-0.065 (0.006)***
Fianarantsoa	0.011 (0.005)**	0.020 (0.006)***
Toamasina	-0.059 (0.006)***	-0.054 (0.006)***
Mahajanga	-0.115 (0.006)***	-0.116 (0.006)***
Toliara	-0.046 (0.005)***	-0.042 (0.006)***
Constant	0.430 (0.041)***	-5.356 (0.765)***
Observations	1117	1117
R-squared	0.53	0.55

Source: See text.

Note: Standard errors in parentheses. * significant at 10%; ** significant at 5%. *** significant at 1%.
Excluded groups are Antsiranana and % population age 15-49.

In urban Ecuador (Table 5b) the relatively low inequality in the Costa region is again observed. Relative to the zonas in the capital Quito, inequality in all zona*s* of the costa region tends to be significantly lower. Other urban areas in the Sierra are again not noticeably less or more equal than Quito. In urban areas, in contrast to rural areas, population size of the zona does not appear to be significantly correlated with its inequality level.[17] Also in contrast to rural areas, conditioning on mean consumption levels does not add much explanatory power: there is no evidence that poorer zonas are also more equal.

[17] Although zonas vary less in population size than parroquias, they still range between 800-1,900 households.

Zonas with large dependency ratios (irrespective of whether these are due to many young children or of a large proportion of elderly) are associated with higher inequality levels, irrespective of controlling for consumption.

Tables 5c and 5d provide analogous results for Madagascar. The broad conclusions are quite similar to those found in Ecuador. As in rural Ecuador, in rural Madagascar population size is positively associated with inequality, and the larger the percentage of elderly in the firaisana the more unequal the community. As in Ecuador, inequality rises with mean consumption (in the Madagascar case the inverted U curve is more clearly discernable) and geography is strongly and independently significant. Relative to the population share aged 15-50, the higher the share of children and the share of population aged 50-59 the more equal the community, whether or not one controls for consumption. In Madagascar it seems that communities with large population shares of children are not markedly more heterogeneous in household size. For rural Madagascar the simple specification employed here yields an R^2 as high as 0.55 when all variables are included.

Table 5d: Correlates of mean log deviation (GE0) in urban Madagascar firaisana-level regression (131 firaisanas)

	Basic Regression	+ expenditure
Log population	-0.014 (0.005)***	-0.011 (0.005)**
% aged 0-5	-1.253 (0.202)***	-1.053 (0.243)***
% aged 6-11	0.166 (0.464)	0.147 (0.465)
% aged 12-14	-0.965 (0.777)	-0.551 (0.826)
% aged 50-59	-2.602 (0.882)***	-2.543 (0.882)***
% aged 60+	1.183 (0.396)***	1.355 (0.417)***
Log mean per capita expenditure		0.117 (0.143)
(Log mean per capita expenditure)^2		-0.004 (0.013)
Provinces		
Antananarivo	0.079 (0.015)***	0.080 (0.015)***
Fianarantsoa	0.059 (0.014)***	0.065 (0.015)***
Toamasina	-0.012 (0.014)	-0.007 (0.015)
Mahajanga	-0.025 (0.014)*	-0.027 (0.014)*
Toliara	0.117 (0.013)***	0.125 (0.014)***
Constant	0.717 (0.106)	-0.270 (2.245)
Observations	131	131
R-squared	0.78	0.79

Source: See text.

Note: Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.
Excluded groups are Antsiranana and % population age 15-49.

In urban Madagascar the explanatory power is even greater (Table 5d). Here, unlike rural areas, population size is significantly negatively associated with inequality. As in rural areas, the larger the percentage of children the lower is inequality. As in urban Ecuador, mean per capita consumption is not significantly associated with inequality—there is no presumption that a poorer urban firaisana is more homogeneous than a rich one. Geographic variables remain independently significant, with urban areas in Antananarivo (the capital province), Fianarantsoa, and Toliara more unequal than the urban areas in the rest of the country. Table 5e confirms that in Mozambique, too, geographic variables are key indicators of local-level inequality, controlling for population characteristics, mean expenditure levels, and urban-rural differences. Compared with Maputo city, the rest of the country has significantly less inequality. There is more inequality in urban areas, an increasing association with mean consumption (but no Kuznets curve), and areas with higher percentage of 17-30 year-olds seem to have higher inequality.

Table 5e: Correlates of mean log deviation (GE0) in Mozambique: administrative post-level regression (464 administrative posts)

	Basic Regression	+ expenditure	+ urban
Pct aged 0-5	-0.002 (0.004)	0.000 (0.003)	0.001 (0.003)
Pct aged 6-10	0.017 (0.005)**	0.015 (0.004)**	0.014 (0.004)**
Pct females aged 11-16	0.027 (0.009)**	0.020 (0.008)**	0.015 (0.008)
Pct males aged 11-16	-0.000 (0.009)	0.001 (0.008)	0.002 (0.008)
Pct females aged 17-30	0.016** (0.004)	0.012 (0.003)**	0.011 (0.003)**
Pct males aged 17-30	0.015 (0.005)**	0.010 (0.004)*	0.009 (0.004)*
Pct females aged 31-60	0.005 (0.006)	0.005 (0.006)	0.007 (0.006)
Pct males aged 31-60	0.007 (0.004)	0.005 (0.004)	0.004 (0.004)
Log (population of posto)	0.001 (0.004)	-0.003 (0.003)	-0.005 (0.004)
Niassa	-0.200 (0.036)**	-0.138 (0.034)**	-0.136 (0.033)**
Cabo Delgado	-0.204 (0.034)**	-0.163 (0.031)**	-0.158 (0.031)**
Nampula	-0.204 (0.035)**	-0.143 (0.032)**	-0.143 (0.032)**
Zambézia	-0.215 (0.035)**	-0.154 (0.032)**	-0.149 (0.032)**
Tete	-0.212 (0.036)**	-0.133 (0.033)**	-0.127 (0.033)**
Manica	-0.135 (0.035)**	-0.095 (0.032)**	-0.089 (0.032)**
Sofala	-0.118 (0.035)**	-0.005 (0.032)	-0.006 (0.032)
Inhambane	-0.178 (0.035)**	-0.088 (0.032)**	-0.090 (0.032)**
Gaza	-0.189 (0.035)**	-0.136 (0.032)**	-0.135 (0.031)**
Maputo Province	-0.088 (0.036)*	-0.045 (0.032)	-0.044 (0.032)
Log (mean expenditure)		-0.406 (0.216)	-0.324 (0.217)
Log (mean expenditure)^2		0.031 (0.013)*	0.025 (0.013)*
Urban			0.037 (0.014)**
Constant	-0.504 (0.321)	0.856 (0.962)	0.605 (0.960)
Observations	424	424	424
R-squared	0.465	0.595	0.601

Source: See text.

Note: Standard errors in parentheses. * significant at 5%; ** significant at 1%. Excluded groups are Maputo city and % persons older than 60 years.

We have not attempted here to identify the best possible set of correlates of local inequality for each of the three countries we are examining. We have chosen to employ a parsimonious, and broadly similar, specification in the three countries in order to ask whether there are any common patterns across countries which in other respects resemble each other very little (particularly the comparison between Ecuador and the two sub-saharan African countries). We have indeed found that in all three countries we consider, in both rural and urban areas, geographic location is a good predictor of local-level inequality, even after controlling for some basic demographic and economic characteristics of the communities. With respect to other characteristics, there appear to be clear differences between urban and rural areas (best seen in the models for Ecuador and Madagascar). In rural areas inequality tends to be higher in communities with larger populations, a higher share of the elderly in the total population, and in communities with higher mean consumption levels. In urban areas, mean consumption is not independently correlated with inequality, and inequality is not typically higher in communities with larger populations. High population shares of elderly are clearly associated with higher inequality, but the correlation with population shares of children depends on the country.

8 Conclusions

This paper has taken three developing countries, Ecuador, Madagascar and Mozambique, and has implemented in each a methodology to produce disaggregated estimates of inequality. The countries are very unlike each other—with different geographies, stages of development, quality and types of data, and so on. The methodology works well in all three settings and produces valuable information about the spatial distribution of poverty and inequality within those countries—information that was previously not available.

The methodology is based on a statistical procedure to combine household survey data with population census data, by imputing into the latter a measure of economic welfare (consumption expenditure in our examples) from the former. Like the usual sample based estimates, the inequality measures produced are also *estimates* and subject to statistical error. The paper has demonstrated that the mean consumption, poverty and inequality estimates produced from census data match well the estimates calculated directly from the country's surveys (at levels of disaggregation that the survey can bear). The precision of the inequality estimates produced with this methodology depends on the degree of disaggregation. In all three countries considered here our inequality estimators allow one to work at a level of disaggregation far below that allowed by surveys.

We have decomposed inequality in our three countries into progressively more disaggregated spatial units, and have shown that even at a very high level of spatial disaggregation the contribution to overall inequality of *within*-community inequality is very high (75 percent or more). We have argued that such a high within-group component does not necessarily imply that there are no between-group differences at all and that all communities in a given country are as unequal as the country as a whole. We have shown that in all three countries, there is a considerable amount of variation in inequality across communities. Many communities are rather more equal than their respective country as a whole, but there are also many communities that are not clearly more homogeneous than society as a whole, and may even be considerably more unequal.

We have explored some basic correlates of local-level inequality in our three countries. We have found consistent patterns across all three countries. Geographic characteristics are strongly correlated with inequality, even after controlling for demographic and economic conditions. The correlation with geography is observed in both rural and urban areas. In rural areas, population size and mean consumption at the community level are positively associated with inequality, while in urban areas that is not the case. In both rural and urban areas, populations with large shares of the elderly tend to be more unequal. In Madagascar, populations with large shares of children and large shares of individuals aged 50-59 are consistently more equal. In Ecuador this is true only in rural areas.

References

Atkinson, A.B. and F. Bourguignon (2000) 'Introduction: Income Distribution and Economics', in A.B. Atkinson and F. Bourguignon (eds) *Handbook of Income Distribution* Vol.1, North Holland: Amsterdam.

Bardhan, P. and D. Mookherjee (1999) 'Relative Capture of Local and Central Governments' (mimeo), Boston University: Boston.

Bourguignon, F. (1979) 'Decomposable Income Inequality Measures', *Econometrica* 47:901-920.

Cowell, F. (1980) 'On the Stucture of Additive Inequality Measures' *Review of Economic Studies* 47:521-31.

Cowell, F. (2000) 'Measurement of Inequality' in A.B. Atkinson and F. Bourguignon (eds) *Handbook of Income Distribution* Vol.1, North Holland: Amsterdam.

Deaton, A. (1997) *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, The Johns Hopkins University Press for the World Bank: Washington DC.

Deaton. A. and C. Paxson (1995) 'Savings, Inequality and Aging: an East Asian Perspective', *Asia-Pacific Economic Review* 1(1):7-19.

Demombynes, G., C. Elbers, J.O. Lanjouw, P. Lanjouw, J. Mistiaen and B. Özler (2002) 'Producing an Improved Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries', *WIDER Discussion Papers* 2002/39, UNU/WIDER: Helsinki.

Elbers, C., J.O. Lanjouw, P. Lanjouw (2000) 'Welfare in Villages and Towns: Micro level Estimation of Poverty and Inequality', *Tinbergen Institute Working Paper*s 029/2, Amsterdam.

Elbers, C., J.O. Lanjouw, P. Lanjouw (2002) 'Micro-Level Estimation of Welfare' *World Bank Policy Research Working Paper*s 2911, Development Research Group, World Bank: Washington DC.

Elbers, C., J.O. Lanjouw, P. Lanjouw (2003) 'Micro-Level Estimation of Poverty and Inequality', forthcoming, *Econometrica*.

Elbers, C., J.O. Lanjouw, P. Lanjouw and P. Leite (2001) 'Poverty and Inequality in Brazil: New Estimates from Combined PPV-PNAD Data' mimeo, DECRG-World Bank: Washington DC.

Hentschel, J. and P. Lanjouw (1996) 'Constructing an Indicator of Consumption for the Analysis of Poverty: Principles and Illustrations with Reference to Ecuador', *LSMS Working Papers* 124, DECRG-World Bank: Washington DC.

Hentschel, J., J.O. Lanjouw, P. Lanjouw and J. Poggi (2000) 'Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador', *World Bank Economic Review* 14(1)147-65.

Kanbur, R. (2000) 'Income Distribution and Development' in A.B. Atkinson and F. Bourguignon (eds) *Handbook of Income Distribution* Vol.1, North Holland: Amsterdam.

Lanjouw, P. and N. Stern (1998) *Economic Development in Palanpur Over Five Decades*, Oxford: Oxford University Press: Oxford.

Mistiaen, J., B. Özler, T. Razafimanantena and J. Razafindravonona (2002) 'Putting Welfare on the Map in Madagascar' mimeo, DECRG-World Bank: Washington DC.

Ravallion, M. (1999) 'Is More Targeting Consistent with Less Spending?', *International Tax and Public Finance* 6:411-19.

Ravallion, M. (2000) 'Monitoring Targeting Performance when Decentralized Allocations to the Poor are Unobserved', *World Bank Economic Review* 14(2):331-45.

Sen, A., and J. Foster (1997) 'Technical Annexe', in A. Sen (ed.) *On Economic Inequality*, Clarendon Press: Oxford.

Shorrocks, A. (1980) 'The Class of Additively Decomposable Inequality Measures', *Econometrica* 48:613-25.

Simler, K., and V. Nhate (2002) 'Poverty, Inequality and Geographic Targeting: Evidence from Small-Area Estimates in Mozambique' mimeo, International Food Policy Research Institute: Washington DC.

Tendler, J. (1997) *Good Government in the Tropics*, The Johns Hopkins University Press: Baltimore.

World Bank (1996) 'Ecuador Poverty Report', *World Bank Country Study Reports* 16087, Ecuador Country Department, World Bank: Washington DC.