View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DSpace at VU

TI 2000-029/2
Tinbergen Institute Discussion Paper

Welfare in Villages and Towns

Chris Elbers
Jean O. Lanjouw
Peter Lanjouw

Tinbergen Institute
The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Most TI discussion papers can be downloaded at
http://www.tinbergen.nl

DRAFT: COMMENTS WELCOME
Reference as: TINBERGEN INSTITUTE WORKING PAPER, 2000

Welfare in Villages and Towns: Micro-Measurement of Poverty and Inequality

Chris Elbers, Jean O. Lanjouw, and Peter Lanjouw[1]

March 25, 2000

Abstract

We construct and derive the properties of an estimator of welfare which takes advantage of the detailed nature of information about living standards available in small household surveys and the comprehensive coverage of a census. By combining the strengths of each, our estimator can be used at a disaggregated level. It has a clear interpretation; it can be expanded in a consistent way to any welfare measure; and can be assessed for reliability using standard statistical theory. Because unit record census data present some computational hurdles, we explore simulation and numerical integration approaches, as well as the use of distributional approximations. For non-separable inequality measures we derive specific formulas to allow the use of 'shortcut' computational methods. Using data from Ecuador we obtain estimates of welfare measures which are very reliable for populations of 5,000 households, 'town's, and in many cases for those as small as 500. We provide simple illustrations of their use. In the longer run, such estimates open up the possibility of estimating and testing, at a more convincing intra-country level, the many recent models relating welfare distributions to growth and a variety of socioeconomic and political outcomes.

Keywords: Welfare Measurement, Inequality, Poverty, Targetting, Decentralization, Ecuador

[1]We are very grateful to Ecuador's Instituto Nacional de Estadistica y Censo (INEC) for making its 1991 unit-record census data available to us. We thank Don Andrews, François Bourguignon, Andrew Chesher, Denis Cogneau, Angus Deaton, Jean-Yves Duclos, Francisco Ferriera, Jesko Hentschel, Michiel Keyzer, Steven Ludlow, Berk Ozler, Piet Rietveld, John Rust and Chris Udry for comments and useful discussions, as well as seminar participants at the Vrije Universiteit and ENRA, Paris. Uwe Deichmann kindly produced the maps in Figure 1. Financial support was received from the World Bank, RPO # 682-73. However, the views presented in this paper are our own and should not be taken to reflect the views of the World Bank or any of its affiliates. All errors are our own.

Authors:

Chris Elbers
Department of Economics
Vrije Universiteit
De Boelelaan 1105
1081 HV Amsterdam, NL
email: celbers@econ.vu.nl

Jean O. Lanjouw
Department of Economics
Yale University
P.O. Box 208264
New Haven, CT 06520-8264
email: jean.lanjouw@yale.edu

Peter Lanjouw
The World Bank
1818 H. Street
Washington, DC 20433
email: planjouw@worldbank.org

1 Introduction

Recent theoretical advances have brought income and wealth distributions back into a prominent position in growth and development theories.[2] Distributions of well-being are also considered determinants of many socio-economic outcomes, such as health or levels of violence. Empirical testing of the importance of these relationships, however, has been held back by the poor quality of distributional data, even at the country level. Time series data are sparse, constraining most econometric analyses to a cross-section of countries. Not only may these data be non-comparable, but such estimations require strong assumptions about the stability of structural relationships across large geographical areas and political units.[3] Further, many of the hypothesized relationships are more obviously relevant for smaller groups or areas. For example, as noted by Deaton (1999), while it is not clear why country-wide inequality should directly affect an individual's health, a link could be made to the degree of inequality within his reference group.

The difficulty confronted is that the detailed household surveys which include reasonable measures of income or consumption are samples which are rarely representative at lower levels of aggregation, nor do they cover a sufficient number of households to yield statistically reliable estimates when disaggregated. At the same time, census data either have no information about income or consumption, or measure these variables poorly.[4] This paper presents a fully-developed statistical

[2]The models in this growing literature describe a wide variety of linkages between distributions and growth. For example, inequality (or poverty) limits the size of markets which slows growth when there are scale economies (Murphy, Shleifer and Vishny, 1989); with imperfect capital markets, greater inequality limits those able to make productive investment and occupational choices (Galor and Zeira, 1993; Banerjee and Newman, 1993). Aghion and Bolton (1997) endogenize inequality, with growth having a feedback effect on the distribution of wealth via its affect on credit, or labour, markets. Political economy models such as Alesina and Rodrik (1994) and Persson and Tabellini (1994) suggest that, in democratic regimes, inequality will lead to distortionary redistributive policies which slow growth.

[3]The state-of-the-art data set for this purpose, compiled by Deininger and Squire (1996), goes a long way towards establishing comparability but the critique by Atkinson and Brandolini (1999) shows it remains very far from ideal. (See also Fields, 1989 and 2000, on data.)

Bruno, Ravallion and Squire (1998) give examples of country-level estimation of growth models. Although they do not include distributional variables, Barro and Sala-i-Martin estimate a growth model using U.S. state-level data where the fact that it is a better controlled situation is emphasized (see Comments and Discussion in Barro and Sala-i-Martin, 1991). Ravaillion (1997) points out that aggregation alone can bias estimates of the relationship between asset inequality and income growth derived from country-level data, and demonstrates this using *county*-level panel data from China. For a more general identification critique of cross-country models see Banerjee and Duflos (1999).

[4]For example, a single question regarding individuals' incomes in the 1996 South African cen-

procedure to combine the two data sources, which takes advantage of the detailed information available in household sample surveys and the comprehensive coverage of a census. Using the survey to impute missing information in the census, we generate reliable estimates of poverty and inequality at a very disaggregated level. We estimate, for instance, headcount rates of poverty for 'towns' of 5,000 households with 95% confidence bounds of just ± 0.012, or 2% of the point estimate. Our estimates of the inequality within 'towns' are similarly reliable. With accurate welfare measures for groups the size of towns or villages, even neighborhoods, researchers should be able to test hypotheses at an appropriate level of disaggregation, where assumptions about a stable underlying structure are more tenable.

In addition to econometric testing, better local measures of poverty and inequality should also be useful to those attempting to target development assistance or understand the tradeoffs involved in decentralizing these spending decisions. While it is beneficial to take advantage of local information about community needs and priorities, if local inequalities are large and decisions are taken by the elite, projects may not benefit the poorest. Local level inequality measures, together with data on project choices, make it possible to shed light on this potential cost of decentralization.

We measure poverty and inequality at a disaggregated level based on a household per-capita measure of expenditure, y_h. The idea is straightforward. First a model of y_h is estimated using the sample survey data, restricting explanatory variables to those common to both sets of data. 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 well-being, such as per-capita expenditure adjusted by equivalence scales, or to estimate inequalities in the distribution of household characteristics other than expenditure, such as assets or unemployment. It could also be readily extended to large sample census-type datasets. In addition to estimates of welfare, we also want an understanding of their reliability. Properties of the estimators are derived for a general case in the following section and for specific, commonly used, measures in Section 4.

Datasets have been combined to fill in missing information or avoid sampling biases in a variety of contexts. Recent examples include Arellano and Meghir (1992) who estimate a labour supply model combining two samples. They use the

sus generates an estimate of national income just 83% the size of the national *expenditure* estimate derived from a representative household survey, and a per-capita poverty rate 25% higher, with discrepancies systematically related to characteristics such as household location (Alderman, *et.al.*,2000).

UK Family Expenditure Survey (FES) to estimate first-stage models of wages and other income conditioning on variables common across the two samples. Hours and job search information from the much larger Labour Force Survey is then supplemented by predicted financial information. In a similar spirit, Angrist and Krueger (1992) combine data from two U.S. censuses. They estimate a model of educational attainment as a function of school entry age, where the first variable is available in only in one census and the second in another, but an instrument, birth quarter, is common to both. Lusardi (1996) applies this two-sample IV estimator in a model of consumption behaviour. Hellerstein and Imbens (1999) estimate weighted wage regressions using the U.S. National Longitudinal Survey, but incorporate aggregate information from the U.S. census by constructing weights which force moments in the weighted sample to match those in the census. In an example where, as here, census and sample data are combined to estimate disaggregated poverty rates, Bramley and Smart (1996) take the oppposite tack. Having assumed that local income distributions are lognormal, they relate the distributional parameters to census-based community characteristics and choose the parameters of that relationship so as to fit the implied national distribution of income to that derived from the UK FES, a representative sample.

Although the idea behind our approach is straightforward its execution is not. Censuses run to millions of observations and the computational demands are significant. To make these estimators of practical use, we propose and test a variety of simplifying approximations which can speed up the required calculations by orders of magnitude at little cost in accuracy. These enable all of the calculations to be performed on a standard laptop computer. Section 3 outlines four different computational options that are straightforward to implement and in Section 4 we investigate how each performs empirically for the various measures. Section 5 considers extensions, in particular generalizing the first-stage model of expenditure given below. In Section 6 we illustrate the use of our disaggregated measures in a number of settings. Section 7 concludes.

2 Definitions and Properties of Estimators

In this section we begin by defining the first-stage model of expenditure, the welfare indicators and our estimators. We then discuss the properties of these estimators.

2.1 Definitions

The basis of the approach is that household expenditure is related to a set of observable household characteristics about which the same questions have been asked in both the sample survey and the census. Thus we model observed per-capita household expenditure, y_h, as a function of a vector of variables $\mathbf{x}_h$ found in both the survey and census:[5]

$$\ln y_h = \mathbf{x}_h' \boldsymbol{\beta} + \varepsilon_h, \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of k parameters and ε_h is a disturbance term which is assumed to be normally distributed across households in the population: $\varepsilon_h \sim \text{iid } \mathcal{N}\left(0, \sigma^2\right)$. It is straightforward to relax this functional form assumption. We discuss this in Section 5, where we also consider modifications to allow less restrictive assumptions regarding heteroscedasticity and spatial autocorrelation.

The model in (1) is estimated using sample data for s households. We are interested in using these estimates to calculate welfare for a population for which we do not have expenditure information. Although the disaggregation may be along any dimension - not necessarily geographic - for convenience we will refer to our target populations as 'villages'. There are M_v households in village v and household h has m_h family members. The characteristics $\mathbf{x}_h$ and the family size m_h of each household are drawn independently from a village-specific constant distribution function $G_v(\mathbf{x}, m)$.[6]

While the unit of observation for expenditure in these data is typically the household, we are more often interested in poverty and inequality measures based on individuals. These measures depend on both household per-capita expenditures

[5]The explanatory variables are observed values and thus need to have the same degree of accuracy in addition to the same definitions across data sources. Comparing distributions of reponses at a level where the survey is representative is a check that we have found to be important in practice.

From the point of view of mapping poverty and inequality it does not matter whether these variables are exogeneous. However, if the resulting poverty or inequality estimates are to be used in subsequent analyses, consideration should be given to the relationship between potential $\mathbf{x}$ variables and the disturbances in those estimations. For example, suppose that one would like an estimate of village inequality as an explanatory variable in a household level regression for education, and that education is a potential $\mathbf{x}$ variable. One could choose not to use it in the estimation of inequality and lose whatever precision it would have contributed to those estimates. Alternatively, one could construct, for each household, "leave-one-out" village inequality estimates which do not include that element for the given household.

[6]The characteristics $\mathbf{x}_h$ and the size m_h of each household are fixed and given for each village and we calculate the various statistics conditional on them. However, when analyzing asymptotics for increasingly large populations, we interpret the actual values as independent draws from a village-specific but constant distribution function $G_v(\mathbf{x},m)$.

and the number of individuals in each household. Thus we write $W(\mathbf{m}_v, \mathbf{X}_v, \boldsymbol{\beta}, \boldsymbol{\varepsilon}_v)$, where $\mathbf{m}_v$ is an M_v vector of household sizes in village v, $\mathbf{X}_v$ is a $M_v \times k$ matrix of observable characteristics and $\boldsymbol{\varepsilon}_v$ is an M_v-vector of disturbances.[7]

Because the vector of disturbances for the target population, $\boldsymbol{\varepsilon}_v$, is always unknown, we consider estimating the expected value of the indicator given the village households' observable characteristics and the model of expenditure in (1). We denote this expectation as

$$\mu_v = E[W|\mathbf{m}_v, \mathbf{X}_v, \boldsymbol{\zeta}], \tag{2}$$

where $\boldsymbol{\zeta}$ is the $k+1$ vector of parameters $\{\boldsymbol{\beta},\ \sigma^2\}$. For most poverty measures, including all of those considered below, the independence of the ε_h across households implies that W can be written as an additively separable function of household poverty rates, $w(\mathbf{x}_h,\ \boldsymbol{\beta}, \varepsilon_h)$, and that μ_v can be written:

$$\mu_v = \frac{1}{N_v} \sum_{h \in H_v} m_h \int_{\varepsilon_h} w_h(\mathbf{x}_h, \boldsymbol{\beta}, \varepsilon_h) d\mathcal{N}(\varepsilon_h), \tag{3}$$

where H_v is the set of all households in village v and $N_v = \sum_{h \in H_v} m_h$ is the total number of individuals. When W is an inequality measure, however, the contribution of one household depends on the level of well-being of other households and W is no longer separable. Then we need the more general form,

$$\mu_v = \int_{\varepsilon_1} ... \int_{\varepsilon_{M_v}} W(\mathbf{m}_v, \mathbf{X}_v, \boldsymbol{\beta}, \boldsymbol{\varepsilon}_v)\, d\mathcal{N}(\varepsilon_{M_v}) ... d\mathcal{N}(\varepsilon_1), \tag{4}$$

where $\varepsilon_1 ... \varepsilon_{M_v}$ are the disturbance terms for the M_v households in village v.

In constructing an estimator of μ_v we replace the unknown vector $\boldsymbol{\zeta}$ with consistent estimators, $\widehat{\boldsymbol{\zeta}}$, from the first-stage expenditure regression. This yields $\widehat{\mu}_v = \mathrm{E}[W \mid \mathbf{m}_v, \mathbf{X}_v, \widehat{\boldsymbol{\zeta}}]$. Because this expectation is analytically intractable we consider various methods of computation in Section 3, each giving us an estimator denoted $\widetilde{\mu}_v$.

[7] Our target is the level of welfare that could be calculated if we were fortunate enough to have *observations* on expenditure for all households in a population. Clearly because expenditures are measured with error this may differ from a measure based on true expenditures. See Chesher and Schluter (1999) for methods to estimate the sensitivity of welfare measures to mismeasurement in y.

2.2 Properties

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

$$W - \widetilde{\mu} = (W - \mu) + (\mu - \widehat{\mu}) + (\widehat{\mu} - \widetilde{\mu}). \tag{5}$$

(The subscript v is suppressed here and below). Thus the prediction error has three components: the first due to the presence of a disturbance term in the first-stage model which causes households' actual expenditures to 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 last due to using an inexact method to compute $\widehat{\mu}$ (computation error). We consider the properties of each:

Idiosyncratic Error - $(W - \mu)$ Given the parameters of the model, we only know the per-capita expenditure of each household up to an unobserved component, ε_h. The actual value of the indicator for a village deviates from its expected value, μ, as a result of the realizations of the unobserved component of expenditure in that village.

When W is a separable measure, it is a weighted sum of household contributions. Thus, we can write:

$$(W - \mu) = \frac{1}{\overline{m}_M} \frac{1}{M} \sum_{h \in H_v} m_h \left[w(\mathbf{x}_h, \boldsymbol{\beta}, \varepsilon_h) - \int_{\varepsilon_h} w(\mathbf{x}_h, \boldsymbol{\beta}, \varepsilon_h) d\mathcal{N}(\varepsilon_h) \right], \tag{6}$$

where $\overline{m}_M = N/M$ is the mean household size among M village households. As the village population size increases, new values of $\mathbf{x}$, m and ε are drawn from the constant distribution functions $G_v(\mathbf{x}, m)$ and $\mathcal{N}(0,\ \sigma^2)$. Assuming that $\overline{m}_M$ converges in probability to $\mathrm{E}[m]$,

$$\sqrt{M}(\mu - W) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathrm{I}}) \quad \text{as } M \to \infty, \tag{7}$$

where

$$\Sigma_{\mathrm{I}} = \frac{1}{(\mathrm{E}[m])^2} \mathrm{E}[m_h^2 \mathrm{Var}(w|\mathbf{x}_h, \beta)]. \tag{8}$$

When W is a non-separable inequality measure we note that, for each measure below save the Gini coefficient, there is some pair of functions f and g, such that W may be written in the general form

$$W = f(\overline{y}, \overline{g}), \tag{9}$$

where $\overline{y} = \frac{1}{N}\sum_{h\in H_v} m_h y_h$ and $\overline{g} = \frac{1}{N}\sum_{h\in H_v} m_h g(y_h)$ are means of independent random variables. The latter may be written

$$\overline{g} = \frac{1}{\overline{m}_M}\frac{1}{M}\sum_{h\in H_v} m_h g(y_h) \tag{10}$$

which is the ratio of means of M iid random variables $g_h = m_h g(y_h)$ and m_h. Assuming that the second moments of g_h exist, $\overline{g}$ converges to its mean and is asymptotically normal. The same remark holds for $\overline{y}$. Thus, non-separable measures of welfare also converge as in (7). Making use of the fact that f is a differentiable function of sample averages, we can use the delta method to obtain an estimate of Σ_{I}.

That the idiosyncratic component of the variance in our estimator, $\mathrm{V_I} = \Sigma_{\mathrm{I}}/M$, falls approximately proportionately in M is a feature we will observe in the empirical results below. It is the fact that this component of the error in our estimator of welfare *increases* as one focuses on smaller target populations that limits the degree of disaggregation possible. How quickly this part of the error variance becomes unacceptably large depends on the explanatory power of the $\mathbf{x}$ variables in the expenditure model and, correspondingly, the importance of the remaining idiosyncratic component of expenditure.

Model Error - $(\mu - \widehat{\mu})$ $\widehat{\mu}$ is a continuous and differentiable function of $\widehat{\boldsymbol{\zeta}}$, which are consistent estimators of the first-stage parameters. Thus $\widehat{\mu}$ is a consistent estimator of μ and:

$$\sqrt{s}(\mu - \widehat{\mu}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathrm{M}}) \quad \text{as } s \to \infty. \tag{11}$$

We use the delta method to calculate the variance Σ_{M}, taking advantage of the fact that μ admits of continuous first-order partial derivatives with respect to $\boldsymbol{\zeta}$. In some cases they have a simple analytical form and can be calculated directly (see Section 4.1 Headcount, below), and in others we use numerical approximations. The vector $\bigtriangledown = [\partial\mu\ /\partial\boldsymbol{\zeta}]|_{\widehat{\zeta}}$ is a consistent estimator of the $(k+1)$-dimensional derivative vector. We use this to estimate the model variance, $\mathrm{V_M} = \Sigma_{\mathrm{M}}/s \approx \bigtriangledown^T \mathrm{V}(\widehat{\boldsymbol{\zeta}})\bigtriangledown$, where $\mathrm{V}(\widehat{\boldsymbol{\zeta}})$ is the asymptotic variance-covariance matrix of the first-stage parameter estimators.

Because this component of the error in our estimator of welfare is determined by the properties of the first-stage estimators, it does not increase or fall systematically as the population size changes. Its magnitude depends, in general, only on the fit of the first-stage model of expenditure and the sensitivity of the indicator to deviations in household expenditure. For a given village v its magnitude will also depend on the distance of the explanatory x variables for households in that village from the levels of those variables in the sample data.

Computation Error - $(\widehat{\mu} - \widetilde{\mu})$ The distribution of this component of the prediction error depends, of course, on the method of computation used. In the following section we describe several approaches, and discuss this error in detail there. In cases using simulation, this error has a variance, denoted $\mathrm{V_C}$, and it is straightforward to give a general form for the asymptotic distribution of this component of the total prediction error. In other cases we instead explore the importance of the computation error for various indicators using empirical examples.

Unlike the previous two sources of error, with sufficient computational resources or time this error can be made as small as desired.

The computational error in calculating the expectation, $\widehat{\mu}$, is uncorrelated with both the error in the first-stage estimators, and therefore the model error, and the idiosyncratic component of expenditure. There may be some correlation between the model error, which arises from the disturbances in the sample survey data, and the idiosyncratic error, which arises from disturbances in the census. Any sampled households which are in the target population will necessarily be in the census, so this will certainly be the case for data collected for the same time period, or when the survey and census are separated in time but disturbances in y are autocorrelated. However, the approach described here is necessary precisely *because* the number of sampled households which are part of the target population is very small. It follows that there will be many census households in the target population which are outside of the sample and, further, that there will be many households in the sample survey which are not in the target population. Thus we assume that the correlation between the model and idiosyncratic error is negligible.

The variance of the prediction error, $W - \widetilde{\mu}$, can now be written

$$\mathrm{E}[(W - \widetilde{\mu})^2] = \mathrm{V_M} + \mathrm{V_I} + \eta. \tag{12}$$

where η is the contribution due to computation error. We use both Monte Carlo simulation and numerical integration methods to calculate $\widetilde{\mu}$. When using simulation, at least one part of the last term is the variance $\mathrm{V_C}$. Other approximations introduce non-stochastic elements. Because the size of η is often in the control of the analyst it is usually assumed to be negligible and ignored. It may not be small, however. Since one of our goals is to determine what steps are necessary to make this a valid assumption, we examine this component of the variance explicitly.

3 Computational Strategies

In this section we consider strategies to calculate: $\widehat{\mu}$, the expected value of our poverty or inequality measure conditional on the first-stage model of expenditure;

V_I, the variance in W due to the idiosyncratic component of household expenditures; and, for use in determining the model variance, the gradient vector $\bigtriangledown = [\partial\mu/\partial\boldsymbol{\zeta}]|_{\widehat{\boldsymbol{\zeta}}}$. Each strategy will entail different types of computation error.

For convenience in the following discussion we define $\widehat{\mathbf{t}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$, the predicted levels of log per-capita expenditure of the M village households. In constructing our estimator of expected welfare, $\widetilde{\mu}$, we condition either on the actual distribution of $\widehat{t}$ in the village or on an approximate frequency distribution. We either simulate or use numerical integration to estimate the effect of the idiosyncratic component of expenditure on the expected level of the indicator W. The two treatments of $\widehat{\mathbf{t}}$ and $\boldsymbol{\varepsilon}$ can be combined giving four different approaches to obtaining $\widetilde{\mu}$.

Monte Carlo simulation of the distribution of W conditional on the actual vector of log per-capita expenditures, $\widehat{\mathbf{t}}$, is both straightforward and requires no approximations. However, the alternative methods of calculation become important when the number of households included in the *set* of villages is large. In most cases we require numerical derivatives and therefore need to evaluate $\widetilde{\mu}$ not just once per target population to obtain the point estimate but an additional $(k+1)$ times (or $2(k+1)$ times if a central difference gradient estimator is used). With a (reasonable) vector of 38 first-stage $\mathbf{x}$ variables, 1000 villages, and 300 simulation draws per village this would imply twelve million evaluations of W. While possible, taking this approach might be both costly and unnecessary - we show below that alternatives result in little loss in accuracy.

The best approach to use in any given situation will depend on the size and number of target populations, desired accuracy, and the programming skills and computational resources available. It is also possible, and sometimes useful, to use different approaches when calculating a point estimate and its variance. Because it is very sensitive to specific datasets and resources, we do not systematically investigate the computational time associated with different approaches. However, at various points in the following section we discuss briefly our own experience with specific measures.[8]

[8] Census data are rarely in a form amenable to analysis at the unit record level. When estimating local welfare measures for a range of countries it has been our experience that by far the most time-consuming part of the procedure comes at the stage of managing and combining the census and survey data. Below we use a combination of SAS and Mathematica, taking advantage of their respective strengths (large database management for SAS and sophisticated mathematical procedures for Mathematica).

3.0.1 Approach 1: True $\widehat{\mathbf{t}}$ and simulated distribution of ε.

Let the vector $\mathbf{e}_r$ be the rth random draw from our estimated disturbance distribution. It is constructed by taking a random draw from an M_v-variate standard normal distribution and multiplying the draw by $\widehat{\sigma}$. With each vector of simulated disturbances we construct a value for the indicator, $\widehat{W}_r = W(\mathbf{m}, \widehat{\mathbf{t}}, \mathbf{e}_r)$. The simulated expected value for the indicator is the mean over the R replications:

$$\widetilde{\mu} = \frac{1}{R}\sum_{r=1}^{R}\widehat{W}_r. \tag{13}$$

An asymptotic estimator of the variance of W around its expected value μ due to the idiosyncratic component of expenditures can be calculated in a straightforward manner using the same simulated values:

$$\widetilde{\mathrm{V}}_{\mathrm{I}} = \frac{1}{R}\sum_{r=1}^{R}(\widehat{W}_r - \widetilde{\mu})^2. \tag{14}$$

Simulated numerical gradient estimators are constructed as follows: We make a positive perturbation to a parameter estimate, say $\widehat{\boldsymbol{\beta}}_k$, by adding $\delta|\widehat{\boldsymbol{\beta}}_k|$, and then calculate $\widehat{\mathbf{t}}^+$, followed by $\widehat{W}_r^+ = W(\mathbf{m}, \widehat{\mathbf{t}}^+, \mathbf{e}_r)$, and $\widetilde{\mu}^+$. A negative perturbation of the same size is used to obtain $\widetilde{\mu}^-$. The simulated central distance estimator of the derivative $\partial\mu/\partial\beta_k|_{\widehat{\zeta}}$ is $(\widetilde{\mu}^+ - \widetilde{\mu}^-)/(2\delta|\widehat{\beta}_k|)$. To construct an estimator of $\partial\mu/\partial\sigma|_{\widehat{\zeta}}$ we make similar forward and backward perturbations to $\widehat{\sigma}$ and calculate $\mathbf{e}_r^+$ and $\mathbf{e}_r^-$ again using the same standard normal draws as those used in the construction of $\mathbf{e}_r$. $\widehat{W}_r^+ = W(\mathbf{m}, \widehat{\mathbf{t}}, \mathbf{e}_r^+)$ is used to calculate $\widetilde{\mu}^+$ and similarly for $\widetilde{\mu}^-$. As we use the same simulation draws in the calculation of $\widetilde{\mu}$, $\widetilde{\mu}^+$and $\widetilde{\mu}^-$, these gradient estimators are consistent as long as δ is specified to fall sufficiently rapidly as $R \to \infty$. Having thus derived an estimate of the gradient vector $\bigtriangledown = [\partial\mu/\partial\zeta]|_{\widehat{\zeta}}$, we can calculate $\widetilde{\mathrm{V}}_{\mathrm{M}} = \bigtriangledown^T \mathrm{V}(\widehat{\boldsymbol{\zeta}})\bigtriangledown$.

Here, using Approach I, $\widetilde{\mu}$ is a sample mean of R independent random draws from the distribution of $(W\,|\mathbf{m}, \widehat{\mathbf{t}}, \widehat{\sigma})$, so the central limit theorem implies that

$$\sqrt{R}(\widetilde{\mu} - \widehat{\mu}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathrm{C}}) \quad \text{as } R \to \infty, \tag{15}$$

where $\Sigma_{\mathrm{C}} = \mathrm{Var}(W|\mathbf{m}, \widehat{\mathbf{t}}, \widehat{\sigma})$.[9]

[9] Efficiency can be improved using a minimum discrepancy simulation estimator, where draws are made systematically from the disturbance distribution. In experiments estimating the headcount measure of poverty, we found that, for $R < 100$, $\sqrt{\mathrm{V}_{\mathrm{C}}}$ for this estimator was 74-78% of its value for Monte Carlo simulation. 50 Monte Carlo simulation draws were needed to obtain the same compuational variance as that of the mimimum discrepancy estimator based on 30 draws.

3.0.2 Approach 2: Approximate $\widehat{\mathbf{t}}$ and simulated distribution of ε.

This approach, and (4) below, are based on an approximation of the vector of expected log per-capita expenditures, $\widehat{\mathbf{t}}$. This has great advantages in reducing computation time. Rather than calculating over vectors where each household enters individually, the households are aggregated into a far smaller number of classes. This approximation requires that W be additively separable. So as to be able to take advantage of this useful shortcut when calculating (non-separable) inequality measures, we have determined specific, very accurate, separable approximations for measures discussed in Section 4. (See also the Appendix.)

In this approach the range of $\widehat{t}$ is divided into D equally spaced intervals with boundaries $b = \{b_0, ..., b_D\}$.[10] The dth interval includes households with $b_{d-1} \leq \widehat{t}_h < b_d$. The log per-capita expenditure of each household in the dth interval is approximated by the midpoint of the interval, τ_d, plus a common stochastic component, ε_d, distributed iid $\mathcal{N}(0,\, \sigma^2)$. The D-vector of these midpoint values is denoted $\boldsymbol{\tau}_D$. The total number of individuals in households falling in the dth interval is denoted m_d, with D-vector $\mathbf{m}_D$, and the sum of *squared* household sizes in the dth interval is denoted $[m^2]_d$.

Simulation proceeds as in Approach 1. Again $\mathbf{e}_{r,D}$, now a $D-$vector, is the rth random draw from the distribution $N(0,\, \widehat{\Sigma})$ where $\widehat{\Sigma}$ is a $D \times D$ diagonal matrix with diagonal entries $\widehat{\sigma}^2$. With each vector of simulated disturbances we calculate the value of the indicator $\widehat{W}_{r,D} = W(\mathbf{m}_D, \boldsymbol{\tau}_D, \mathbf{e}_{r,D})$. The simulated expected value for the indicator is calculated as in equation (13).

The idiosyncratic variance for truly separable measures can be estimated using equation (14), although not directly. To use equation (14) would amount to assuming that all households in a given interval have the same unobserved component of log per-capita expenditure. In fact these differ so that there is an idiosyncratic variance among households in each interval. $\mathrm{V_I}$ can be estimated for poverty measures as a weighted average of simulated within interval variances:

$$\widetilde{\mathrm{V}}_{\mathrm{I}} = \frac{1}{RN^2} \sum_{d=1}^{D} [m^2]_d \sum_{r=1}^{R} (w(\tau_d, e_{r,d}) - \widetilde{\mu}_d)^2, \tag{16}$$

where $\widetilde{\mu}_d = 1/R \sum_r w(\tau_d, e_{r,d})$ is the expected value of the indicator w for a household with $t_h = \tau_d$. The expected value of $\widetilde{\mathrm{V}}_{\mathrm{I}}$ approaches $\mathrm{V_I}$ as $D \to \infty$ and, as the examples below demonstrate, it gives good approximations at levels of D which

[10] The width 2λ of the intervals is determined by $2\lambda = (\max[\widehat{t}_h | h \in H_v] - \min[\widehat{t}_h | h \in H_v\}/(D-1)$. The boundaries are given by $b_i = \min[\widehat{t}_h | h \in H_v] + (2i-1)\lambda$, for $i = 0, \ldots, D$. This implies that the lower and upper bounds are λ below and above the lowest and highest values of $\widehat{t}$, respectively.

are practical to implement. The separating approximations of the non-separable inequality measures used below, while they give accurate point estimates, cannot be used in (16) to obtain an estimate of $\mathrm{V_I}$. Thus we have derived specific variance formulas for those measures.

Simulated numerical gradient estimators can be constructed as described under Approach 1, making the obvious changes in notation. We hold constant the set of boundary values b across perturbations so that changes in the intervals, and therefore frequencies and midpoint values τ_d, are not driven by changes to single endpoint values of $\widehat{t}$. As a result, the perturbed values $\widetilde{\mu}^+$ and $\widetilde{\mu}^-$ differ from $\widetilde{\mu}$ only as households move across interval boundaries. Given this, for the accuracy of the numerical derivative vector to improve with a decrease in the size of the perturbation factor δ requires a concommitant increase in the number of intervals D.

For the same number of simulation draws, R, the expected computational error using an approximate distribution for $\widehat{t}$ is equal or greater than that in Approach 1, but its distribution approaches that in equation (15) as $D \to \infty$. As we will see, the loss of precision is small.

3.0.3 Approach 3: True $\widehat{\mathbf{t}}$ and numerical integration of ε

With this approach $\widetilde{\mu}$ is again described by equations (3) and (4) above, but with the integrals approximated numerically. Calculating non-separable measures directly using this approach is complex because of the nested nature of the integrals so we again make use of the separable approximations given below.

A consistent estimator of the idiosyncratic variance of a separable poverty measure is

$$\widetilde{\mathrm{V}}_{\mathrm{I}} = \frac{1}{N^2} \sum_{h \in H_v} m_h^2 \int (w_h(\widehat{\mathbf{t}}_h, \varepsilon_h) - \widetilde{\mu}_h)^2 d\mathcal{N}(\varepsilon_h) \tag{17}$$

(see equation (8)). As in Approach 2, however, if a separating approximation is used to calculate the point estimate of a non-separable measure, (17) should not be used to obtain an estimate of $\mathrm{V_I}$.

When elements of the gradient vector $\bigtriangledown$ have a simple analytical form they can be estimated directly using numerical integration. Otherwise we use a numerical gradient estimator as described under Approach 1.

The size of the computation error depends on the quality of the numerical integration algorithm. We have opted for the default algorithm of Mathematica 3.0 which over samples regions of rapid change in the integrand. It is easy both to use and performs well (see examples in Section 4).

3.0.4 Approach 4: Approximate $\widehat{\mathbf{t}}$ and numerical integration of $\boldsymbol{\varepsilon}$.

An approximate distribution for $\widehat{t}$ is constructed as in Approach 2. Then the point estimator of the expected value of W is calculated as:

$$\widetilde{\mu} = \frac{1}{N} \sum_{d=1}^{D} \int_{\varepsilon_d} m_d w(\tau_d, \varepsilon_d) d\mathcal{N}(\varepsilon_d), \tag{18}$$

with, for poverty measures,

$$\widetilde{\mathrm{V}}_{\mathrm{I}} = \frac{1}{N^2} \sum_{d=1}^{D} [m^2]_d \int_{\varepsilon_d} (w(\tau_d, \boldsymbol{\varepsilon}_d) - \widetilde{\mu}_d)^2 d\mathcal{N}(\varepsilon_d). \tag{19}$$

When using this approach to estimate poverty measures we also investigate a simpler variance formula – one which bounds V_{I}– obtained from replacing the household-level expected values of the indicator, $\widetilde{\mu}_d$,with the population expected value, $\widetilde{\mu}$. We find that it is often very close.

As in Approach 3, gradient estimators may be constructed by numerically integrating an explicit analytical form or using numerical difference methods. The computation error is the same or greater than that in Approach 3 when using the same numerical integration algorithm.

When using Approach 4 one can take advantage of the discrete Fourier transform to speed computations dramatically. This technique is used to create a convolution of the two distributions, of t and ε, and integration is then over the resulting (single) unconditioned distribution for y. It is commonly used in other fields and sub-routine software is readily available.

4 The Measures – with Examples.

This section presents estimates of commonly used poverty and inequality measures using data from Ecuador. The household survey is the Ecuadorian *Encuesta Sobre Las Condiciones de Vida*, 1994, and the census data are from the closest available year, 1991. The analysis in this section uses only data from the rural Costa region. The R^2 of the first-stage log per-capita expenditure regression was 0.53 (see Appendix Table A.1 and Hentschel, Lanjouw, Lanjouw and Poggi, 2000, for further details).

We randomly draw four nested populations, with 50 to 50,000 households, from the census data. Each of Tables 1 to 7 gives results for a given measure W, using each computational approach with the different populations. In all cases we

adjust for outliers. Although poverty measures automatically give zero weight to expenditure levels above the poverty line, and are not very sensitive to variations in log-expenditure at negative values, inequality measures are sometimes very sensitive to outlying values. In standard situations, where the analyst has direct information about y, it is common to have outliers in that variable due to mismeasurement, inputing errors, etc. The problem is typically dealt with by discarding suspect observations. Here we have an analoguous problem with respect to the $\mathbf{x}$ variables used to infer expenditure levels, and we deal with it in the usual way.[11] In addition to the standard "dirty data" problem, having assumed a normal distribution for ε there is also a certain (in the case of numerical integration) or at least non-zero (in the case of simulation) probability of getting at least one very large positive or negative disturbance, and therefore value for y_h, in the process of estimating μ. This problem is resolved by using a truncated normal to describe the distribution of ε. Since it is the best information we have, we use the minimum and maximum residuals from our first-stage log-expenditure regression as truncation points.[12] For notational simplicity we do not, however, indicate this adjustment in the formulas which follow.

The first rows of each table give values for the expected value of the indicator; the estimated standard deviation of the total predication error; and the estimated standard deviation of each component. These are calculated using Approach 1 with a sufficient number of simulation draws to ensure that the standard deviation of the variance due to computation is less than 0.001. This is at most 1% of the value of the estimate $\widetilde{\mu}$, and usually far less. The estimates resulting from these simulations are then treated as the 'truth' (conditional on $\widehat{\boldsymbol{\zeta}}$). Subsequent rows give the point estimates for $\widetilde{\mu}$ and $\mathrm{V_I}$ obtained using each of the four methods of computation.[13] Where applicable, values for the idiosyncratic variance using upper bound formulas are given.

Thus, looking across columns one can see how the variance of the estimator changes with the size of the target population. Looking down the rows one can compare the results of the various computational approaches.

[11] We delete households with predicted per-capita expenditure, $\widehat{t}_h$, outside the range of observed per-capita expenditure in the household survey, losing less than 0.5% of our total census observations as a result.

[12] Although they are in line with common practice, both steps of this procedure are admittedly somewhat *ad hoc*. Addressing the standard problem of mismeasurement in y_h, Cowell and Victoria-Feser (1996) suggest leaving suspected outliers in the data when estimating inequality and using weighting to lessen their importance. A similar approach could be taken here.

[13] Because the point estimates are very close, calculations of $\mathrm{V_M}$ using the various approaches are also very close. In the tables we provide the values for $\mathrm{V_M}$ calculated only using Approach 1, or Approach 4 in the case of the Gini coefficient.

4.1 Poverty Measure 1: Headcount Ratio

This measure of the incidence of poverty is defined as

$$W = \frac{1}{N} \sum_{h \in H_v} m_h \mathbb{1}(y_h < z), \tag{20}$$

where z is a poverty line defined in per-capita expenditure terms and $\mathbb{1}(\ )$ is an indicator function taking on the value of one if the expression inside of the brackets is true and zero otherwise. Given our distribution for ε_h,

$$\widehat{\mu} = \frac{1}{N} \sum_{h \in H_v} m_h \Phi((\ln z - \widehat{t}_h)/\widehat{\sigma}), \tag{21}$$

where $\Phi(.)$ is the standard normal distribution function. Denote by $\widehat{\mu}_h$ the expected welfare indicator of household h with expected log per-capita expenditure $\widehat{t}_h$. Then,

$$\widehat{\mathrm{V}}_{\mathrm{I}} = \frac{1}{N^2} \sum_{h \in H_v} m_h^2 \widehat{\mu}_h (1 - \widehat{\mu}_h). \tag{22}$$

In this case there is an analytic form for the estimator of the gradient vector:

$$\bigtriangledown^T = -\frac{1}{N} \sum_{h \in H_v} \frac{m_h}{\widehat{\sigma}} \phi \left(\frac{\ln z - \widehat{t}_h}{\widehat{\sigma}} \right) \left[\mathbf{x}_h^T, \frac{\ln z - \widehat{t}_h}{\widehat{\sigma}} \right] \tag{23}$$

Table 1 gives results for the headcount measure. Looking across the rows marked 'Truth' we see that the standard error of the headcount estimator is reasonable even for the population of just 500 households and very small for that of 5,000 households (about 4 and 1 percent of the point estimate, respectively). The piece of the standard error due to the idiosyncratic component of population expenditures is substantial for the population of fifty households but it drops off quickly as one adds households. The piece due to the first-stage model error, while constant, is small. Although large numbers of simulation draws were taken to ensure an accurate 'Truth', $R = 30$ suffices to give a computation error which contributes only negligibly to the overall standard error in the estimator. The results for Approach (2) indicate that there is almost no loss in accuracy associated with moving from the true distribution of predicted expenditures, $\widehat{t}$, to an approximate frequency distribution with 128 intervals, either in the calculation of the point estimate (and hence V_{M}) or in the calculation of the idiosyncratic variance, V_{I}. The results for Approaches (3) and (4) indicate that evaluating integrals numerically with a

standard software routine gives accurate estimates. We found, however, that it was very slow when the evaluation had to be done at a large number of t values. Thus approximating the distribution of t, which works well with a population of 500 households, is the feasible approach for larger populations.

For Approaches (2) and (4) we include results for the population of 50 households even though it may not seem reasonable to approximate the distribution of their expenditures with 128 frequency intervals. In the case of Approach (4), if the fast Fourier transform is used to obtain an approximate distribution for y this method is actually faster than basing calculations on the actual distribution even with just 50 households. On the other hand, it would never be sensible to use Approach (2) if calculations were being done only for a single population of 50 households. However, most often estimates will be required for many populations, some possibly large and others small, and it may be easiest to use the same approach for all of them.

4.2 Poverty Measure 2: FGT(c) Measures

This measure of the severity of poverty is defined for each choice of $c > 0$ as

$$W_c = \frac{1}{N} \sum_{h \in H_v} m_h (1 - \frac{y_h}{z})^c \mathbb{1}(y_h < z). \tag{24}$$

Thus

$$\widehat{\mu}_c = \frac{1}{N} \sum_{h \in H_v} m_h \int_{-\infty}^{\ln z - \widehat{t}_h} (1 - e^{\widehat{t}_h + \varepsilon_h - \ln z})^c d\mathcal{N}(\varepsilon_h). \tag{25}$$

When using approaches which involve numerical integration (3 and 4), one can use the formulas for the idiosyncratic variance, $\mathrm{V_I}$, given in equations (17) and (19). However, $\mathrm{V_I}$ is bounded by the simpler formula

$$\mathrm{V_I} \leq \frac{1}{N} (\widetilde{\mu}_{2c} - \widetilde{\mu}_c^2), \tag{26}$$

which is derived from replacing household with population expected poverty as discussed following equation (19).

Results for FGT(1) and FGT(2) are in Tables 2 and 3. Again the estimated standard errors of the poverty measures are reasonable for the population of 500 households and are small once the population has reached 5,000 households. Again, too, all approaches to calculating the estimator work equally well. Calculations of the bound on $\mathrm{V_I}$, defined in equation (26), are in the final row of each table labelled 'Upper Idiosyncratic'. For both measures the bound is very close.

4.3 Inequality Measure 1: Variance of Log Expenditure

This commonly used measure of inequality is:

$$W = \frac{1}{N} \sum_{h \in H_v} m_h (\ln y_h - \overline{\ln y})^2 = \tag{27}$$

$$\frac{1}{N} \sum_{h \in H_v} m_h (t_h - \overline{t})^2 + \frac{1}{N} \sum_{h \in H_v} m_h (\varepsilon_h - \overline{\varepsilon})^2 + \frac{2}{N} \sum_{h \in H_v} m_h (t_h - \overline{t})(\varepsilon_h - \overline{\varepsilon}),$$

where village means $\overline{t}$ and $\overline{\varepsilon}$ are weighted by household size. We denote these variances and covariances in the village population as

$$W = \mathrm{v}(t) + \mathrm{v}(\varepsilon) + 2\mathrm{cv}(t, \varepsilon) \tag{28}$$

and note that the last two terms are random variables because they depend on realizations of the vector $\boldsymbol{\varepsilon}$. Because the expected value of the covariance term is zero

$$\widehat{\mu} = \mathrm{v}(\widehat{t}) + \widehat{\sigma}^2 \left[1 - \frac{1}{N^2} \sum_{h \in H_v} m_h^2 \right]. \tag{29}$$

This can be obtained directly so there is no computation error and $\widetilde{\mu} = \widehat{\mu}$.

When defining the idiosyncratic variance we use Var() to indicate that we are taking expectations over the theoretical distribution of $\boldsymbol{\varepsilon}$. Likewise, here and below $\mathrm{E}[.|t]$ represents $\mathrm{E}[.|\ t, m, \sigma^2]$. Then

$$\mathrm{V_I} = \mathrm{Var}(W|\mathbf{t}) = \mathrm{E}[W^2|\mathbf{t}] \text{ - } \mu^2. \tag{30}$$

The first term can be written

$$\mathrm{E}[W^2|\mathbf{t}] = \mathrm{v}(t)^2 + \mathrm{E}[\mathrm{v}(\varepsilon)^2] + 4\mathrm{E}[\mathrm{cv}(t, \varepsilon)^2|\mathbf{t}] + 2\mathrm{v}(t)\mathrm{E}[\mathrm{v}(\varepsilon)]. \tag{31}$$

Writing μ =v(t)+E[v(ε)] and substituting this and (31) into (30), it follows that

$$\mathrm{V_I} = \mathrm{Var}(\mathrm{v}(\varepsilon)) + 4\mathrm{E}[\mathrm{cv}(t, \varepsilon)^2|\mathbf{t}] = \mathrm{Var}(\mathrm{v}(\varepsilon)) + \frac{4\sigma^2}{N^2} \sum_{h \in H_v} m_h^2 (t_h - \overline{t})^2.$$

It can be shown that

$$\mathrm{Var}(\mathrm{v}(\varepsilon)) = (\mathrm{E}[\varepsilon_h^4] - \sigma^4) \left\{ \sum_{h \in H_v} \frac{m_h^2}{N^2} - 2 \sum_{h \in H_v} \frac{m_h^3}{N^3} + \sum_{h \in H_v} \frac{m_h^4}{N^4} \right\}. \tag{32}$$

In practice, the last two terms are of order M^{-2} and M^{-3}, respectively and can be dropped for moderate values of M. Doing so, and with ε_h distributed normally, we have the simple formula

$$\mathrm{Var}(\mathrm{v}(\varepsilon)) \approx 2\sigma^4 \sum_{h \in H_v} \frac{m_h^2}{N^2}. \tag{33}$$

Results for this measure are in Table 4. Compared to the poverty estimators, the estimator of the variance of log expenditure is somewhat less accurate at low levels of disaggregation. Nevertheless, good estimates are obtained for "towns" of 5,000 households. Further, because a smaller share of the total standard error of this estimator is due to the idiosyncratic component of expenditures in the population, there is less to be lost from disaggregating to that level.

4.4 Inequality Measure 2: Atkinson Index

The Atkinson measure with inequality aversion parameter $c \geq 0$ ($\neq 1$) is defined as:[14]

$$W_c = 1 - \left\{ \frac{1}{N} \sum_{h \in H_v} m_h (\frac{y_h}{\overline{y}})^{1-c} \right\}^{\frac{1}{1-c}}, \tag{34}$$

where, again, the village mean expenditure, $\overline{y}$, is weighted by household size. This measure clearly does not share the separability property of the poverty measures, nor does its expectation have a convenient analytical form like the variance of log expenditure. Thus, when approximating the distribution of t or using numerical integration we use the following separable approximation to the expectation μ. It is based on replacing the reciprocal of average per-capita expenditure, $1/\overline{y}$, by an approximation to its conditional expectation, $1/\mathrm{E}[\overline{y}|\mathbf{t}]$:[15]

$$\mu_{\text{sep}} = 1 - \left\{ \frac{1}{N} \frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]} \sum_{h \in H_v} m_h (\mathrm{E}[y^{1-c}|t_h]) \right\}^{\frac{1}{1-c}}. \tag{35}$$

[14] As usual with CES-type functions, the singularity at $c = 1$ is overcome by (smoothly) switching to the Cobb-Douglas function. The assumption $c \neq 1$ is made for editorial rather than technical reasons.

[15] This expectation is calculated per draw when simulation is used. We experimented with replacing $\frac{1}{\overline{y}}$ with the second-order Taylor series approximation $\frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]} + \left[\frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]}\right]^3 \mathrm{Var}([\overline{y}|\mathbf{t}])$. For some measures this refinement led to lower computation error and for others the error increased. Thus, until further testing with other datasets we would recommend use of the simpler approximation.

In the Appendix, Part A, we show that

$$\mathrm{E}[\mu_{\mathrm{sep}} - \mu] = O\left[\frac{\sigma^2}{M}\right]. \tag{36}$$

Thus the error associated with this approximation falls rapidly in M and, as we shall see below, it is very accurate for population sizes likely to be used in practice.

With ε_h distributed normally, an estimator of the idiosyncratic variance for this measure is

$$\widetilde{\mathrm{V}}_{\mathrm{I}} = \left[\frac{(1-\mu_{\mathrm{sep}})^c}{(1-c)N}\right]^2 e^{(1-c)^2\sigma^2}\left[e^{(1-c)^2\sigma^2} - 1\right]\sum_{h\in H_v}^{M} m_h^2\left(\frac{e^{t_h}}{\mathrm{E}[\overline{y}|\mathbf{t}]}\right)^{2(1-c)}. \tag{37}$$

This approximation is very accurate for values of $c \geq 1.5$ (for $c < 1.5$ it gives an overestimate and so is conservative). Again, see the Appendix, Part A, for the derivation, and a similar formula for cases using an approximate frequency distribution for t.

Table 5 gives results for the Atkinson measure. One can see from Approaches (2)-(4) that the separable approximation given in equation (35) gives point estimates which are very precise for target populations of 500 households or more and within about one standard deviation of the true value for the 50 household population. It is slightly more accurate when using numerical integration. The estimator for the idiosyncratic variance is similarly precise across approaches. As with the poverty measures, the estimator gives reasonable results for the Atkinson index even for villages as small as 500 households (with a standard error of about 4% of the point estimate).

4.5 Inequality Measure 3: Gini Coefficient

The Gini coefficient can be written

$$W = 1 - \frac{2}{N\overline{y}}\sum_{i=1}^{N} q_{(i)}\left(1 - \frac{i}{N}\right), \tag{38}$$

where q_i denotes the per-capita expenditure of person i such that $q_i = y_h$ if person i is in household h, and $q_{(i)}$ denotes the ith order statistic, *i.e.*, $q_{(i)}$ comes i^{th} if the sample is sorted in increasing order (with individuals having the same q_i randomly ranked). The expectation of W is, like the other inequality measures, non-separable. We derive in the Appendix, Part B, the simpler separatable approximation:

$$\mu_{\mathrm{sep}} = 1 - \frac{2}{N\mathrm{E}[\overline{y}|\mathbf{t}]}\sum_{h\in H_v} m_h y_h(1 - \mathrm{E}[H(y_h|\mathbf{t})]). \tag{39}$$

where $\mathrm{E}[H(y_h|\mathbf{t})]$ is the cumulative distribution of per-capita expenditure obtained from a convolution of the distribution of ε and the actual distribution of t. While $\mathrm{E}[H(y_h|\mathbf{t})]$ can be estimated using simulation, it is very time consuming so we do not use Approach (2) for this measure.

Results for the Gini Coefficient are in Table 6. The separable approximation to the Gini used in Approaches (3) and (4) again works very well for even moderately sized values of M. This is a particularly useful finding with this measure, as the repeated sorting of the data required when using simulation is very slow. The first approach becomes unattractive for any but small populations. (For this reason we use Approach (4) in calculating the 'truth' point estimates and model variance estimates for our populations of 5,000 households or greater.) Note that, as with all of the inequality measures, the separable approximation used to calculate the point estimate cannot be used to estimate the idiosyncratic variance. In this one case we have not found an acceptable approximation, so Approach (1) must be used to obtain an estimate of $\mathrm{V_I}$. However, because the prediction error is typically used only to assess 'how many digits' of a welfare estimate are significant, the degree of accuracy needed in its calculation, and therefore the idiosyncratic component, is typically much lower than that desired in the point estimate and will thus require far fewer simulation draws.

4.6 Inequality Measure 4: General Entropy Class

The general entropy measure with parameter $c > 0$ ($\neq 1$) is defined as

$$W_c = \frac{1}{c(1-c)} \left\{ 1 - \frac{1}{N} \sum_{h \in H_v} m_h (\frac{y_h}{\overline{y}})^c \right\}. \tag{40}$$

As with the Atkinson measure, simply replacing $\overline{y}$ by $\mathrm{E}[\overline{y}|\mathbf{t}]$ makes the measure separable and yields an approximation which rapidly converges to μ as the size of the population grows.

A good approximation of $\mathrm{V_I}$ for $0.1 \leq c \leq 0.9$ is obtained by using a Taylor approximation of the *true* general entropy measure, taken around the vector of expected expenditure values, $\mathrm{E}[\mathbf{y}|\mathbf{t}]$. This gives:

$$\begin{aligned} \widetilde{\mathrm{V}}_{\mathrm{I}} &= \sum_{h \in H_v} \mathrm{Var}(y_h|t_h) \times \\ &\left(\frac{m_h}{c(1-c)N} \left[\frac{1}{\mathrm{E}[y_h|t_h]} \left(\frac{\mathrm{E}[y_h|t_h]}{\mathrm{E}[\overline{y}|\mathbf{t}]} \right)^c - \frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]N} \sum_{k \in H_v} m_k (\frac{\mathrm{E}[y_k|t_k]}{\mathrm{E}[\overline{y}|\mathbf{t}]})^c \right] \right)^2 . \end{aligned} \tag{41}$$

See the Appendix, Part C, for details, and a similar formula for the cases using an approximate frequency distribution for t.

Results for the general entropy measure with $c = 0.5$ are in Table 7. We again obtain very accurate estimates of both μ and $\mathrm{V_I}$ using the separable approximations for Approaches (2) - (4). Like the variance of log expenditure measure, reliable estimates for this measure are obtained for populations beginning at about 5,000 households.

5 Model Extensions

Obtaining an accurate model of the conditional distribution of per-capita expenditure from the first-stage estimations is clearly crucial. This can be tackled in many ways. The one we have taken here uses classical methods which will be familiar to most economists. We specify a parametric linear regression model with disturbances which have a convenient form. (Tests for the normality and homoskedasticity of the residuals in the first-stage regressions could not reject either hypothesis - see notes Table A.1). This approach has the large advantage of being easily implemented. However, there are alternatives or extensions which allow more flexibility in the model estimation, flexibility which may sometimes be important. These include:

Semi-Parametric Estimation - Non-Normality: When using simulation to approximate the distribution of the disturbances one can avoid making any assumption about its functional form by taking simulation draws from the actual first-stage residuals. In our case, because the residuals are very close to normally distributed, it of course makes little difference. Estimating the Headcount and Atkinson (2) measures for 50 households in this way we obtain values within two percent of those presented above, and the idiosyncratic variance estimates are the same.

In the same spirit, when using numerical integration, the assumed normal distribution can be replaced by an empirical frequency distribution derived from the residuals. This would, of course, increase the amount of programming necessary because one could no longer rely on the numerical integration algorithms provided in many statistical software packages.

Non-Parametric Estimation Rather than impose a log-linear, or other, parametric form on the conditional distribution of per-capita household expenditure given the observables, $\mathbf{x}_h$, one could consider estimating the conditional expectation $\mathrm{E}(y|x)$ or the conditional density $\mathrm{p}(y|x)$ directly using kernel density estimation. Estimating expenditure for each household in the population of interest (perhaps

millions) based on a vector of say thirty observed characteristics, the main difficulty to confront is in devising a method of weighting to minimize the computational burden. (See Keyzer and Ermoliev, 2000, for an example of this strategy and further discussion.) For separable measures, another practical approach might be along the lines of Hellerstein and Imbens (1999). An estimate $\widetilde{\mu}$ for each population could be calculated from the expenditure information in weighted sample data, with, in each case, weights constructed to fit a limited number of sample moments to census moments. However, this approach would still be computationally complex with a large number of target populations.

Spatial autocorrelation We assume, in equation (1) that the idiosyncratic component of expenditures is distributed independently across households. However, if the available data do not allow one to control for community characteristics in the estimations, the omission could cause a substantial degree of spatial autocorrelation in the conditional distribution of expenditures. Even with good controls for community characteristics one might still expect some degree of correlation to remain. Expected poverty rates calculated under an (incorrect) assumption of independence would remain be unbiased, although their idiosyncratic variance would be underestimated. However, ignoring the fact that a component of the disturbance is shared within groups would cause expected inequality estimates to be biased upward.

Given the problem at hand, the independence assumption is a difficult one to relax. It is not possible to estimate or test general forms of spatial autocorrelation in expenditure at local levels precisely because the problem is insufficient data or, for some localities, no data on expenditure. With some types of surveys it may be possible to test, in a regression framework, the more restricted hypothesis that there is a common community (village, region) correlation in the disturbances. If so, and if a community random effect is found, the disturbance could then be decomposed into community and household components in a straightforward way when using either simulation or numerical integration. Not all survey data have appropriate identifiers to do this very satisfactorily, however. The Ecuadorian survey data used here, for example, include only cluster identifiers - allowing us to put households into 'neighborhoods' but not into larger groups. Thus we can only test whether there is a correlation among households in very close proximity - where one would have the highest expectation of finding a relationship. Given this, it is remarkable that, for eight different regional estimations, we cannot reject that cluster random effects are zero in all rural areas and both rural and urban Oriente. It appears that our set of explanatory variables, while they are all household indicators, are

adequate controls for common community characteristics.[16] We do, however, get significant (and sizable) cluster-level random effects in the other urban areas, which may or may not indicate that there is a significant degree of autocorrelation at the levels of aggregation at which we calculate our welfare measures. It suggests caution and we return to this below in Section 6.

Incorporation of Multiple Data Sources In lieu of being able to estimate its size, it is useful to minimize the magnitude of spatial autocorrelation in the disturbances by including an informative set of community-level variables, or household level proxies, in the expenditure estimations. Census data are particularly limited in this regard. However, with appropriate identifiers, other sources of information could be merged with both census and survey datasets. For example, geographic information system (GIS) databases allow a multitude of environmental and community characteristics (for example, the density of road networks or agroclimatic characteristics) to be geographically defined both comprehensively and with great precision. The increasing availability of a variety of GIS, and other, data holds great promise as a way to enrich household survey and census data and thereby minimize the potential for spatial autocorrelation.

Ancillary data could be integrated into the analysis in other ways as well. As an example, suppose that there was no information about schooling in the survey or census, but that one had, from the analysis of other data, a model of the relationship between expenditure levels and schooling decisions. Suppose further that one had information about *aggregate* community schooling levels. These data could be combined with the model of schooling decisions to provide additional information about the likelihood of different realizations of the vector $\boldsymbol{\varepsilon}$, and therefore $\mathbf{y}$, for each community.

6 Putting the Indicators to Work – Illustrations

In the following subsections we use estimates of distributional measures in several types of applications. The measures have been calculated for all *parroquias* of Ecuador using the full census.[17] These are the lowest adminstrative units and over 95% (out of 1326 in total) have more than 100 households, a level of disaggregation at which we obtain reasonably accurate estimates. The calculations are based on

[16] For example, indicators of household access to various types of infrastructure are likely to be highly correlated with community supplies.

[17] Within the metropolitan areas of Quito and Guayaquil we disaggregate to the level of the *zona*. For a poverty map of South Africa based on the same methodology see Alderman, *et. al.* (2000).

eight separate regional first-stage consumption models (estimation results available from the authors on request).

6.1 Geographical Maps of Welfare

A useful way of understanding the geographical spread of poverty or inequality is to contruct a map using GIS data. Figure 1 provides an example. Comparisons between the Costa, the coastal region of Ecuador, and the Sierra, the central mountainous region, feature highly in popular political debate in Ecuador.[18] The top two maps in Figure 1 depict the spatial distribution of poverty across cantons[19] on the basis of two common measures: the headcount and the poverty gap, FGT(1). The bottom two maps in Figure 1 indicate those instances where the two alternative poverty measures differ in their ranking of cantons. The map on the lower left shows that in the Costa a number of cantons are ranked poorer under the headcount criterion than under the poverty gap. In contrast, in the Sierra, numerous cantons are ranked more poor under the poverty gap criterion than under the headcount. Thus, it is clear that views about the relative poverty of the regions will be affected by the measure of poverty employed. Further, it is also clear that, irrespective of poverty measure used, all cantons in the eastern part of Ecuador are particularly poor.

This type of map could be used for targetting development efforts, or for exploring relationships between welfare indicators and other variables. For example, a poverty or inequality map could be overlaid with maps of other types of data, say on agro-climatic or other environmental characteristics. The visual nature of the maps may highlight unexpected relationships that would escape notice in a standard regression analysis.

The map shows differences in the point estimates for expected poverty across regions, say A and B. One can test whether differences across populations are statistically significant using the statistic

$$\frac{(\widetilde{\mu}_A - \widetilde{\mu}_B)^2}{\mathrm{Var}[(\widetilde{\mu}_A - W_A) - (\widetilde{\mu}_B - W_B)]}, \tag{42}$$

which is distributed asymptotically $\chi^2(1)$ under the null hypothesis $\mathrm{H}_0 : W_A = W_B$. The parts of the variance in the prediction error for populations A and B due to computation and the idiosyncratic component of W are independent. However, if the same first-stage model estimates are used to estimate t_h for households in both

[18] See, for example, "Under the Volcano", *The Economist*, November 27, 1999, p. 66.

[19] For visibility we have disaggregated only to the level of the *canton*, the administrative level just above a *parroquia*.

populations, then the model component of the prediction error will be correlated across populations. Let $\boldsymbol{\psi}$ be a vector of all of the parameters used in the estimation of either $\widetilde{\mu}_A$ or $\widetilde{\mu}_B$, and let $\mathbf{q}$ be a vector of the partial derivatives $[\partial(\widetilde{\mu}_A - \widetilde{\mu}_B)/\partial\boldsymbol{\psi}]|_{\widehat{\boldsymbol{\psi}}}$. Then,

$$\mathrm{Var}[(\widetilde{\mu}_A - W_A) - (\widetilde{\mu}_B - W_B)] \approx \mathbf{q}^T \mathrm{V}\left(\widehat{\boldsymbol{\psi}}\right)\mathbf{q} + \mathrm{V}_{\mathrm{I}}^A + \mathrm{V}_{\mathrm{I}}^B + \mathrm{V}_{\mathrm{C}}^A + \mathrm{V}_{\mathrm{C}}^B. \tag{43}$$

If the first-stage parameter estimates used to estimate household expenditure differ across the two regions then the first term is simply $\mathrm{V}_{\mathrm{M}}^A + \mathrm{V}_{\mathrm{M}}^B$.

6.2 Are Neighbors Equal?: A Decomposition of Inequality

An important issue in the area of political economy and public policy is to determine the appropriate level of government to give responsibility for public services and their financing. The advantage of decentralizing 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 (and poverty) as determinants 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 has been held back by a lack of empirical measures for most variables.[20] Our community-level welfare estimates can help to address this problem.

We can answer, first, any number of questions about the level and heterogeneity of welfare at different levels of government. For example, here we decompose inequality in rural Ecuador into between- and within-group components and examine how within-group inequality evolves at progressively lower levels of regional disaggregation. 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 (say, a village or neighbourhood) differences within groups are small, and most of overall inequality is due to differences between groups?

[20] Ravallion and Galasso (2000), which compares the inter- vs intra-district targetting of schooling in Bangladesh, uses village-level inequality measures, but is limited to those sampled in the household expenditure survey.

We employ the General Entropy (0.5) inequality measure which can be straightforwardly decomposed. If N individuals are placed in one of J groups subscripted by j, and the proportion of the population in the jth group, denoted f_j, has weighted mean per-capita expenditure $\overline{y}_j$ and inequality ω_j, then

$$W_{0.5} = 4\left\{1 - \sum_{j=1}^{J} f_j(\frac{\overline{y}_j}{\overline{y}})^{0.5}\right\} + \sum_{j=1}^{J} w_j f_j(\frac{\overline{y}_j}{\overline{y}})^{0.5}, \tag{44}$$

where the first term is the inequality between groups and the second is within groups (Cowell, 1995). In stages we disaggregate the country down to the *parroquia* level. Table 8 illustrates that even at a very high degree of spatial disaggregation, more than 85% of overall rural inequality can still be attributed to differences within groups.[21]

Thus, as often suggested by anecdotal evidence, even within local communities there exists a considerable heterogeneity of living standards. In addition to affecting the likelihood of political capture, this may have implications for the feasibility of raising revenues locally, as well as for the extent to which residents of such communities can be viewed as having similar demands and priorities.

Put together with either survey data on attitudes towards government or on the allocation of public spending, disaggregated inequality estimates could be used to directly assess the influence of welfare distributions on the political process. We plan to explore this further in the context of the targetting of social fund programs.

6.3 Revisiting the Kuznets Curve

One of the classic questions in development economics concerns the relationship between economic development and the distribution of well-being. (See Fields, 2000, for an exhaustive survey of empirical research on this topic.) It has been postulated that inequality first rises with economic development, and then falls (a pattern often called the "Kuznets' Inverted U-Curve" in association with Simon Kuznets' pioneering work in the 1950s and early 1960s). Empirical investigations of this relationship have, to date, been cross-country regressions or longitudinal studies of individual countries. As noted in the introduction, the first type of analysis implies the unattractive assumption that the countries can be viewed as draws out of a common data generating process. More practically, cross-country regressions require comparable data, and the degree to which such data exist is still

[21] We have confined our attention to rural areas where there is no evidence of spatial autocorrelation in ε. Results using all of Ecuador were very similar.

debated. Longitudinal studies also raise concerns about data comparability and few countries possess sufficiently rich historical detail to allow statistical investigations of the Kuznets relationship.

Here we have constructed a cross-sectional dataset of inequality outcomes and average expenditure levels at the level of communities. The indicators are constructed in an identical way avoiding completely problems of comparability. Moreover, the underlying assumption of a common data generating process is rather less objectionable in this context. Figure 2 depicts a scatterplot of *parroquia*-level inequality and average per-capita expenditure estimates for the Sierra region of Ecuador. A quadratic regression curve of inequality on expenditure is overlaid (with 95% confidence bounds), as well as a non-parametric regression curve based using a Normal weight kernel fit. A number of interesting observations follow: i) there does appear to be some evidence of inequality first rising with expenditure and declining, or at least levelling off, although inequality begins to decline as expenditure increases only at a fairly high level of expenditure; ii) even confining attention to a single region there is considerable heterogeneity.

This latter point is confirmed in Table 9 where, in the basic model, only about 12% of the variation in *parroquia*-level inequality is explained by per-capita expenditure and its squared term. When we include demographic variables and a full set of regional interaction terms the inverted U relationship remains and the R^2 improves substantially in all regions.[22] Deaton and Paxson (1994, 1997) and Higgins and Williamson (1999) have noted the importance of including population demographics in this type of analysis. Demographic effects differ across regions, although it does appear that the larger the population share in the higher working age group (40 to 60 years) the lower is inequality. A possible mechanism driving this finding is that in areas with large numbers of older, more experienced workers, there is a lower

[22]Because the regressor 'average per-capita expenditure' is correlated with its own estimation error (mismeasurement), and with that in the dependent variable, OLS estimators are biased. However, we have seen that the magnitude of the estimation errors is very small. We examine the size of the coefficent bias arising from the idiosyncratic component of the prediction error by re-estimating the model with 30 simulated populations and find that each coefficient estimate presented is within five percent of the mean of the thirty corresponding estimates. Future work will further explore 'mismeasurement' biases induced by the estimation of welfare indicators, in particular relative to those induced by standard data mismeasurement.

All of the Sierra is included in the analysis even though there was some evidence of spatial autocorrelation in ε among neighboring households in the urban sector. If this were a source of bias in our inequality estimates it would be upward - and work against finding a Kuznets relationship. To check robustness we re-estimated the inequality measures, decomposing ε into *parroquia* and household components with the share of the variance due to the first assumed (conservatively) to be that estimated for clusters. The results hardly differ. We also estimate the original model but restricted to the rural sector and find that, qualitatively, the results remain.

labor market premium on experience, resulting in lower aggregate inequality. While age composition is statistically related to inequality, it is interesting to note that community size is not. As in the previous subsection, there is again no support for the view that smaller communities are likely to be less unequal than larger ones.

6.4 The Elasticity of Poverty with Respect to Growth and Redistribution

The cross-sectional data set employed above can also be used to ask how sensitive different poverty measures are to changes in average consumption and overall inequality. In Table 10 we regress (the log of) three different expected poverty measures on *parroquia*-level (log) expected average per-capita expenditure, (log) expected inequality, demographic characteristics, and regional dummies.[23]

The elasticity of all three poverty measures with respect to average per-capita consumption is significantly less than negative one: the poor in Ecuador are thus not left behind in overall growth. As one would expect, an increase in inequality at a given level of average consumption also increases poverty. Moving from FGT(0) to FGT(2), such that the poverty measure becomes more sensitive to distance below the poverty line, the elasticity of poverty with respect to changes in both 'growth' and inequality increases in absolute value. Further, if the focus is on the poorest, as captured by FGT(2), reducing inequality is as effective as raising average expenditure.

7 Concluding Comments

In constructing disaggregated estimates of welfare we have explored a straightforward idea. We use detailed household survey data to estimate a model of per-capita expenditure and then use the resulting parameter estimates to weight the census-based characteristics of a target population in determining its expected welfare level. While others have taken weighted combinations of variables in the census to estimate household poverty, this merging of data sources has the advantage of yielding an estimator with a clear interpretation via its link to household expenditure; one which can be expanded in a consistent way to any welfare measure; and, perhaps most importantly, can be assessed for reliability using standard statistical theory.

Dealing with unit record census data presents some computational hurdles. Thus one of our goals was to determine methods of calculation which are accurate as

[23] As in Table 9, the estimated regressors (average per-capita consumption and inequality) and the dependent variable will have correlated estimation errors, but again these errors are very small.

well as feasible for researchers with standard software and a laptop computer. We explore simulation and numerical integration approaches, as well as distributional approximations. When calculating point estimates and variances for non-separable inequality measures we derive specific formulas to allow the use of 'shortcut' computational methods. All of the computational options, including those which make use of several types of approximation, are shown to give very accurate results, so in deciding which to use the most important consideration is their relative ease given the resources available.

What is quite remarkable is how well this method of estimating welfare measures works in practice. In our examples using Ecuadorian data we find that estimates of all poverty and inequality measures are very reliable for populations as small as 5,000 households, a 'town', and in many cases they are quite good even for populations as small as 500 households. This is only one country, of course, and one could imagine that we were particularly lucky in the extent of overlap in census and survey variables on which to base the estimations. However, experience using this method to calculate headcount measures in South Africa and Panama suggest that Ecuador is not an unusual case (see Alderman, *et.al.*,2000). And given these promising initial results there is also no reason to be passive consumers of existing data sets. Governments and surveying bodies can be encouraged to design both census and survey instruments to correspond more closely for this purpose – and they are doing so. (Nicaragua and Jamaica are two of which we are aware.)

So now that we have estimates of poverty and inequality in thousands of 'towns' or other groups, what can we do with them? The possibilities seem almost endless. For many questions, intra-regional cross-town analysis could considerably enrich the existing results of cross-country studies - our Kuznets' regressions are just one example. At the micro-level increasing attention is being paid to ways in which welfare distributions within groups relate to socioeconomic and political outcomes. Of the resulting multitude of theories, most remain to be tested. Again, our findings regarding the level and heterogeneity of well-being at different levels of government, features which have been linked *in theory* to political capture and the targetting of public resources, are just one illustration of what is possible. Merging these measures with data on crime, education, health, voting patterns, unemployment, and so on, will open up many promising avenues for further research.

8 References

Aghion, Phillipe and Patrick Bolton (1997) "A Theory of Trickle Down Growth

and Development," *Review of Economic Studies.* Vol. 64, no. 2, pp. 151-72.

Alderman, Harold, Miriam Babita, Jean O. Lanjouw, Peter Lanjouw, Nthabiseng Makhatha, Amina Mohamed, Berk Ozler,, and Olivia Qaba (2000) "Is Census Income an Adequate Measure of Household Welfare? Combining Census and Survey Data to Construct a Poverty Map of South Africa." Mimeo. The World Bank, Washington, D.C.

Alesina, Alberto and Dani Rodrik (1994) "Distributive Politics and Economic Growth," *Quarterly Journal of Economics.* Vol. 109, pp. 465-90.

Angrist, Joshua D. and Alan B. Krueger (1992), "The Effect of Age of School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples", *Journal of the American Statistical Association.* Vol. 87, pp. 328-36.

Arellano, Manuel and Costas Meghir (1992), "Female Labour Supply and on the Job Search: an Empirical Model Estimated using Complementary Data Sets," *Review of Economic Studies.* Vol. 59, pp. 537-59.

Atkinson, Anthony B. and Andrea Brandolini (1999) "Promise and Pitfalls in the Use of Secondary Datasets: Income Inequality in OECD Countries." Mimeo. Oxford University.

Banerjee, Abhijit and Esther Duflo (1999) "Inequality and Growth: What Can the Data Say?" Mimeo. MIT.

Banerjee, Abhijit and Andrew Newman (1993) "Occupational Choice and the Process of Development," *Journal of Political Economy.* Vol 101, no. 1, pp. 274-98.

Bardhan, Pranab and Dilip Mookherjee (1999) "Relative Capture of Local and Central Governments," Mimeo. Boston University.

Barro, Robert and Xavier Sala-i-Martin (1991) "Convergence Across States and Regions," *Brookings Papers on Economic Activity.* no. 1, pp. 107-82.

Bramley, G. and G. Smart (1996), "Modelling Local Income Distributions in Britain," *Regional Studies.* Vol. 30, pp. 239-55.

Bruno, Michael, Martin Ravaillion and Lyn Squire (1998) "Equity and Growth in Developing Countries: Old and New Perspectives on the Policy Issues," in Vito Tanzi and Ke-Young Chu, eds., *Income Distribution and High-Quality Growth* (Cambridge: MIT Press).

Chesher, Andrew and Christian Schluter (1999) "Welfare Measurement and Measurement Error". Mimeo. University College London.

Clarke, George R.G. (1995) "More Evidence on Income Distribution and Growth," *Journal of Development Economics.* Vol. 47, p. 403-28.

Cowell, Frank and Maria-Pia Victoria-Feser (1996) "Robustness Properties of Inequality Measures," *Econometria.* Vol. 64, no. 1, pp. 77-101.

Cowell, Frank (1995) *The Measurement of Inequality.* 2nd ed. (Hemel Hempstead: Prentice Hall/Harvester Wheatsheaf).

Deaton, Angus (1999) "Inequalities in Income and in Health," NBER Working paper no. 7141.

Deaton, Angus and Christina Paxson (1994) "Intertemporal Choice and Inequality," *Journal of Political Economy.* Vol. 102, no. 3, pp. 437-67.

Deaton, Angus and Christina Paxson (1997) "The Effects of Economic and Population Growth on National Saving and Inequality," *Demography.* Vol. 34, no. 1, pp 97-114.

Deininger, Klaus and Lyn Squire (1996) "A New Data Set Measuring Income Inequality," *The World Bank Economic Review.* Vol. 10, 565-91.

Fields, Gary (1989) "A Compendium of Data on Inequality and Poverty for the Developing World," Cornell University. Mimeo.

Fields, Gary (2000) "Economic Growth and Inequality: A Review of the Empirical Evidence," Chapter 3 in Distribution and Development: A New Look at the Developing World. (Russel Sage Foundation and MIT Press).

Galor, Oded and Joseph Zeira (1993) "Income Distribution and Macroeconomics," *Review of Economic Studies.* Vol. 60, pp. 35-52.

Galasso, Emanuela and Martin Ravallion (2000) "Distributional Outcomes of a Decentralized Welfare Program," Mimeo. The World Bank. Washington, DC.

Hellerstein, Judith and Guido Imbens (1999) "Imposing Moment Restrictions from Auxiliary Data by Weighting," *Review of Economics and Statistics.* Vol 81, no 1, pp. 1-14.

Hentschel, Jesko, Jean O. Lanjouw, Peter Lanjouw and Javier Poggi (2000) "Combining Census and Survey Data to Study Spatial Dimensions of Poverty: A Case Study of Ecuador," *The World Bank Economic Review* (forthcoming).

Higgins, Matthew and Jeffrey Williamson (1999) "Explaining Inequality the World Round: Cohort Size, Kuznets Curves, and Openness," NBER Working Paper no. 7224.

Keyzer, Michiel and Yuri Ermoliev (2000) "Reweighting Survey Observations by Monte Carlo Integration on a Census," Stichting Onderzoek Wereldvoedselvoorziening. Staff Working Paper no. 00.04, the Vrije Universiteit, Amsterdam.

Lusardi, Annamaria (1996) "Permanent Income, Current Income and Consumption: Evidence from Two Panel Data Sets," *Journal of Business and Economic Statistics.* Vol. 14, no. 1.

Murphy, Kevin M., Shleifer, Andrea and R.C. Vishny (1989) "Income Distribution, Market Size and Industrialization," *Quarterly Journal of Economics.* Vol.

104, pp. 537-64.
Persson, Torsten and Guido Tabellini (1994) "Is Inequality Harmful for Growth," *American Economic Review.* Vol. 84, pp. 600-21.
Ravallion, Martin (1997) "Does Aggregation Hide the Harmful Effects of Inequality on Growth?" Mimeo. The World Bank, Washington, D.C.

9 Appendix

9.0.1 Part A: The Separation Approximation to the Atkinson Inequality Measure and its Idiosyncratic Variance

The Atkinson inequality measure defined in equation (34)

$$W_c = 1 - \frac{1}{\overline{y}} \left\{ \frac{1}{N} \sum_{h \in H_v} m_h y_h^{1-c} \right\}^{\frac{1}{1-c}}, \tag{31}$$

is not additively separable. However, the transformation

$$B = [\overline{y}(1 - W_c)]^{1-c} = \frac{1}{N} \sum_{h \in H_v} m_h y_h^{1-c} \tag{45}$$

is separable. Thus our computational strategy is to use the expected value of B (given households' expected log per-capita expenditure, $\mathbf{t}$) as the basis of our separable approximation to μ,

$$\mu_{\text{sep}} = 1 - \frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]} \left(\mathrm{E}[B|\mathbf{t}]\right)^{\frac{1}{1-c}}. \tag{46}$$

$\mu_{\text{sep}} \neq \mu$ because of the non-linearity of W_c, but the error is small since both $\bar{y}$ and B are averages and therefore have highly concentrated distributions. As M$\rightarrow \infty$, W_c and μ_{sep} tend to μ (that is, $\text{plim}(W_c - \mu) = \text{plim}(\mu_{\text{sep}} - \mu) = 0$). In fact, the approximation error ($\mu_{\text{sep}} - \mu$) falls as $O(\sigma^2/M)$. To prove this, note that for a twice continuously differentiable function $W(\mathbf{y})$ it follows from a second-order Taylor approximation that

$$\mathrm{E}[W(\mathbf{y})|\mathbf{t}] \approx W(\mathrm{E}[\mathbf{y}|\mathbf{t}]) + \frac{1}{2} \sum_{h \in H_v} \frac{\partial^2 W}{\partial y_h^2}|_{\mathrm{E}[\mathbf{y}|\mathbf{t}]} \mathrm{Var}(y_h|t_h). \tag{47}$$

Applying this approximation of the expectation consecutively to the functions

$$a(\mathbf{y}) = \left(\frac{1}{\overline{y}} - \frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]} \right) \left\{ \frac{1}{N} \sum_{h \in H_v} m_h \, y_h^{1-c} \right\}^{\frac{1}{1-c}} \tag{48}$$

and

$$b(\mathbf{z}) = \frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]}\left\{\frac{1}{N}\sum_{h\in H_v} m_h\, z_h\right\}^{\frac{1}{1-c}} - \frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]}\left\{\frac{1}{N}\sum_{h\in H_v} m_h\, \mathrm{E}[z_h|t_h]\right\}^{\frac{1}{1-c}}, \tag{49}$$

(where $z_h = y_h^{1-c}$) one sees that both $\mathrm{E}[a(y)]$ and $\mathrm{E}[b(z)]$ are $O(\sigma^2/M)$. Thus, so is $(\mu_{\text{sep}} - \mu) = \mathrm{E}[a(y)|\mathbf{t}] + \mathrm{E}[b(z)|\mathbf{t}]$. Take, for instance, $a(\mathbf{y})$:

$$\begin{aligned}\left[\frac{\partial^2 a}{\partial y_k^2}\right]|_{\mathrm{E}[\mathbf{y}|\mathbf{t}]} &= \frac{2}{(\mathrm{E}[\,\bar{y}|\mathbf{t}])^2}\frac{m_k^2}{N^2}\left(\frac{1}{N}\sum_{h\in H_v} m_h(\mathrm{E}[\,y_h|t_h])^{1-c}\right)^{\frac{c}{1-c}} \times \\ &\quad \left\{\frac{1}{\mathrm{E}[\bar{y}\,|\mathbf{t}]}\left(\frac{1}{N}\sum_{h\in H_v} m_h(\mathrm{E}[\,y_h|t_h])^{1-c}\right) - (\mathrm{E}[\,y_k|t_k])^{-c}\right\}\end{aligned} \tag{50}$$

hence

$$\begin{aligned}\mathrm{E}[a(\mathbf{y})|\mathbf{t}] &\approx \frac{e^{\sigma^2}(e^{\sigma^2}-1)}{M}\frac{1}{(\mathrm{E}\,[\overline{y}|\mathbf{t}])^2}\left(\frac{M}{N}\frac{1}{M}\sum_{h\in H_v} m_h\mathrm{E}[\,y_h|t_h]^{1-c}\right)^{\frac{c}{1-c}}\left\{\frac{M^2}{N^2}\right\} \\ &\quad \frac{1}{M}\sum_{k\in H_v} e^{2\,t_k} m_k^2\left\{\frac{1}{\mathrm{E}[\overline{y}\,|\mathbf{t}]}\left(\frac{M}{N}\frac{1}{M}\sum_{p\in H_v} m_p(\mathrm{E}[\,y_p|t_p])^{1-c}\right) - (\mathrm{E}[\,y_k|t_k])^{-c}\right\} \\ &= O(\frac{\sigma^2}{M}),\end{aligned} \tag{51}$$

since the term in braces is bounded (a.s.) given that both $\mathbf{x}$ and m are bounded (a.s.).

An approximation for $\mathrm{V_I}$ which works very well for $c \geq 1.5$ can be derived as follows. Note that

$$\mathrm{V_I} = \mathrm{Var}\left(\frac{1}{\overline{y}}\left\{\frac{1}{N}\sum_{h\in H_v} m_h y_h^{1-c}\right\}^{\frac{1}{1-c}}|\mathbf{t}\right). \tag{52}$$

For a general random variable v,the variance of $v^{\frac{1}{1-c}}$ can be approximated by the first-order Taylor approximation:

$$\mathrm{Var}\left(v^{\frac{1}{1-c}}\right) \approx \left[\frac{\mathrm{E}[v]^{\frac{c}{1-c}}}{1-c}\right]^2 \mathrm{Var}(v). \tag{53}$$

Thus

$$\mathrm{V_I} \approx \left[\frac{\left(\frac{1}{N} \sum_{h \in H_v} m_h \mathrm{E}[\frac{y_h}{\overline{y}} | \mathbf{t}]^{1-c} \right)^{\frac{c}{1-c}}}{1-c} \right]^2 \mathrm{Var} \left(\frac{1}{N} \sum_{k \in H_v} m_k \left[\frac{y_k}{\overline{y}} \right]^{1-c} | \mathbf{t} \right). \tag{54}$$

Replacing $\overline{y}$ by $\mathrm{E}[\overline{y}|\mathbf{t}]$, and substituting,

$$\mathrm{V_I} \approx \left[\frac{(1-\mu_{\mathrm{sep}})^c}{1-c} \right]^2 \mathrm{Var} \left(\frac{1}{N} \sum_{h \in H_v} m_h \left[\frac{y_h}{\mathrm{E}[\overline{y}|\mathbf{t}]} \right]^{1-c} | \mathbf{t} \right). \tag{55}$$

The variance on the RHS is

$$\mathrm{Var} \left(\frac{1}{N} \sum_{h \in H_v} m_h \left[\frac{y_h}{\mathrm{E}[\overline{y}|\mathbf{t}]} \right]^{1-c} | \mathbf{t} \right) = \frac{1}{N^2} \sum_{h \in H_v} m_h^2 \left[\frac{e^{t_h}}{\mathrm{E}[\overline{y}|\mathbf{t}]} \right]^{2(1-c)} \mathrm{Var} \left(e^{\varepsilon(1-c)} \right). \tag{56}$$

To arrive at the formula given in the text (equation 37), note that when ε_h is normally distributed, $\mathrm{Var}\left(e^{\varepsilon(1-c)}\right) = e^{(1-c)^2\sigma^2} \left[e^{(1-c)^2\sigma^2} - 1 \right]$.

Using the notation of section 3.0.2, when D frequency classes are used to approximate the distribution of t_h the formula analogous to (55)becomes:

$$V_{\mathrm{I}} \approx \left[\frac{(1-\mu_{\mathrm{sep}})^c}{1-c} \right]^2 \frac{1}{N^2} \sum_{d=1}^{D} [m^2]_d \left[\frac{e^{\tau_d}}{\mathrm{E}[\overline{y}|\boldsymbol{\tau}_D, \mathbf{m}_D]} \right]^{2(1-c)} Var \left(e^{\varepsilon(1-c)} \right). \tag{57}$$

9.0.2 Part B: The Separation Approximation to the Gini Inequality Measure

The Gini coefficient in equation (38)

$$W = 1 - \frac{2}{N\overline{y}} \sum_{i=1}^{N} q_{(i)} \left(1 - \frac{i}{N} \right), \tag{35}$$

can be rewritten as

$$W = -1 + \frac{2}{N\overline{y}} \sum_{i=1}^{N} q_{(i)} \frac{i}{N}. \tag{58}$$

The contribution of household h to the Gini is

$$w_h = \frac{2}{N\overline{y}} y_h \sum_{i=i_h+1}^{i_h+m_h} \frac{i}{N} = \frac{2y_h}{N\overline{y}} \frac{m_h(2i_h + m_h + 1)}{2N}, \tag{59}$$

where $i_h = \sum_k m_k \mathbb{1}(y_k < y_h)$ is the number of individuals poorer than members of household h.[24] For large populations this expression is approximately equal to

$$w_h \approx \frac{2}{N\overline{y}} y_h m_h \frac{i_h}{N}, \tag{60}$$

since $(m_h + 1)/N \to 0$, as $M \to \infty$. Define

$$H(y) = \frac{1}{N} \sum_{h \in H_v} m_h \mathbb{1}(y_h < y). \tag{61}$$

Then $i_h = NH(y_h)$, and

$$W \approx -1 + \frac{2}{N\overline{y}} \sum_{h \in H_v} y_h m_h H(y_h). \tag{62}$$

Because $H(y)$ is an average of independent random variables, it concentrates on its expectation:

$$\mathrm{E}[H(y)|\mathbf{t}] = \frac{1}{N} \sum_{h \in H_v} m_h P\{y_h \leq y \,|t_h\}. \tag{63}$$

Accordingly, we use the following separable approximation to the expected Gini coefficient:

$$\mu_{\text{sep}} = 1 - \frac{2}{N\mathrm{E}[\overline{y}|\mathbf{t}]} \sum_{h \in H_v} m_h y_h (1 - \mathrm{E}[H(y_h)|\mathbf{t}]). \tag{64}$$

To use the approximation in practice one must compute $\mathrm{E}[H(y)|\mathbf{t}]$ and an attractive way of doing this is based on the observation that it can be interpreted as

$$\mathrm{E}[H(y)|\mathbf{t}] = P\{\varrho + \varepsilon \leq \ln y\}, \tag{65}$$

where ϱ and ε are independent random variables, with ϱ taking values t_h with probability m_h/N. Hence $\mathrm{E}[H(y)|\mathbf{t}]$ is actually the cumulative distribution function obtained from a convolution of the distributions of ϱ and ε, which can be conveniently computed using Fourier transforms.

[24] Since y has a continuous distribution, two households have different per capita expenditures with probability one.

9.0.3 Part C: Idiosyncratic Variance of the General Entropy Measure

A good approximation of the idiosyncratic variance, $\mathrm{V_I}$,of the general entropy measure is obtained by using a Taylor approximation of the *true* formula, in equation (40),

$$W_c = \frac{1}{c(1-c)} \left\{ 1 - \frac{1}{N} \sum_{h \in H_v} m_h \left(\frac{y_h}{\overline{y}} \right)^c \right\}, \tag{37}$$

taken around the vector of expected expenditure values, $\mathrm{E}[\mathbf{y}|\mathbf{t}]$. Letting $\nabla \mathbf{W}$ represent the vector of partial derivatives evaluated at $\mathrm{E}[\mathbf{y}|\mathbf{t}]$ with elements

$$\begin{aligned} \frac{\partial W_c}{\partial y_h}|_{\mathrm{E}[y_h|t_h]} &= \frac{-m_h}{c(1-c)N} \times \\ & \left[\frac{1}{\mathrm{E}[y_h|t_h]} \left(\frac{\mathrm{E}[y_h|t_h]}{\mathrm{E}[\overline{y}|\mathbf{t}]} \right)^c - \frac{1}{\mathrm{E}[\overline{y}|\mathbf{t}]} \frac{1}{N} \sum_{k \in H_v} m_k \left(\frac{\mathrm{E}[y_k|t_k]}{\mathrm{E}[\overline{y}|\mathbf{t}]} \right)^c \right], \end{aligned} \tag{66}$$

we have

$$\mathrm{V_I} \approx \mathrm{E}[(\mathbf{y} - \mathrm{E}[\mathbf{y}|\mathbf{t}])^T \nabla \mathbf{W}^T \nabla \mathbf{W} (\mathbf{y} - \mathrm{E}[\mathbf{y}|\mathbf{t}])] = \sum_{h \in H_v} \left(\frac{\partial W_c}{\partial y_h}|_{\mathrm{E}[y_h|t_h]} \right)^2 \mathrm{Var}(y_h|t_h), \tag{67}$$

where the last equality holds under the assumption that the y_h are independent.

When D frequency classes are used to approximate the distribution of t the formula becomes:

$$\mathrm{V_I} \approx \frac{1}{N^2} \sum_{d=1}^{D} \frac{[m^2]_d}{(1-c)^2} \times$$

$$\left[\frac{1}{\mathrm{E}[y_d|\tau_d]} \left(\frac{\mathrm{E}[y_d|\tau_d]}{\mathrm{E}[\overline{y}|\boldsymbol{\tau}_D, \mathbf{m}_D]} \right)^c - \frac{1}{\mathrm{E}[\overline{y}|\boldsymbol{\tau}_D, \mathbf{m}_D]} \frac{1}{N} \sum_{k=1}^{D} m_k \left(\frac{\mathrm{E}[y_k|\tau_k]}{\mathrm{E}[\overline{y}|\boldsymbol{\tau}_D, \mathbf{m}_D]} \right)^c \right]^2 \mathrm{Var}(y_d|\tau_d). \tag{68}$$

This approximation works well for $0.1 \leq c \leq 0.9$.

Table 1: Headcount Measure of Poverty

Approach[b]	Estimates	Number of households in population[a]			
		50	**500**	**5,000**	**50,000**
(1) "Truth"	No. Draws *R*	4000	400	300	100
	$\hat{\mu}$	**0.537**	**0.522**	**0.515**	**0.516**
	Estimated Standard Error	**0.064**	**0.019**	**0.006**	**0.003**
	Due to:[c] Model	0.003	0.002	0.002	0.002
	Idiosyncratic	0.064	0.019	0.006	0.002
	Computation	<0.001	<0.001	<0.001	<0.001
(1): true $\hat{t}$ sim ε	No. Draws *R*	30	30	30	30
	Computation	0.011	0.004	0.001	<0.001
(2): approx $\hat{t}$ sim ε	$\hat{\mu}$	**0.534**	**0.524**	**0.516**	**0.518**
	Idiosyncratic Error	0.064	0.019	0.006	0.002
(3): true $\hat{t}$ NI ε	$\hat{\mu}$	**0.535**	**0.521**		
	Idiosyncratic Error	0.064	0.019		
(4): approx $\hat{t}$ NI ε	$\hat{\mu}$	**0.535**	**0.521**	**0.515**	**0.516**
	Idiosyncratic Error	0.064	0.019	0.006	0.002

Table Notes – Tables 1 to 7:

[a] These are household groups drawn randomly from the same sampling frame without replacement. Smaller 'population' samples are subsets of the larger 'populations'.

[b] Approach (1) uses the actual $\hat{t}$ distribution and simulation for ε.

Approach (2) uses an approximate distribution for $\hat{t}$ with D = 128, and simulation for ε with R=500.

Approach (3) uses the actual $\hat{t}$ distribution and numerical integration for ε.

Approach (4) uses an approximate distribution for $\hat{t}$ with D=128, and numerical integration for ε.

[c]These are the estimated standard deviations for each separate piece of the total variance, V_M, V_I and V_C.

Table 2: FGT(1) Measure of Poverty[a]

Approach	Estimates	Number of households in population			
		50	**500**	**5,000**	**50,000**
(1) "Truth"	No. Draws *R*	1000	300	300	100
	$\hat{\mu}$	**0.192**	**0.193**	**0.193**	**0.194**
	Estimated standard error	**0.031**	**0.009**	**0.003**	**0.002**
	Due to:[c] Model	0.002	0.002	0.002	0.002
	Idiosyncratic	0.031	0.009	0.003	<0.001
	Computation	<0.001	<0.001	<0.001	<0.001
(1): true $\hat{t}$ sim ε	No. Draws *R*	30	30	30	30
	Computation	0.006	0.002	<0.001	<0.001
(2): approx $\hat{t}$ sim ε	$\hat{\mu}$	**0.193**	**0.192**	**0.192**	**0.194**
	Idiosyncratic Error	0.030	0.009	0.003	0.001
(3): true $\hat{t}$ NI ε	$\hat{\mu}$	**0.190**	**0.192**		
	Idiosyncratic Error	0.030	0.009		
(4): approx $\hat{t}$ NI ε	$\hat{\mu}$	**0.190**	**0.192**	**0.193**	**0.194**
	Idiosyncratic Error	0.030	0.009	0.003	0.001
	Upper S.E.	0.033	0.011	0.003	0.001

Notes: a,b,c) See Table notes following Table 1.

Table 3: FGT(2) Measure of Poverty[a]

Approach	Estimates	Number of households in population			
		50	**500**	**5,000**	**50,000**
(1) "Truth"	No. Draws *R*	1000	300	300	100
	$\hat{\mu}$	**0.089**	**0.095**	**0.095**	**0.096**
	Estimated standard error	**0.019**	**0.006**	**0.003**	**0.002**
	Due to: [c] Model	0.002	0.001	0.002	0.002
	Idiosyncratic	0.019	0.006	0.002	<0.001
	Computation	<0.001	<0.001	<0.001	<0.001
(1): true $\hat{t}$ sim ε	No. Draws *R*	30	30	30	30
	Computation	0.004	0.001	<0.001	<0.001
(2): approx $\hat{t}$ sim ε	$\hat{\mu}$	**0.089**	**0.093**	**0.095**	**0.096**
	Idiosyncratic error	0.019	0.006	0.002	0.001
(3): true $\hat{t}$ NI ε	$\hat{\mu}$	**0.089**	**0.094**		
	Idiosyncratic Error	0.020	0.006		
(4): approx $\hat{t}$ NI ε	$\hat{\mu}$	**0.089**	**0.094**	**0.095**	**0.096**
	Idiosyncratic Error	0.020	0.007	0.002	<0.001
	Upper S.E.	0.020	0.011	0.002	<0.001

Notes: a,b,c) See Table notes following Table 1.

Table 4: Variance of log per-capita expenditure[a]

Estimates	Number of households in population			
	50	**500**	**5,000**	**50,000**
$\hat{\mu}$	**0.569**	**0.522**	**0.561**	**0.567**
Estimated standard error	**0.087**	**0.030**	**0.017**	**0.015**
Due to: Model	0.016	0.014	0.015	0.015
Idiosyncratic	0.086	0.027	0.009	0.003

Notes: a) see Table notes following Table 1.

Table 5: Atkinson (2) Measure of Inequality[a]

Approach	Estimates	Number of households in population			
		50	**500**	**5,000**	**50,000**
(1) **"Truth"**	No. Draws *R*	3000	300	300	100
	$\hat{\mu}$	**0.449**	**0.414**	**0.440**	**0.442**
	Estimated standard error	**0.051**	**0.018**	**0.009**	**0.006**
	Due to [c]: Model	0.007	0.005	0.006	0.006
	Idiosyncratic	0.051	0.017	0.006	0.002
	Computation	<0.001	<0.001	<0.001	<0.001
(1): true $\hat{t}$ sim ε	No. Draws *R*	30	30	30	30
	Computation	0.010	0.003	0.001	<0.001
(2): approx $\hat{t}$ sim ε	$\hat{\mu}$	**0.448**	**0.411**	**0.436**	**0.440**
	Idiosyncratic Error	0.047	0.017	0.005	0.002
(3): true $\hat{t}$ NI ε	$\hat{\mu}$	**0.455**	**0.414**	**0.440**	**0.442**
	Idiosyncratic Error	0.047	0.017	0.005	0.002
(4): approx $\hat{t}$ NI ε	$\hat{\mu}$	**0.455**	**0.414**	**0.440**	**0.443**
	Idiosyncratic Error	0.047	0.017	0.005	0.002

Notes: a,b,c) see Table notes following Table 1.

Table 6: The Gini Coefficient[a]

Approach	Estimates[b]	Number of households in population			
		50	**500**	**5,000**	**50,000**
(1) "Truth"	No. Draws *R*	1600	500		
	$\hat{\mu}$	**0.464**	**0.419**	**0.441**[d]	**0.443**
	Estimated standard error	**0.039**	**0.013**	**0.008**	**0.007**
	Due to[c]: Model	0.009	0.006	0.006	0.007
	Idiosyncratic	0.038	0.011	0.005	0.001
	Computation	<0.001	<0.001		
(1): true $\hat{t}$ sim ε	No. Draws *R*	30	30	30	30
	Computation	0.007	0.002	<0.001	<0.001
(3): true $\hat{t}$ NI ε	$\hat{\mu}$	**0.472**	**0.420**	**0.441**	**0.443**
(4): approx $\hat{t}$ NI ε	$\hat{\mu}$	**0.472**	**0.420**	**0.441**	**0.443**

Notes: a,b,c) see Table notes following Table 1.
d) The point estimate and model variance component are calculated using Approach (4) for populations of 5,000 and 50,000.

Table 7: General Entropy (0.5) Measure of Inequality[a]

Approach	Estimates[b]	Number of households in population			
		50	**500**	**5,000**	**50,000**
(1) **"Truth"**	No. Draws *R*	7,000	700	300	100
	$\hat{\mu}$	**0.394**	**0.301**	**0.338**	**0.339**
	Estimated standard error	**0.078**	**0.021**	**0.014**	**0.011**
	Due to[c]: Model	0.016	0.009	0.011	0.011
	Idiosyncratic	0.077	0.019	0.008	0.003
	Computation	<0.001	<0.001	<0.001	<0.001
(1): true $\hat{t}$ sim ε	No. Draws *R*	30	30	30	30
	Computation	0.015	0.004	0.001	<0.001
(2): approx $\hat{t}$ sim ε	$\hat{\mu}$	**0.392**	**0.302**	**0.336**	**0.337**
	Idiosyncratic Error	0.077	0.018	0.008	0.003
(3): true $\hat{t}$ NI ε	$\hat{\mu}$	**0.399**	**0.302**	**0.338**	**0.339**
	Idiosyncratic Error	0.077	0.018	0.008	0.003
(4): approx $\hat{t}$ NI ε	$\hat{\mu}$	**0.398**	**0.301**	**0.337**	**0.339**
	Idiosyncratic Error	0.077	0.018	0.008	0.003

Notes: a,b,c) see Table notes following Table 1.

Table 8
Decomposition of Inequality in Rural Ecuador by Regional Sub-Group
General Entropy (0.5)

Level of Decomposition	No. of sub-groups	Within-Group (%)	Between-Group (%)
National	1	100	0
Sector and: Region (Costa, Sierra, Oriente)	3	96.5	3.5
Province	21	95.7	4.3
Canton	195	93.6	6.3
Parroquia	**915**	**87.0**	**13.0**
Household	955,985	0	100

Table 9
Kuznets Curve Regression:
Explaining Community (*Parroquia*) Level Inequality

Dependent Variable: Atkinson (2) Inequality Measure

***Parroquia*-Level:**	**Basic Model (t-value)**	**Augmented Model[a] (t-value)**		
		Costa	**Sierra**	**Oriente**
Average Per-Capita Consumption (hundreds of sucres)	0.18 (6.82)	0.50 (12.20)	0.40 (11.60)	0.69 (6.55)
Squared Per-Capita Consumption	-0.052 (-3.74)	-0.181 (-8.90)	-0.150 (-9.84)	-0.453 (-5.35)
% Population 0-10 years	-	0.87 (14.75)	-0.10 (-1.26)	-0.48 (-3.19)
% Population 10-20 years	-	0.98 (13.20)	-0.30 (-3.75)	0.05 (0.28)
% Population 40-60 years	-	-0.33 (-2.80)	0.14 (1.15)	-0.64 (-3.17)
% Population 60+ years	-	1.26 (11.14)	-0.64 (-6.81)	-0.71 (-2.48)
No. of Households (thousands)	-	-0.0013 (-1.60)	-0.0004 (-0.53)	-0.0002 (-0.03)
Constant	0.30 (29.54)	-0.36 (-8.48)	0.36 (6.88)	0.46 (4.33)
Adjusted R^2	0.12	0.64		
No. of Observations	1325	1325		

Note: a) Coefficient estimates are for the variable in the row interacted with a dummy variable for the region indicated at the top of the column. T-values are in parentheses.

Table 10
Elasticity of Poverty with Respect to Consumption and Inequality[a]

***Parroquia*-Level:**	**Dependent Variable**		
	Headcount FGT(0)	**FGT(1)**	**FGT(2)**
(log) Average Per-Capita Consumption	-1.51 (-76.22)	-1.93 (-98.82)	-2.19 (-109.55)
(log) Atkinson (2) Inequality	1.20 (35.27)	1.74 (51.98)	2.16 (62.88)
% Population 0-10 years	-0.75 (-4.45)	-0.37 (-2.24)	-0.18 (-1.06)
% Population 10-20 years	1.14 (6.42)	0.90 (5.13)	0.80 (4.48)
% Population 40-60 years	-2.70 (-11.75)	-1.96 (-8.66)	-1.44 (-6.18)
% Population in 60+ years	1.46 (5.97)	1.10 (4.58)	0.79 (3.20)
No. of Households (thousands)	0.0009 (0.50)	-0.00005 (-0.03)	-0.0015 (-0.80)
Adjusted R^2	0.93	0.96	0.97
No. of Observations	1325	1325	1325

Notes: a) (Unreported) coefficient estimates on region and sector dummy variables included in all estimations were either insignificant or small. T-values are in parentheses.

Table A.1.
First-Stage Estimates for Per-Capita Expenditure: Rural Costa[a]

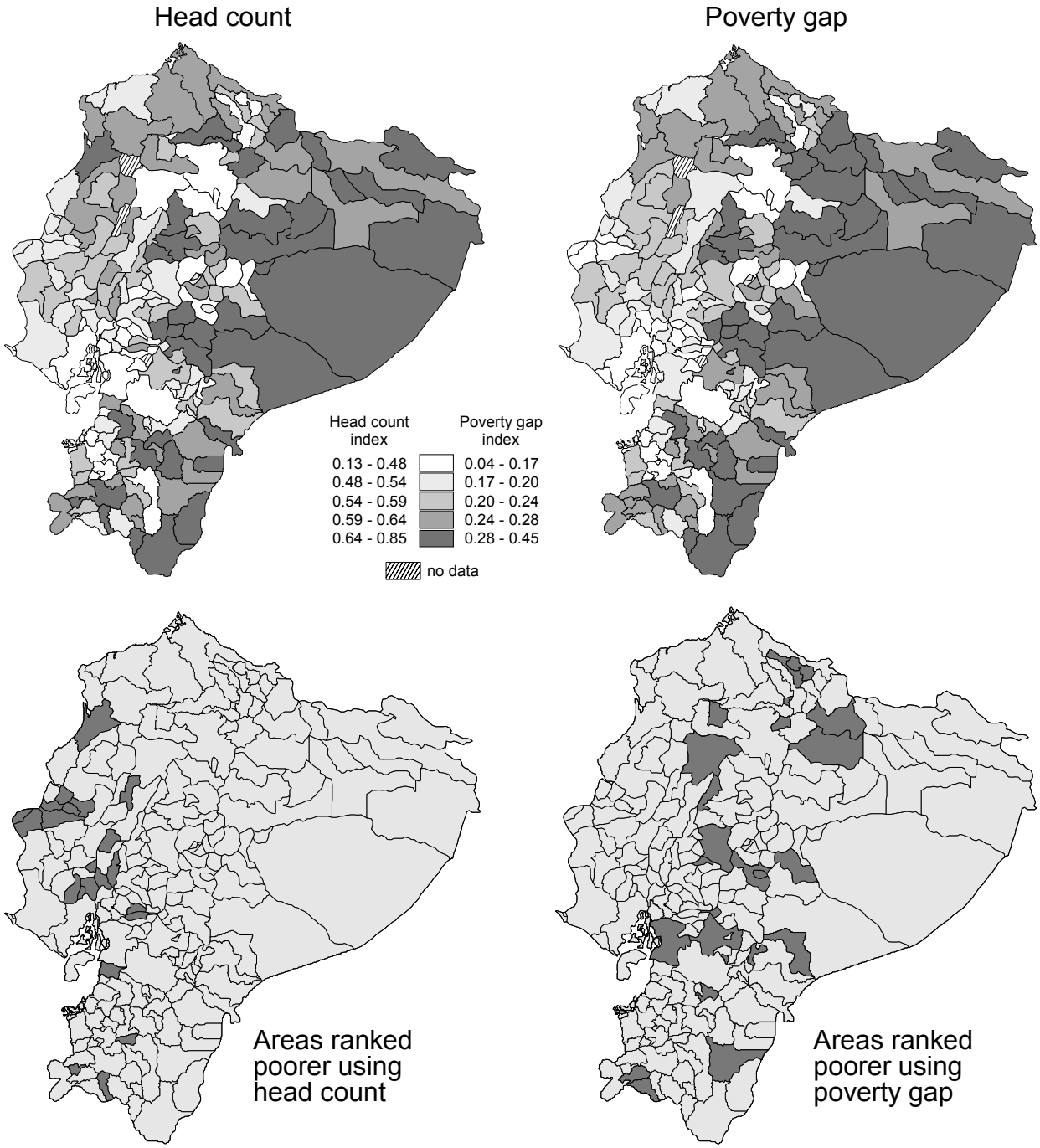
	Parameter estimate	P-value
Years of schooling of household head	0.0106	0.33
Age of household head	-0.0024	0.39
Years of schooling of spouse of head	0.0306	0.03
Age of spouse of head	-0.0015	0.48
Years of schooling of: eldest child (0 otherwise)	0.0040	0.73
2nd child (0 otherwise)	-0.0068	0.72
3rd child (0 otherwise)	-0.0094	0.72
4th child (0 otherwise)	-0.0054	0.87
5th child (0 otherwise)	0.0154	0.67
6th child (0 otherwise)	0.1615	0.37
7th child (0 otherwise)	-0.7031	0.13
Age of: Eldest child (0 otherwise)	0.0113	0.01
2nd child (0 otherwise)	0.0010	0.88
3rd child (0 otherwise)	0.0028	0.80
4th child (0 otherwise)	-0.0113	0.37
5th child (0 otherwise)	-0.0121	0.48
6th child (0 otherwise)	-0.0466	0.24
7th child (0 otherwise)	0.1978	0.11
Number of family members: employed in agriculture	0.0679	0.08
Employed in low-productivity non-agriculture	-0.0501	0.36
Employed in high productivity non-agriculture	0.1534	0.02
Persons per bedroom	-0.1197	0.34
Publically provided waste collection	0.0062	0.03
Burn household waste	0.0011	0.13
Own connection to networked sewage	0.0007	0.81
Shared connection to networked sewage	0.0001	0.98
Own latrine	0.00027	0.76
Networked water connection	-0.0015	0.47
Water from well	-0.0007	0.28
Water delivered by truck	0.0002	0.87
Electricity connection	0.1805	0.01
Telephone connection	0.0120	<0.01
Walls of brick	0.0009	0.36
Walls of wood	-0.0013	0.31
Cooking on gas fire	0.0041	0.17
Cooking with woodfuel	0.0045	0.14
Own shower/bath	0.0008	0.68
Rented home	-0.0011	0.62
Owned home	0.0022	0.01
Indigenous language spoken	0.0030	0.12
Gender of household head	0.0020	0.16

Family size	-0.6718	<0.01
Use of septic tank	0.0008	0.80
Use of "blind well" for waste	0.0003	0.91
Family size squared	0.0662	<0.01
Family size cubed	-0.0023	0.02
Persons per bedroom squared	0.0207	0.36
Persons per bedroom cubed	-0.0010	0.40
R^2	0.532	
Number of observations	483	

Notes:

a) Shapiro Wilk *W* test for normality of the disturbances: P-value: 0.28 (after dropping two observations with extremely high squared residuals). We also do not reject homoskedasticity on the basis of a White test: w = 53.5 ~χ(47).

Figure 1
Rural Poverty by Canton: Headcount and Poverty Gap

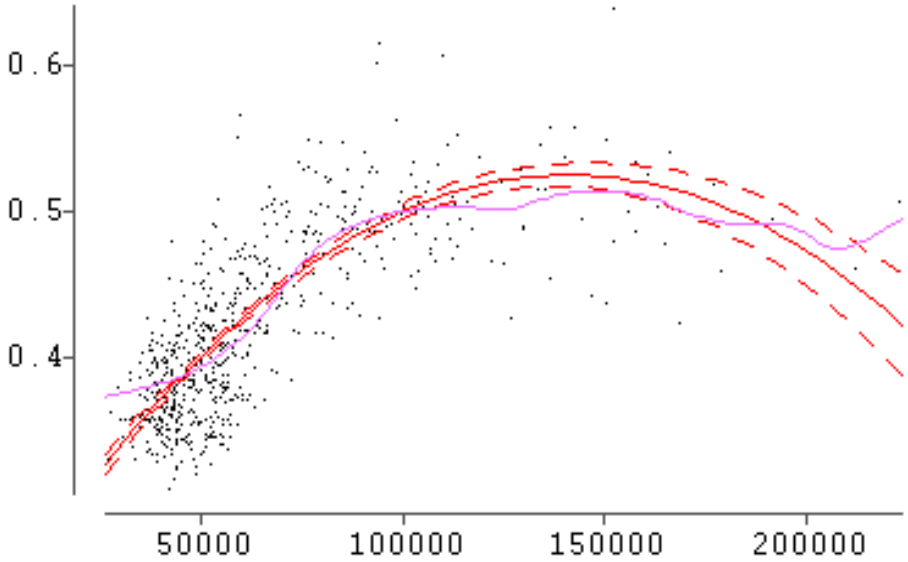


Notes:

1) The top two maps illustrate the geographical distribution of rural poverty across cantons based on respectively, the headcount measure of poverty and the poverty gap index.
2) The shaded regions in the bottom two maps highlight those cantons where the rankings in the top two maps are not the same. The map on the left highlights those cantons that are ranked lower (more poor), according to the headcount measure, than they would be according to the poverty gap index. The map on the right highlights those cantons that are ranked lower according to the poverty gap index, than they would be according to the headcount measure.

Figure 2
The Kuznets Curve in Ecuador's Sierra Region
Kernel and Quadratic (with 95% C.I) Regression Models

Atkinson 2



Average Per Capita Expenditure Per Parroquia (1994 sucres/month)