THE WORLD BANK ECONOMIC REVIEW, VOL. 14, NO. 1: 147-65

Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador

Jesko Hentschel, Jean Olson Lanjouw, Peter Lanjouw, and Javier Poggi

Poverty maps provide information on the spatial distribution of living standards. They are an important tool for policymakers, who rely on them to allocate transfers and inform policy design. Poverty maps are also an important tool for researchers, who use them to investigate the relationship between distribution within a country and growth or other economic, environmental, or social outcomes. A major impediment to the development of poverty maps has been that needed data on income or consumption typically are available only from relatively small surveys. Census data have the required sample size but generally do not have the required information. This article uses the case of Ecuador to demonstrate how sample survey data can be combined with census data to yield predicted poverty rates for the population covered by the census. These poverty rates are found to be precisely measured, even at fairly disaggregated levels. However, beyond a certain level of spatial disaggregation, standard errors rise rapidly.

Poverty maps provide a detailed description of the spatial distribution of poverty within a country. They can be extremely valuable to governments, nongovernmental organizations, and multilateral institutions that want to strengthen the impact that their spending has on poverty. For example, many developing countries use poverty maps to guide the division of resources among local agencies or administrations as a first step in reaching the poor.

Poverty maps also can be an important tool for research. Researchers typically look for the empirical relationship between poverty or inequality and indicators of development, such as economic growth, using a cross-country regression frame-

Jesko Hentschel is with the Poverty Division of the Poverty Reduction and Economic Management Network at the World Bank, Jean Olson Lanjouw is with Yale University and Free University, Amsterdam, Peter Lanjouw is with the Development Research Group at the World Bank and Free University, Amsterdam, and Javier Poggi is with Superintendencia Banco y Seguros, Lima. The e-mail addresses of the first three authors are *jhentschel@worldbank.org*, *planjouw@worldbank.org*, and *lanjouw@econ.yale.edu*. The authors are grateful to the Instituto Nacional de Estadística y Censos in Ecuador for access to the unit record data of the 1990 census. They benefited from discussions with Uwe Deichmann, Stefan Dercon, Joanna Gomulka, Vassilis Hajivassiliou, Jeffrey Hammer, Berk Ozler, and Martin Ravallion, and received useful comments from three anonymous referees, participants at a seminar at the Suntory and Toyota International Centres for Economics and Related Disciplines (STICERD) at the London School of Economics in May 1997, and participants at the World Bank workshop Geographical Targeting for Poverty Reduction and Rural Development at the World Bank in November 1997. The authors also thank Jim Shafer for helpful assistance.

work.[1] It is difficult, however, to control for the enormous heterogeneity across countries, and heterogeneity may mask true relationships. Further, there is limited country experience that we can use to understand the determinants and effects of the distribution of welfare. Moving toward more microeconomic studies—relying on the distributional variation across communities within a single country—may be a way forward.

However, we are hampered in developing poverty maps by the scarcity of disaggregated data. For example, indicators based on income or expenditures often are favored, but the information required for a finely disaggregated map based on income or expenditures is not generally available for a sufficient number of households. The World Bank's Living Standards Measurement Study (LSMS) surveys, which have been fielded in many developing countries, collect the information needed to construct comprehensive measures of income and consumption. But they are too small to allow for disaggregation beyond a simple rural-urban breakdown within broad regions of a country. Census data do not suffer from small-sample problems, but they typically contain limited information.

Many Latin American, African, and Asian countries have constructed welfare indexes designed to rank regions by combining basic census information, such as access to public services and level of education, and have used these indexes to build poverty maps.[2] The indexes sometimes are labeled "basic needs indicators." They generally are constructed in a fairly ad hoc manner and are restricted to the limited qualitative (and not quality-adjusted) data available in a census. Such indicators may be poor proxies for household consumption. Using detailed household survey data for Ecuador, we show that a crude basic needs indicator and a comprehensive consumption measure yield markedly different welfare rankings. We explore how census-based maps can be improved when using an income- or consumption-based indicator of welfare.

In some situations one may want to consider a notion of welfare that reflects not just access to resources but nonincome components as well. For example, in evaluating education programs, one might want to account for the intrinsic value of education in addition to its influence on income or consumption. An appropriate welfare indicator might give greater weight to education than it would receive implicitly from a consumption or income indicator. But if, for example, one wants to figure out how to compensate households for a general change in price levels, a welfare measure based on a fairly narrow measure of consumption might be more suitable.

In general, we can construct a basic needs indicator, choosing weights reflecting the contribution of each variable to total household income as well as any

1. Deininger and Squire (1996) compile a large international database for this purpose. Bruno, Ravallion, and Squire (1998) use this database to explore the relationship between economic growth and inequality. See also Alesina and Rodrik (1994) and Fields (1989).

2. For recent descriptions of the derivation of such maps, see World Bank (1996) for Ecuador, Government of El Salvador (1995) for El Salvador, and FONCODES (1995) for Peru. Other Latin American countries using such poverty maps include Colombia, Honduras, and Republica Bolivariana de Venezuela.

direct contribution to welfare not captured in income. We do not expect rankings based on such an indicator to correspond to those based on consumption. However, we do want a broader measure to also reflect actual consumption. In this article we demonstrate how this may be done more reliably at a disaggregated level. When one is concerned with a broader measure of welfare, these more accurate predictions of poverty based on consumption simply need to be combined in some manner with the other indicators considered relevant given the policy issue of concern.

Using an LSMS data set for Ecuador, we estimate models of consumption expenditures, restricting the set of explanatory variables to those that also are available in Ecuador's most recent census. We apply the parameter estimates from these models to the census data to predict the probability that a given household, taken from the census, is poor. We evaluate our approach by estimating the incidence of poverty in six broad regions and comparing these rates with rates estimated from the LSMS data alone.

We consider some of the statistical issues that arise because the poverty figures are predicted. Our approach yields unbiased estimates of the incidence of poverty from the census data, so that the expected prediction errors are zero. However, the poverty estimates do have standard errors, which must be calculated along with the poverty rates. These standard errors are small at levels of regional disaggregation likely to be of practical relevance, which is encouraging, but the errors become sizable when the poverty rates are calculated over very small groups of households. We thus are warned against disaggregating the poverty map excessively.

I. How Well Does a Basic Needs Indicator Target Poverty?

In this section we examine how effectively a basic needs indicator is able to identify the poor when the poverty indicator is consumption expenditures.[3] The basic needs indicator that we consider was developed in 1994 by the National Statistical Institute of Ecuador (Instituto Nacional de Estadística y Censos, INEC) in response to the government's request for a directory of poor households. The government was considering eliminating its gas subsidy and wanted the directory to target poor households for a compensatory transfer. It happens that the government did not remove the subsidy, and we do not want to imply that INEC regarded the basic needs indicator as anything other than a fairly crude measure

3. Consumption is an imperfect measure of the standard of living, but a comprehensive measure of consumption expenditures comes reasonably close to capturing a household's *achievement* of well-being in accordance with its own chosen bundle of goods and services. The choice between income and consumption also merits attention. For developing countries probably the most compelling argument in favor of consumption is that it typically is easier than income to measure accurately. Its relative smoothness across seasons or even years may make it a better indicator of long-term living standards than a measure of current income (see, however, Chaudhuri and Ravallion 1994). For more discussion see Atkinson (1989), Ravallion (1994), and Sen (1984). Hentschel and Lanjouw (1996) and Lanjouw and Lanjouw (1997) discuss the attraction of a comprehensive measure of consumption expenditures as an indicator of welfare.

Table 1. *Points Assigned to Services Included in* INEC's *Basic Needs Indicator*

Level	*Water*[a]	*Sanitation*[b]	*Waste*[c]	*Education*[d]	*Crowding*[e]
1	100	100	100	100	100
2	50	50	50	50	75
3	25	25	25	25	50
4	0	0	0	0	25
5	n.a.	n.a.	n.a.	n.a.	0

n.a. Not applicable.

Note: INEC is the National Statistical Institute of Ecuador.

a. Level 1 = public network; 2 = water truck; 3 = well; 4 = other.

b. Level 1 = in house, flush; 2 = in house, no flush; 3 = shared; 4 = other.

c. Level 1 = collection by truck; 2 = burned or buried; 3 = discarded; 4 = other.

d. Level 1 = household head has tertiary education; 2 = secondary; 3 = primary or literate: 4 = none or unknown.

e. Level 1 = one person or fewer to a bedroom; 2 = between one and two; 3 = between two and three; 4 = between 3 and 4; 5 = more than four.

Source: INEC (1996).

developed to meet an urgent government request at short notice. However, INEC's approach to constructing its indicator does resemble one that many countries have followed and therefore provides a useful example.

INEC constructed its basic needs indicator at the household level. It consists of a weighted composite of five variables capturing access to water, access to sanitation services, access to waste disposal services, education (of the household head), and a crowding index (the number of people per bedroom).[4] Each service was assigned a certain number of points according to its availability and its type or level (table 1). INEC assigned these points based on its own judgment, not as the result of our empirical analysis. For each household the value of the basic needs indicator was taken simply as the sum of points across services. The lower the sum, the poorer the household.

Using data from a recent household survey, the Ecuador Encuesta sobre las Condiciones de Vida (ECV), we can judge how well the basic needs indicator is able to identify households that are poor, as measured by consumption expenditures. The ECV for 1994 is a nationally representative household survey modeled closely on the World Bank's LSMS surveys. It provides detailed information on a wide range of topics, including food consumption, nonfood consumption, labor activities, access to services such as education and health, agricultural practices, and entrepreneurial activities. The survey was fielded by the Servicio Ecuatoriana de Capacitación (SECAP) in Ecuador during June–September 1994. It covered more than 4,500 households; after cleaning the data and checking for consistency, 4,391 households were left.

The survey design incorporated both clustering and stratification on the basis of the country's three main agroclimatic zones and a rural-urban breakdown. It also oversampled Ecuador's two main cities, Quito and Guayaquil, and covered

4. In other poverty-mapping exercises INEC has experimented with a wider range of variables. In El Salvador the government has constructed a poverty map using 12 variables (Government of El Salvador 1995).

Table 2. *Poverty Incidence under Alternative Definitions of Welfare*

Region	*Per capita consumption*	*Basic needs indicator*
Costa	0.35	0.39
Urban	0.26	0.18
Rural	0.49	0.76
Sierra	0.33	0.28
Urban	0.22	0.04
Rural	0.43	0.50
Oriente	0.59	0.65
Urban	0.20	0.03
Rural	0.67	0.76
National	0.35	0.35
Urban	0.25	0.13
Rural	0.47	0.62

Note: Figures are based on two alternative welfare indicators applied to the ECV household survey data.
Source: Authors' calculations.

about 1,374 rural households in total. Household expansion factors were added to the data set so that inferences could be made about population aggregates. World Bank (1996) analyzes the ECV data set as part of its study of poverty in Ecuador. Hentschel and Lanjouw (1996) construct consumption totals for each household, and World Bank (1996) bases all comparisons of welfare across households on their totals. Expenditures also have been adjusted to take into account regional variation in the cost of living based on a Laspeyres food price index reflecting the consumption patterns of the poor.

We compare poverty by region and area using the basic needs and consumption indicators (table 2). Since no poverty line was developed specifically for the basic needs indicator, we must infer poverty rates. We do this by equating the national incidence of poverty as measured by the basic needs indicator with the headcount rate, which we obtain using per capita consumption and the consumption poverty line of 45,476 sucres per person per fortnight (approximately $1.50 per person per day) developed in World Bank (1996). Hence we are asking how the regional ranking of poverty differs when poverty is defined using these two different indicators, but holding constant the fraction of the population identified as poor. We distinguish only between rural and urban areas and among the country's three main agroclimatic zones.

At this level of aggregation we derive the same rankings from the two alternative definitions of welfare. But regional differences are much more accentuated using the basic needs indicator. With the basic needs indicator rural areas appear more poor and urban areas appear less poor than with the consumption indicator. Among the rural areas, Oriente and Costa are poorer than Sierra, and among the urban areas, Costa is poorest, followed by Sierra and then Oriente. As World Bank (1996) emphasizes, under the consumption criterion the rankings of rural and urban areas between Costa and Sierra are highly unstable and easily over-

turned, depending on where one draws the poverty line and if one chooses to work with a poverty measure other than the headcount ratio. The basic needs criterion gives the impression that differences in well-being across regions are unambiguous.

Finally, we compare the performance of the two indicators at the household level. We follow the design of the planned intervention by taking the bottom 20 percent of households as the intended beneficiaries. First, we compute the total number of households represented in the ECV data and calculate that just fewer than 450,000 represent 20 percent of all households. Next, we calculate the total number of points for each household according to INEC's basic needs criterion and select the 450,000 households that score lowest. Last, we calculate the percentage of beneficiary households that fall into each household per capita expenditure quintile (column 2 in table 3). Since the target group is the first quintile, the percentage of beneficiaries in the first quintile indicates how well the basic needs indicator performs. Also, if all households were to receive the same amount of money, the percentage of beneficiaries in the first quintile represents the percentage of resources that would reach the targeted group. Only 41 percent of households identified by the basic needs criterion as constituting the poorest quintile are, in fact, among the bottom 20 percent according to the consumption criterion. Thus the leakage from an allocation based on the basic needs criterion would be very high: 60 percent of resources would go to households that are not in the lowest quintile, with almost 10 percent going to the top two quintiles.[5]

II. Predicting Poverty

To give our weighting scheme a more analytical basis, we consider developing a poverty map by imputing household consumption levels using census data.[6] We can do this only if certain data requirements are met. A household survey, such as the ECV in Ecuador, must be available and should correspond roughly to the same period covered by the census. In addition, unit (household)-level census data must be available. We were fortunate to have been granted access to the 1990 census data for Ecuador, covering roughly 2 million households. Although the 1994 ECV data were collected four years after the census, the 1990–94 period was one of relatively slow growth and low inflation in Ecuador, so it is reasonable to assume relatively little change.

The underlying intention of the method we propose here is similar to that of small-area and synthetic estimation procedures used in demographic and area

5. The basic needs indicator might not perform so poorly if the targeting scheme were aimed at, say, only urban areas or if an alternative cutoff point in the distribution were used.

6. Although imputing missing observations within a sample is a common procedure (see, for example, Paulin and Ferraro 1994), imputing values from a combination of different data sets (out-of-sample imputation) is less usual. Bramley and Smart (1996) combine the British Family Expenditure Survey with census information to estimate local income distributions. However, Bramley and Smart do not have access to unit-level data from both sources. They derive local income distributions, not from predicted household income, but from estimates of mean income and distributions of household characteristics.

Table 3. *Distribution of Beneficiaries Identified Using the Basic Needs Criterion and Predicted Consumption across Actual Consumption Expenditure Quintiles*

Quintile (as measured by per capita consumption)	*Percentage of beneficiary households (based on a basic needs indicator)*	*Percentage of beneficiary households (based on predicted consumption)*	
		Within-sample[a]	*Outside-sample*[b]
1 (poorest)	41.4	59.9	51.0
2	29.5	22.0	27.0
3	19.5	13.8	13.1
4	8.0	3.9	8.0
5	1.6	0.2	0.9

Note: Beneficiary households are those in the poorest quintile as measured by the basic needs indicator or by predicted consumption.

a. The within-sample exercise derives predicted household consumption from models estimated using the full household survey, applying the parameter estimates again to the full sample.

b. The outside-sample exercise consists of estimating the models for a subsample of the ECV and using the resulting parameter estimates to predict consumption for the remaining sample.

Source: Authors' calculations.

statistics.[7] In those procedures the interest is in deriving (unobserved) local-area attributes, such as a mean or total, often in the form of proportions (Farrell, MacGibbon, and Tomberlin 1997). For example, if we know population changes for a large area, we can use small-area estimation techniques to calculate population changes at lower geographic levels based on postulated functional relationships. The method we use here works in the opposite direction: we predict our variable of interest (consumption) at the unit (household) level and base aggregate statistics on those predictions.[8]

Estimating Models of Consumption

To impute expenditures using census data, we must first estimate a model of consumption using household survey data. Of course, the only variables that we can use to predict consumption are those that also are available in the census. In the case of Ecuador this set consists of demographic variables, such as the household's size and its age and sex composition; the education and occupation of each family member; the quality of housing (materials, size); access to public services, such as electricity and water; principal language spoken in the household; and location of the household. (See the appendix for comparative summary statistics from the two data sets.) The total number of explanatory variables, including dummy variables, interaction terms, and higher-order terms, is 48.

7. See Purcell and Kish (1980) for an overview. Isaki (1990) uses small-area estimation to obtain economic statistics.

8. The issue of combining information from different data sets has sparked a recent interest in the literature (see, for example, Arellano and Meghir 1992; Angrist and Krueger 1992; Lusardi 1996; and Imbens and Hellerstein 1999). These studies generally combine several household surveys, rather than census and survey data, and do not address spatial poverty estimation.

We estimate separate models for each region (Costa, Sierra, and Oriente), and within each region we distinguish between urban and rural areas. We also obtain separate estimates for Guayaquil and Quito, since the ECV oversampled these two cities.[9] The dependent variable in each regression is the logarithm of per capita consumption expenditures for household i, $\ln y_i$:

$$\ln y_i = \mathbf{X}_i'\boldsymbol{\beta} + \varepsilon_i \tag{1}$$

where $\mathbf{X}_i$ is a vector of independent variables common to the ECV and the census, and ε_i is a random disturbance term that is distributed i.i.d. $N(0, \sigma^2)$. We estimate the models with weighted least squares, using household sampling weights. The explanatory power of the rural models ranges from an R^2 of 0.46 for rural Sierra to an R^2 of 0.74 for rural Oriente. The explanatory power of the urban models ranges from an R^2 of 0.55 for Quito to an R^2 of 0.64 for urban Sierra.

We do not report here full sets of parameter estimates, standard errors, and diagnostics from the eight regression models for reasons of space, but these are available on request. It is important to correctly specify the precise functional form of the disturbance term in the consumption regression when calculating the second-stage poverty estimates. Thus we test the normality assumption made in equation 1. In three of the eight regions we cannot reject normality based on Shapiro-Wilk and joint skewness and kurtosis tests (all p-values > 0.15). Closer inspection of the residuals reveals that, in the other regions, we can reject normality only because of a few outliers in one or both of the tails. The outliers may be the result of mismeasurement. (For example, in one case the highest value of consumption expenditures is six times larger than the second-highest value.) After eliminating these observations—a total of only 13 out of 4,365—we cannot reject normality at conventional levels of significance in any region. Such small deviations from the assumed normality of the disturbance term should have a negligible effect on the accuracy of our results that follow. Further, with the exception of Guayaquil, we cannot reject (at the 10 percent level) the null hypothesis of homoskedasticity against the alternative of heteroskedasticity for the full set of independent variables.

9. It is of interest to consider how the basic needs weights in table 1 compare to those implied by the regression coefficients. Quito is a typical example. In the basic needs classification a decrease from four to three people per bedroom is associated with a welfare improvement equivalent to that from an increase in the education of the household head from primary to secondary school. The point estimates from the consumption regression suggest that an increase in the education of the household head from primary to secondary school is associated with an increase in consumption of 30 percent, while a reduction from four to three people per bedroom is associated with an increase in consumption of just 6.7 percent. An increase in education from secondary to tertiary also is associated with an increase in consumption of 30 percent. However, a decrease from three people to one person per bedroom, which has an equivalent welfare effect as moving from a secondary to a university education according to basic needs weights, is associated with a much larger increase in consumption: 47.6 percent. The same pattern holds across regions. Thus at high levels of crowding and low levels of education, the basic needs system gives more weight to reductions in crowding than to increases in education than would be appropriate given their relationship to consumption. If the basic needs weights are intended to reflect the relationships of both variables to overall consumption as well as an adjustment for their intrinsic value, then the weights seem to suggest a value judgment that being literate or attending primary school is less important than reducing crowded bedrooms.

Before moving to the second step, in which we apply the models to the census data, we test to see if predicting consumption (on the basis of the survey) improves targeting relative to the basic needs indicator. Although we obtain reasonable fits for cross-sectional regressions (as reported above), the coefficients of determination remain significantly lower than 1. To assess the performance of the model, we compare the basic needs indicator to actual consumption, as in the exercise reported in column 2 of table 3. We find that prediction models are better at identifying the poorest households—poorest in terms of consumption expenditures—than the basic needs indicator (columns 3 and 4 of table 3). In the first test we use the full household sample in the prediction models and apply the parameter estimates to the full sample (column 3). Targeting efficiency improves by almost 50 percent, with almost 60 percent of the bottom quintile as designated by predicted consumption also found in the bottom quintile as designated by actual consumption. The second test is considerably more demanding (column 4). Here, we (randomly) split the household survey in half and estimate the model of consumption using only half of the survey data. We then predict consumption for the other half of the sample (an out-of-sample prediction). As expected, the improvement over the basic needs indicator is less dramatic with this test. Nevertheless, if our goal is to target the bottom 20 percent of the population, this approach still improves targeting efficiency from 41.4 percent (basic needs) to 51.0 percent.

Predicting Poverty

We now proceed to the second step of the imputation exercise and apply to the census data the parameter estimates from the regressions (using the full household sample). For each household in the census we multiply its characteristics by the parameter estimates from the applicable regression (determined by the location of residence) in order to obtain an imputed value for the log of per capita consumption expenditures. We then estimate the household's probability of being poor, taking into account that the model does not perfectly explain consumption (the R^2 values never equal 1) and that predicted consumption is based on sample data. Finally, we calculate the incidence of poverty as the mean of the household-specific estimates for the population in a given region of the census.[10]

More formally, given a poverty line, z, the indicator of poverty, P_i, for each household i is

$$P_i = 1 \text{ if } \ln y_i < \ln z;\ P_i = 0 \text{ otherwise.} \tag{2}$$

10. Our discussion relies on a single poverty criterion—the incidence of poverty—and a single poverty line. One could, however, rank regions according to a range of poverty or inequality measures and experiment with a range of poverty lines (see Elbers, Lanjouw, and Lanjouw 2000). Also, our study examines the incidence of poverty among households. To calculate the incidence of poverty at the level of individuals, it is necessary to weight each household-level observation by the corresponding household size. The poverty figures provided in the tables are such weighted totals.

Using the model of consumption in equation 1, the expected poverty of household i with observable characteristics $\mathbf{X}_i$ is

$$E\left(P_i \mid \mathbf{X}_i', \boldsymbol{\beta}, \sigma\right) = \Phi\left[\frac{\ln z - \mathbf{X}_i'\boldsymbol{\beta}}{\sigma}\right] \tag{3}$$

where Φ is the cumulative standard normal distribution. Given that we are dealing with the headcount poverty indicator (equation 2), equation 3 is simply the probability that a household with observable characteristics $\mathbf{X}_i$ is poor.[11] From equation 1 we obtain estimates of $\hat{\boldsymbol{\beta}}$, the vector of coefficients, and $\hat{\sigma}$. Thus our estimator of the expected poverty of household i in the census is

$$P_i^* = E\left(P_i \mid \mathbf{X}'=_i, \hat{\boldsymbol{\beta}}, \hat{\sigma}\right) = \Phi\left[\frac{\ln z - \mathbf{X}_i'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right], \tag{4}$$

which, as a continuous function of consistent estimators is, itself, a consistent estimator of $E(P_i)$. P, regional poverty, is

$$P = \frac{1}{N}\sum_{i=1}^{N} P_i, \tag{5}$$

where N is the number of households in the region, and expected poverty is

$$E(P \mid \mathbf{X}, \boldsymbol{\beta}, \sigma) = \frac{1}{N}\sum_{i=1}^{N} E(P_i \mid \mathbf{X}_i, \boldsymbol{\beta}, \sigma). \tag{6}$$

The predicted incidence of poverty, P^*, given the estimated model of consumption, is thus

$$P^* = E(P \mid \mathbf{X}, \hat{\boldsymbol{\beta}}, \hat{\sigma}) = \frac{1}{N}\sum_{i=1}^{N} \Phi\left[\frac{\ln z - \mathbf{X}_i'\hat{\boldsymbol{\beta}}}{\hat{\sigma}}\right]. \tag{7}$$

We calculate the incidence of poverty as the mean of households' probability of being poor rather than simply count households whose predicted expenditures are below the poverty line. The latter approach would give biased estimates of poverty rates.[12] Because of the random component of consumption, ε, no household has a zero probability of being poor or nonpoor, given its observed characteristics.

11. That is, if we were to take infinite draws from a population of households, the resulting poverty rate among households with observable characteristics $\mathbf{X}_i$ would be that given in equation 3. This value is not, in general, the same as the actual poverty rate, which is a *sample* from this infinite population, and depends on the particular realizations of ε_i.

12. This problem has been noted in the context of measuring the welfare of individuals, in which the bias arises because of inequality in intrahousehold distribution (Haddad and Kanbur 1990). See also Ravallion (1988). The Peruvian statistical institute—INEI (1996)—develops a model very similar to the one used here but derives poverty rates by directly estimating the headcount rate, not the predicted probability of being poor.

For each geographic region we compare the estimated incidence of poverty from the census data, using our imputed consumption values, with the rates obtained from the ECV household survey, using the consumption figures actually in the data (table 4). The incidence of poverty estimated from the ECV data in Ecuador as a whole is 35 percent. In general, poverty rates in the survey are reasonably close to, although somewhat lower than, those from the census (except in rural Oriente, for which the figures are the same in the two data sources). The differences are likely a result of changes in the exogenous variables underpinning the consumption regressions between the 1990 census and the 1994 ECV survey. For example, reductions in poverty are most apparent for Sierra, the region in which mean years of schooling of the household head appear to have risen most sharply between 1990 and 1994 (see the appendix). At the regional level, standard errors on the poverty rates calculated from the census are remarkably low.[13]

The two data sources do not rank the eight regions identically, but both clearly identify rural areas as poorer than urban areas, with rural Oriente emerging as the poorest region. World Bank (1996) indicates that orderings of regions, based on the ECV data, generally are not robust in that alternative poverty lines and poverty rates produce different rankings. The only exception is the rural versus urban ranking, which is found to be highly robust (first-order stochastic dominance held, with rural Ecuador being consistently poorer than urban Ecuador). The comparison of regional rankings based on the ECV and census data is consistent with these dominance results.

Standard errors on the ECV poverty rates (table 4) are such that we cannot reject the hypothesis that within sectors (urban and rural) poverty rates across regions are the same (although we can distinguish statistically between urban

13. For poverty incidence calculated from census data, the standard error of our indicator around the true poverty rate can be calculated as follows (see Elbers, Lanjouw, and Lanjouw 2000 for details):

$$P^* = \sum_{i=1}^{N} \frac{m_i}{M}\left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}}\right),$$

$$Var(P^*) \approx \left(\frac{\partial P^*}{\partial \hat{\beta}}\right)' Var(\hat{\beta}) \frac{\partial P^*}{\partial \hat{\beta}} + \left(\frac{\partial P^*}{\partial \hat{\sigma}^2}\right)^2 \frac{2\hat{\sigma}^4}{n-k-1} + \sum_{i=1}^{N} \frac{m_i^2 P_i^*(1-P_i^*)}{M^2}$$

where n is the sample size for the consumption model with k parameters, estimated using the ECV survey; N is the number of households in the census population in the region of interest; m_i is the number of individuals in household i; and M is the total number of individuals in the census population.

$$\frac{\partial P^*}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} \frac{m_i}{M}\left(\frac{-x_{ij}}{\hat{\sigma}}\right)\phi\left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}}\right),$$

where ij indicates the jth element of the vector of explanatory variables for the ith household, and

$$\frac{\partial P^*}{\partial \hat{\sigma}^2} = -\frac{1}{2}\sum_{i=1}^{N} \frac{m_i}{M}\left(\frac{\ln z - X_i'\beta}{\hat{\sigma}^3}\right)\phi\left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}}\right).$$

Table 4. *Regional Poverty Rates for Ecuador*

Region	*ECV data*	*Census data*
Rural Oriente	0.67	0.67
	(0.02)	(0.004)
Rural Costa	0.50	0.52
	(0.02)	(0.002)
Rural Sierra	0.43	0.53
	(0.02)	(0.001)
Guayaquil	0.29	0.35
	(0.02)	(0.002)
Quito	0.25	0.33
	(0.02)	(0.002)
Urban Costa	0.25	0.29
	(0.02)	(0.002)
Urban Oriente	0.20	0.25
	(0.02)	(0.009)
Urban Sierra	0.19	0.29
	(0.02)	(0.003)

Note: Estimated standard errors are in parentheses.

a. Rates from the census are calculated using imputed expenditures based on a model calibrated from the ECV survey.

Source: Authors' calculations.

and rural sectors). Our estimates from the census data are sufficiently precise to permit meaningful comparisons across regions within sectors.[14]

III. Tracking Poverty at the Provincial Level

Using the methodology we have outlined, we can construct a poverty map, based on consumption expenditures, at a level of disaggregation below the eight broad regions for which the ECV is suitable. For example, there are nearly 400 cantons in Ecuador, each with some degree of local autonomy and administration, and these cantons can themselves be divided into more than 1,000 *parroquias* (parishes). Working with the census data, we can easily calculate expected poverty rates at the canton or parish level to determine where poverty is concentrated. In fact, as we have seen in the example described in section I, we can, in principle, use the census data to identify poor households and to target transfers to these households directly.

However, the standard errors on poverty estimates are a function of the degree of disaggregation of the poverty map (see the final term in the third equation of note 13). This warns us against attempting to use our methodology to identify, say, individual households that are poor.[15] Moreover, these objections come in addi-

14. Because the eight regions that we are comparing are based on different regression models in the ECV, the parameter estimates underlying the predicted expenditures are independent across regions. Therefore we can test for statistical significance of the difference in poverty rates between region r and region u based on the formula:

$$Var(P_r^* - P_u^*) \approx Var(P_r^*) + Var(P_u^*).$$

15. Suppose that the predicted probability of poverty for a given household is 48 percent. For a single household a lower-bound estimate of the standard error on that household's poverty rate would be $0.49 \cong \sqrt{0.48(1-0.48)}$.

tion to the well-known arguments against targeting in this way, which focus on the impact that such policies could have on the behavior of potential beneficiaries.[16]

Despite the dangers of micro-targeting, it may be desirable to develop a poverty map that is more disaggregated than broad regions. Ultimately, the optimal degree of disaggregation will depend on a number of factors. One is the precise purpose of the poverty map. Will it, for example, be used to identify government administrative areas so that the desired level of disaggregation is some level of local government? Or will it be used to identify poor villages or neighborhoods so that community-level projects (such as public infrastructure projects) can be better targeted? A second important consideration is whether we can assume that the parameter estimates from a regression model estimated, say, at the regional level, apply at subregional levels. Throughout this exercise we implicitly assume that, within a region, the model of consumption is the same for all households irrespective of the province, county, or community in which they reside. We cannot test this assumption, and at very fine levels of disaggregation it might be less appealing. (See Elbers, Lanjouw, and Lanjouw 2000 for a discussion of the implications of spatial autocorrections).

The desired degree of disaggregation also will depend on the availability of other sources of information, possibly local sources, on the poverty of individuals. Finally, other methods of local targeting, such as self-targeting, will become more important and effective at certain levels of disaggregation. Constructing a poverty map thus is likely to be a sequential process of gradual disaggregation until one reaches the point at which it seems there is no further insight to be gained.

Breaking down the headcount poverty rate by province, we see that poverty rates vary considerably (table 5). We also see that the standard errors on the poverty rates remain low, so that disaggregating to the level of provinces has not come at a significant cost in terms of statistical precision. A poverty map would have to be highly disaggregated before the standard errors would increase significantly because of small populations. In fact, only when the parish population falls well below 500 households do the corresponding standard errors rise enough to compromise comparisons (figure 1).[17]

16. Van de Walle and Nead (1995) provide a clear and thorough discussion of these issues.

17. We would calculate the standard error on the difference in poverty rates between two parishes in different regions as described earlier. However, because the parameter estimates determining the imputed expenditure figures are the same for all parishes within a given region, the standard error on the difference between two parishes in a given region is:

$$Var(P_1^* - P_2^*) = \left(\frac{\partial(P_1^* - P_2^*)}{\partial\hat{\beta}}\right)' Var(\hat{\beta})\left(\frac{\partial(P_1^* - P_2^*)}{\partial\hat{\beta}}\right) + \left(\frac{\partial(P_1^* - P_2^*)}{\partial\hat{\sigma}^2}\right)^2 \frac{2\hat{\sigma}^4}{n-k-1} + \sum_{i=1}^{N_1} \frac{m_i^2 P_{1i}^*(1-P_{1i}^*)}{M_1^2} + \sum_{k=1}^{N_2} \frac{m_k^2 P_{2k}^*(1-P_{2k}^*)}{M_2^2}.$$

$$\frac{\partial(P_1^* - P_2^*)}{\partial\hat{\beta}_j} = \sum_{i=1}^{N_1} \frac{m_i}{M_1}\left(\frac{-x_{ij}}{\hat{\sigma}}\right)\phi\left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}}\right) - \sum_{k=1}^{N_2} \frac{m_k}{M_2}\left(\frac{-x_{kj}}{\hat{\sigma}}\right)\phi\left(\frac{\ln z - X_k'\hat{\beta}}{\hat{\sigma}}\right)$$

where N, M, and m are defined as in footnote 13 for parishes 1 and 2, which are subscripted by i and k, respectively, the subscript j indicates the jth element of the given vector, and

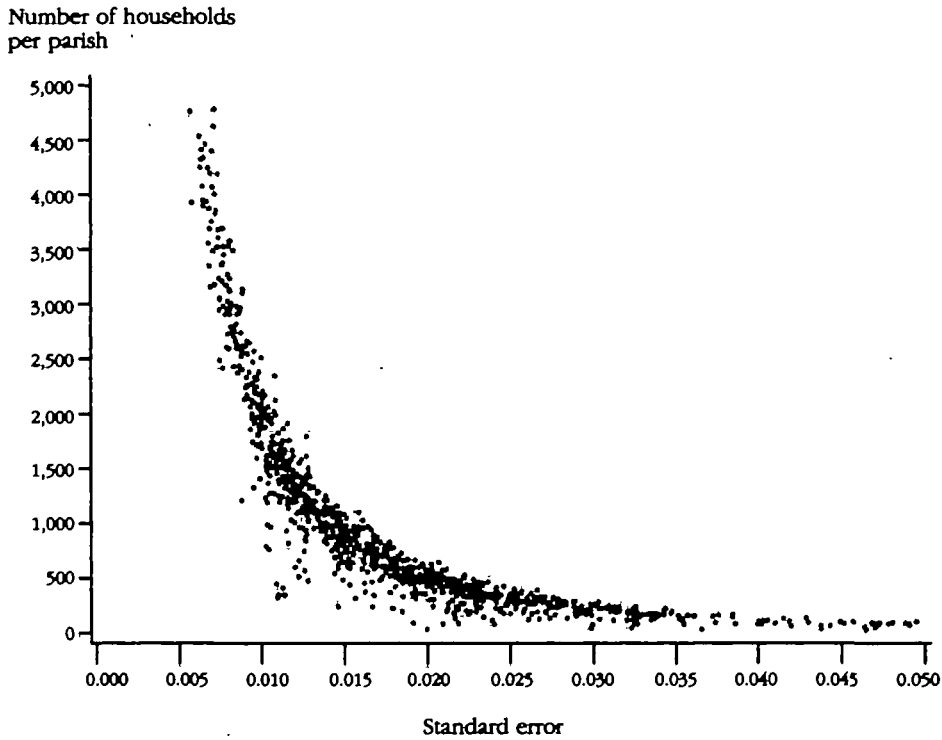
$$\frac{\partial(P_1^* - P_2^*)}{\partial\hat{\sigma}^2} = -\frac{1}{2}\sum_{i=1}^{N_1} \frac{m_i}{M_1}\left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}^3}\right)\phi\left(\frac{\ln z - X_i'\hat{\beta}}{\hat{\sigma}}\right) + \frac{1}{2}\sum_{k=1}^{N_2} \frac{m_k}{M_2}\left(\frac{\ln z - X_k'\hat{\beta}}{\hat{\sigma}^3}\right)\phi\left(\frac{\ln z - X_k'\hat{\beta}}{\hat{\sigma}}\right).$$

Table 5. *Poverty Map of Ecuador*

Region and province	*Expected poverty rate*	*Standard error*
Rural Oriente	0.67	0.004
Pastaza	0.65	0.005
Sucumbios	0.65	0.005
Morona Santiago	0.66	0.005
Zamora Chinchipe	0.67	0.005
Napo	0.69	0.004
Rural Sierra	0.53	0.001
Tungurahua	0.45	0.002
Pichincha	0.46	0.002
Azuay	0.50	0.002
Canar	0.52	0.003
Bolivar	0.55	0.003
Imbabura	0.56	0.003
Loja	0.57	0.003
Carchi	0.58	0.004
Chimborazo	0.59	0.003
Cotopaxi	0.63	0.003
Rural Costa	0.52	0.002
El Oro	0.45	0.003
Guayas	0.48	0.002
Los Rios	0.55	0.002
Manabi	0.56	0.002
Esmeraldas	0.59	0.003
Galapagos	0.14	0.008
Urban Costa		
Guayaquil	0.35	0.002
Urban Sierra		
Quito	0.33	0.002
Costa other urban	0.29	0.002
El Oro	0.24	0.003
Esmeraldas	0.27	0.004
Manabi	0.29	0.003
Guayas	0.30	0.003
Los Rios	0.32	0.003
Sierra other urban	0.29	0.003
Azuay	0.23	0.003
Tungurahua	0.25	0.004
Chimborazo	0.25	0.004
Cotopaxi	0.28	0.004
Loja	0.31	0.004
Canar	0.31	0.006
Imbabura	0.33	0.004
Carchi	0.33	0.005
Pichincha	0.33	0.003
Urban Oriente	0.25	0.009
Pastaza	0.24	0.011
Zamora Chinchipe	0.24	0.013
Morona Santiago	0.28	0.013

Source: Authors' calculations.

Figure 1. *Standard Errors on Headcount Rates and Population Disaggregation in Ecuador, Parish-Level Estimates*



Source: Authors' calculations.

IV. Concluding Remarks

In many developing countries poverty maps play an important role in guiding the allocation of public spending for alleviating poverty. A poverty map is essentially a geographic profile of poverty, indicating in which parts of a country poverty is concentrated and thus in which locations policies might be expected to have the greatest impact on poverty. A poverty map is most useful if it can be constructed at a fine level of geographic disaggregation.

To achieve such fine levels of disaggregation, it is essential to work with very large data sets. However, it is rare to find survey data that are both large in sample size and detailed in terms of household welfare. In general, there is a trade-off between size and quality because both goals are costly in financial and administrative terms.

In this article we have explored the possibility of combining the best parts of two different sources of data in order to construct a disaggregated poverty map that is based on an income or consumption measure of welfare. Constructing a poverty map based on census data, but using an ad hoc weighting scheme, may not be a good way to target those households deemed poor on the basis of their

consumption. Transfer programs based on such a map might reach only a subset of the intended beneficiaries and might entail considerable leakage to the nonpoor.

We suggest an alternative approach. Using household data from a high-quality, but small, living standards survey for Ecuador (SECAP 1994), we directly model consumption as a function of explanatory variables that also are present in the census. Because even the relatively few explanatory variables common to both the census and the ECV explain much of the variation in household consumption in the ECV, the incidence of poverty calculated from the census, based on this imputed consumption figure, is close to that calculated from the ECV. In Ecuador the poverty rates derived in the census generally are calculated with a high level of statistical precision. This precision declines as the degree of spatial disaggregation increases: although one might be tempted to use the methodology developed here to identify individual poor households, we demonstrate that such an application would be inappropriate. Our approach can be used at a high level of disaggregation but should be supplemented with complementary sources of information.

The most useful practical application of this methodology is probably in making comparisons with regional patterns of other indicators of well-being, opportunity, and access. For example, one could overlay our poverty map on a map documenting regional patterns of access to primary health care centers. Such an exercise could help policymakers decide where to prioritize efforts to expand access to primary health centers. It also could help policymakers decide how to expand such access—they might want to subsidize access in poor areas but experiment with cost-recovery methods in less-needy areas. Furthermore, a close correlation between, say, regional patterns of rural poverty and road access also might offer clues as to possible causes of poverty. This type of exercise could be undertaken for a wide range of indicators: levels of health and education, ethnicity and indigeneity, access to infrastructure and other public services, land quality and ecology, environmental conditions, and so on.

Finally, the ability to construct finely disaggregated poverty maps also might inform broader research questions. One direction would be to examine how the relationship between distributional outcomes and economic performance varies spatially within a country, in a manner analogous to cross-country analysis. This approach may avoid some of the methodological concerns arising with cross-country analysis. Other research questions also could be tackled. For example, underlying some of the current arguments in favor of decentralizing poverty programs is a notion that local communities themselves are best placed to identify the kinds of interventions that would be most beneficial to the poor. This position hinges somewhat on the contention that at the community level a subset of nonpoor households is less likely to capture public resources. This assumption probably is linked to the degree of inequality at the community level, something that traditionally has been difficult to measure. With the methodology presented here, household consumption inferred from the census could be analyzed to assess the extent of inequality within the community.

Appendix. *Comparative Descriptive Statistics from the 1994 LSMS and the 1990 Census*

	Rural Sierra		*Urban Sierra*		*Quito*		*Rural Costa*	
Indicator	*LSMS*	*Census*	*LSMS*	*Census*	*LSMS*	*Census*	*LSMS*	*Census*
Years of schooling of household head	4.48	4.33	8.75	8.11	10.67	9.52	3.63	4.12
	(3.36)	(4.18)	(5.15)	(5.19)	(5.25)	(5.23)	(3.16)	(4.01)
Male household head	0.84	0.78	0.81	0.77	0.82	0.79	0.96	0.87
Persons per bedroom	3.04	3.28	2.42	2.59	2.21	2.45	3.74	3.73
	(1.78)	(2.05)	(1.49)	(1.66)	(1.32)	(1.52)	(1.98)	(2.29)
Connection to public water network	0.31	0.52	0.94	0.89	0.90	0.83	0.08	0.21
Garbage collection by truck	0.25	0.19	0.80	0.81	0.89	0.88	0.05	0.12
Flush toilet	0.37	0.24	0.69	0.68	0.79	0.68	0.27	0.33
Telephone connection	0.04	0.07	0.31	0.27	0.43	0.36	0.00	0.03

	Urban Costa		*Guayaquil*		*Rural Oriente*		*Urban Oriente*	
Indicator	*LSMS*	*Census*	*LSMS*	*Census*	*LSMS*	*Census*	*LSMS*	*Census*
Years of schooling of household head	6.64	6.91	8.88	8.65	5.82	5.16	8.27	8.29
	(4.42)	(4.98)	(4.89)	(4.96)	(3.94)	(4.08)	(4.50)	(4.79)
Male household head	0.83	0.80	0.77	0.78	0.89	0.85	0.83	0.78
Persons per bedroom	3.16	3.12	3.01	2.99	3.54	3.49	2.54	2.64
	(2.08)	(1.96)	(1.87)	(1.92)	(1.89)	(2.26)	(1.50)	(1.63)
Connection to public water network	0.55	0.71	0.72	0.62	0.16	0.29	0.92	0.87
Garbage collection by truck	0.57	0.56	0.75	0.54	0.10	0.20	0.81	0.83
Flush toilet	0.66	0.71	0.76	0.75	0.23	0.18	0.66	0.60
Telephone connection	0.12	0.13	0.25	0.23	0.01	0.03	0.24	0.14

Note: Figures given are means. Standard deviations are in parentheses. All variables except for years of schooling and persons per bedroom are dummy variables taking the value of 1 for a positive response and 0 otherwise.

Source: Authors' calculations.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Alesina, Alberto, and Dani Rodrik. 1994. "Distributive Politics and Economic Growth." *Quarterly Journal of Economics* 109(2):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* 87(418):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* 59(3):537–59.

Atkinson, Anthony. 1989. *Poverty and Social Security.* Hemel Hempstead, U.K.: Harvester Wheatsheaf.

Bramley, Glen, and Gavin Smart. 1996. "Modelling Local Income Distributions in Britain." *Regional Studies* 30(3):239–55.

Bruno, Michael, Martin Ravallion, 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, Mass.: MIT Press.

Chaudhuri, Shubham, and Martin Ravallion. 1994. "How Well Do Static Welfare Indicators Identify the Chronically Poor?" *Journal of Public Economics* 53(3):367–94.

Deininger, Klaus, and Lyn Squire. 1996. "A New Data Set Measuring Income Inequality." *The World Bank Economic Review* 10(3):565–91.

Elbers, Chris, Jean Olson Lanjouw, and Peter Lanjouw. 2000. "Welfare in Villages and Towns: Micro-Measurement of Poverty and Inequality." Tinbergen Institute Working Paper. (forthcoming).

El Salvador, Government of. 1995. "Priorización de municipios a partir de datos censales." Dirección General de Política Económica y Social, Ministerio de Coordinación del Desarollo Económico y Social. San Salvador.

Farrell, Patrick, Brenda MacGibbon, and Thomas J. Tomberlin. 1997. "Empirical Bayes Estimation Using Logistic Regression Models and Summary Statistics." *Journal of Business and Economic Statistics* 15(1):101–08.

Fields, Gary. 1989. "Changes in Poverty and Inequality in Developing Countries." *The World Bank Research Observer* 4(2):167–85.

FONCODES. 1995. "Mapa de pobreza en Perú." Lima. Processed.

Haddad, Lawrence, and Ravi Kanbur. 1990. "How Serious Is the Neglect of Intra-Household Inequality?" *Economic Journal* 100(402):866–81.

Hentschel, Jesko, and Peter Lanjouw. 1996. "Constructing an Indicator of Consumption for the Analysis of Poverty: Principles and Illustrations with Reference to Ecuador." LSMS Working Paper 124. Development Research Group, World Bank, Washington, D.C. Processed.

Imbens, Guido, and Judith Hellerstein. 1999. "Imposing Moment Restrictions from Auxiliary Data by Weighting." Review of Economics and Statistics 81(1):1–4.

INEI (Instituto Nacional de Estadística e Investigaciones). 1996. "Metodología para determinar el ingreso y la proporción de hogares pobres." Lima. Processed.

Isaki, C. T. 1990. "Small-Area Estimation of Economic Statistics." *Journal of Business and Economic Statistics* 8(4):435–41.

Lanjouw, Jean Olson, and Peter Lanjouw. 1997. "Poverty Comparisons with Non-compatible Data: Theory and Illustrations." Policy Research Working Paper 1709. Policy Research Department, World Bank, Washington, D.C. Processed.

Lusardi, Annemaria. 1996. "Permanent Income, Current Income, and Consumption: Evidence from Two Panel Data Sets." *Journal of Business and Economic Statistics* 14(1):81–90.

Paulin, G. D., and D. L. Ferraro. 1994. "Imputing Income in the Consumer Expenditure Survey." *Monthly Labor Review* 117(12):23–31.

Purcell, N. J., and Leslie Kish. 1980. "Postcensal Estimates for Local Areas (or Domains)." *International Statistical Review* 48(1):3–18.

Ravallion, Martin. 1988. "Expected Poverty under Risk-Induced Welfare Variability." *Economic Journal* 98(393):1171–82.

———. 1994. *Poverty Comparisons*. Chur, Switzerland: Harwood Academic Publishers.

SECAP (Servicio Ecuatoriana de Capacitación). 1994. *Encuesta sobre las condiciones de vida*. Quito.

Sen, Amartya K. 1984. *Resources, Values, and Development*. Oxford: Blackwell.

van de Walle, Dominique, and Kimberly Nead, eds. 1995. *Public Spending and the Poor: Theory and Evidence*. Baltimore, Md.: Johns Hopkins University Press.

World Bank. 1996. "Ecuador Poverty Report." World Bank Country Study. Ecuador Country Department, World Bank, Washington, D.C. Processed.