WPS 1928

POLICY RESEARCH WORKING PAPER 1928

Combining Census and Survey Data to Study Spatial Dimensions of Poverty

Jesko Hentschel
Jean Olson Lanjouw
Peter Lanjouw
Javier Poggi

Combining sample survey data and census data can yield predicted poverty rates for all households covered by the census. This offers a means to construct detailed poverty maps. But standard errors on the estimated poverty rates are not negligible.

The World Bank
Development Research Group
and
Poverty Reduction and Economic Management Network
Poverty Division
June 1998

Summary findings

Poverty maps, providing information on the spatial distribution of living standards, are an important tool for policymaking and economic research.

Policymakers can use such maps to allocate transfers and inform policy design. The maps can also be used to investigate the relationship between growth and distribution inside a country, thereby complementing research using cross-country regressions.

The development of detailed poverty maps is difficult because of data constraints. Household surveys contain data on income or consumption but are typically small. Census data cover a large sample but do not generally contain the right information. Poverty maps based on census data but constructed in an ad-hoc manner can be unreliable.

Hentschel, Lanjouw, Lanjouw, and Poggi demonstrate how sample survey data and census data can be combined to yield predicted poverty rates for all households covered by the census. This represents an improvement over ad hoc poverty maps. However, standard errors on the estimated poverty rates are not negligible, so additional efforts to cross-check results are warranted.

This paper — a joint product of the Development Research Group and the Poverty Reduction and Economic Management Network, Poverty Division — is part of a larger effort in the Bank to study the spatial distribution and determinants of poverty. Copies of the paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Peter Lanjouw, room MC3-555, telephone 202-473-4529, fax 202-522-1153, Internet address planjouw@worldbank.org. Jesko Hentschel may be contacted at jhentschel@worldbank.org. June 1998. (31 pages)

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

Jesko Hentschel
World Bank

Jean Olson Lanjouw
Yale University

Peter Lanjouw
World Bank

Javier Poggi
Superintendencia Banco y Seguros, Lima, Peru

We are very grateful to the Instituto Nacional de Estadistica y Censos (INEC) in Ecuador for access to the unit record data of the 1990 Census. We have benefited from discussions with Uwe Deichmann, Stefan Dercon, Jean-Yves Duclos, Joanna Gomulka, Vassilis Hajivassiliou, Jeffrey Hammer, Berk Ozler and Martin Ravallion. Useful comments were received from seminar participants at the London School of Economics and Université Laval, Quebec, and from participants at a workshop on Geographical Targeting for Poverty Reduction and Rural Development at the World Bank in November 1997. Thanks also to Jim Shafer for helpful assistance.

I. Introduction

Poverty maps provide a detailed description of the spatial distribution of poverty within a country. They can be of considerable value to governments, non-governmental organizations and multilateral institutions interested in strengthening the poverty alleviation impact of their spending. For example, they can be used to guide the division of resources among local agencies or administrations as a first step in reaching the poor. Poverty maps are currently being used in many developing countries for this purpose.

Poverty maps can also be an important tool for research. The empirical relationship between poverty or inequality and indicators of development, such as economic growth, is typically examined in a cross-country regression framework.[1] It is difficult, however, to control for the enormous heterogeneity which exists across countries; heterogeneity which may mask true relationships. There is, too, a limited universe of country experience to bring to bear on the question. Moving to more micro studies--using variation across communities within a single country--offers an attractive way forward.

But the development of poverty maps is often hampered by the sparsity of disaggregated data. The welfare indicator on which a finely disaggregated poverty map is based should be one which enjoys widespread acceptance. Carefully measured income or expenditure indicators are often favored for this purpose. However, the information required for a finely disaggregated

[1] Deininger and Squire (1996) have recently compiled a large international database for this purpose. Bruno, Ravallion and Squire (1998) utilize this database to explore the relationship between economic growth and inequality. For other recent contributions see Alesina and Rodrik (1994), Anand and Kanbur (1993), Fields (1989), and Persson and Tabellini (1994).

map based on income or expenditure, is not generally available for sufficient numbers of households. For example, the World Bank's Living Standard Measurement Surveys (LSMS), variants of which have been fielded in many developing countries, collect the requisite information but are too small to allow for disaggregation beyond a simple rural/urban breakdown within broad regions of a given country.

Census data do not suffer from small sample problems, but typically contain little direct information on household resources. The lack of income or expenditure information in such datasets has prompted policy makers to explore alternative welfare indicators on which to base their poverty maps. In many Latin American countries, but also in Africa and Asia, poverty maps used to rank regions have been based on indices of welfare constructed by combining, in some manner, different pieces of information available in the census, such as access to public services, education levels, etc.[2] This type of welfare indicator, occasionally labeled a "Basic Needs" (BN) indicator, is generally constructed in a fairly *ad-hoc* manner and is limited by the essentially qualitative nature of much of the information available in the census. Using detailed household survey data for Ecuador, we show in this paper that a crude BN indicator and a comprehensive consumption measure yield markedly different welfare rankings of households. With consumption as a yardstick, we illustrate that sizable targeting errors can result from the use of a crude BN indicator.

We then explore to what extent census-based maps can be improved if a household survey

---

[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 in which such maps are used to guide the allocation of public social sector spending include Colombia, Honduras, and Venezuela.

containing income or expenditure data is also available. Using such a dataset for Ecuador, we estimate models of consumption expenditure, restricting the set of explanatory variables to those which are also available in the most recent census for Ecuador. We apply the parameter estimates from these models to the census data to predict the probability that a given household in the census is in poverty. We check the performance of our approach by estimating the incidence of poverty in six broad regions and comparing these with rates estimated from the household survey alone. The poverty rates coincide closely across datasets.

We consider some of the statistical issues which arise from the fact that the poverty figures have been predicted. The approach described above yields estimates of the incidence of poverty from the census which are unbiased, so that, in expectation, prediction errors are zero. However the poverty estimates do have standard errors associated with them and we illustrate that these are not small in size. Although confidence intervals do not widen further as a function of the degree of spatial disaggregation, they remain large enough that only a minority of pairwise comparisons of poverty are significantly different from each other at the regional or provincial level. These considerations imply that a poverty map based on the approach described here should be regarded as a useful first step only and that maps derived in this way would benefit from further validation.

In the next section we describe the data for Ecuador used to illustrate the analysis. We then consider what targeting errors would be implied by an allocation based on a Basic Needs indicator, relative to an allocation based on consumption expenditures. In Section III we estimate models of expenditure and then predict the probability of poverty for each household

in the census. From this we estimate aggregate poverty rates and compare these with rates obtained from the household survey. Section IV develops a simple province-level poverty map for Ecuador and illustrates that, while poverty in Ecuador does vary markedly across provinces and between rural and urban areas, only a minority of differences between provinces are statistically significant. Section V offers concluding remarks and suggestions for further research.

II. Targeting Errors With a Basic Needs Indicator: An Illustration from Ecuador

In this section we examine how effectively a BN indicator performs as an indicator of welfare - judged in terms of its ability to identify the poor defined in terms of consumption expenditures. Consumption is itself an imperfect indicator of the standard of living, and not an entirely non-controversial one. However, a comprehensive measure of consumption expenditure comes reasonably close to the goal of capturing a household's *achievement* of well-being; its own chosen bundle of goods and services - the outcome of its own utility maximization calculation.[3]

The BN indicator we are considering was developed in 1994 by the National Statistical Institute of Ecuador (INEC) in response to a specific request from government to develop a directory of poor households. This directory was to be used to target compensatory transfers

[3] The choice between income and consumption is also one which merits attention. For developing countries probably the most compelling argument in favor of consumption is that it is typically easier to measure accurately. Additional arguments in favor of consumption typically point to the relative smoothness of consumption across seasons or even from year to year, making consumption a better indicator of long-term living standards than a measure of current income (although see Chaudhuri and Ravallion, 1994). For more discussion see Atkinson (1989), Ravallion (1994) and Sen (1984). Hentschel and Lanjouw (1996) and Lanjouw and Lanjouw (1997) also discuss further the attractions of a *comprehensive* indicator of consumption expenditures as an indicator of welfare

to poor households for a gas price increase that would result if the government were to remove its gas subsidy. In the event, this program was not implemented, and we do not wish to imply that the BN indicator was regarded by INEC as anything other than a fairly crude measure developed to meet an urgent government request at short notice. However, the approach taken by INEC in constructing their BN indicator does resemble that which has been followed in many countries, and therefore provides a useful example.

INEC's BN indicator was constructed at the household level and consisted of a weighted composite of 5 variables capturing access to water, access to sanitation and waste disposal services, education (of the head of household) 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 quality. The points assigned to each service were arbitrary and are presented in Table 1.

[4] In other poverty mapping exercises for Ecuador, INEC has experimented with wider ranges of variables. In El Salvador, the government has constructed a poverty map using 12 different variables (Government of El Salvador, 1995).

Table 1
Points By Services Included in the INEC BN Indicator

service/level	water	sanitation	waste	education	crowding
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	-	-	-	-	0

Key:

Water:	1=public network; 2=water truck; 3=well; 4=other.
Sanitation:	1=flush, in house; 2=in house, no flush; 3=shared; 4=other.
Waste:	1=collection by truck; 2=burned or buried; 3=discarded; 4=other.
Education: (of head)	1=tertiary; 2=secondary; 3=primary or literate; 4=none or unknown.
Crowding: (persons per bedroom)	1=one or less; 2=between one and two; 3=between 2 and three; 4=between 3 and 4; 5=more than four.

For each household, the BN value was simply taken as the total sum of points across services. The lower the value of points per household, the poorer it was designated.

Using the data from a recent household survey we can examine how well the BN indicator performs in identifying the poor. The *Ecuador Encuesta Sobre Las Condiciones de Vida* (ECV) for 1994 is a nationally representative household survey modeled closely on the World Bank's Living Standards Measurement Surveys. It provides detailed information for each household on a wide range of topics including food consumption, non-food items, labor

activities, access to services such as education and health, agricultural practices and household entrepreneurial activities. The survey was fielded by the Servicio Ecuatoriana de Capacitacion (SECAP) in Ecuador during the period June-September, 1994. Over 4,500 households were surveyed in total and after cleaning and data consistency checks, information on 4,391 households is available for analysis.[5] The ECV dataset has been analyzed in a detailed study of poverty in Ecuador by the World Bank (World Bank, 1996). Hentschel and Lanjouw (1996) constructed consumption totals for each household, and all comparisons of welfare across households in the World Bank Poverty Report (World Bank, 1996) were based on that criterion.[6]

In Table 2 we compare poverty by region and area using the BN and consumption indicators. As no poverty line was developed specifically for the BN indicator, we must infer poverty rates. We do this by equating the national incidence of poverty using this indicator with the headcount rate which obtains using per capita consumption and the consumption poverty line 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 total fraction of the population identified as poor. We distinguish only between rural and urban areas, and the three main agro-climatic zones of the country.

---

[5] The survey design incorporated both clustering and stratification on the basis of the three main agro-climatic zones of the country and a rural/urban breakdown. The survey design also included an oversampling of Ecuador's two main cities, Quito and Guayaquil. Some 1374 rural households were surveyed in total. Household expansion factors were added to the data set so that inferences can be made about population aggregates.

[6] Expenditures have also been adjusted to take into account regional cost of living variation based on a Laspeyres food price index reflecting the consumption patterns of the poor.

Table 2

Poverty Incidence Under Alternative Welfare Definitions

	Per Capita Consumption	Indicator of "Basic Needs"
Costa		
Urban	0.26	0.18
Rural	0.49	0.76
Total	**0.35**	**0.39**
Sierra		
Urban	0.22	0.04
Rural	0.43	0.50
Total	**0.33**	**0.28**
Oriente		
Urban	0.20	0.03
Rural	0.67	0.76
Total	**0.59**	**0.65**
National		
Urban	0.25	0.13
Rural	0.47	0.62
Total	**0.35**	**0.35**

Note: Calculations are based on two alternative welfare indicators applied to the ECV household survey data.

At this level of aggregation the rankings which derive from the two alternative definitions of welfare are the same but regional differences are much more accentuated using the BN indicator. Rural areas appear more poor using the BN indicator than with consumption, and

urban areas look less poor. Within rural areas, rural Oriente and rural Costa are more poor than rural Sierra and within urban areas, the Costa region is poorest, followed by the Sierra and then Oriente. As has been emphasized in World Bank (1996), under the consumption criterion the rankings within rural and urban areas between Costa and Sierra are highly unstable and easily overturned depending on where one draws the poverty line, and whether one chooses to work with some alternative poverty measure than the headcount ratio. Under the BN criterion the impression gained is that differences in well-being across regions are unambiguous.

Finally, we looked beyond regional comparisons to also compare the performance of the two indicators at the household level. For this purpose we followed the design of the planned intervention by taking the bottom 20% of households as the intended beneficiaries of the program. We conducted the following experiment: we computed the total number of households represented by the ECV data and calculated that just under 450,000 households represented one fifth of all households. Next, we calculated the total number of points for each household according to INEC's BN criterion and selected the 450,000 households with the lowest points. Finally we calculated the percentage of the beneficiary households falling into each household per-capita expenditure quintile. Since the intended target group is the first quintile, the percentage of beneficiaries in the first quintile indicates how well the BN indicator identifies this particular target group. Also, if all households were to receive the same amount of money, the percentage of beneficiaries also represents the percentage of resources that would reach the targeted group. The results of this experiment are presented in Table 3.

Table 3
Distribution of Bottom 20% Under the BN Criterion
Across Consumption Expenditure Quintiles

Quintile of True Per Capita Consumption	Percentage of Beneficiary Households (Based on a BN Indicator)	
	Percentage	Cumulative
Poorest 20%	41.4	41.4
20-40%	29.5	70.9
40-60%	19.5	91.4
60-80%	8.0	98.4
Richest 20%	1.6	100.0

From Table 3 we can see that only 41.4% of households identified under the BN criterion as constituting the bottom 20% of all households are, in fact, among the bottom 20% under a consumption criterion. Thus, the leakage of resources from an allocation based on this criterion would be very high - 60% of resources would go to non-intended beneficiaries, with almost 10% going to the top two quintiles.

III. Predicting Poverty

Given the arbitrariness of a poverty map based on an indicator such as the BN indicator described above, we consider here the possibility of imputing a household consumption levels using census data to form the basis of a poverty map.[7] This course of action can be pursued only if certain data requirements are met. First, a household survey such as the ECV in

[7] While within sample imputation of missing observations is a quite common procedure (e.g. Paulin and Ferraro 1994), out of sample imputation which combines different datasets is less frequent. In a recent article, Bramley and Smart (1996) combine the British Family Expenditure Survey with Census information to estimate local income distributions. However, Bramley and Smart did not have access to unit level data from both data sources. They derived local income distributions not from predicted household incomes but from estimates of mean incomes and distribution characteristics.

Ecuador must be available, and should correspond roughly to the same period as covered by the census. Second, unit record level census data must be available for analysis. We have been fortunate to have been granted access to the 1990 census data for Ecuador, covering roughly 2.5 million households, for the purpose of this analysis. Although the 1994 ECV data were collected four years after the census, this period was one of relatively slow growth and low inflation in Ecuador so that problems of comparability are possibly less pressing.

The underlying intention of the method proposed here is similar to that of small-area and synthetic estimation procedures applied in demography and area statistics.[8] There, the interest is with the derivation of (unobserved) local area attributes such as a mean or total, often in the form of proportions (Farrell *et al*, 1997). For example, if population changes are known for a large area, small-area estimation techniques allow one to calculate population changes at lower geographic levels based on postulated functional relationships. An important difference in the method proposed here is that we predict our variable of interest (consumption) at the unit (household) level and base aggregate statistics on these predictions.[9]

Estimating Models of Consumption

To impute expenditures using the census, the first step is to estimate a model of consumption using household survey data. Of course, the only variables which can be used to

[8] See Purcell and Kish (1980) for an overview and Isaki (1990) for an application of small-area estimation to obtain economic statistics.

[9] The combination of information of different datasets has sparked a recent interest in the literature (e.g. Arellano and Meghir 1992, Angrist and Krueger 1992, Lusardi, 1996). These studies generally combine several household surveys, rather than census and survey data, and have not addressed spatial poverty estimation.

predict consumption are variables which are also available in the census. In the case of Ecuador this set of potential predictors consisted of various demographic variables such as household size and its age/sex composition; education and occupation information for each family member; housing quality data (materials, size); access to public services such as electricity and water; principal language spoken in the house; and location of residence. After defining various dummy variables, interaction terms and higher-order terms, the total number of explanatory variables available for the regressions was 48.

Separate models were estimated for each region (Costa, Sierra and Oriente) and, within these, distinguishing between urban and rural areas. Separate estimates were also obtained for Guayaquil and Quito as the ECV oversampled these two cities. The dependent variable in each regression was the logarithm of per-capita consumption expenditure. The models were estimated using weighted least squares with household sampling weights as weights. The explanatory power of the eight models ranged from an R^2 of 0.46 for the rural Sierra, to an R^2 of 0.74 for the rural Oriente. The R^2's for the urban models ranged from 0.55 (Quito) to 0.64 (Urban Sierra).[10]

Before moving on to the second step, which involves applying the models to the census data, we tested whether predicting consumption (on the basis of the survey) would improve targeting as compared to the more *ad hoc* Basic Needs Indicator discussed above. Although we obtained quite reasonable fits for cross-sectional regressions (as reported above), the coefficients of determination remained significantly lower than one. To assess the performance of the

[10]. Parameter estimates, standard errors, and diagnostics from the eight regression models are not reported here for reasons of space, but are available from the authors upon request.

model, we performed an analogous exercise to the one reported in Table 3, in which the Basic Needs Indicator was compared with actual consumption. Table 4 shows the results of comparing predicted with actual consumption levels.

Table 4
Distribution of Bottom 20% Using Predicted Consumption Across Actual Consumption Expenditure Quintiles

Quintile of True Per Capita Consumption	Percentage of Beneficiary Households (Based on Predicted Consumption)		BN Indicator
	Within-Sample[1]	Out-of-Sample[2]	(from Table 3)
Poorest 20%	59.9	51.0	41.4
20-40%	22.0	27.0	29.5
40-60%	13.8	13.1	19.5
60-80%	3.9	8.0	8.0
Richest 20%	0.2	0.9	1.6

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

2 The outside-sample exercise consisted of estimating the models for a sub-sample of the LSMS and then using the resulting parameter estimates to predict consumption for the remaining sample.

From Table 4 we see that prediction models do indeed perform better in identifying the poorest households than the Basic Needs Indicator. The first test consisted of using the full household sample in the prediction models and applying the parameter estimates to the full sample. In Table 4 we can see an improvement in the targeting efficiency by almost fifty percent with 59.9 percent of the bottom quintile according to predicted consumption also being found in the bottom quintile according to actual consumption. The second test was considerably more demanding. Here we split the household survey in half (randomly), estimated the model of consumption using only one half of the survey data and predicted consumption for the other

half (an out of sample prediction). As expected, the improvement over the Basic Needs indicator is less dramatic with this test. Nevertheless, if one's goal is to target the bottom 20% of the population, this approach still improves the targeting efficiency from 41.4 percent (Basic Needs) to 51.0 percent.

Predicting Poverty

We now proceed to the second step in the imputation exercise, which consisted of applying the parameter estimates from the regressions (using the full household sample) to the census data. For each household in the census, the parameter estimates from the applicable regression (determined by the location of residence) were multiplied by the household's characteristics in order to obtain an imputed value for (log) per capita consumption expenditure. We then estimated the household's probability of being poor taking into account the fact that consumption was not perfectly explained by the model (the R^2's were never 1) and that predicted consumption was based on sample data. Finally, the incidence of poverty was calculated as the mean, over all households in a given region of the census, of the household-specific estimates.[11]

More formally, we model log per capita expenditures for household i, lny_i, as a function of a vector of explanatory variables, X_i, common to the ECV and the census, and a random disturbance term, ϵ_i:

[11] Our discussion will be in terms of a single poverty criterion - the incidence of poverty - and a single poverty line. One could, however, rank regions on the basis of a large range of alternative poverty measures, and experiment with a range of possible poverty lines. The poverty map is not likely to remain invariant to such alternatives.

$$lny_i = X_i'\beta + \epsilon_i \qquad \epsilon_i \sim N(0,\sigma^2) \tag{1}$$

Given a poverty line, z, the indicator of poverty P_i for each household i is

$$P_i = 1 \text{ if } lny_i < lnz; \; P_i = 0 \text{ otherwise.} \tag{2}$$

The expected poverty of a household i with observable characteristics X_i is thus

$$E[P_i|X_i,\beta,\sigma] = \Phi[\frac{lnz - X_i'\beta}{\sigma}] \tag{3}$$

where Φ is a cumulative standard normal distribution. Given that we are dealing with the headcount poverty indicator (2), the value in (3) is simply the probability that a household with observable characteristics X_i is poor.[12] We estimate (1) to obtain estimates of $\hat{\beta}$ the vector of coefficients, and $\hat{\sigma}$. Thus our estimator of the expected poverty of household i in the census is

[12] That is, if one were to take infinite draws from a population of households, the resulting poverty rate among households with observables X_i would be that given in (3). Note that this value will not, in general, be the same as the headcount rate in any particular year among households with observables X_i, since the latter can be seen as a *sample* from this infinite population, and will be a function of the particular realizations of ϵ_i in that year.

$$\hat{E}[P_i|X_i,\hat{\beta},\hat{\sigma}] = \Phi(\frac{lnz - X_i'\hat{\beta}}{\hat{\sigma}}] \tag{4}$$

which, as a continuous function of consistent estimators is, itself, a consistent estimator of $E[P_i]$.

P, regional poverty, is

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

and expected regional poverty

$$E[P|X_i,\beta,\sigma] = \frac{1}{N}\sum_{i=1}^{N} E[P_i|X_i,\beta,\sigma] \ . \tag{6}$$

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

$$[P^*|X_i,\hat{\beta},\hat{\sigma}] = \frac{1}{N}\sum_{i=1}^{N} \Phi[\frac{lnz - X_i'\hat{\beta}}{\hat{\sigma}}]. \tag{7}$$

Note that we calculate the expected incidence of poverty as a mean of household-level probabilities of being poor, rather than simply counting up those households whose predicted expenditure is below the poverty line, z. The latter approach would give biased estimates of

poverty rates.[13] Given the random component of consumption, ϵ, no household has a zero probability of being poor or non-poor given its observed characteristics.

In Table 5 we report the estimated expected incidence of poverty from the census data, using our imputed consumption values, for each of the eight geographic regions. We compare these rates with those obtained from the ECV household survey using the consumption figures actually in the data. In the ECV data the estimated incidence of poverty in Ecuador as a whole is 35%.

[13] This has been noted in the context of errors in individual welfare measurement due to inequality in intra-household distribution (Haddad and Kanbur, 1990). See also Ravallion (1988). The Peruvian statistical institute INEI (1996) has developed a model very similar to the one described above but derived poverty rates from direct estimation of the headcount rate and not from the predicted poverty probabilities.

Table 5
Regional Poverty Rates For Ecuador

Comparing Rates from the 1994 ECV to Rates from the Census Using Imputed Expenditures Based on a Model Calibrated from the ECV Survey

	Poverty Incidence (standard errors in brackets)		
ECV Ranking	*ECV*	*Census Rank*	
1. Rural Oriente	0.67	0.53 (0.04)	(1)
2. Rural Costa	0.50	0.42 (0.04)	(3)
3. Rural Sierra	0.43	0.47 (0.03)	(2)
4. Guayaquil	0.29	0.28 (0.07)	(4)
5. Quito	0.25	0.28 (0.08)	(5)
6. Urban Costa	0.25	0.24 (0.09)	(6)
7. Urban Oriente	0.20	0.21 (0.11)	(8)
8. Urban Sierra	0.19	0.23 (0.09)	(7)

The poverty rates calculated on the basis of consumption imputed from the census data are quite

close to those based on the survey. With the exception of the rural Oriente, which appears substantially poorer in the ECV than in the census, one cannot reject that the poverty rates are the same across the two datasets.[14]

Rankings of the eight regions are not identical across the two data sources, but in both cases rural areas are clearly identified as poorer than urban areas, with rural Oriente emerging as clearly the poorest of all regions. World Bank (1996) indicated that orderings of regions, based on the ECV data, were generally non-robust in the sense that the use of alternative poverty lines and poverty rates often resulted in re-rankings of regions. The only exception in this regard was the rural versus urban ranking, which was found to be highly robust (first-order stochastic dominance held with rural Ecuador consistently poorer than urban Ecuador). The comparison of regional rankings across the ECV and census data is thus consistent with these dominance results.

In terms of statistical significance, using the census-based estimates it is not possible to

---

[14] Hentschel and Lanjouw (1996) and Lanjouw and Lanjouw (1996) report standard errors on the ECV poverty rates of around 0.02. For the census poverty incidence, the standard error can be calculated using a Taylor's approximation of the variance:

$$Var(P^*) = \left(\frac{\partial P^*}{\partial \hat{\beta}}\right)' Var(\hat{\beta}) \frac{\partial P^*}{\partial \hat{\beta}} + \left(\frac{\partial P^*}{\partial \hat{\sigma}}\right)^2 \hat{\sigma}^2 .$$

$$\frac{\partial P^*}{\partial \hat{\beta}_j} = \frac{1}{N} \Sigma_{i=1}^{N} \left(-\frac{x_{ij}}{\hat{\sigma}}\right)\left(\frac{lnz - X_i'\hat{\beta}}{\hat{\sigma}}\right),$$

where N is the number of census households and ij indicates the jth element of the vector of explanatory variables for the ith household.

$$\frac{\partial P^*}{\partial \hat{\sigma}} = \frac{1}{N} \Sigma_{i=1}^{N} -\left(\frac{lnz - X_i'\beta}{\hat{\sigma}^2}\right)\phi\left(\frac{lnz - X_i'\hat{\beta}}{\hat{\sigma}}\right).$$

reject that the incidence of poverty across rural regions is the same, and similarly that across urban areas the incidence is the same. It is possible to establish, at a 95% confidence level, that poverty in rural Oriente and in rural Sierra is lower than in urban regions. And rural Costa is poorer than urban areas (except Quito) with 90% confidence.[15]

IV. Province-Level Poverty in Ecuador: an Illustration

The purpose of the methodology outlined in the previous sections is to allow us to 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 themselves can be divided into a total of well over 1000 parroquias ("parishes"). Working with the census data, one could easily calculate canton-level or parroquia-level poverty rates to determine where poverty is concentrated. In fact, as we have seen in the example described in Section II, the census-level information can, in principle, be used to identify poor households and to target transfers to these households directly.

While the standard errors on poverty estimates are not a function of the degree of disaggregation of the poverty map, they are clearly large enough to caution strongly against

[15] Because the eight regions being compared are based on different regression models in the ECV, the parameter estimates underlying the predicted expenditures are independent across regions. In this case one can test for statistical significance of the difference in poverty rates between region r and region u based on the formula:

$$Var(P_r^* - P_u^*) = Var(P_r^*)^2 + Var(P_u^*)^2 .$$

attempts to identify poor households for targeting purposes directly from the census data.[16] Either a large number of non-poor households would have to be included among the recipients of transfers or one would have to accept that a large proportion of poor households would be excluded. Moreover, these objections are in addition 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.[17]

Despite the caution against micro-targeting, it may well be useful to develop a poverty map at a degree of disaggregation below broad regions. Ultimately, the optimal degree of disaggregation will depend on a number of factors. One is the precise purpose that the poverty map is expected to fulfill. Is it, for example, intended to identify government administrative areas so that the desired level of disaggregation is some level of local government? Or is it intended to identify poor villages or neighborhoods so that community-level project interventions (such as public infrastructure) can be better targeted? A second important consideration is whether the parameter estimates from a regression model estimated, say, at the regional level, can be assumed to apply to sub-regional breakdowns. Throughout this exercise we are implicitly assuming that, within a region, the model of consumption is the same for all households

---

[16] The standard errors for the estimates of expected poverty, which are found in footnote 14, are around the true value of *expected* poverty. If one is specifically concerned with the headcount, then one might be more interested in the variance of estimated expected poverty around the headcount rate. To calculate this, the variance given must be increased by the variance of the headcount, say h, around the true value of expected poverty. This is $h(1-h)/N$. Thus, the variance of our poverty estimator around the *headcount* rate does decline as the sample size increases, because the headcount becomes a better measure of the expected poverty of households with given observable characteristics.

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

irrespective of which province, county or community they reside in.[18] This is an assumption we cannot test, and at very fine levels of disaggregation it might be less appealing. The desired degree of disaggregation will also depend on the availability of other sources of information on the poverty of individuals which might become available locally. Finally, other methods of local targeting, such as self-targeting, will become more important and effective at certain levels of disaggregation. The process of constructing a poverty map is thus likely to be a sequential process of gradual disaggregation until it seems there is no further insight gained from further disaggregation. At all stages it will be very important to keep in mind the purpose of the poverty map.

In Table 6 we present a breakdown of the headcount rate of poverty in Ecuador by province, distinguishing between rural and urban areas in each. It is clear from Table 6 that poverty rates across provinces can vary considerably. Figure 1, which also shows the confidence intervals around the calculated headcount rates, illustrates this further. Once again poverty in urban areas appears markedly lower than in rural areas.

While we cannot reject that urban poverty rates are the same across provinces, it is

[18] Partly this depends on whether ϵ_i is viewed primarily as a household fixed effect or whether most variation is idiosyncratic shocks to income. We assume that ϵ_i has mean zero at the level of estimation. Moving to subgroups, however, this will, in general, no longer be true if ϵ_i is a household fixed effect. In this case, households in one subgroup may have relatively high incomes, given their observable characteristics, compared to those in another subgroup with similar characteristics. The expected poverty measure would then tend to be biased, understating the wellbeing of the first group and overstating that of the second.

possible to distinguish statistically between the incidence of poverty of some rural provinces.[19] In fact, of the 210 possible pairwise comparisons across rural provinces, 85 (approximately 41%) are significantly different at a 90% confidence level. This is worth emphasizing as we were not able to distinguish poverty rates between rural areas at a higher level of aggregation (see Table 5). Disaggregation has led to a more sharply defined poverty profile -- something we would expect as larger areas are not homogenous and composed of pockets of poverty and wealth. With the differences in poverty becoming more accentuated at lower levels of aggregation our ability to distinguish poverty rates in statistical terms has improved as well.

---

[19] The standard error on the difference in poverty rates between two provinces in different regions is calculated in the same manner as was described earlier. However, because the parameter estimates determining the imputed expenditure figures are the same for all provinces within a given region, the standard error on the difference in the incidence of poverty between two provinces in a given region is

$$Var(P_1^*-P_2^*) = \left(\frac{\partial(P_1^*-P_2^*)}{\partial\hat{\beta}}\right)' Var(\hat{\beta})\left(\frac{\partial(P_1^*-P_2^*)}{\partial\hat{\beta}}\right) + \left(\frac{\partial(P_1^*-P_2^*)}{\partial\hat{\sigma}}\right)^2\hat{\sigma}^2 .$$

$$\frac{\partial(P_1^*-P_2^*)}{\partial\hat{\beta}_j} = \frac{1}{N_1}\Sigma_{i=1}^{N_1}\left(-\frac{x_{ij}}{\hat{\sigma}^2}\right)\phi\left(\frac{lnz-X_i'\hat{\beta}}{\hat{\sigma}}\right) - \frac{1}{N_2}\Sigma_{k=1}^{N_2}\left(-\frac{x_{kj}}{\hat{\sigma}}\right)\phi\left(\frac{lnz-X_k'\hat{\beta}}{\hat{\sigma}}\right),$$

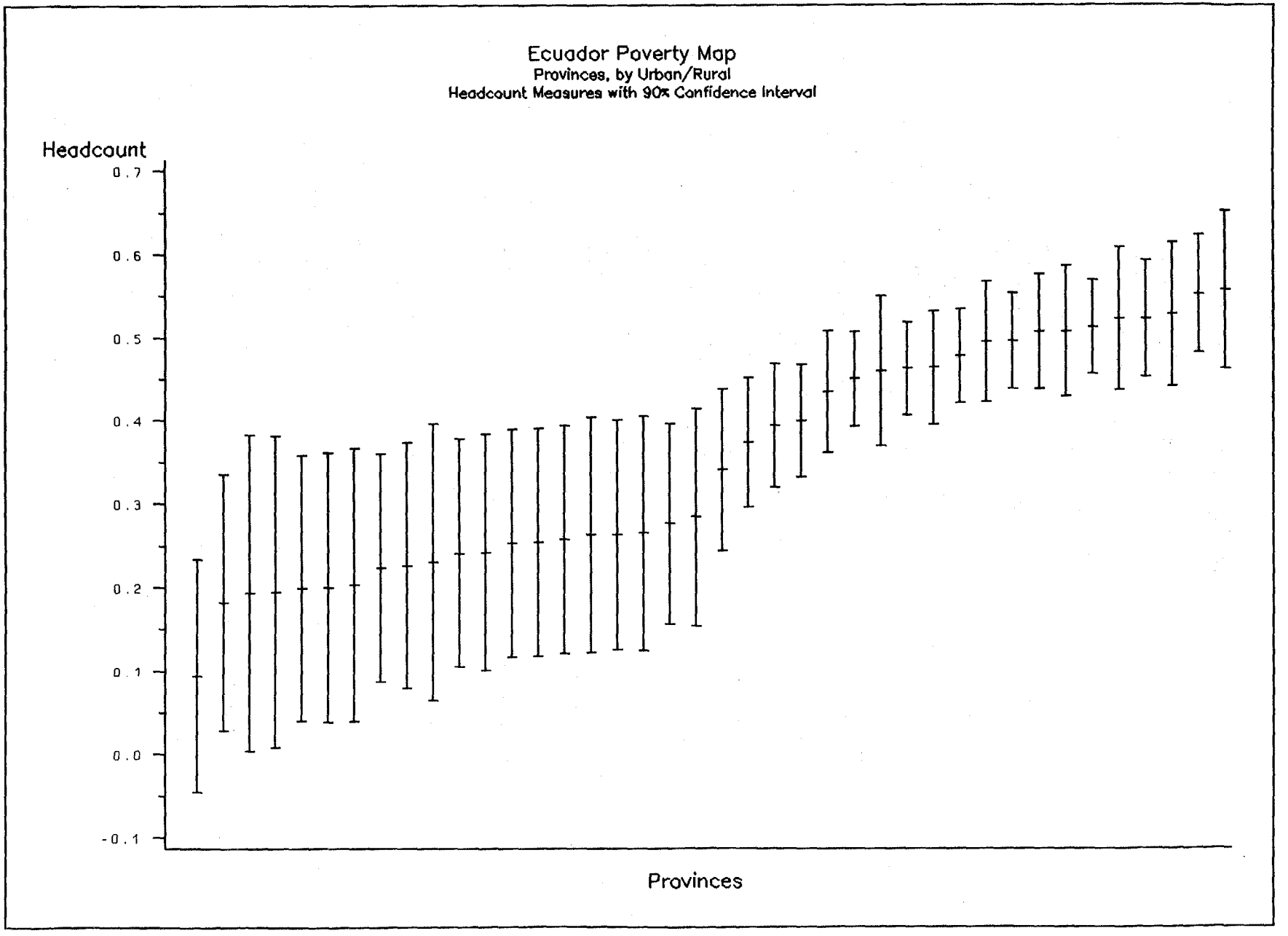
where N_1 and N_2 are the number of census households in province 1 and 2, respectively, and the subscript j indicates the *jth* element of the given vector, and

$$\frac{\partial(P_1^*-P_2^*)}{\partial\hat{\sigma}} = \frac{1}{N_1}\Sigma_{i=1}^{N_1}\left(-\frac{lnz - X_i'\hat{\beta}}{\hat{\sigma}^2}\right)\phi\left(\frac{lnz-X_i'\hat{\beta}}{\hat{\sigma}}\right) - \frac{1}{N_2}\Sigma_{k=1}^{N_2}\left(-\frac{\ln z - X_k'\hat{\beta}}{\hat{\sigma}^2}\right)\phi\left(\frac{lnz-X_k'\hat{\beta}}{\hat{\sigma}}\right) .$$

Ecuador Poverty Map: Urban and Rural Provinces

REGION	PROVINCE	Headcount	standard error
Costa	rural Galapagos	0.0937	0.0850
Sierra	urban Azuay	0.1817	0.0932
Oriente	urban Pastaza	0.1934	0.1150
Oriente	urban Zamora Chinchipe	0.1947	0.1131
Sierra	urban Chimborazo	0.1984	0.0965
Sierra	urban Tungurahua	0.2002	0.0980
Costa	urban El Oro	0.2032	0.0995
Costa	urban Esmereldas	0.2235	0.0831
Sierra	urban Cotopaxi	0.2262	0.0893
Oriente	urban Morona Santiage	0.2307	0.1007
Costa	urban Manabi	0.2411	0.0829
Sierra	urban Loja	0.2418	0.0859
Sierra	urban Canar	0.2534	0.0829
Costa	urban Los Rios	0.2538	0.0830
Costa	urban Guayas	0.2572	0.0831
Sierra	urban Imbabura	0.2627	0.0855
Sierra	urban Pichincha	0.2628	0.0834
Sierra	urban Carchi	0.2649	0.0855
Costa	urban Guayaquil	0.2760	0.0730
Sierra	urban Quito	0.2840	0.0790
Costa	rural El Oro	0.3409	0.0591
Costa	rural Guayas	0.3738	0.0474
Sierra	rural Tungurahua	0.3937	0.0450
Sierra	rural Pichincha	0.3995	0.0413
Costa	rural Los Rios	0.4343	0.0446
Sierra	rural Azuay	0.4498	0.0346
Costa	rural Manabi	0.4591	0.0550
Sierra	rural Canar	0.4622	0.0338
Costa	rural Esmereldas	0.4630	0.0412
Sierra	rural Bolivar	0.4769	0.0342
Sierra	rural Loja	0.4937	0.0439
Sierra	rural Imbabura	0.4950	0.0350
Sierra	rural Carchi	0.5061	0.0418
Oriente	rural Sucumbios	0.5066	0.0477
Oriente	rural Pastaza	0.5119	0.0344
Oriente	rural Zamora Chinchipe	0.5217	0.0521
Oriente	rural Morona Santiago	0.5224	0.0426
Sierra	rural Chimborazo	0.5273	0.0524
Oriente	rural Napo	0.5517	0.0429
Sierra	rural Cotopaxi	0.5572	0.0575

Figure 1


Ecuador Poverty Map
Provinces, by Urban/Rural
Headcount Measures with 90% Confidence Interval
Headcount
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
Provinces

V. Concluding Remarks

In many developing countries poverty maps play an important role in guiding the allocation of public spending for poverty alleviation purposes. A poverty map is essentially a geographical 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 geographical disaggregation.

To achieve such fine levels of disaggregation it is essential to be able to work with very large data sets. However, it is rare to find survey data which are both large in sample size but detailed in terms of the information they collect on household welfare. In general, there is a tradeoff between size and quality because both goals are costly in financial and administrative terms.

In this paper we have explored the possibility of combining the best of two different sources of data in order to construct a disaggregated poverty map which is also based on a sound indicator of welfare. We illustrated, first, that constructing a poverty map using census data, but based on an *ad-hoc* welfare indicator can be quite risky. Transfer programs to alleviate poverty, based on such a map, might reach only a subset of the intended beneficiaries and might entail considerable leakages to the non-poor.

We then suggest an alternative approach: using household data in a high-quality, but small, living standards survey for Ecuador (ECV 1994), we directly model consumption as a function of explanatory variables which are also present in the census. Because even the relatively few explanatory variables common across the census and the ECV were able to 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, was quite close to that calculated from the ECV.

At the same time, we also show that confidence intervals on the poverty rates calculated from the census are not small. In our example, we found that in only a fraction of comparisons across provinces in Ecuador were differences in poverty rates *statistically* significant. While differences at lower levels of aggregation may be more accentuated, with a higher percentage statistically significant, it remains of utmost importance to calculate such confidence intervals in the application of the methodology to determine what conclusions may be drawn. The magnitude of the calculated standard error here also implies that, where possible, other sources of information and targeting possibilities should complement the investigation. Certainly, it would seem highly unadvised to push this methodology to the extreme of deriving a directory which purports to identify individual poor households.

Probably the most useful practical application to which this methodology can be devoted lies in comparisons against regional patterns of *other* indicators of well-being, opportunity, and access. For example, one could overlay a map documenting, say, regional patterns of access to primary health care centers against our map illustrating where poverty is concentrated. Such an exercise could be of considerable use to policy makers for a number of reasons. It might help policy makers decide *where* to prioritize efforts to expand access to primary health centers. It could also help in thinking about *how* one might want to expand access to primary health -- one might want to subsidize access in poor areas, but experiment with cost-recovery methods in the less poor ones. Furthermore, a close correlation between, say, regional patterns of rural poverty and road access, might also offer clues as to possible *causes* of poverty. This type of exercise could be undertaken with respect to a wide range of indicators: levels of health and education; ethnicity and indigeneity; access to infrastructure and other public services; land quality and ecology; environment, and so on.

Finely, as mentioned in the introduction, an ability to construct finely disaggregated poverty maps

might also inform broader research questions. One direction is to analyze the spatially varying relationship between distributional outcomes and economic performance within a country, in a manner analogous to the cross-country analysis which currently receives much attention among researchers. This approach may well avoid some of the methodological concerns which the cross-country approach raises. There are also other research questions which could be tackled. For example, underlying some of the current arguments in favor of decentralizing poverty alleviation efforts is a notion that local communities themselves are best placed to identify the kinds of interventions which are most beneficial to the poor within those communities. This position hinges somewhat on the contention that at the local community level public resources are less likely to be captured by a subset of non-poor households. This is probably linked to the degree of inequality at the community level; something which has traditionally not been easy to investigate. With the methodology presented here, household level consumption inferred from the census could be analyzed to assess the extent of inequality within smaller geographic areas.

References

Alesina, A. and Rodrik, D. (1994) 'Distributive Policies and Economic Growth' *Quarterly Journal of Economics*, 109: 465-490.

Anand, S. and Kanbur, R. (1993) 'The Kuznets Process and the Inequality-Development Relationship'*Journal of Development Economics* 40:25-52.

Angrist, J.D. and A.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, pp. 328-336.

Arellano, M. and C.Meghir (1992), 'Female Labour Supply and on the Job Search: an Empirical Model Estimated using Complementary Data Sets'. *Review of Economic Studies* 59, pp. 537-559.

Atkinson, A. (1989) Poverty and Social Security (Hemel Hempstead: Harvester Wheatsheaf).

Bramley, G. and G. Smart (1996), 'Modelling Local Income Distributions in Britain', *Regional Studies* 30, pp. 239-255.

Bruno, M., Ravallion, M. and Squire, L. (1998) 'Equity and Growth in Developing Countries: Old and New Perspectives on the Policy Issues', in Tanzi, V. and Chu, K. (eds) (1998) Income Distribution and High Quality Growth (Cambridge: MIT Press).

Chaudhuri, S. and Ravallion, M. (1994) 'How Well Do Static Welfare Indicators Identify the Chronically Poor?', *Journal of Public Economics*, 53(3): 367-394.

Deininger, K. and Squire, L. (1996) 'A New Data Set for Measuring Income Inequality' *World Bank Economic Review* 10:565-592.

Farrell, P., B. MacGibbon and T. J. Tomberlin (1997), 'Empirical Bayes Estimation Using Logistic Regression Models and Summary Statistics', *Journal of Business and Economic Statistics* 15, pp. 101-108.

Fields, B. (1989) 'Changes in Poverty and Inequality in Developing Countries' *World Bank Research Observer* 4:167-185.

FONCODES (1995), Mapa de Pobreza en Peru, Lima.

Government of El Salvador (1995) 'Priorizacion de Municipios a Partir de Datos Censales', Direccion General de Politica Economica y Social, Ministerio de Coordinacion del Desarollo Economico y Social.

Haddad, L. and Kanbur, R. (1990) 'How Serious is the Neglect of Intra-Household Inequality?' *The*

Economic Journal, 100: 866-881.

Hentschel, J. and Lanjouw, P. (1995) 'Perfil de la Pobreza en Ecuador', *Cuestiones Economicas*, Banco Central del Ecuador.

Hentschel, J. and Lanjouw, P. (1996) 'Constructing an Indicator of Consumption for the Analysis of Poverty: Principles and Illustrations with Reference to Ecuador', LSMS Working Paper No. 124, Policy Research Department, the World Bank.

INEI (1996), Metodologia Para Determinar el Ingreso y la Proporcion de Hogares Pobres, Lima.

Isaki, C.T. (1990), 'Small-Area Estimation of Economic Statistics', *Journal of Business and Economic Statistics* 8, pp.435-441.

Lanjouw, J. and Lanjouw, P. (1997) 'Poverty Comparisons with Noncompatible Data: Theory and Illustrations' Policy Research Working Paper 1709, Policy Research Department, the World Bank.

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

Paulin, G. D. and D.L. Ferraro (1994), 'Imputing Income in the Consumer Expenditure Survey', *Monthly Labor Review*, December, pp. 23-31.

Persson, T. and Tabellini, G. (1994) 'Is Inequality Harmful for Growth?' *American Economic Review* 84: 600-621.

Purcell, N.J. and L. Kish (1980), 'Postcensal Estimates for Local Areas (or Domains)', *International Statistical Review* 48, pp.3-18.

Ravallion, M. (1988) 'Expected Poverty Under Risk-Induced Welfare Variability' *The Economic Journal*, 98:1171-1182.

Ravallion, M. (1994) Poverty Comparisons (Chur: Harwood Academic Publishers).

Sen, A.K. (1984) Resources, Values and Development (Oxford: Blackwell).

van de Walle, D. and Nead, K. (eds) (1995) Public Spending and the Poor: Theory and Evidence (Baltimore: Johns Hopkins University Press).

World Bank (1996) 'Ecuador Poverty Report' World Bank Country Study, the World Bank.

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1911	The Internationalization of Financial Services in Asia	Stijn Claessens Tom Glaessner	April 1998	R. Vo 33722
WPS1912	Pay and Grade Differentials at the World Bank	Deon Filmer Margaret Grosh Elizabeth King Dominique van de Walle	April 1998	C. Bernardo 31148
WPS1913	The 1994 Currency Crisis in Turkey	Oya Celasun	April 1998	K. Labrie 31001
WPS1914	Distinguishing between Types of Data and Methods of Collecting Them	Jesko Hentschel	April 1998	PREM Advisory 87736
WPS1915	Distortionary Effects of State Trading in Agriculture: Issues for the Next Round of Multilateral Trade Negotiations	Merlinda Ingco Francis Ng	April 1998	M. Fernandez 33766
WPS1916	The Size, Origins, and Character of Mongolia's Informal Sector during the Transition	James H. Anderson	May 1998	P. Sintim-Aboagye 37656
WPS1917	Financial Liberalization and Financial Fragility	Asli Demirgüç-Kunt Enrica Detragiache	May 1998	P. Sintim-Aboagye 37656
WPS1918	How Does Foreign Entry Affect the Domestic Banking Market?	Stijn Claessens Asli Demirgüç-Kunt Harry Huizinga	May 1998	R. Vo 33722
WPS1919	The Empirical Effects of Performance Contracts: Evidence from China	Mary Shirley Lixin Colin Xu	May 1998	P. Sintim-Aboagye 38526
WPS1920	Education and Earnings in a Transition Economy (Vietnam)	Peter R. Moock Harry Anthony Patrinos Meera Venkataraman	May 1998	M. Christian 36736
WPS1921	Making Voice Work: The Report Card on Bangalore's Public Service	Samuel Paul	May 1998	C. Bernardo 31148
WPS1922	Regional Groupings Among Microstates	Soamiely Andriamananjara Maurice Schiff	May 1998	L. Tabada 36896
WPS1923	When Vintage Technology Makes Sense: Matching Imports to Skills	Giorgio Barba Navaretti Isidro Soloaga Wendy Takacs	May 1998	L. Tabada 36896
WPS1924	Voucher Privatization with Investment Funds: An Institutional Analysis	David Ellerman	May 1998	M. Murray 36095

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1925	Half a Century of Development Economics: A Review Based on the *Handbook of Development Economics*	Jean Waelbroeck	May 1998	J. Sweeney 31021
WPS1926	Do Budgets Really Matter? Evidence from Public Spending on Education and Health in Uganda	Emmanuel Ablo Ritva Reinikka	June 1998	K. Rivera 34141
WPS1927	Revenue-productive Income Tax Structures and Tax Reforms in Emerging Market Economies: Evidence from Bulgaria	Fareed M. A. Hassan	June 1998	A. Panton 85433